



Rapport
2022:18

Kartlegging av nordisk forskning på eksamen

Elisabeth Hovdhaugen, Fride Flobakk-Sitter, Sabine Wollscheid,
Lone Wanderås Fossum og Lene Korseberg

NIFU

Rapport
2022:18

Kartlegging av nordisk forskning på eksamen

Elisabeth Hovdhaugen, Fride Flobakk-Sitter, Sabine Wollscheid,
Lone Wanderås Fossum og Lene Korseberg

Rapport 2022:18

Utgitt av Nordisk institutt for studier av innovasjon, forskning og utdanning (NIFU)
Adresse Postboks 2815 Tøyen, 0608 Oslo. Besøksadresse: Økernveien 9, 0653 Oslo.

Prosjektnr. 21306

Oppdragsgiver Utdanningsdirektoratet
Adresse Postboks 9359 Grønland, 0135 Oslo

Fotomontasje NIFU

ISBN 978-82-327-0568-9
ISSN 1892-2597 (online)



Copyright NIFU: CC BY 4.0

www.nifu.no

Forord

Tema for denne rapporten er sluttvurdering i de nordiske land. Rapporten er to-delt, der del 1 beskriver utdanningssystemet og vurderingssystemet i de fem nordiske land, i tillegg til å vise i hvilken grad sluttvurdering og karakterer brukes for opptak til høyere utdanning, mens del 2 er en systematisk kunnskapskartlegging om eksamen og sluttvurdering i de nordiske land.

Rapporten er skrevet på oppdrag for Utdanningsdirektoratet (Udir), med Elisabeth Hovdhaugen som prosjektleder. I arbeidet med rapporten har Elisabeth Hovdhaugen sammen med Lene Korseberg hatt ansvar for landsammenligningen, mens Sabine Wollscheid og Fride Flobakk-Sitter har hatt felles ansvar for kunnskapskartleggingen. Lone Wanderås Fossum har hatt ansvar for utforming av literatursøk og innhenting av artikler. I tillegg har Fossum bidratt i de tidlige stadiene av utvalgsprosessen i gjennomføringen av kartleggingen.

Kapittel 1 er skrevet av Korseberg og Hovdhaugen. Korseberg, Hovdhaugen og Flobakk-Sitter har alle bidratt i kapittel 2, mens kapittel 3 er ført i penn av Korseberg. Wollscheid og Fossum har skrevet kapittel 4, og Flobakk-Sitter og Wollscheid har sammen skrevet kapittel 5 og 6. Alle forskerne har vært involvert i avslutningskapitlet.

For å sikre at landbeskrivelsene er korrekt beskrevet har vi brukt NIFUs internasjonale nettverk. Professor Juhani Rautopuro ved University of Jyväskylä har svart på spørsmål om det finske systemet. Vi vil rette en takk til seniorforsker Jens-Peter Thomsen, Vive, professor Jón Torfi Jonasson, University of Iceland og professor Christian Lundahl, Örebro universitet, som har lest og kommentert de respektive beskrivelsene av Danmark, Island og Sverige. Rapporten er lest og kommentert av forsker Berit Lødding, NIFU, samt av en anonym reviewer i Udir.

Oslo, september 2022

Vibeke Opheim
direktør

Mari Elken
forskningsleder

Innhold

| | |
|---|-----------|
| Sammendrag | 7 |
| 1 Innledning..... | 10 |
| 1.1 Tidligere kartlegginger av forskning på eksamen | 10 |
| 1.2 Studiens omfang, problemstillinger og metode..... | 11 |
| 1.2.1 Del 1: Landsammenligningen | 12 |
| 1.2.2 Del 2: Den systematiske kunnskapskartleggingen | 13 |
| 1.2.3 Avgrensninger..... | 14 |
| 1.3 Gangen i rapporten..... | 15 |
| DEL 1 LANDSAMMENLIGNING | 16 |
| 2 Sluttvurdering i de nordiske landene | 17 |
| 2.1 Norge..... | 17 |
| 2.1.1 Vurdering..... | 19 |
| 2.1.2 Opptak til høyere utdanning | 25 |
| 2.2 Danmark | 26 |
| 2.2.1 Vurdering..... | 27 |
| 2.2.2 Opptak til høyere utdanning | 32 |
| 2.3 Finland..... | 34 |
| 2.3.1 Vurdering..... | 35 |
| 2.3.2 Opptak til høyere utdanning | 39 |
| 2.4 Island..... | 40 |
| 2.4.1 Vurdering..... | 40 |
| 2.4.2 Opptak til høyere utdanning | 44 |
| 2.5 Sverige | 44 |
| 2.5.1 Vurdering..... | 46 |
| 2.5.2 Opptak til høyere utdanning | 51 |
| 3 Diskusjon av landssammenligningen - kvalitetssikring av sluttvurdering | 53 |
| 3.1 Validitet og reliabilitet | 54 |
| 3.1.1 Validitet i lys av sluttvurderingens formål..... | 55 |

| | | |
|---|---|------------|
| 3.1.2 | Sluttvurderingens form og innhold..... | 57 |
| 3.1.3 | Tolkningsfellesskap | 58 |
| 3.2 | Rettferdighet..... | 59 |
| 3.2.1 | Valg av agenter og «safeguards» i sluttvurderingsprosessen..... | 60 |
| 3.2.2 | Mulighet for å klage på sluttvurderingen | 62 |
| DEL 2 KUNNSKAPSKARTLEGGING | | 65 |
| 4 | Systematisk kartlegging - metode | 66 |
| 4.1 | Om metoden..... | 66 |
| 4.2 | Metodiske styrker og begrensninger..... | 72 |
| 5 | Funn fra kartleggingen..... | 74 |
| 5.1 | Oversikt over de inkluderte studiene..... | 74 |
| 5.2 | Oppsummering av de inkluderte studiene..... | 77 |
| 5.2.1 | Studier knyttet til et overordnet systemperspektiv på eksamen og sluttvurderinger | 77 |
| 5.2.2 | Studier knyttet til institusjonelle faktorer og organisering av eksamen og sluttvurdering..... | 84 |
| 5.2.3 | Studier knyttet til gjennomføring, praksiser og erfaringer knyttet til eksamen og sluttvurdering..... | 97 |
| 5.3 | Kunnskapsklynger..... | 109 |
| 5.3.1 | Individuelle faktorer som påvirker eksamen | 111 |
| 5.3.2 | Eksterne og institusjonelle faktorer som påvirker eksamen | 112 |
| 5.3.3 | Omtale av andre vurderingsformer | 112 |
| 6 | Diskusjon av kunnskapskartleggingen | 114 |
| 6.1 | Kunnskapslandskapet..... | 114 |
| 6.1.1 | Distinkte forskningsområder..... | 114 |
| 6.1.2 | Områder med lite forskning | 115 |
| 6.1.3 | Diskusjon..... | 116 |
| 7 | Konklusjon og mulige forskningsimplikasjoner..... | 118 |
| Referanser..... | | 122 |
| Vedlegg 1..... | | 130 |
| Vedlegg 2..... | | 131 |

Sammendrag

Tema for denne rapporten er sluttvurdering i grunnskole og videregående opplæring i de nordiske landene. Rapporten er todelt og består av en landsammenligning (del 1) og en systematisk kartlegging (del 2).

Del 1 gir en beskrivelse av landenes utdanningssystem og deres system for sluttvurdering og hvilke konsekvenser dette har for opptak til høyere utdanning. Basert på dokumentstudier og landbeskrivelsene ser vi også nærmere på hvordan de ulike landene forholder seg til validitet, reliabilitet og rettferdighet i sluttvurderingen. Denne delen av rapporten dekker følgende problemstillinger: omfang og utforming av sluttvurdering, om sluttvurderingen er standardisert og/eller sentralisert, og hvilke konsekvenser sluttvurderingen får for overgang til neste utdanningsnivå.

Del 2 av rapporten er en systematisk kartlegging av forskning på organisering av sluttvurdering, og de to følgende problemstillingene har guidet analysene:

Hva foreligger av kunnskap om systemet rundt og organiseringen av sluttvurdering og eksamen i de nordiske landene? Hvilke kunnskapshull kan identifiseres i forskningslitteraturen?

Hovedfunnet i del 1 er at, selv om utdanningssystemene i stor grad er veldig like på tvers av de nordiske landene, har man valgt å utforme sluttvurderingen forskjellig. Hovedfunn i del 2 er at vi, til tross for over 3700 identifiserte referanser i de systematiske litteratursøkene, kun fant 19 bidrag som var av høy relevans for kartleggingen. Resultatet viser kunnskapshull særlig knyttet til det overordnede systemet, organiseringen og gjennomføringen av sluttvurdering i de nordiske landene.

Kun Norge og Danmark har eksamen

Landsammenligningen viser at det kun er Norge og Danmark som har eksamen som del av sluttvurderingen i skolen. I Sverige og på Island settes alle karakterer av læreren og det er kun standpunktkarakterer på vitnemålet. I Finland er det ingen eksamen i ordinær skole, alle karakterer settes av lærer, men de som ønsker

studiekompetanse, må ta Studentexamen, det vil si et ekstra sett med prøver som utgjør opptaksgrunnlaget for høyere utdanning.

Landforskjellene har også implikasjoner for landenes tilnærming til validitet og reliabilitet i sluttvurderingen. På den ene siden er det stort samsvar mellom landene, men det er også noen vesentlige forskjeller. Disse er blant annet knyttet til hvorvidt man tar utgangspunkt i sentralt eller lokalt gitte kompetansemål, samt hvorvidt man har standardisert og ekstern rettet eksamen eller lokalt bestemt sluttvurdering med standpunktkarakter satt av faglærer.

Landene har videre ulike tilnærming for å sikre at sluttvurderingen gjøres på en korrekt måte, såkalte «safe guards». I Norge og Danmark brukes eksamen som en form for ekstern sluttvurdering, der elevene fullfører samme prøve og vurderes etter de samme kriteriene. Dette kan tenkes som en form for ekstern kvalitetssikring. I Sverige fyller de nasjonale prøvene en tilsvarende funksjon, ved at de bidrar til en likeverdig vurdering av elevenes måloppnåelse på tvers av de ulike faglærerne. I tillegg åpner alle de nordiske landene opp for muligheten for å klage på sluttvurderingen, selv om denne muligheten i stor grad begrenser seg til formelle og prosessuelle feil når det gjelder standpunktkarakter. En ytterligere forskjell er at muligheten til å klage på eksamen i Danmark begrenser seg til såkalte «rettslige feil», mens man i Norge kan be om en ny innholdsmessig vurdering av eksamenskarakteren på skriftlig eksamen.

Få studier av eksamen

Kunnskapskartleggingen har rettet oppmerksomheten mot forskning som tar for seg selve systemet for, og organiseringen av sluttvurdering i de nordiske landene. I kartleggingen finner vi noe forskning knyttet til dette området, men en nærmere gjennomgang viser at kunnskapslandskapet har tydelige kunnskapsklynger og kunnskapshull.

Vi identifiserte 3765 forskningsbidrag i de systematiske litteratursøkene, og 19 av disse var av høy relevans for kartleggingen. Dette er et relativt lite antall, og vi valgte derfor å gå utover en ren kartlegging og oppsummerte de inkluderte studiene. For å belyse noen sentrale sider ved sluttvurderingssystemet på tvers av de nordiske landene, kategoriserte vi studiene i tre hovedområder (makro-, meso- og mikronivå). I tillegg identifiserte vi tre relaterte kunnskapsklynger, av temaer som påvirker eksamen, men som ikke tok opp selve organiseringen eller systemet rundt sluttvurdering. Disse dekker forskning på individuelle faktorer ved eleven som kan påvirke eksamen, forskning på eksterne faktorer som påvirker eksamen, og forskning på andre vurderingsformer. Disse kunnskapsklyngene bidrar til å belyse ulike variabler som kan påvirke elevens sluttkarakterer, samt belyse spesifikke sider ved vurdering eller eksamen.

Kunnskapshullene vi finner er særlig knyttet til det overordnede systemet, organiseringen og gjennomføringen av sluttvurdering generelt (inkludert eksamen) i de nordiske landene – hvilket nettopp er fokuset for vår systematiske kartlegging. Dette inkluderer manglende forskning på den helhetlige prosessen fra systemnivå til det mer praksisnære nivået i grunnskolen og videregående opplæring. Videre mangler det også gode sammenligninger mellom landene, og det er relativt lite forskning som knytter sluttvurderingssystemet i Norden opp mot kvalitet- og rettferdighetsperspektivet.

Mulige fremtidige forskningstemaer - behov for mer forskning

Til slutt har vi forslått noen mulige fremtidige forskningstemaer som kommer ut av den systematiske kartleggingen og landsammenligningen.

- Studier som tar for seg elevperspektivet på sluttvurdering.
- En mer helhetlig tilnærming i forskningen på eksamen og sluttvurderinger i skolen, dvs. studier som ser vurdering og sluttvurdering på ulike nivåer i forhold til hverandre.
- Studier som tar for seg rettferdighetsperspektivet i organisering av sluttvurdering, og hvordan individuelle faktorer påvirker sluttvurderingens reliabilitet.
- Internasjonale sammenligninger av sluttvurdering i ulike land; organisering, ressursbruk, involvering av ulike aktører, prinsipper for utforming av systemer og ordninger.
- Mer inngående studier og sammenligninger av de ulike sluttvurderingssystemene i Norden, for eksempel hvorfor noen land har eksamen og andre ikke, og hva forskjellene betyr for organisering, ressursbruk, elever og andre aktører.

1 Innledning

Denne rapporten presenterer en systematisk kunnskapskartlegging av internasjonal forskning på eksamen, med fokus på forskning fra Norden – forstått som Norge, Sverige, Danmark, Finland og Island. Den inkluderer også en gjennomgang og sammenligning av sluttvurderingssystemene i de nordiske landene. Hensikten med rapporten er å sammenfatte hva som finnes av litteratur og forskning om eksamen og avsluttende prøvers rolle i sluttvurderingssystemet og belyse sentrale funn og utfordringer. Studier og utredninger som er publisert nylig (se blant annet Blömeke et al., 2019a, 2019b), samt innføring av nye læreplaner og videreutvikling av eksamen, har avdekket flere kunnskapshull når det gjelder denne typen forskning i Norge. Prosjektet har hatt som mål å sammenfatte kunnskapsgrunnlaget om sluttvurdering i de nordiske landene, samt identifisere hvilke kunnskapshull som fortsatt finnes.

1.1 Tidligere kartlegginger av forskning på eksamen

Ludvigsen-utvalget anbefalte i hovedutredning *Fremtidens skole. Fornyelse av fag og kompetanser* at det burde nedsettes et ekspertutvalg som skulle «vurdere dagens eksamenssystem og se på hvordan standpunktvurdering og eksamen samlet sett kan gi pålitelig og relevant informasjon om elevenes kompetanse» (NOU 2015:8, s. 84). I forlengelse av dette ble det pekt på behovet for å sammenstille kunnskapsgrunnlaget vi har om eksamen i dag. Kunnskapsdepartementet nedsatte en ekspertgruppe – Eksamensgruppa – for nettopp dette formålet. Den skulle se eksamen i lys av revideringen av læreplanen (Fagfornyelsen) og dens utvidede kompetansebegrep, i tillegg til utviklingen av digital teknologi som åpner for andre og alternative prøveformer.

Samtidig var det mange aspekter ved det gjeldende eksamenssystemet som ikke skulle vurderes av Eksamensgruppa. Både standpunkt- og eksamenskarakter skulle fortsatt være en del av den individuelle sluttvurderingen, og omfanget det hadde da arbeidet til gruppen startet, skulle i stor grad videreføres. Videre måtte også et fremtidig eksamenssystem kunne gjennomføres innenfor den samme

ressursrammen som i dag. Ekstern vurdering og klagerett slik det har vært tidligere, skulle videreføres.

Utvalgets arbeid resulterte i en kunnskapsoversikt som blant annet tok for seg utviklingen av dagens eksamenssystem, sett i et langt historisk perspektiv. Utvalget sporer dagens eksamenssystem helt tilbake til fremveksten av lærerprofesjonen på 1700- og 1800-tallet, utbyggingen av enhetsskolen gjennom 1800- og 1900-tallet, samt etterkrigstidens diskusjoner om karaktersetting og reformer i utdanningssystemet de siste 30 årene (fra 1990-tallet og frem til innføringen av Kunnskapsløftet). Gjennomgangen viser hvordan synet på karakterer og eksamen har endret seg over tid. For eksempel er en av de sentrale debattene de siste partårene diskusjonen om vurderingsprinsipper, om karakterer skal være normbaserte, dvs. bygge på en relativ skala, eller mål-/kriteriebaserte, der skalaen forholder seg til oppsatte kriterier eller mål. Innføringen av Kunnskapsløftet markerer en full overgang til karakterer som tar utgangspunkt i elevenes måloppnåelse, der eksamen ses på som sentralt for å kunne oppnå en upartisk vurdering. I tillegg diskuterte Eksamensgruppa formålet med eksamen, samt utformingen av eksamen og betydningen av validitet, reliabilitet og rettferdighet, i det norske skolesystemet.

Selv om Eksamensgruppa i all hovedsak fokuserte på nasjonal forskning, inkluderte de også internasjonal forskning der det var formålstjenlig. Som de påpeker, også her finnes det generelt «lite forskning om eksamen. De studiene som finnes, har stort sett blitt gjennomført fra to ulike perspektiver: et målingsperspektiv med vekt på studier og klassiske kvalitetskriterier (...), eller et skoleperspektiv med vekt på studier om hvordan skolepraksisen blir berørt av eksamen» (Blömeke et al., 2019a, s. 7-8). I tillegg påpekte Eksamensgruppa gjennom sitt arbeid at det fortsatt var mange aspekter ved den norske eksamensordningen som ikke er godt nok dokumentert gjennom forskning. Dette inkluderer elevperspektivet og rettferdighetsbegrepet, forskjeller i vurderingspraksis og bruk av karakternivåer. I tillegg mangler det kunnskap om fagenes egenart, hvilke konsekvenser dette kan ha for eksamen og sammenligning av ulike eksamensformer, for å nevne noe.

Dette er bakgrunnen til at Utdanningsdirektoratet ønsket å få gjennomført en kartlegging av forskning på eksamen i Norge og sammenlignbare land, her forstått som Norden. I denne rapporten har NIFU bygd videre på funnene presentert i Eksamensgruppas kunnskapsoversikt, ved å supplere informasjonen presentert der med informasjon om sluttvurderingssystemene i de fem nordiske land

1.2 Studiens omfang, problemstillinger og metode

Formålet med dette prosjektet har vært å sammenfatte forskning på sluttvurdering, med fokus på eksamen i Norden og belyse sentrale funn, utfordringer og

virksomheter på tvers av landene. Metodene som er benyttet for å gjennomføre prosjektet er en systematisk kartlegging av internasjonal forskning på feltet, samt en dokumentgjennomgang av informasjon knyttet til organiseringen av sluttvurdering og eksamen. Fokus har vært på sluttvurderingens organisering og gjennomføring, heller enn på sluttvurdering og eksamen som utfallsmål. Dette har vært med på å strukturere analysen som presenteres i rapporten, samt avgrense omfanget av kunnskapstilfanget.

I utgangspunktet skulle fokus for denne evalueringen være forskning på *eksamen* i de nordiske landene. Som beskrevet nærmere i kapittel 2, er det imidlertid kun Norge og Danmark – og til dels Finland¹ – som benytter eksamen som vurderingsform i elevenes sluttvurdering. I både landssammenligningen og i den systematiske kunnskapskartleggingen har vi derfor valgt å fokusere på de ulike landenes sluttvurderingssystem generelt og inkludert informasjon om eksamen spesielt der dette er relevant. I resten av rapporten kommer vi derfor til å bruke begrepet «sluttvurdering» om eksamen og andre avsluttende vurderinger, inkludert standpunkt karakter, og vi vil kun referere spesifikt til eksamen dersom dette er særlig relevant for tematikken som diskuteres.

Denne rapporten er delt inn i to konkrete deler. I del 1 presenteres en landsammenligning av utdannings- og sluttvurderingssystemet i de fem nordiske landene, mens del 2 presenterer funnene fra den systematiske kartleggingen.

1.2.1 Del 1: Landsammenligningen

I rapportens første del redegjør vi for utdannings- og sluttvurderingssystemene i de nordiske landene. Hensikten har vært å identifisere rammene for sluttvurdering i landene. Studien bygger primært på en komparativ tilnærming, med fokus på likheter og forskjeller i sluttvurdering mellom landene. Samtidig mener vi at en forutsetning for å gjøre gode sammenligninger av vurderingssystemene i ulike land er en grundig gjennomgang av hvordan hvert enkelt system fungerer og hvilken kontekst det operer innenfor. Kapitlene i del 1 tar derfor for seg henholdsvis det norske, danske, finske, islandske og svenske systemet (kapittel 2), for deretter å presentere en sammenligning basert på studiens problemstillinger i kapittel 3.

Følgende problemstillinger har ligget til grunn for beskrivelsen og sammenligningen av utdannings- og sluttvurderingssystemet i de nordiske landene:

¹ I det finske systemet må elever som ønsker å begynne på høyere utdanning, ta en Studenteksamen, som vurderer om de er kvalifisert til høyere utdanning. Denne eksamen er imidlertid ikke en del av den formelle sluttvurderingen i grunnskolen eller den videregående opplæringen, se kapittel 2.3.

1. Hva er omfanget og utformingen av sluttvurdering, inkludert eksamen, i de nordiske landene?
 - a. Hvordan vurderes sammenhengen mellom eksamen/sluttvurdering og læreplan/opplæring?
 - b. Hvordan vurderes forholdet mellom eksamenskarakterer og andre typer sluttvurdering?
2. Hvordan fremstilles formålet/hensikten med sluttvurdering i de ulike landene?
3. I hvilken grad er sluttvurderingen standardisert og/eller sentralisert, og hvorfor er denne tilnærmingen valgt?
4. Hvordan gjennomføres sensur/sluttvurdering, og hvilke klagerettigheter har elevene?
5. Hvilke konsekvenser kan sluttvurderingssystemet få for videre utdanning og valg og muligheter underveis i grunntutdanningen i de ulike landene på kort og lang sikt?
6. Hvordan forstås kvalitetsbegrepene validitet, reliabilitet, rettferdighet og håndterbarhet i de ulike landene?

Del 1 baserer seg på en gjennomgang av relevant dokumentasjon, herunder lover, forskrifter, annen offentlig dokumentasjon og sekundærlitteratur, fra de ulike landene. Funnene har blitt analysert basert på lovtekstens ordinære språklige betydning, og i lys av det overordnede formålet med sluttvurdering beskrevet i lovverket og ellers. Funnene fra denne dokumentgjennomgangen har blitt kvalitetssikret med referanse til de relevante landbeskrivelse i Eurydice.² I tillegg har forskere fra Danmark, Sverige og Island lest igjennom utkast til oversikten for sitt land og dermed bidratt til å kvalitetssikre informasjonen. For Finland har vi hatt tilgang på en utdanningsforsker som vi har kunnet stille spørsmål på engelsk til, for å sikre at vi har oppfattet ulike aspekter ved det finske sluttvurderingssystemet korrekt.

1.2.2 Del 2: Den systematiske kunnskapskartleggingen

I del 2 av denne rapporten presenteres en oversikt over internasjonal forskning på eksamen og sluttvurdering i de nordiske landene. I vår oversikt vektlegges systemet rundt sluttvurderinger og eksamen, samt sluttvurderinger og eksamens organisering og gjennomføring i grunnskolen og videregående opplæring i Norden. I oversikten benytter vi forskningsmetoden systematisk kartlegging («systematic mapping review»). Siden denne metoden ikke har som hensikt å besvare spesifikke spørsmål, men heller sikter på å kartlegge, beskrive og katalogisere bakenforliggende kunnskap relatert til et bestemt tema, er problemstillingene relativt

² Eurydice er en nettside administrert av EU-kommisjonen, med beskrivelser av hvordan de europeiske utdanningssystemene er organisert og hvordan de fungerer, se: [National Education Systems | Eurydice \(europa.eu\)](https://eurydice.eu)

åpent formulert. Følgende problemstillinger var ledende for den systematiske kartleggingen:

1. Hva foreligger av kunnskap om systemet rundt og organiseringen av sluttvurdering og eksamen i de nordiske landene?
2. Hvilke kunnskapshull kan identifiseres i forskningslitteraturen?

En systematisk kartlegging tilstreber å fremskaffe kunnskap gjennom objektive kriterier og transparente prosedyrer. Tydeliggjøring av fremgangsmåte, operasjonalisering og inklusjons- og eksklusjonskriterier står dermed sentralt. En detaljert beskrivelse av metoden, operasjonaliseringen og metodiske styrker og begrensninger presenteres derfor i begynnelsen av del 2, i kapittel 4.

1.2.3 Avgrensninger

Særlig to avgrensninger er viktige for den følgende kartleggingen. Den første gjelder språk. I både dokumentgjennomgangen for landsammenligningene (del 1) og i den systematiske kartleggingen (del 2), er kunnskapsgrunnlaget hentet fra litteratur på norsk, svensk, dansk eller engelsk. Siden ingen i prosjektgruppen behersker finsk eller islandsk, faller dokumenter skrevet på disse språkene utenfor kartleggingen, selv om de i utgangspunktet kunne ha vært relevante for tematikken som diskuteres. Relevante dokumenter for arbeidet i del 1 med det norske, danske og svenske systemet er lest av prosjektgruppen på originalspråket. For dokumenter som omhandler det finske systemet, har vi basert oss på de svenske versjonene av lovverket, der disse finnes. Når det gjelder det islandske systemet er de mest sentrale lovtekstene oversatt til engelsk, og det er denne informasjonen som ligger til grunn for beskrivelsen i kapittel 2.4. I del 2 fremkommer de samme språkbarrierene, ved at forskningsbidrag på norsk, engelsk, dansk og svensk er inkludert i den systematiske kartleggingen. Forskningsbidrag på finsk eller islandsk har derimot blitt ekskludert. Dette kan ha resultert i et noe skjevt kunnskapsgrunnlag for henholdsvis Finland og Island.

Den andre viktige avgrensningen gjelder dokumentgrunnlaget som landssammenligningen bygger på (del 1). Som nevnt tidligere, har vi i denne studien gjennomgått relevante lover, forskrifter og annen offentlig dokumentasjon fra de ulike landene. Selv om disse står sentralt for å forstå hvordan sluttvurderingssystemene formelt fungerer i Norden, inneholder de begrenset informasjon om hvorfor de ulike tilnærmingene er valgt og hvilken prosess og eventuell diskusjon som ledet frem til det endelige resultatet man ser i lovverket. De gir heller ikke en beskrivelse av hvordan lovverket og ulike styringsdokumenter tolkes og brukes i praksis, særlig på det lokale nivået. For en slik kartlegging behøves en gjennomgang av annen type dokumentasjon og litteratur, inkludert dokumentasjon fra de

parlamentariske prosessene i de ulike landene, diskusjoner i media, politiske utredninger (ofte fra lengre tid tilbake) og innhenting av nye data. I tillegg til at en gjennomgang, og eventuelt innhenting av denne typen kilder er utenfor denne studiens omfang, ville det ikke ha vært mulig å gjennomgå disse for Island og Finland, da disse ofte skrives på landets språk. Spørsmål om prosess og bakgrunn er dermed i stor grad utelukket fra del 1, med mindre det har kommet opp i det gjennomgåtte dokumentgrunnlaget. Noen av disse temaene har imidlertid blitt fanget opp gjennom den systematiske forskningskartleggingen i del 2.

1.3 Gangen i rapporten

Dette første kapittel presenterer rammen for prosjektet og problemstillingene som de to delene av prosjektet skal dekke, i tillegg til å diskutere begrensninger i dataene. Kapittel 2 og 3 utgjør del 1, landsammenligningen. Kapittel 2 er en gjennomgang av utdanningssystemet og vurderingssystemet i Norge, Danmark, Finland, Island og Sverige. I tillegg tar dette kapitlet også opp hvordan karakterer blir brukt for opptak til høyere utdanning. I kapittel 3 sammenligner og diskuterer vi de fem landene og fokuserer særlig på validitet, reliabilitet og rettferdighet i sluttvurdering. Del 2 omfatter den systematiske kunnskapskartleggingen. Kapittel 4 gjør rede for metoden, ved å si noe om litteratursøk, inklusjons- og eksklusjonskriterier og hvordan artiklene har blitt screenet og analysert. Kapittel 5 er et oversiktskapittel som viser bidragene som er inkludert i kartleggingen, mens kapittel 6 diskuterer funnene i kartleggingen, både med hensyn på hvordan kunnskapslandskapet ser ut, hva det finnes forskning om og hva som er områder med lite forskning. Til slutt oppsummeres alle funnene i et konklusjonskapittel.

DEL 1 LANDSAMMENLIGNING

2 Sluttvurdering i de nordiske landene

Dette kapitlet presenterer de fem nordiske landenes utdanningsystem, hvordan vurdering gjøres og hvordan systemet for sluttvurdering er bygget opp i det enkelte land. I tillegg blir det forklart i hvilken grad karakterer som kommer ut av sluttvurdering brukes i videre opptak til høyere utdanning.

Siden rapporten er norsk, tar vi først utgangspunkt i det norske systemet og presenterer deretter de andre nordiske landene i alfabetisk rekkefølge³. Hensikten med kapitlet er å gi en god bakgrunn og kontekst for kunnskapsoppsummeringen som blir presentert i rapportens andre del. Strukturen på gjennomgangen er følgende: først gir vi en generell oversikt over utdanningsystemet, deretter tar vi for oss ulike aspekter av vurdering og til slutt forklares hvordan opptak til høyere utdanning skjer.

2.1 Norge

Rammeverket for norsk grunnskole- og videregående opplæring er spesifisert i Opplæringslova og tilhørende forskrifter, inkludert læreplanverket. Formålet med opplæringen defineres i paragraf 1-1 og inkluderer en rekke ulike elementer. Blant annet skal opplæringen, i samarbeid med hjemmet, «opne dører mot verda og framtida og gi elevane og lærlingane historisk og kulturell innsikt og forankring». Den skal i tillegg «fremje demokrati, likestilling og vitskapleg tenkjemåte» og gi elever og lærlinger mulighet til å «utvikle kunnskap, dugleik og holdningar for å kunne meistre liva sine og for å kunne delta i arbeid og fellesskap i samfunnet» (Opplæringslova, § 1-1).

I Norge er det obligatorisk tiårig grunnskoleutdanning for alle mellom 6 til 16 år. Barn og unge har dermed rett og plikt til grunnskoleutdanning. Offentlige grunnskoler er gratis og drives av kommunen. Grunnskolen er delt inn i et barnetrinn (barneskole, 1.-7. klasse) og et ungdomstrinn (ungdomsskole, 8.-10. klasse). Etter endt grunnskole kan elever fortsette i videregående opplæring. Hovedregelen er at ungdom som har fullført grunnskolen eller tilsvarende opplæring, har lovfestet rett til tre års gratis videregående opplæring i Norge. Videregående opplæring bygger på den tiårige grunnskoleutdanningen og gir enten

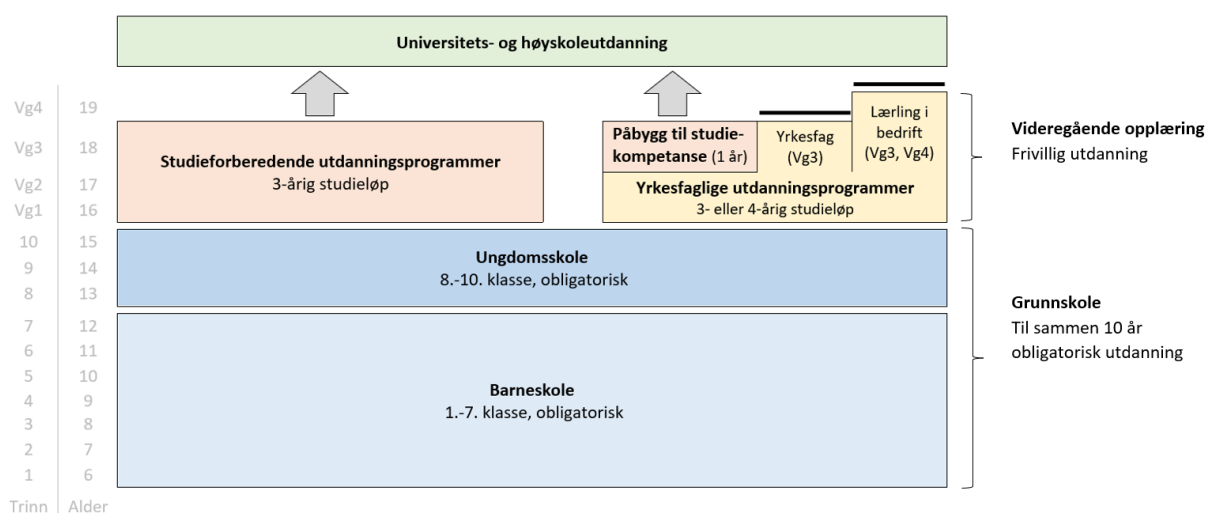
³ I denne rapporten er begreper hovedsakelig skrevet på norsk, med eventuelle nordiske begreper oppgitt i parentes. Dette er gjort for å forenkle sammenligningen mellom landene, ettersom det brukes ulike begreper og termer i tilknytning til utdanningsystemene i de nordiske landene.

studiekompetanse, yrkeskompetanse eller grunnkompetanse⁴. Eleven velger selv hvilket videregående opplæringsprogram han eller hun ønsker å fortsette på, og har rett til plass på ett av tre program vedkommende har søkt på. Den videregående opplæringen deles i to hovedretninger: studieforberedende utdanningsprogram og yrkesfaglig utdanningsprogram.

Studieforberedende utdanningsprogram er treårige studieforløp. Elever kan velge mellom fem ulike program: i) studiespesialiserende utdanningsprogram, ii) idrettsfag, iii) musikk, dans og drama, iv) kunst, design og arkitektur og v) medier og kommunikasjon. Studieforberedende utdanningsprogram gir generell studiekompetanse og gir mulighet for å søke opptak til de fleste studier ved universiteter og høyskoler.

Hovedformen for yrkesfaglige utdanningsprogram er et fireårig forløp, der eleven går to år på skole (vg1 og vg2) og deretter har to års opplæring i bedrift. Dette løpet avsluttes med fag-/svennebrev. Noen yrkesfag avsluttes med vg3 i skole og gir da yrkeskompetanse i det faget. Elever kan i dag velge mellom ti ulike yrkesfaglige utdanningsprogram, som i sin tur leder frem til mange ulike yrker. Sluttkompetanse gjennom yrkesfag, enten som fag-/svennebrev eller yrkeskompetanse, gir i seg selv ikke adgang til høyere utdanning.

Elever som har fullført vg2 på yrkesfag, yrkeskompetanse eller fag- og svenneprøve, kan ta påbygning til studiekompetanse. Påbygning gir eleven de fagene som mangler for å få generell studiekompetanse, og dermed også mulighet for opptak til høyere utdanning.



Figur 2.1 Utdanningssystemet i Norge.

⁴ Grunnkompetanse er kompetanse på et lavere nivå enn full studie- eller yrkeskompetanse.

2.1.1 Vurdering

Formålet med vurdering i det norske utdanningssystemet er å «fremje læring og bidra til lærelyst undervegs, og å gi informasjon om kompetanse undervegs og ved avslutninga av opplæringa i faget» (Forskrift til opplæringslova, § 3-3). I Norge er vurdering av elevenes sluttkompetanse på alle nivåer regulert i forskrift til opplæringslova, kapittel 3. I tillegg må vurderingskapitlet sees i sammenheng med de ulike forskriftsfestede læreplanene, blant annet ved at grunnlaget for vurdering er kompetansemålene i læreplanen i faget.

Forholdet mellom sluttvurdering og læreplan

Ifølge forskrift til opplæringsloven § 3-3 er «grunnlaget for vurdering i fag (...) kompetansemåla i læreplanen i faget. Kompetansemåla skal forståast i lys av teksten om faget i læreplanen. Elevar, lærlingar, lære kandidat og praksisbrevkandidat skal vere kjende med læreplanen i faget». Dette gjelder både for elevenes sluttvurdering (§ 3-14) og deres eksamenskarakterer (§ 3-22).

Forskriftens *Overordnet del* (Utdanningsdirektoratet [Udir], 2017) utgjør, sammen med læreplanene i konkrete fag, det norske læreplanverket. Den overordnede delen skal gi retning for opplæring i de ulike fagene, mens kompetansemålene legger føringer for hva elevene skal lære i hvert enkelt fag. I forlengelse av dette er lærerne forpliktet til å tilrettelegge og gjennomføre opplæringen, inkludert sluttevaluering, i samsvar med de relevante delene av læreplanverket (Opplæringslova, § 2-3 og § 3-4). Med andre ord er hele læreplanverket «grunnlaget for opplæringen, og de ulike delene henger tett sammen og må brukes sammen» (Udir, 2017, s. 2). I Norge er det dermed en tydelig sammenheng mellom sluttvurdering (inklusive eksamen), innholdet i læreplanen i konkrete fag og kompetansemålene i hver enkelt læreplan.

Sluttvurdering

I Norge skiller lovverket mellom underveisvurdering og sluttvurdering. Mens underveisvurdering «skal brukast til å fremje læring, tilpasse opplæringa og auke kompetansen i fag», er formålet med sluttvurdering å gi «informasjon om kompetansen til eleven, lærlingen, lære kandidaten eller praksisbrevkandidaten ved avslutninga av opplæringa i fag» (Forskrift til opplæringslova, § 3-10, § 3-14).

Sluttvurderinger i grunnskolen gis som standpunkt- og eksamenskarakterer, mens sluttvurderinger i videregående opplæring kan gis som standpunkt- og eksamenskarakterer, samt karakterer til fag-/svenneprøve, praksisbrevprøve og kompetanseprøve. Alle karakterer skal føres på elevenes eller kandidatens vitnemål. Standpunktkarakterer og eksamenskarakterer er ulike uttrykk for elevens

sluttkompetanse, og rammene for de to karakterene er ulike. Dermed er karakterene heller ikke direkte sammenlignbare.

Eksamenskarakterer er et uttrykk for den kompetansen eleven viser på eksamen. Det er et sentralt poeng at «eksamen skal gi eleven eller privatisten høve til å vise sin kompetanse i så stor del av faget som mogleg ut frå eksamensforma» (Forskrift til opplæringslova, § 3-22). Hvor stor del av læreplanen en eksamensoppgave dekker, vil derimot kunne variere ut ifra blant annet eksamensform, tidsramme og hvordan kompetansemålene er formulert i læreplanen for faget. Læreplanene for hvert fag fastsetter følgende aspekter ved eksamen: i) om det skal være eksamen, ii) når i opplæringsløpet eksamen skal avholdes, iii) om eksamen er obligatorisk eller elever skal trekkes ut, iv) om eksamensformen skal være skriftlig eller muntlig og v) om eksamen skal være lokalt eller sentralt gitt. Departementet fastsetter hvor mange eksamener det skal være i grunnskolen og på hvert årstrinn i videregående opplæring.

Standpunktkarakterer skal være et uttrykk for elevens samlede kompetanse i faget ved avslutningen av opplæringen. Standpunktkarakterene settes av faglæreren, og fastsettes ut fra lærerens vurdering av elevens samlede kompetanse ut fra kompetansemålene i læreplanen for faget. Det går frem at «eleven skal ha mulighet til å vise kompetanse på flere og varierte måter for at læreren skal få tilstrekkelig kunnskap om elevens samlede kompetanse» (Udir, 2021; Forskrift til opplæringslova, § 3-15). Standpunktkarakteren er derfor basert på et bredere vurderingsgrunnlag over lengre tid enn det som er tilfelle for eksamenskarakterer. Elevenes orden, oppførsel, fravær og forutsetninger skal ikke inngå i vurderingen i et fag, og elevens innsats er kun en del av vurderingen hvor dette står spesifikt i læreplanen for faget. Standpunktkarakterene utgjør størsteparten av karakterene på vitnemålet fra både grunnskolen og videregående skole. Disse har dermed stor betydning for elevenes inntak til videre utdanningsforløp, samt framtidig arbeidsliv.

Sluttvurdering i grunnskolen

Elever i ungdomsskolen får sluttvurdering i form av standpunktkarakterer (primært i 10. klasse, selv om noen avgangskarakterer settes i 8. eller 9. klasse) ved avslutningen av opplæringen i faget og eksamenskarakterer (i 10. klasse) (Udir, 2021). Eksamen gjennomføres i 10. klasse i grunnskolen, og alle elevene trekkes ut til eksamen i to fag. Dette er én skriftlig sentralt gitt eksamen (enten i matematikk, norsk eller engelsk) og én lokalt gitt muntlig eller muntlig-praktisk eksamen (Udir, 2022). Utdanningsdirektoratet har ansvar for organisering, gjennomføring og oppgaveproduksjon for sentralt gitte skriftlige eksamener, og oppgavene er felles for hele landet. Det er skoleeier som har ansvar for organisering og oppgaveproduksjon av lokalt gitte skriftlige og muntlige eksamen. Det betyr at oppgavene kan være ulike avhengig av hvor elevene er bosatt.

Både standpunktkarakterer og eksamenskarakterer føres på vitnemålet for grunnskolen. Ettersom alle har gjennomført grunnskolen har rett til videregående opplæring i Norge, vil alle få utdelt vitnemål uavhengig av om de har bestått fagene eller ikke. Med andre ord er det mulig å både ha karakteren 1 (som tilsvarer stryk) eller IV for 'ikke vurdert' på vitnemålet. Disse karakterene teller med som 1 respektive 0 i karaktergjennomsnittet. Karaktergjennomsnittet vil kunne ha betydning for hvilket videregående utdanningsprogram (og på noen steder hvilken skole) eleven kan komme inn på.

Sluttvurdering i videregående skole

Sluttvurdering i videregående opplæring i Norge består også av standpunktkarakterer og eksamenskarakterer. Det er faglærer som setter standpunktkarakter, og elever får standpunktkarakter i alle fag ved avslutningen av hvert fag i videregående opplæring. Elever trekkes også ut til eksamen i et utvalg av fagene ved avslutningen av hvert fag i videregående opplæring. Antall eksamener en elev trekkes ut til, vil variere ut fra hvilket utdanningsprogram eleven tar (Udir, 2021).

I studieforbereende utdanningsprogram⁵, skal om lag 20 prosent av elevene i det første året (vg1) trekkes ut til eksamen i ett fag. Dette kan være skriftlig, praktisk, muntlig eller muntlig-praktisk eksamen. I det andre året (vg2) skal alle elever trekkes ut til eksamen i ett fag. Dette kan også være skriftlig, praktisk, muntlig eller muntlig-praktisk eksamen. I det siste året (vg3) i studieforbereende utdanningsprogram, skal alle elever opp til obligatorisk skriftlig eksamen i sitt hovedmål (norsk hovedmål eller samisk førstespråk). I tillegg blir alle vg3-elever trukket ut til tre andre eksamener. Hvilke fag dette kan være, samt eksamensformen disse vil ha, varierer ut ifra hvilket utdanningsprogram eller programområde eleven tar⁶. Elever på studiespesialiserende program trekkes ut til to skriftlige eksamener (fordypningsfag eller sidemål), og én muntlig (eller praktisk, muntlig-praktisk) eksamen. I resten av de studieforbereende utdanningsprogrammene skal minst én av de tre trekk-eksamenene være innenfor studieprogrammets programområde (Udir, 2022). Dermed vil elever som har fullført studieforbereende utdanningsprogram vanligvis ha fem eksamenskarakterer på sitt vitnemål. Hvis eleven ble trukket ut til eksamen i vg1, vil han eller hun ha seks eksamenskarakterer på vitnemålet.

I yrkesfaglige utdanningsprogram vil om lag 20 prosent av elevene på vg1 og vg2 trekkes ut til eksamen i ett fag. Andelen på 20 prosent skal sees over en toårsperiode. Alle vg2-elever skal i tillegg ha en obligatorisk tverrfaglig eksamen i

⁵ Studieforbereende utdanningsprogram har fem ulike retninger: studiespesialiserende utdanningsprogram; idrettsfag; musikk, dans og drama; kunst, design og arkitektur; medier og kommunikasjon.

⁶ Elever som har forsert et fag (tatt fag på høyere nivå) skal være med i trekk til eksamen i faget det året opplæringen i faget avsluttes. Dette gjelder både for elever som forserer fag mens de går i grunnskolen og i videregående opplæring.

programfag. Elever kommer også opp i flere spesifikke eksamener i vg3 (og eventuelt i vg4). Hvilke eksamener en elev kan bli trukket ut i ved disse årene, vil være avhengig av hvilket yrkesfaglige utdanningsprogram eleven tar⁷.

Karakterskalaen

Norske elever får vurdering uten karakterer i barneskolen (1.-7. klasse), og vurderinger med og uten tallkarakterer i ungdomsskolen (8.-10. klasse) og videregående opplæring (vg1-vg3/vg4). I 8. og 9. klasse får elever kun standpunktkarakterer i fag, og i 10. klasse og videregående skole får de både standpunktkarakterer og eksamenskarakterer.

Det norske karaktersystemet for grunnskolen og videregående opplæring er fastsatt i forskrift til opplæringslova § 3-5. Det benyttes en 6-trinns karakterskala i ungdomsskolen og videregående skole, som spenner fra karakter 1 til 6. Karakteren 6 er best. Kun hele tallkarakterer benyttes, og de enkelte karaktergradene har følgende innhold:

- Karakter 1: eleven har svært lav kompetanse i faget.
- Karakter 2: eleven har lav kompetanse i faget.
- Karakter 3: eleven har nokså god kompetanse i faget.
- Karakter 4: eleven har god kompetanse i faget.
- Karakter 5: eleven har meget god kompetanse i faget.
- Karakter 6: eleven har svært god kompetanse i faget.

Ettersom alle elever har rett til videregående opplæring i Norge, vil alle få utdelt et vitnemål etter fullført ungdomsskole uavhengig av om de har oppnådd karakter i fagene. I videregående opplæring regnes derimot karakter 1 som stryk, og elever må få karakter 2 eller bedre for å bestå et fag. Fag med karakter 1 i standpunktkarakter er bestått i faget, dersom eksamenskarakteren er 2 eller bedre (Forskrift til opplæringslova, § 3-5). Dersom en elev har vært oppe til eksamen i et fag hvor det ikke foreligger standpunktkarakter, for eksempel ved at eleven ikke hadde tilstrekkelig vurderingsgrunnlag, skal eksamenskarakteren bli annullert (§ 3-34).

Nivå av standardisering og sentralisering

Som allerede nevnt er «grunnlaget for vurdering i fag (...) kompetansemåla i læreplanen i faget» (Forskrift til opplæringslova, § 3-3). Kompetansemålene for hvert enkelt fag er fastsatt av Kunnskapsdepartementet i nasjonale forskrifter (Opplæringslova, § 2-3). Disse kompetansemålene gjelder for alle elever i Norge, noe som betyr at grunnlaget for vurdering er standardisert for alle elever på konkrete årstrinn. Samtidig varierer det noe i hvor stor grad den faktiske sluttvurderingen,

⁷ Eksamenene ved de ulike yrkesfaglige programmene blir ikke nærmere beskrevet her.

og innholdet og gjennomføringen av denne, er standardisert, dvs. den samme på tvers av skoler og kommuner/fylkeskommuner.

Når det gjelder standpunktkarakterer er det faglærer i hvert enkelt fag som setter denne. Selv om dette skal gjøres med utgangspunkt i de nasjonale kompetansemålene, vil de kontinuerlige vurderingsformene som standpunktkarakteren bygger på, variere fra skole til skole og fra lærer til lærer.

Når det gjelder eksamen, avhenger nivå av standardisering av hvorvidt eksamen er lokalt eller sentralt gitt. Det er læreplanene i hvert enkelt fag som fastsetter dette. For lokalt gitte eksamener, er det skoleeier, dvs. kommunene for grunnskoleopplæringen og fylkeskommunen for den videregående opplæringen, som har ansvar for å utarbeide eksamensoppgaver og organisere sensur (Forskrift til opplæringslova, § 3-23). Utarbeidelse av eksamensoppgaver, samt retningslinjer for sensur, inkludert sensorveiledning, vil dermed variere fra kommune til kommune og fylkeskommune til fylkeskommune. Selv i lokalt gitte eksamener skal eksamensoppgavene imidlertid ta utgangspunkt i kompetansemålene, og retningslinjene som ligger til grunn i læreplanverket. Dette er med på å sikre en viss konsistens på tvers.

Når det gjelder sentralt gitte eksamener, bestemmes eksamensoppgavene og organiseringen av sensur sentralt av Utdanningsdirektoratet. Dette betyr at innholdet og organiseringen av disse eksamenene, inkludert formulering av eksamensoppgaver og retningslinjer for sensur, vil være standardisert på tvers av skoler, kommuner og fylkeskommuner. For begge typer eksamener er det skoleeier som har ansvar for den faktiske gjennomføringen av eksamen, samt uttrekking av elever (Forskrift til opplæringslova, § 3-23).

Et eksempel på hvordan nivå av sentralisering og standardisering avhenger av hvorvidt eksamen er sentralt eller lokalt gitt, finner vi i norskfaget for 10. trinn: her kan eleven trekkes ut til to dagers skriftlig eksamen, samt en muntlig eksamen. Mens den skriftlige eksamen blir utarbeidet og sensurert sentralt, blir den muntlige eksamen utarbeidet og sensurert lokalt (Udir, 2020). På generelt grunnlag vil skriftlig eksamener ofte være standardisert og bli gitt sentralt, mens muntlige eksamener fastsettes og gjennomføres på lokalt nivå.

Sensur og mulighet for å klage på sluttvurdering

Sluttvurdering som fører til standpunktkarakter gjennomføres, som allerede nevnt, av elevenes faglærer. Dersom det er usikkerhet knyttet til om reglene for fastsetting av standpunktkarakter ble fulgt i et konkret tilfelle, kan rektor kreve at faglærer gjør en ny faglig vurdering før standpunktkarakterene blir offisielt fastsatt (Forskrift til opplæringslova, § 3-15).

Ifølge forskrift til opplæringslova § 3-28 skal alle prøvesvar på eksamen vurderes av to sensorer. Ved sentralt gitte eksamener skal begge disse sensorene være

eksterne, dvs. ikke ansatt på skolen elevene går på. Dersom disse to er uenige, oppnevnes det en oppmann som avgjør den endelige karakteren. Ved lokalt gitte eksamener skal minst en av sensorene være ekstern. Dersom de to sensorene ikke er enige i karakteren, er det den eksterne sensoren som avgjør.

I Norge har elevene rett til å klage på henholdsvis standpunktkarakter, eksamenskarakter, karakterer til fag-/svenneprøver, kompetanseprøve, og realkompetansevurdering. De har også rett til å klage på et eventuelt vedtak om å ikke sette standpunktkarakter (Forskrift til opplæringslova, § 5-1). Klagen skal sendes til skolen der den som klager er elev. Privatister sender sin klage til den skolen som arrangerte eksamen, mens klager som gjelder fag- og svenneprøver, samt kompetanseprøver, sendes til fylkeskommunen som var ansvarlig for prøven.

Ved klage på standpunktkarakter kan klageinstansen kun vurdere om læreren fulgte de gjeldende retningslinjer for vurdering i faget. Med andre ord gjøres det ikke en ny, innholdsmessig vurdering av elevens prestasjon. For grunnskolen er gjeldende klageinstans statsforvalteren i hvert fylke, men klageinstansen i videregående opplæring er det gjeldende fylkestinget (Forskrift til opplæringslova, § 5-12).

Ved klage på karakter til skriftlig eksamen, både i grunnskolen og i den videregående opplæringen, opprettes en klagenemd. Medlemmene av denne nemden kan ikke ha vært en av sensorene i førstegangsvurderingen, og bør – så langt det er praktisk mulig – ikke være ansatt på samme skole som eleven som valgte å klage. I motsetning til klage på standpunktkarakter, vurderer her nemda alle siden ved karakterfastsettingen. Dersom nemda avgjør at karakteren er urimelig i forhold til elevenes eksamensprestasjon, skal de sette en ny karakter (Forskrift til opplæringslova, § 5-9).

Når det gjelder muntlig eksamen, muntlig-praktisk eksamen og praktisk eksamen er det kun mulig å klage på formelle feil som kan ha hatt påvirkning på resultatet. Ved en slik klage henter rektor inn informasjon om vurderingen fra sensor og eksaminator. Disse blir da sendt, sammen med klagen, til statsforvalteren (hvis eleven er i grunnskolen) eller til fylkestinget (dersom eleven er i videregående opplæring). Dersom eleven får medhold i sin klage, blir den opprinnelige karakteren annullert og eleven får mulighet til å gjennomføre den muntlige eksamen på nytt. I slike tilfeller skal det oppnevnes en ny sensor. Merk at dersom klagen gjelder en prøve i et trekkfag, skal eleven trekkes ut til eksamen på nytt. Den nye muntlige eksamen gjennomføres deretter på ordinær måte (Forskrift til opplæringslova, § 5-10).

Dersom klagen gjelder en vurdering av «ikke bestått» i prøvearbeidet knyttet til fag- og svenneprøve, samt praksisbrevprøve og kompetanseprøve i videregående opplæring, skal klagen sendes til den aktuelle fylkeskommunen. Fylkeskommunen innhenter informasjon fra de personene som gjennomført den

opprinnelige vurderingen, og sender dette videre til en oppnevnt klagenemd. Klagenemda kan omgjøre resultatet fra ikke bestått til bestått, de kan avvise klagen, eller de kan annullere den opprinnelige vurderingen og be om at det gjennomføres en ny prøve. Det siste må alltid gjøres dersom klagenemda påviser formelle feil ved vurderingen (Forskrift til opplæringslova, § 5-14).

2.1.2 Opptak til høyere utdanning

Ved å fullføre et av de studieforbereende utdanningsprogrammene oppnår eleven generell studiekompetanse. Elevenes vitnemål fra videregående opplæring danner grunnlaget for opptak til høyere utdanning. Standpunkt karakterene utgjør størsteparten av karakterene på vitnemålet. Likevel har eksamens karakterene fått en større betydning enn standpunkt karakterene i videregående opplæring, ettersom eksamens karakterene overstyrer standpunkt karakterene (Blömeke et al., 2019b). Dersom en elev får 2 i standpunkt (bestått) og 1 på eksamen (ikke bestått), er faget ikke bestått. Eleven må da ta en ny eksamen for å bestå faget. Får derimot en elev 1 i standpunkt (ikke bestått) og 2 eller høyere på eksamen (bestått), er faget bestått. I rapporten fra Eksamensgruppa blir det påpekt at det kan oppfattes som urimelig at eksamens karakteren skal ha den avgjørende betydningen i slike tilfeller, ettersom eksamen er et mer begrenset uttrykk for elevers kompetanse enn standpunkt karakteren (Blömeke et al., 2019a).

Hovedregelen er at en sluttvurdering skal følge en elev frem til vitnemål eller kompetansebevis. Dette betyr at en elev som har fått sluttvurdering og bestått et fag (fått karakteren 2 eller bedre), ikke kan bli elev i faget på nytt. Dersom eleven ønsker å forbedre karakteren i faget, kreves en privatisteksamen. Tidligere standpunkt- og eksamens karakter vil da slettes, og det vil kun være karakteren fra privatisteksamen som føres på vitnemålet (Udir, 2021).

Opptak til studier ved universiteter og høyskoler i Norge krever generell studiekompetanse⁸. Studiekompetanse gir i seg selv ingen garanti for at søkeren kommer inn, ettersom eleven også må oppfylle eventuelle karakterkrav som stilles for opptak. For opptak til høyere utdanning blir elevens standpunkt- og eksamens karakterer lagt til grunn for poengberegning i form av skole- eller konkurransepoeng. Her er det også avgjørende hvilken kvote søkeren kan søke opptak i. Studieplassene i høyere utdanning er fordelt på to kvoter: 50 prosent av alle studieplasser gis i kvote for førstegangsvitnemål og 50 prosent gis i ordinær kvote. I kvoten for førstegangsvitnemål konkurrerer kun elever som er 21 år eller yngre, mens i ordinær kvote konkurrerer alle søkere. Med andre ord konkurrerer de under 21

⁸ Man kan søke opptak til enkelte universitets- og høyskolestudier uten generell studiekompetanse – et forløp kalt Y-veien (yrkesfaglig vei). Y-veistudier er vanligvis innen ingeniøruddanning, og forutsetter at elevene har relevant fagbrev, svennebrev eller yrkeskompetanse fra videregående skole.

år og som ikke har forbedret karakterer om studieplassene i begge kvoter, mens søkere som er 22 år eller eldre kun konkurrerer om plasser i ordinær kvote. I førstegangsvitnemålskvoten konkurrerer søkerne kun på bakgrunn av sine skolepoeng. Skolepoeng regnes ut gjennom å ta samlet karaktergjennomsnitt x 10, samt at man får medregne tilleggspoeng for realfag eller språk (maksimalt 4 poeng), og eventuelle kjønns-poeng. I den ordinære kvoten konkurrerer søkerne på bakgrunn av konkurransepoeng, der man også medregner eventuelle andre tilleggspoeng (for enten folkehøgskole, militærtjeneste, siviltjeneste, fagskole eller høyere utdanning), alderspoeng, nye fag eller fag søkeren har forbedret karakter i etter videregående skole (Hovdhaugen et al., 2020).

Det er altså søkerens skole- og konkurransepoeng, samt fagkombinasjonen fra videregående opplæring, som avgjør om de får plass på studiet eller ikke. Dette betyr at små marginer på vitnemålet kan få store utslag for en søkers studiemuligheter. Det blir derfor viktig at karakterene på vitnemålet gir et valid, pålitelig og rettferdig bilde av elevens kompetanse. I tillegg må poengberegningen for opptak til høyere utdanning være rettferdig og pålitelig utformet (Blömeke et al., 2019b).

2.2 Danmark

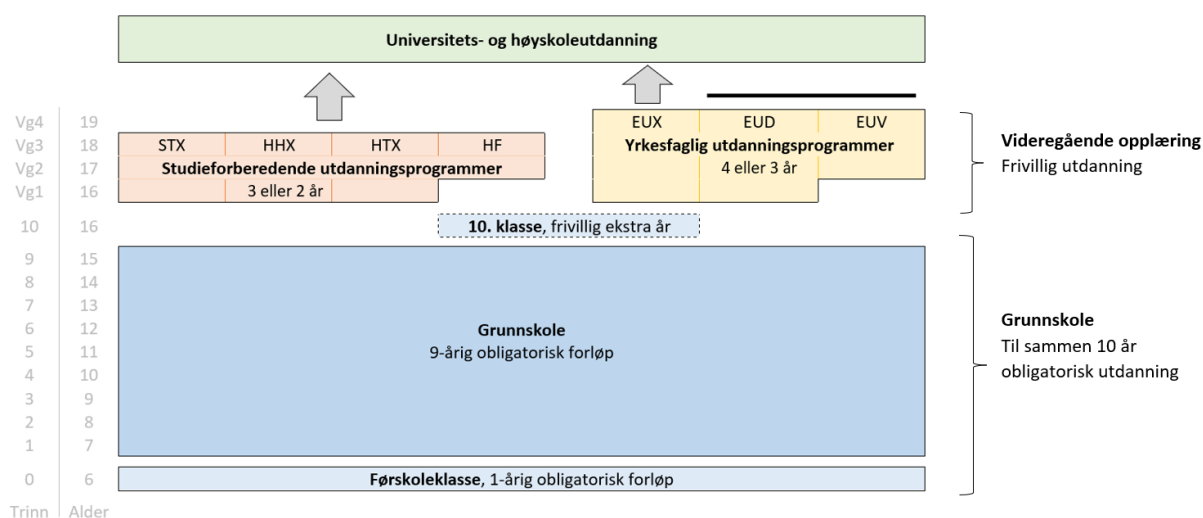
Formålet til det danske utdanningssystemet er å sørge for at alle barn og unge tilegner seg kunnskap og kompetanse som vil kvalifisere dem til å ta en aktiv rolle i det danske kunnskaps-samfunnet og bidra til dets videre utvikling (Folkeskoleloven, 2021, § 1; Lov om de gymnasiale uddannelser, 2022, § 1). Dansk grunnlov gir alle barn fra 6 til 16 år rett til gratis utdanning i offentlige skoler. Den danske grunnskolen (*Folkeskolen*) omfatter en ettårig førskoleklasse (*0. klasse* eller *børnehaveklasse*), som siden 2009 har vært obligatorisk for alle barn i seksårsalderen. Deretter følger en niårig grunnskole (1.-9. klasse). I tillegg er det mulig å forlenge det obligatoriske utdanningsforløpet med en frivillig 10. klasse. Den frivillige 10. klasse er primært for elever som trenger å forbedre sitt kunnskapsgrunnlag for å bestå grunnskolen.

Etter gjennomført obligatorisk grunnskoleutdanning, kan elever fortsette med videregående opplæring (*ungdomsuddannelse*). Den videregående opplæringen deles i to hovedretninger: studieforberevende utdanningsprogram (*gymnasiale uddannelser*) og yrkesfaglige utdanningsprogram (*erhvervsuddannelser*).

Studieforberevende utdanningsprogram er delt inn i fire retninger: allmenn studenteksamen (STX), høyere handels/merkantileksamen (HHX), høyere teknisk eksamen (HTX) og høyere forberedelseseksamen (HF). De tre førstnevnte programmene er treårige utdanningsløp, mens høyere forberedelseseksamen er et to-årig utdanningsløp. Alle retningene er allmenndannende og gir studiekompetanse

for opptak til høyere utdanning (Børne- og undervisningsministeriet, 2022b; European Commission, 2022a).

Yrkesfaglige utdanningsprogram har gjerne en varighet på 3-4 år, og elevene skifter mellom å gå på skole og være lærling i en bedrift. Det yrkesfaglige utdanningsprogrammet har tre spor: EUD (for unge), EUV (for voksne) og EUX (en kombinasjon med gymnasiale eksamen). En avsluttet yrkesfaglig utdanning gir tilgang til arbeidsmarkedet som faglært, mens et EUX-løp også gir generell studiekompetanse og mulighet til å søke inn på høyere utdanning (Børne- og Undervisningsministeriet, 2022d).



Figur 2.2 Utdanningssystemet i Danmark.

2.2.1 Vurdering

Formålet med vurdering i det danske utdanningssystemet er å «dokumentere, i hvilken grad eksaminanden oppfyller de mål og krav, der er fastsat for faget og utdannelsen» (Almen eksamensbekendtgørelsen, 2016, § 1)⁹. Målene er nedfelt i en rekke lovtekster, og gir en beskrivelse av felles mål, herunder henholdsvis ferdighets- og kunnskapsmål, for bestemte klassetrinn, samt hvilke typer kompetanse elevene skal ha tilegnet seg i hvert fag. Det er Børne- og undervisningsministeren som har ansvar og myndighet til å fastsette disse (Folkeskoleloven, 2021, § 10).

Forholdet mellom sluttvurdering og læreplan

Det danske utdanningssystemet har vært kriteriestyrt siden 2006 (Lundahl et al., 2016). Imidlertid ble det i 2017 vedtatt en lovendring som reduserte antall

⁹ En tilsvarende lov – Prøvebekendtgørelsen (2021) – gjelder for grunnskolen, men inneholder ingen eksplisitt beskrivelse av formålet med vurderinger.

bindende mål i læreplanverket (Børne- og Undervisningsministeriet, 2022c)¹⁰. De fleste målene er nå veiledende, heller enn obligatoriske, noe som betyr at den enkelte lærer og skole har større frihet når det kommer til å bestemme innhold og undervisningsform. Hvert enkelt fag må imidlertid planlegges og gjennomføres i lys av den overordnede formålsparagrafen for den danske grunnskolen (folkeskolen), samt fagets egen formålsbeskrivelse. All vurdering må også ta utgangspunkt i kompetansemålene som er satt for hvert fag. Dette resulterer, som beskrevet av Lundahl et al., (2016, s. 34), i en situasjon hvor lærere og sensorer er pålagt å gjennomføre en vurdering basert på målene fastsatt i læreplanverket, samtidig som lærerne ikke er pålagt å legge opp undervisningen ut fra disse målene.

I de konkrete læreplanene legges det vekt på at oppgaver og prøveformer som brukes i vurderingen skal, i «al væsentlighed tilsammen dække de faglige mål, kernetoffet og det supplerende stof» som ligger til grunn for faget (Børne- og Undervisningsministeriet, 2017). Dette er en måte å sikre vurderingens validitet, ved at elevene får mulighet til å vise sine ferdigheter innenfor hele fagets bredde.

Sluttvurdering

Sluttvurdering i grunnskolen

I dansk grunnskole består sluttvurderingen av standpunkt karakterer og eksamens karakterer. Standpunkt karakterer settes av faglærer, og gis ut fra lærerens vurdering av i hvor høy grad eleven oppfyller fagets kompetansemål innenfor ferdighets- og kunnskapsområder i faget. I utgangspunktet inngår ikke elevenes arbeidsinnsats eller oppførsel i grunnlaget for karaktersettingen. Ifølge folkeskoleloven (2021, § 13, stk. 6) skal standpunkt karakter settes to ganger om året i 8., 9. og 10. klasse, og det gis standpunkt karakterer i alle fag der det avholdes eksamen. De siste standpunkt karakterene gis umiddelbart før de skriftlige avgangseksamenene i fagene og skal uttrykke elevenes faglige standpunkt på dette tidspunktet. Det er kun disse som overføres til elevens vitnemål for grunnskolen. Alle elever får ti standpunkt karakterer, fordelt på henholdsvis dansk (fire karakterer), matematikk (tre karakterer), språk (to karakterer) og andre fag (en karakter).

Prøver og eksamener er en sentral del av sluttvurderingen i grunnskolen. Disse skal sikre enhet og sammenheng mellom utdanningens formål, læreplan og vurderingen (Børne- og undervisningsministeriet, 2022b). I grunnskolen gjennomføres tellende slutteksamen, kalt *Folkeskolens avgangseksamen*. Avgangseksamen består av én valgfagsprøve i 8. klasse (i enten håndverk og design, bildekunst, musikk eller matkunnskap), samt syv obligatoriske eksamener i 9. klasse. I 9. klasse er fem

¹⁰ Unntatt fra dette er imidlertid de obligatoriske fagene og valgfagene på 10. trinn, samt det obligatoriske emnet «utdanning og jobb». I disse fagene er lærerne pålagt å bruke eksisterende læreplaner og undervisningsveiledninger.

eksamener obligatoriske og to er uttrekkseksamener. De obligatoriske eksamenene består av dansk (muntlig og skriftlig), matematikk (skriftlig) og engelsk (muntlig), samt en fellesprøve i fysikk/kjemi, biologi og geografi (muntlig). Det danske Børne- og Undervisningsministeriet trekker de to uttrekksprøvene. Disse består av ett fag fra den humanistiske fagbolken (engelsk, tysk/fransk, historie, samfunnskap og kristendoms-kunnskap) og ett fag fra den naturfaglige fagbolken (biologi, geografi, matematikk, idrett, fysikk/kjemi). De obligatoriske prøvene i 9. klasse er alltid like, mens uttrekksprøvene varierer fra år til år. Mens oppgavene til de skriftlige avgangseksamener er gitt fra nasjonalt hold, er det den enkelte skoles leder som har ansvaret for oppgavene til de øvrige prøvene (Prøvebekendtgørelsen, 2021, § 2). For alle eksamener er det pålagt å bruke eksterne sensorer (§ 49).

Den danske grunnskolens avgangseksamen er laget for å teste elevenes faglige kunnskapsnivå, og sluttkarakterene brukes for å vurdere om den enkelte elev er egnet til videre opptak til videregående opplæring (*ungdomsuddannelserne*). For å bestå grunnskolens avgangseksamen, må eleven ha et karaktergjennomsnitt på minimum 2,0 på de åtte obligatoriske prøvene. Vitnemål skal utstedes for elever i 9. klasse som har bestått grunnskolens avgangseksamen, eller elever i 10. klasse (frivillig år) som har avlagt eksamen i én eller flere av de fastsatte prøvene på 9. klassenivå. Vitnemålet inneholder blant annet elevenes seneste standpunktkarakterer, karakterer fra grunnskolens obligatoriske 8.-klasseeksamen og karakterer fra grunnskolens 9.-klasseeksamen.

Opptak til studieforberedende utdanningsprogram (STX, HHX og HTX) krever et karaktergjennomsnitt på minst 5,0 i de avsluttende standpunktkarakterene i 9. klasse. I tillegg må eleven oppfylle en rekke sosiale, personlige og praksisfaglige forutsetninger. Hvis en elev ikke oppfyller disse kravene og dermed blir vurdert som «ikke-uddannelsesparat», kan de likevel bli tatt opp til de studieforberedende utdanningsprogrammene ved å avlegge en opptaksprøve og gjennomfører en opptakssamtale (Børne- og undervisningsministeriet, 2022c). Hvis elever ønsker opptak til yrkesfaglige studieprogram (*erhvervsuddannelse*), er det elevens gjennomsnitt i henholdsvis dansk (minst 2,0 i skriftlig og muntlig) og matematikk (minst 2,0 skriftlig) som er viktig (Børne- og undervisningsministeriet, 2022c). Dersom elever ikke består grunnskolens avgangseksamen i 9. klasse, kan de prøve å forbedre resultatet i det frivillige 10. klasseforløpet.

Sluttvurdering i studieforberedende utdanningsprogram

Studieforberedende utdanningsprogram (gymnasiale uddannelser, STX, HHX og HTX) i dansk videregående opplæring inkluderer avsluttende standpunktkarakterer (kalt '*årskarakterer*') og eksamens-karakterer. Vurderingsgrunnlaget og vurderingssituasjonen er forskjellig for disse to (Danmarks evalueringsinstitutt, 2016).

Skolen skal gi sine elever standpunktkarakter minst en gang i løpet av skoleåret som en del av elevenes underveisevaluering (Bekendtgørelse om de gymnasiale uddannelsene, 2017, § 22). Standpunktkarakter gis ved avslutningen av de enkelte fagene. Det er disse karakterene som føres på elevenes endelige vitnemål. Standpunktkarakterer fastsettes av elevenes lærer, og karakteren skal være et uttrykk for elevens aktuelle faglige nivå i forhold til de krav som stilles i faget. Standpunktkarakterer gis i alle fag ved skoleårets slutt i de tre årene for studieforberedende utdanningsprogram (1.g, 2.g og 3.g.).

Eksamenskarakterer (også kalt '*prøvekarakterer*') settes på bakgrunn av elevenes eksamener. Det avholdes muntlige og skriftlige eksamener ved semesterslutt i alle de tre årene. I løpet av det treårige forløpet skal elever i de studieforberedende utdanningsprogrammene STX, HHX og HTX opp i minst ti skriftlige eller muntlige eksamener. Dette inkluderer én skriftlig prøve i dansk og et studieprosjekt/studieområdeprosjekt. Elever i HTX går også opp i én eksamen i teknikkfag. De resterende eksamenene er uttrekkseksamener. Dette betyr at elever i STX og HHX blir trukket opp i åtte uttrekkseksamen, og elever i HTX blir trukket opp til syv uttrekkseksamener. Hvis elever tar ekstra A-nivåfag, vil de trekkes ut til et tilsvarende ekstra antall eksamener (Børne- og undervisningsministeriet, 2021b).

Et sentralt prinsipp ved trekkordningen i Danmark, er at den enkelte elev ikke på forhånd kan resonere seg frem til hvilke prøver han eller hun ikke kan trekkes ut til – verken ved den aktuelle termin eller kommende terminer (Børne- og undervisningsministeriet, 2021b). Det er det danske undervisningsministeriet som bestemmer hvilke fag den enkelte elev skal avlegge eksamen i, men det er skolen som gjør elevene kjent med hvilke fag de kommer opp i. Dette skjer hvert år ved offentliggjørelse av eksamensplanen i mai.

Fullført studieforberedende utdanningsprogram (*gymnasiale uddannelser*) gir generell studiekompetanse. I tillegg utstedes et vitnemål. Både standpunktkarakterer (*årskarakterer*) og eksamenskarakterer (*prøvekarakterer*) blir oppført på elevens vitnemål, og begge karakterene teller i det samlede karaktergjennomsnittet for faget. Hvis en elev ikke tar eksamen i et fag, er det kun elevens standpunktkarakter som vises på vitnemålet. Har en elev hatt et fag på flere nivå, er det kun karakteren for det høyest gjennomførte nivået som blir oppført på vitnemålet. De ulike fagene blir også vektet ulikt, avhengig av hvilket nivå man har tatt faget på.

Det regnes også ut et samlet poengsnitt fra videregående skole, og dette er et gjennomsnitt av to karaktersnitt: Det første er et vektet karaktergjennomsnitt for elevens eksamenskarakterer fra de obligatoriske eksamenene. Det andre karaktersnittet er et vektet gjennomsnitt av standpunktkarakterene, som er fastsatt av lærerens vurdering av elevens faglige kompetanse (Hvidman og Sievertsen, 2021). Poengsnittet fra vitnemålet danner grunnlag for opptak til høyere utdanning.

Karakterskalaen

I 2006/07 gikk Danmark bort fra en 13-trinns karakterskala, og innførte i stedet en syvtrinnskala basert på European Credit Transfer and Accumulation System (ECTS-skalaen). Den nye karakterskalaen ble blant annet innført på bakgrunn av et ønske om å forenkle sammenligningsgrunnlaget mellom danske og utenlandske karakterer (Uddannelses- og Forskningsministeriet, 2021).

Karakterskalaen har syv mulige karakterer, der fem av disse er ståkarakterer: 12, 10, 7, 4 og 02 (tilsvarende henholdsvis A, B, C, D og E). I tillegg er det to stryk-karakterer: 00 og -3 (tilsvarende Fx og F). Sammenhengen mellom den enkelte karakter og de faglige målene står sentralt i beskrivelsene av alle de syv karakterene. Syvtrinnskalaen brukes både i den danske grunnskolen og i videregående opplæring. I tillegg finnes forskjellige «verbale skalaer», som benyttes ved svenneprøver innenfor yrkesfaglige studieprogram.

Nivå av sentralisering og standardisering

Fastsettelse av standpunktkarakter i både grunnskolen og den videregående opplæringen i Danmark gjøres av elevens faglærer i hvert enkelt fag, med utgangspunkt i kompetansemålene for faget. Vurderingen gjøres dermed lokalt, basert på sentralt gitte og standardiserte mål. Når det gjelder prøver i dansk grunnskole (folkeskolen), skal oppgavene til de skriftlige prøvene utarbeides av Styrelsen for Undervisning og Kvalitet. Disse er dermed sentralt gitt og standardisert for alle elever. De ulike skolens ledere har ansvar for å utarbeide oppgavene til øvrige prøver, inkludert muntlige og praktiske prøver (Prøvebekendtgørelsen, 2021, § 2).

For studieforberedende utdanningsprogram bestemmes standpunktkarakterene av elevenes faglærer, med utgangspunkt i læreplanens kompetansemål. Det er det danske undervisningsministeriet som bestemmer hvilke fag den enkelte elev skal avlegge eksamen i.

Sensur og mulighet til å klage på sluttvurdering

For skriftlige eksamener med sentralt gitte oppgaver rettes eksamenen av to eksterne sensorer, men for de resterende eksamenene vurderes prestasjonen av en intern eksaminator, gjerne faglærer, og én ekstern sensor (Almen eksamenbekendtgørelsen, 2016, § 31, stk. 2). Det er et krav om at sensorer skal «medvirke til og påse, at eksaminanderne får en ensartet og retfærdig behandling, og at deres præstationer får en pålidelig bedømmelse, der er i overensstemmelse med reglene om karaktergivning og øvrige regler for uddannelsen» (§ 31-3).

Når det gjelder klage, har danske elever rett til å klage på vurderinger gitt på prøver og eksamener i henholdsvis grunnskolen (*folkeskolen*) og studieforberedende utdanningsprogram (*gymnasiale uddannelser*) (hhv.

Prøvebekendtgjørelsen, 2021, kap. 10; Almen eksamensbekendtgjørelsen, 2016, kap. 10). Dette gjelder med andre ord både skriftlige, muntlige og praktiske eksamener. Enhver klage må leveres skriftlig, og spesifisere de faktiske omstendighetene ved gjennomføringen eller vurderingen av prøven som eleven mener berettiger en ny vurdering. Klagen kan vedrøre eksaminasjonsgrunnlaget, herunder prøvens oppgave og dens forhold til utdannelsens mål og krav; prøveforløpet; eller den faktiske vurderingen av prøven. Prosedyren for vurdering av en klage er som følger: skolen gjør først en vurdering av om klagesaken skal fremmes, eller avvises som åpenbart grunnløs. Dersom skolen vurderer at klagen kan være berettiget, skal den innhente en faglig uttalelse fra hver av de opprinnelige sensorene. På bakgrunn av klagen, samt disse uttalelsene, gjør skolen en vurdering og kommer med sin avgjørelse.

Det er tre mulige utfall av en klagesak. Skolen kan i) gi eleven tilbud om en ny vurdering av den opprinnelige prøven (ikke mulig for muntlige prøver), ii) gi eleven tilbud om å gjennomføre prøven på nytt, eller iii) avvise klagen. Merk at skolen kun skal gi tilbud om ny vurdering/prøve dersom det foreligger «særlige omstendigheter, der giver anledning til berettiget tvivl om bedømmelsen, herunder væsentlige proceduremæssige mangler» (Almen eksamensbekendtgjørelse, 2016, § 51, stk. 2). Dersom prøven skal vurderes eller gjennomføres på nytt, skal nye sensorer oppnevnes. Lovverket spesifiserer at den nye vurderingen kan resultere i en lavere karakter enn det eleven fikk tildelt i utgangspunktet (§ 53, stk. 5).

Elever på de studieforberedende utdanningsprogrammene (gymnasiale utdannelse) har også mulighet til å klage på sin standpunktkarakter (Bekendtgjørelse om de gymnasiale utdannelse, 2017, § 74). En slik klage sendes til Styrelsen for Undervisning og Kvalitet. Elevene har imidlertid kun mulighet til å klage på såkalt «rettslige spørsmål». Dette innebærer at de ikke kan klage på det faglige skjønnnet som den aktuelle læreren utøvde når den satt karakteren. Styret har med andre ord kun anledning til å ta klager til følge dersom det kan fastslås at det foreligger vesentlige rettslige mangler, for eksempel at det viser seg at læreren ikke la kompetansemålene i læreplanen til grunn for vurderingen eller at det var vesentlige prosedurale feil ved vurderingsprosessen. I slike tilfeller kan Styret pålegge skolen å gjennomføre en ny vurdering.

2.2.2 Opptak til høyere utdanning

I Danmark tilbys en rekke høyere utdanningsforløp, heriblant universitetsutdanninger, profesjonsbachelorutdanninger og yrkesakademiutdanninger. Som hovedregel må man ha gjennomført studieforberedende videregående opplæring for å

kunne bli tatt opp til høyere utdanning. Det er dette som tilsvarer generell studiekompetanse i Norge.

Imidlertid er det forholdsvis mange studier som har spesifikke adgangskrav, hvilket innebærer krav om å ha tatt spesifikke fag (eller fag på et bestemt nivå) i videregående opplæring, eller at man har et minimumskaractergjennomsnitt for å kunne søke inn på det studiet. Til forskjell fra Norge får man ikke ekstrapoeng for disse spesifikke fagene, de er kun en forutsetning for opptak. Danmark har et sentralt administrert søknadssystem til høyere utdanning, hvor elevene sender en søknad med inntil åtte utdanningsprogram i prioritert rekkefølge. Det er ikke anledning til å være tatt opp på mer enn ett høyere utdanningsprogram av gangen (Hovdhaugen et al., 2020).

Derimot har det vært og er ulike former for bonuspoeng i det danske opptakssystemet, som påvirker karaktersnittet. Fra 2009 til 2020 fikk de som søkte opptak til høyere utdanning i løpet av de første to årene etter at de hadde fullført videregående opplæring tilleggspoeng, såkalt «Hurtigstartsbonus», som gjorde at deres karaktersnitt ble multiplisert med 1,08. Hensikten med dette var å få flere unge studenter inn i systemet, siden det var en generell oppfatning at man hadde en stor andel eldre studenter i Danmark (Hovdhaugen et al., 2020). Denne ordningen er i dag fjernet og erstattet med en bonusordning for å ha tatt ekstra avanserte fag i videregående opplæring, såkalt «A-niveau» eller «bonus A». Dette er de mest avanserte kursene i videregående, og søkeren får bare bonus for fag de ikke allerede må ta A-nivå i. De som har tatt et ekstra fag på A-nivå kan multiplisere sitt snitt med 1,03, mens de som har tatt to multipliserer med 1,06 (Børne- og undervisningsministeriet, 2022a).

De fleste høyere utdannelsesstedene i Danmark benytter et kvotesystem, der søkerne kan tas opp gjennom to kvoter. Det er universitetene selv som bestemmer fordelingen av plasser mellom de to kvotene på de ulike studieprogrammene. I kvote 1 blir søkeren kun vurdert på bakgrunn av karaktersnitt fra videregående («adgangsgivende gymnasiale eksamen»). Hvis eleven har et annet adgangsgrunnlag enn en fullverdig gymnaseksamen, som for eksempel fra et yrkesfaglig utdanningsprogram (erhvervsuddannelse) eller HF-enkeltfag, må de søke i kvote 2. I kvote 2 blir elevene også vurdert på andre faglige kriterier enn karaktersnitt, som for eksempel opptaksprøve, intervju eller test.

Det er lærestedene selv som bestemmer hvilke kriterier elever som søker i kvote 2, skal vurderes etter. Dette betyr at det kan være store forskjeller på kvote 2 fra utdannelsessted til utdannelsessted. Med andre ord betyr det at kriteriene for opptak i kvote 2, og hvor lett/vanskelig det er å komme inn via kvote 2 varierer fra lærested til lærested. Tidligere opptak har vist at ved de aller fleste universiteter er majoriteten av plassene fordelt via kvote 1, men at dette også i noen grad avhenger av fag. Studiesteder som strever med rekruttering av studenter, fordeler

ofte flere plasser via kvote 2. Dersom et studium har opptaksprøve eller opptaks-samtale, tas alle søkere opp i kvote 2. De to kvotene er gjensidig utelukkende så søkerne kan ikke bli vurdert dobbelt (Hovdhaugen et al., 2020, Børne- og Under-visningsministeriet, 2022b).

2.3 Finland

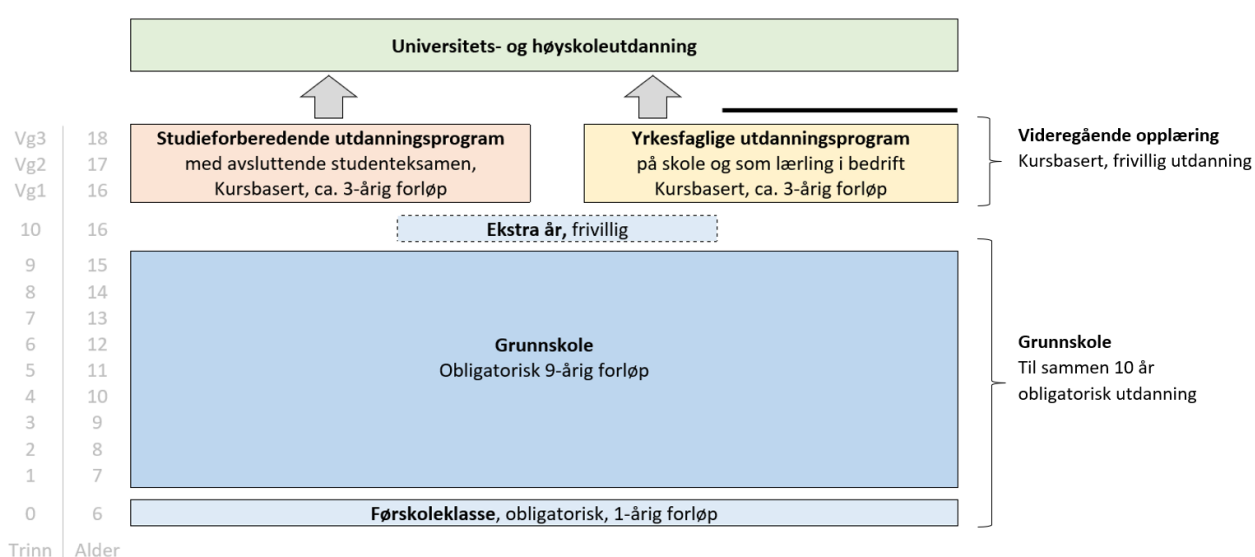
Formålet med det finske utdanningssystemet er å «at stödja de studerandes utveckling till goda, harmoniska och bildade människor och aktiva samhällsmedlemmar samt att ge dem sådana kunskaper, färdigheter och beredskaper som de behöver i arbetslivet, för fritidsintressen och för en allsidig personlighetsutveckling.» (Gymnasielag, 2018, § 2)¹¹. Formålsbeskrivelsen, samt det overordnede rammeverket for det finske utdanningssystemet, er spesifisert i henholdsvis Lov om grunnutdanning (*'Lag om grundläggande utbildning'*) og Lov om videregående utdanning (*'Gymnasielag'*), samt tilhørende forskrifter og lære-planverk.

I Finland består grunnskolen av et førskoleår og niårig obligatorisk grunnskole. Det er ingen eksamener i grunnskolen, og alle karakterer gis av læreren. For å kunne fortsette til neste utdanningsnivå, videregående opplæring, må eleven ha bestått alle kjernefagene i grunnskolen: morsmål, annet nasjonalt språk, fremmed-språk, matematikk, fysikk, kjemi, biologi, geografi, helseutdanning, religion/etikk, historie, samfunnsfag, musikk, kunst, håndverk, gym og hjemkunnskap. Eleven får da et sertifikat fra grunnskolen. Dette sertifikatet gir adgang til neste utdannings-nivå, videregående opplæring. For elever som ikke består grunnskolen finnes det tilbud om et 10. år (European Commission, 2022b; Lundahl et al., 2016).

Videregående er frivillig, men nesten alle elever søker plass og begynner. Fin-land skiller mellom studieforberevende program, som undervises på gymnasier, og yrkesfaglig utdanning, som finnes på yrkesskoler. Begge skoletyper er i ut-gangspunktet treårige løp (Kansanen og Meri, 2015). Elever som har tatt et 10. skoleår søker oftest yrkesfaglig utdanning. Søkning til gymnasiet foregår basert på karakterer fra grunnskolen. Det kan med andre ord være konkurranse for å komme inn på et program. Mer enn halvparten av alle elever begynner i studiefor-beredende videregående opplæring og sikter på å fullføre med finsk studenteksa-men (*matrikulasjonseksamen*) (Kansanen og Meri, 2015). Fullføring av et av de stu-dieforberevende programmene gir tilgang til høyere utdanning. Dette er gitt at eleven også har fullført studenteksamen, hvilket er den eneste eksamenen som ek-sisterer i finsk skole.

¹¹ Dette er formuleringen gitt for den videregående opplæring (*'allmänbildande gymnasieutbildning'*). En tilsvarende beskrivelse finnes for grunnopplæringen (*Lag om grundläggande utbildning*, 1998, § 2).

Grunnskolen er organisert i trinn, der elever kan bli holdt igjen på et trinn dersom de ikke består fagene. Dette gjøres i så fall etter individuell vurdering. Videregående opplæring er organisert etter kurs, der elevene for å fullføre må ta totalt 75 kurs. Av disse er 47-51 obligatoriske, antall er avhengig av om eleven velger grunnleggende eller avansert matematikk. I tillegg tar elever minst to språk i tillegg til morsmål (European Commission, 2022b). Men siden videregående er kursbasert er det mulig for elever å ta studieforbereende videregående på kortere eller lengre tid enn tre år (Kansanen og Meri, 2015).



Figur 2.3 Utdanningssystemet i Finland

2.3.1 Vurdering

Formålet ved vurdering i det finske utdanningssystemet er spesifisert i loven om videregående utdanning. Vurdering skal «leda och sporra studerandena i deras studier och att utveckla deras förutsättningar för självbedömning» (Gymnasielag, 2018, § 37)¹². Dette gjøres på to ulike måter. Først og fremst skal vurderingen støtte og veilede læring. Dette gjøres gjennom tilbakemeldinger gitt i løpet av studiedelen, såkalt formativ vurdering. Tilbakemeldingene skal beskrive elevens fremgang i forhold til målene som er satt. For det andre skal vurderingen vise hvor godt eleven har oppnådd de mål som er satt for det faget eleven blir vurdert i. Dette refereres i læreplanverket til som summativ vurdering. Vurderingen av studentens kunnskaper skal baseres på prestasjoner som studenten har vist og hvor det

¹² En tilsvarende formulering finnes for grunnskoleopplæringen, og er spesifisert i Grunderna för läroplanen för den grundläggande utbildningen (Utbildningsstyrelsen, 2014, s. 48).

fremgår hvor godt studenten har nådd målene spesifisert for faget (Skolstyrelsen, 2014, s. 50; Utbildningsstyrelsen, 2019, s. 46).

Formålet med det nasjonale styringssystemet for den grunnleggende utdanningen er å «garantera jämlikhet och kvalitet i utbildningen samt skapa goda förutsättningar för elevernas växande, utveckling och lärande» (Utbildningsstyrelsen, 2014). Formålet er med andre ord å støtte og veilede undervisning og skolearbeid og fremme en enhetlig grunnopplæring på like vilkår.

Forhold mellom sluttvurdering og læreplan

Som de andre nordiske landene, opererer det finske utdanningssystemet med kriteriestyrt vurdering. Dette skjer gjennom en kombinasjon av en overordnet nasjonal læreplan og lokale, tilpassede læreplaner. Førstnevnte fastsettes av den finske *Utbildningsstyrelsen*. Den inneholder mål og grunnleggende innhold i ulike fag, samt prinsippene for elevvurdering, spesialundervisning, elevvelferd og pedagogisk veiledning. Sistnevnte fastsettes av de ulike skolene selv.

De lokale læreplanene regnes som en viktig del av styrings- og kontrollsystemet for den finske skolen, da den både beskriver og kontrollerer gjennomføringen av de nasjonale målene, men også gir et felles grunnlag og retningslinjer for det daglige skolearbeidet. Vurderingen av læring, arbeid og atferd, samt responsen til elevene, bør alltid ta utgangspunkt i målene som er satt i det grunnleggende delen i læreplanen og spesifisert i den lokale læreplanen. Elever og deres prestasjoner skal ikke sammenlignes med hverandre og elevens personlige egenskaper skal ikke påvirke vurderingen (Utbildningsstyrelsen, 2014, s. 49). Ifølge lovverket skal elevenes arbeid, læring og kompetanseutvikling vurderes på mange og varierte måter. Dette må gjøres for å sikre at enhver elev får en reell mulighet til å vise frem sin kompetanse, og for å forhindre at vurderingen legges opp på en måte som favoriserer visse typer måter å lære på (Gymnasielag, 2018, § 37; Utbildningsstyrelsen, 2014, s. 48).

Sluttvurdering

Det er ingen nasjonale prøver eller eksamener for elever i grunnopplæringen i Finland. Sluttvurdering gjøres i stedet av læreren på alle trinn, basert på de målene som er spesifisert for faget i de lokale læreplanene. Dette gjelder også karakterene i grunnskolebeviset, sluttbeviset gitt ved utgangen av 9. trinn og som bestemmer elevenes opptak til videregående utdanning. For å sikre en viss reliabilitet i vurderingen på tvers av klasser og skoler inneholder den nasjonale læreplanen retningslinjer for vurdering i alle grunnskolens fellesfag. Disse kan benyttes av lærerne når de skal vurdere om hvorvidt elevene har nådd de ulike målene satt for faget (Utbildningsstyrelsen, 2018).

Det er kun ved avslutning av studieforbereidende videregående opplæring at elever må ta en eksamen. Eksamen ved avslutning av et studieforbereidende program i videregående opplæring kalles studenteksamen, og er den eneste eksamen som finnes i finsk skole (Kansanen og Meri, 2015). Studenteksamen består av en obligatorisk eksamen i morsmål (finsk/svensk), og tre eksamener som eleven velger i følgende fire områder: matematikk, det andre nasjonale språket (finsk for svensktalende, svensk for finsttalende), fremmedspråk og generelle fag (real-fag/humaniora/samfunnsfag). I tillegg har elever anledning til å melde seg opp i ytterligere eksamener om de ønsker det. Eksamener kan tas både høst og vår, i eksamensperioden (oktober og mars), og de fire eksamener elevene må ta kan spres over maksimalt tre tilfeller (European Commission, 2022b).

I løpet av de siste tiår har det vært gjort endringer i hvordan sammensetningen av eksamener ser ut i studenteksamen. Tidligere var både eksamen i morsmål og nasjonalt andrespråk (svensk for finskspråklige, finsk for svenskspråklige) obligatoriske eksamener. Fra 2005 er det imidlertid kun morsmål som er obligatorisk, mens eksamen i nasjonalt andrespråk er en av de fire områdene elevene må velge tre eksamener fra. Fra 2006 skjedde ytterligere en endring, ved at den tidligere sammenholdte eksamen i generelle fag (real-fag/humaniora /samfunnsfag) ble delt opp i totalt 10 fag: religion (lutheran), religion (ortodox), etikk og moralhistorie, filosofi, psykologi, historie, samfunnsfag, fysikk, kjemi, biologi, geografi og helseutdanning (Kupianinen et al., 2016). Fra og med våren 2022 må elevene ta minst fem eksamener. Hver eksamen varer i seks timer. De ulike eksamener har litt ulikt format: Språkeksamen består av høreforståelse, leseforståelse, grammatikk og et essay. Matematikk-eksamen består av 15 oppgaver, hvorav minst 10 må være fullført for å bli vurdert som bestått. Eksamen i et av de generelle fagene er ofte i form av essay.

Karakterskalaen

Finland har en karakterskala som går fra 4 til 10, der 5 er lavest beståttkarakter. Som nevnt er all vurdering i finsk utdanning målrelatert, der kun kriteriene for hva som kreves for å få karakteren 8 er spesifisert. Ved vurdering skal elever ikke sammenlignes med andre elever, men heller vurderes etter i hvilken grad de har oppnådd målene i pensum. Dette forholder seg til de lokale og nasjonale målene i læreplanverket.

Kursene i videregående tar også nasjonal og lokal læreplan som utgangspunkt for vurdering. Hvert kurs blir målt når det er fullført, og hensikten er å gi eleven tilbakemelding om i hvilken grad de har møtt kursmålene, samt hvilken utvikling de har hatt underveis. Karakteren for et emne, som består av flere kurs, bestemmes av gjennomsnittskarakteren på de individuelle kursene. Elever som ønsker å

forbedre sine karakterer, må få tilbud eller mulighet til å ta en separat test med økning av karakter som formål.

Det er lærerne som har ansvar for å sette karakterer. Det er obligatorisk å gi tallkarakter fra og med 8. klasse, men det for tidligere klassetrinn er valgfritt for skolen om de gir tilbakemelding til elevene verbalt eller gjennom tallkarakterer. Det er vanligst å gi verbal tilbakemelding. Dersom eleven får sin siste karakter i faget i 7. eller 8. klasse føres dette på vitnemålet. Vitnemålet er dermed en oppsamling av alle karakterer i alle fag, og eleven får dette når den har fått en bestått karakter i alle fag (European Commission, 2022b).

Karakterene elevene får på studenteksamen er annerledes enn de vanlige karakterene. Eksamen rettes etter en skala med syv karakterer med latinske navn, som skal være normalfordelt.

- 5% får *improbatur* (I), som er stryk
- 15% får *approbatur* (A), som betyr bestått
- 20% får *lubenter approbatur* (B), som betyr med stor glede bestått
- 20% får *cum laude approbatur* (C), som betyr godkjent med ros
- 20% får *magna cum laude approbatur* (M), som betyr godkjent med mye ros
- 15% får *eximia cum laude approbatur* (E), er godkjent med eksepsjonell ros
- 5% får *laudatur* (L), som betyr berømmet/utmerket

Med andre ord er fordelingen av karakterene normalfordelt for alle enkeltprøver, uavhengig av hvor mange elever som tar akkurat denne eksamen. Kupianinen, Marjanen og Hautamäki (2016) påpeker at dette kan gi utfordringer i sammenlignbarhet mellom ulike eksamener, ved at man måles mot de andre som tar samme enkelteksamen. Normalfordelingen da vil påvirkes av hvem de andre er som tar prøven. Dette forholdet påvirker alle prøvene, utenom morsmål, siden det er en eksamen som alle elever tar.

Nivå av sentralisering og standardisering

Med unntak av den avsluttende studenteksamen, som settes og vurderes sentralt av Studenteksamensnemnda, gjennomføres all sluttvurdering i finsk grunn- og videregående opplæring lokalt på hver enkelt skole. Til tross for at all vurdering skal ta utgangspunkt i de nasjonalt satte læreplanmålene, er det de lokale læreplanene som spesifiserer hvordan vurderingen skal gjennomføres og hvilke retningslinjer som skal ligge til grunn. Det er også elevenes faglærer, i samarbeid med rektor, som har ansvar for å sette standpunkt-karakter. Det finske sluttvurderingssystemet har derfor en liten grad av både sentralisering og standardisering.

Sensur og mulighet til å klage på sluttvurdering

Ifølge lovverket skal «beslut om bedømmingen av varje enkelt studieavsnitt görs av den studerandes lärare eller, om lärarna är många, av lärarna gemensamt. Beslut om slutbedømmingen för ett läroämne görs av rektorn och den studerandes lärare gemensamt (hhv. Förordning om grundläggande utbildning, 1998, § 13; Gymnasielag, 2018, § 38). Med andre ord er det elevenes faglærer, i samarbeid med rektor, som setter karakter i hvert enkelt fag.

I både grunnskolen og den videregående opplæringen kan eleven klage på denne karakteren. Dette gjøres ved at eleven, eventuelt elevens foreldre, ber rektor om en vurdering innen to måneder etter at han eller hun ble informert om vedtaket. Rektor og elevens lærer foretar sammen en ny vurdering. Dersom elev er uenig i den nye vurderingen som er gjort, kan eleven klage til det regionale forvaltningskontoret. Disse kan beslutte å endre eller oppheve vurderingen, avvise forespørselen om ny vurdering, eller sende saken tilbake til rektor for en ny vurdering (hhv. Lag om grundläggande utbildning, 1998, § 47c; Gymnasielag, 2018, § 53).

Når det gjelder den avsluttende studenteksamen i videregående opplæring, samles elevenes besvarelse i en nettbasert tjeneste hvor læreren kan lese og vurdere dem. I etterkant av lærernes vurdering foretar Studenteksamensnemnda den endelige vurderingen av eksamensprestasjonen. Elevene har rett til å klage på nemndas vedtak. En slik forespørsel koster 50 euro, men dersom anmodningen fører frem blir dette beløpet refundert. Klagen skal inneholde informasjon om hvilken del av vurderingen klagen gjelder, og hva eleven mener er feil med den opprinnelige vurderingen som er gitt. I tillegg skal det komme frem av hvordan eleven vurderes at den påståtte feilen bør rettes (Studentexamensnämnden, 2022).

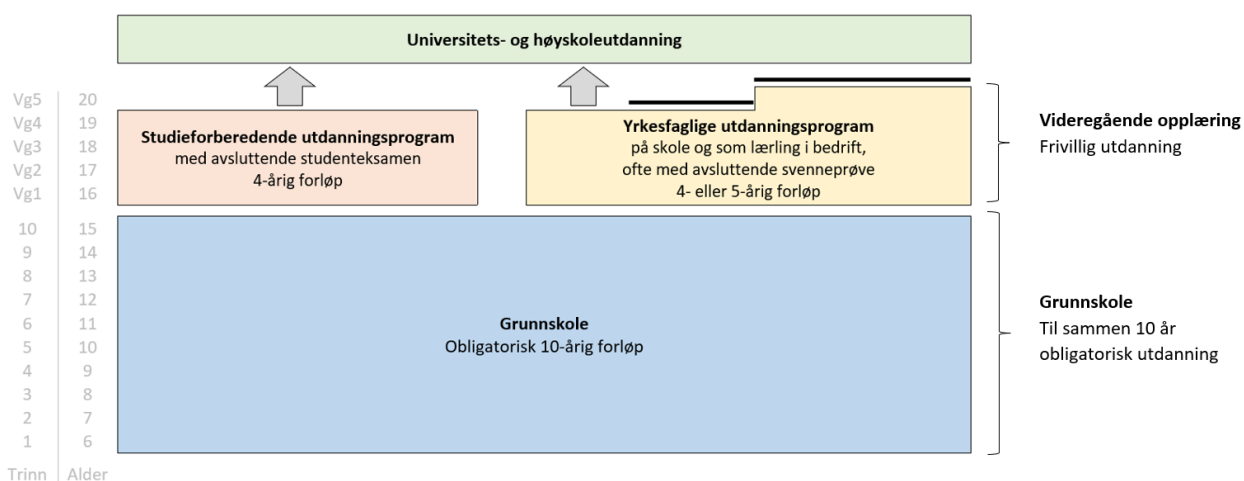
2.3.2 Opptak til høyere utdanning

Formelt sett er det studenteksamen som gir rett til opptak til høyere utdanning i Finland. Men opptak til høyere utdanning, særlig universitetsutdanning, gjøres ikke bare basert på studenteksamen. Lærestedene står i dag mer fritt til å bestemme egne opptakskrav, og de kan derfor velge å bruke andre opptakskrav i tillegg til studenteksamen ved opptak. Eksempler på andre former for krav er karakterersnitt på vitnemålet fra videregående opplæring eller opptaksprøver laget av universitetene. Ved noen læresteder brukes også intervju. Med andre ord er det ikke bare karakterer fra videregående opplæring eller studenteksamen som teller ved opptak til høyere utdanning. Likevel blir studenteksamen sett på som en «high-stakes-test» i Finland, siden dette er det vanligste opptaksgrunnlaget for høyere utdanning.

2.4 Island

Formålet med den islandske grunnsopplæringen er, i samarbeid med hjemmet, å fremme allsidig utvikling for alle elever og deres deltakelse i et demokratisk samfunn i stadig utvikling (Compulsory School Act, 2008, Art. 2). Island har i likhet med Norge en 10-årig grunnskole, der barn begynner det år de fyller seks år og er ferdige det året de blir 16. Alle elever som har gjennomført grunnskolen, uavhengig om de har oppnådd bestått karakter i alle fag eller ikke, har rett til å begynne i videregående skole. Denne rettigheten varer til de er 18 år (European Commission, 2022c).

Videregående opplæring er frivillig, og elevene søker opptak. Det er den enkelte skole som selekterer elever til skolen, basert på krav for opptak som de (skolen) selv bestemmer. Dette gjelder for både studieforberedende og yrkesfaglige program (European Commission, 2022c).



Figur 2.4 Utdanningssystemet på Island

2.4.1 Vurdering

I Grunnskoleloven beskrives formålet med vurdering som følger: «The purpose of the assessment is to evaluate how successfully pupils meet the objectives of the National Curriculum Guide as well as individually determined learning objectives, to encourage pupils to make progress, and to determine who among them may be in need of special support» (Compulsory School Act, 2008)¹³.

¹³ En tilsvarende lov - Upper Secondary Education Act (2008) – finnes for videregående skoler, hvor artikkel 30 omhandler vurdering. Formålet for vurdering i videregående opplæring er imidlertid ikke spesifisert der.

En mer detaljert beskrivelse av hvordan vurdering skal forstås og gjennomføres i praksis finner man i den islandske nasjonale læreplanguiden for henholdsvis grunnskolen og videregående opplæring (Ministry of Education, Science and Culture, 2021a, s. 25; Ministry of Education, Science and Culture, 2021b, s. 25). Formålet med vurdering blir i begge tilfeller beskrevet som følger: «to give guiding information about education and how far its objectives have been achieved». Mer konkret er målet å vurdere hvorvidt de generelle kompetansemålene fastsatt i de nasjonale læreplanene har blitt møtt, hvorvidt elevene har forbedret seg over tid, og eventuelt hvilken støtte elevene trenger for å nå de målene som er satt. På bakgrunn av informasjonen som vurderingen produserer, kan både elever og deres foreldre, samt skolene og skoleeier, bestemme hvilke tiltak som eventuelt bør gjøres for å forbedre måloppnåelse.

I den islandske konteksten spiller dermed vurdering flere roller, både på individuelt og på overordnet nivå. For hver enkelt elev har vurdering både en viktig pedagogisk dimensjon, ved at elevene får kontinuerlig informasjon om hvordan de har prestert så langt og hva de kan gjøre for å forbedre sin prestasjon. På overordnet nivå spiller vurderingsresultatene en rolle i evalueringen og kvalitetssikringen av grunnskolen, ved at skolene får informasjon som de kan bruke i sitt utviklingsarbeid.

Validitet, reliabilitet og rettferdighet blir også trukket frem i de islandske læreplanene. Dersom vurderingen skal oppleves som korrekt og rettferdig, er det viktig at det blir «planned and carried out in such a way that it evaluates what it is supposed to evaluate [validitet] in a reliable manner». Videre skal all vurdering være «reliable, impartial, honest and fair» (Ministry of Education, Science and Culture, 2021a, s. 26; Ministry of Education, Science and Culture, 2021b, s. 26). Særlig viktig er det at læreplanverket i detalj beskriver hvilke krav som stilles til elevene og hvordan disse kravene skal bli vurdert. Dette må gjøres på en måte som skaper en felles forståelse av innholdet og forventningene i disse kravene, slik at både elever, lærere og sensorer har like forutsetninger for å forstå hva som skal til for å nå de spesifiserte målene.

Den nasjonale læreplanguiden legger vekt på at vurdering skal favne så mange ulike lærings- og kunnskapsaspekter som mulig. Derfor er det også nødvendig å bruke en rekke ulike vurderingsformer, inkludert skriftlige, muntlige og praktiske vurderinger, samt vurderinger av ulike lengde og dybde, individuell og gruppevurdering.

Forholdet mellom sluttvurdering og læreplan

Som tidligere nevnt er formålet med vurdering å «evaluate how successfully pupils meet the objectives of the National Curriculum Guide and their individual learning objectives» (Compulsory School Act, 2008, Art. 27). Med andre ord er det de

nasjonale kompetansemålene presisert i den nasjonale læreplanveilederne som skal ligge til grunn for all vurdering.

Samtidig er hver enkelt skole pålagt å utvikle en lokal pensumguide, hvor den nasjonale læreplanen skal utvikles i samsvar med retningslinjene til den aktuelle skolen og skolens faglige vektlegging og særtrekk (Ministry of Education, Science and Culture, 2012, s. 64). I utgangspunktet gir disse hver enkelt skole relativt stor autonomi når det kommer til hvilke kriterier som skal legges til grunn for vurderingen. De lokale læreplanveilederne skal gi en videre utdypning av kompetansemålene i den nasjonale læreplanen, og spesifisere hvilke vurderingsformer og veiledende kriterier som ligger til grunn for vurderingen. Dette er viktig for å sikre at både lærere, elever og foreldre på hver enkelt skole har et felles forståelsesgrunnlag for hvordan vurderingen skal gjennomføres og hvilke kriterier som skal legges til grunn.

Sluttvurdering

Sluttvurdering i grunnskolen

I den islandske grunnskolen avholdes det nasjonale prøver i henholdsvis islandsk og matematikk for 4., 7. og 10. trinn, i tillegg til engelsk i 10. trinn. Formålene med disse prøvene er både å vurdere hvorvidt en elev har nådd de fastsatte kompetansemålene i faget, men også å gi elever, foreldre, skolepersonell og utdanningsmyndigheter landsdekkende informasjon og referanser og elevenes prestasjonsnivå. Utover dette gis alle karakterer som standpunktkarakterer, satt av elevens faglærere.

Sluttvurdering i gymnaset

I den videregående opplæringen gjennomføres det ingen nasjonale eksamener. Standpunktkarakteren skal i stedet bygge på elevenes kontinuerlige prestasjoner i faget, samt deres resultater på terminbaserte prøver. Utover dette er det opp til hver enkelt skole å bestemme hvilke vurderingsformer som skal ligge til grunn for en elevs sluttvurdering i faget. Dette kan innebære lokalt gitte eksamener, men skolen kan også velge å ta utgangspunkt i andre vurderingsformer, for eksempel semesterprøver, mappevurdering og lignende.

Karakterskalaen

Det er kun obligatorisk å sette karakterer i siste klasse av grunnskolen, og karakterer settes utfra gitte kriterier. Karakterskalaen er dermed kriteriebasert, og i grunnskolen finnes det beskrivende kriterier for karaktertrinnene som innebærer at faget er bestått. På Island brukes i grunnskolen en skala fra A til D, der A er best og D er underkjent, mens elever i videregående får karakterer på en skala fra 1 til

10, der 10 er høyest (Lundahl et al., 2016; European Commission, 2022c). Alle karakterene, både i grunnskolen og i videregående opplæring, settes av læreren. I videregående er det prøver i slutten av hvert semester, som gjennomføres av læreren, i samarbeid med hovedlærer eller rektor ved skolen. Karakterene som elevene får i videregående bygger på en sammenstilling av lærerens bedømmelse gjennom semesteret (på ulike oppgaver/prøver) og semesterprøvene. Det finnes ingen nasjonale prøver i videregående skole (Lundahl et al., 2016).

Nivå av standardisering og sentralisering

I det islandske utdanningssystemet er det først og fremst de nasjonale – og dermed sentraliserte – kompetansemålene som er presisert i den nasjonale læreplanveilederen som skal ligge til grunn for all vurdering. Samtidig er hver enkelt skole, som allerede beskrevet, pålagt å utvikle en lokal læreplanguide, hvor den nasjonale læreplanen skal utvikles i samsvar med retningslinjene til den aktuelle skolen og skolens faglige vektlegging og særtrekk. I utgangspunktet står skolene fritt til å bestemme hvilke veiledende kriterier de ønsker at skal være styrende for sine elever, samt hvilken vurderingsform som skal benyttes i disse læreplanene. I praksis varierer det i hvor stor grad skolene benytter seg av denne muligheten, og ofte vil de nasjonale kompetansemålene i stor grad være ledende.

Utover kravet om at vurdering skal ta utgangspunkt i den nasjonale læreplanen og skje gjennom varierte vurderingsmetoder, er det dermed skolene selv som velger når og hvordan vurdering skal skje. Imidlertid er det visse obligatoriske sluttvurderinger som skolene er pålagt å følge. I grunnskolen er det kun i 10. trinn at det er obligatorisk for alle skoler å sette karakter. Ved denne sluttvurderingen skal, som beskrevet over, fire standardiserte evalueringskriterier (A til D) legges til grunn, som spesifiserer i hvilken grad en elev har møtt kompetansemålene i faget. Disse er fastsatt fra nasjonalt hold og er like for alle skoler og elever. Standpunkt-karakter i videregående opplæring er satt av elevenes faglærer.

Det islandske Kunnskapsdepartementet avholder nasjonale prøver i henholdsvis islandsk og matematikk for 4., 7. og 10. trinn, i tillegg til engelsk i 10. trinn. Formålene med disse prøvene er både å vurdere hvorvidt en elev har nådd de fastsatte kompetansemålene i faget, men også å gi elever, foreldre, skolepersonell og utdanningsmyndigheter landsdekkende informasjon og referanser om elevenes prestasjonsnivå. Disse teller imidlertid ikke mot elevenes standpunkt-karakter.

I det islandske skolesystemet er sluttvurderingen med andre ord både desentralisert og lite standardisert. Hver enkelt skole har stor autonomi når det kommer til valg av vurderingsform, når og hvordan vurdering skal gjennomføres og hvordan karaktersetting skal foregå. Dette skal imidlertid presiseres i de lokale læreplanveilederne, slik at skolens elever, deres foreldre og lærere alle kjenner til og har samme oppfatning av utgangspunktet for vurderingen i et fag.

Sensur og mulighet til å klage på sluttvurdering

I både grunnskolen og den videregående opplæringen settes standpunkt karakter av hver enkelt faglærer, basert på hvor godt en elev oppfyller kompetansemålene satt i den nasjonale læreplanen og spesifisert i den lokale læreplanveilederen.

Når det gjelder muligheten for å klage er denne begrenset i grunnskolen. Elever og deres foreldre har rett på en muntlig begrunnelse for elevens karakter og «to the revision by the compulsory school of the results of the assessment» (Compulsory School Act, 2008, art. 27). Dersom en videregående elev får vurderingen ikke bestått i et fag, er det imidlertid visse muligheter for å klage. En slik klage sendes til rektor, som da oppnevner en ny, ekstern sensor til å vurdere den opprinnelige vurderingen. Den eksterne sensorens vurdering er endelig og kan ikke påklages videre (Upper Secondary Education Act, 2008, art. 30).

2.4.2 Opptak til høyere utdanning

Island har i alt syv høyere utdanningsinstitusjoner: fire offentlige læresteder, to som drives av en non-profit-organisasjon og en privat institusjon (European Commission, 2022c). For å ha rett til opptak til høyere utdanning på Island kreves det at søkeren har fullført og bestått et studieforberedende program i videregående opplæring og dermed fått *Studentsprof*, som er hva vitnemål fra videregående kalles på islandsk. Dette er i dag den vanligste formen for kvalifikasjon som ungdom fullfører videregående opplæring med (Jónasson 1997). Det er det enkelte lærested som bestemmer om de ønsker å ha opptakskrav utover fullført studieforberedende videregående opplæring (Stefansson og Karlsdottir 2008).

Opptak gjøres lokalt ved en av de syv høyere utdanningsinstitusjonene på Island, og disse står selv fritt til å ha tilleggskrav til det generelle kravet om fullført studieforberedende videregående opplæring (Jónasson 2004). For program der det er mange flere søkere enn plasser gjelder tilleggskrav. Det er en opptakseksamen (entrance exam) for medisin og fysioterapi. For fag som sykepleie, tannhelse og odontologi holdes konkurranseeksamen i sluttet av første semester, og det er ikke alle som har fått opptak som får anledning til å fortsette. Det samme gjelder for de som studerer radiografi, der er antallet plasser på andre året av studiet begrenset (University of Iceland, u. å.).

2.5 Sverige

Formålet med utdanning i Sverige er «att barn och elever ska inhämta och utveckla kunskaper och värden. Den ska främja alla barns och elevers utveckling och lärande samt en livslång lust att lära. Utbildningen ska också förmedla och förankra

respekt för de mänskliga rättigheterna och de grundläggande demokratiska värderingar som det svenska samhället vilar på» (Skollagen 2010, § 4).

Sverige har niårig grunnskole, i tillegg til et førskoleår, slik at grunnskolen totalt blir 10 år. Barn kan begynne i førskoleåret det året de fyller seks år. Frem til 2018 var det frivillig å sende sitt barn på førskole, men fra og med høsten 2018 er førskoleåret gjort obligatorisk og dermed omfattet av skoleplikten. Siden 2012 har elever i svensk grunnskole fått karakterer fra 6. trinn, mens det tidligere ble gitt karakterer først på 8. trinn¹⁴, dvs. bare for de to siste årene av grunnskolen (European Commission, 2022d). Grunnskoleelever flyttes automatisk opp fra et år til et annet, og det er svært uvanlig å gå et år om igjen (Lundahl et al., 2016).

Videregående opplæring er formelt frivillig, men nesten alle elever som fullfører grunnskolen, fortsetter til videregående opplæring. Ifølge Skollagen (2010) må alle kommuner tilby videregående utdanning til ungdom som har fullført grunnskole. Denne rettigheten er begrenset, slik at elever må begynne i videregående opplæring før de fyller 20 år. For å kunne begynne i videregående opplæring må eleven ha fullstendig vitnemål. Karakter settes i alle fag, inklusive valgbare kurs eleven har tatt, og det er læreren i faget som setter sluttkarakter. Vitnemålet undertegnes av rektor (Lundahl et al., 2016).

Videregående opplæring er organisert i 18 nasjonale program som alle har en varighet på tre år. 12 av programmene er yrkesfaglige, mens seks program er studieforbereende. De studieforbereende programmene gir alle tilgang til høyere utdanning, mens elever som tar et yrkesfaglig program må ta ekstra kurs (i tillegg til kursene som inngår i programmet) for å få tilgang til høyere utdanning (European Commission, 2022d).

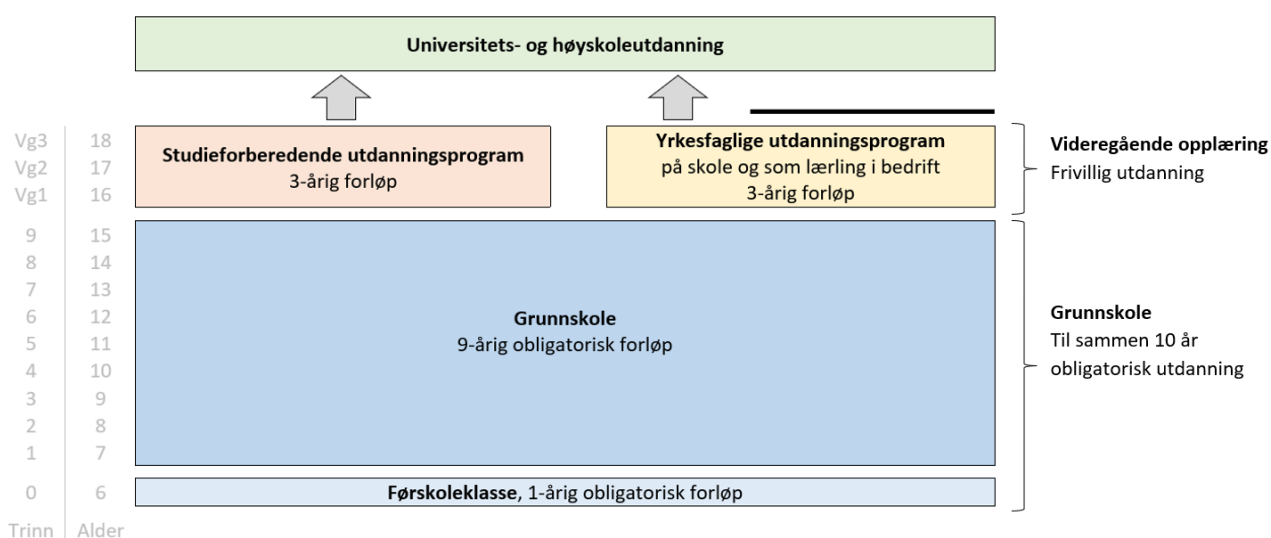
For å kunne søke opptak på et studieforbereende program må eleven ha bestått karakter i fagene svensk, engelsk og matematikk, i tillegg til minst ni andre fag fra grunnskolen. Eleven må derfor ha godkjent karakter i minst 12 fag. For å kunne søke opptak på et yrkesfaglig program må eleven ha en bestått karakter i fagene svensk, engelsk og matematikk, i tillegg til minst fem andre fag fra grunnskolen. Det er med andre ord ulike opptakskrav for å kunne søke studieforbereende og yrkesfaglige program. Elever som ikke oppfyller opptakskravene tilbys et introduksjonsprogram (Skolverket 2022a).

For å fullføre videregående opplæring må elever i studieforbereende program har oppnådd beståttkarakter (A-E) i minst 2250 poeng av de 2500 poeng programmet omfatter¹⁵. I tillegg må eleven ha fått godkjent sitt '*gymnasiearbeite*'. Dette er en oppgave som skal påse at elevene er studieforbereid, og tilsvarer i noen grad

¹⁴ Fra og med 2022 kan skoler, i samråd med skoleeier, velge å sette karakterer på elevene allerede fra 4. trinn, men det er valgfritt å gjøre det. Alle må fortsatt sette karakter fra og med 6. trinn.

¹⁵ Fra og med 2025 vil elevene få karakter i fag i stedet for kurs, og med andre ord blir det dermed færre karakterer på vitnemålet. For mer informasjon, se: <https://www.skolverket.se/undervisning/gymnasieskolan/aktuella-forandringar-pa-gymnasial-niva/amnesbetygsreformen>

norsk særemne. Det er imidlertid en viktig forskjell; mens særemne er knyttet til faget norsk, skal denne oppgaven ('*gymnasiearbeidet*') skrives innenfor fagområdet for programmet eleven går på (f.eks. elever på økonomiprogrammet skriver innen temaet økonomi, innovasjon eller jus) (Skolverket 2022d).



Figur 2.5 Utdanningssystemet i Sverige

2.5.1 Vurdering

For både grunnskolen og den videregående opplæringen er målet med vurdering at hver elev skal «tar ansvar for sitt lærande och sina studieresultat, och kan bedöma sina studieresultat och utvecklingsbehov i förhållande till kraven för utbildningen» (hhv. Läroplan for grundskolan 2019, s. 16; Läroplan for gymnasieskolan 2020, s. 10). Vurdering gjøres på en rekke ulike måter, inkludert gjennom utviklingssamtaler mellom lærer og elev, utarbeiding av individuelle utviklingsplaner som skal understøtte elevenes kunnskapsmessige og sosiale utvikling, samt setting av karakterer. I grunnskolen settes det karakter fra og med 6. trinn, men sluttvurdering gis kun i slutten av 9. trinn (European Commission, 2022d). I videregående gis sluttvurdering når et kurs er gjennomført, dvs. på alle trinn.

Forhold mellom sluttvurdering og læreplan

Sluttvurdering i både grunnskolen og den videregående opplæringen skal uttrykke «i vad mån den enskilda eleven har uppnått de nationella kunskapskrav som finns för respektive ämne» (hhv. Läroplan for grundskolan 2019, s. 16; Läroplan for gymnasieskolan 2020, s. 10). De nasjonale kunnskapskravene er fastsatt i de nasjonale læreplanene for henholdsvis grunnskolen og den videregående

opplæringen, både som overordnede mål og gjennom konkrete kursplaner for hvert enkelt fag. Fra og med høsten 2022 omtales 'kunnskapskrav' som 'karakterkriterier' (*betygskriterier* på svensk).

I grunnskolen presiserer læreplanene hva elevene skal lære på ulike årstrinn, samt hvilke karakterkriterier som kreves for å oppnå en viss karakter. I den videregående opplæringen styres undervisning ut fra de overordnede eksamensmålene, som gjelder generelt for hvert av de nasjonale studieprogrammene og deres respektive '*gymnasiearbeite*'. I tillegg skal de respektive kurs- og emneplanene legges til grunn for vurderingen i hvert enkelt fag. Læreplanverket presiserer at elevene skal vurderes allsidig, slik at de får mulighet til å vise sine kunnskaper og utvikling i lys av læreplanmålene. I tillegg skal faglærerne «vid betygssætningen utnyttja all tillgänglig information om elevens kunskaper i förhållande till de nationella kunskapskraven och göra en allsidig bedömning av dessa kunskaper» (hhv. Läroplan for grundskolan 2019, s. 16).

Det er med andre ord en tett kobling mellom læreplan og sluttvurdering i det svenske utdanningssystemet. Lærerne skal legge de nasjonale læreplanmålene til grunn når de gjennomfører sluttvurderingen i de ulike fagene, og elevene skal få fortløpende informasjon om sin egen måloppnåelse i løpet av året.

Sluttvurdering

Sluttvurdering gjøres i Sverige ved slutten av siste året i grunnskolen og i videregående skole. I Sverige inneholder sluttvitnemålet kun karakterer satt av lærer, det finnes ingen felles eksamen, verken i grunnskolen eller i videregående skole. Derimot finnes det andre elementer som er ment å sikre god og likeverdig karaktersetting, blant annet lærerlegitimasjon og nasjonale prøver.

Lærerlegitimasjon ble innført i 2011, etter et vedtak i Riksdagen. Legitimasjon omfatter alle lærere i grunnskolen og videregående og er ment å sikre at kun lærere som har godkjent utdanning får lov til å sette karakterer. For å få lærerlegitimasjon som lærer må nyutdannede lærere ha et prøveår det første året de jobber, der de har en mer erfaren (legitimert) lærer som mentor. Kun lærere som har fått godkjent legitimasjon (utstedt av Skolverket) som kan få fast ansettelse, og det er også kun disse som får lov til å sette karakter på egen hånd (Lilja 2011).

Nasjonale prøver har eksistert i den svenske skolen siden 1940-tallet, og brukes både i grunnskolen og i videregående opplæring. Bakgrunnen for innføringen av en form for standardiserte prøver som alle elever skulle gjennomføre på likt var at det tidlig på 1940-tallet ble oppdaget at det var variasjon i lærernes bedømming mellom klasser og kommuner, noe som ledet til ulikhet i karaktersetting for samme resultat. Med andre ord var lærerne gode å skille mellom ulike prestasjonsnivå i egen klasse, men da lærerne ikke hadde tilstrekkelig informasjon om prestasjonsnivået i andre klasser eller i landet som helhet, kunne de ikke ta det med i

betraktningen. Hensikten med nasjonale prøver var dermed å skape et instrument som kunne bistå lærerne med informasjon om egne elevers prestasjonsnivå, sammenlignet med andre elever i landet (Gustafsson og Erickson, 2018).

I dag er det nasjonale prøver på 3. trinn, 6. trinn, 9. trinn og i videregående opplæring. Prøvene er obligatoriske i noen fag. På 6. trinn gjelder dette svensk, engelsk og matematikk, mens det på 9. trinn avholdes nasjonale prøver i fem fag: svensk, engelsk, matematikk, en prøve i samfunnsrettede fag (geografi, historie, religion eller samfunnsfag) og en prøve i realfag (fysikk, kjemi eller biologi) (European Commission, 2022d). Disse prøvene påvirker sluttvurdering ved at læreren skal ta spesielt hensyn til dem ved karaktersetting, ved å la dem få større vekt i den samlede karakteren enn andre resultater. I tillegg finnes det frivillige prøver (sentralt utviklede, med sensorveiledning) som lærerne kan velge å bruke dersom de ønsker det.

Sluttvurdering i grunnskolen

Sluttvurdering i grunnskolen gjøres i slutten av 9. trinn, i alle fag eleven har tatt. Dersom et fag avsluttes før 9. trinn, settes sluttvurdering når faget avsluttes. Når sluttvurderingen gjøres, blir denne gjort i forhold til kriteriene for karakterer som gjelder for trinn 9, og elevenes kunnskaper blir da vurdert opp mot samtlige deler av kriteriene. Da gjøres en «sammentagen bedömning», der læreren setter den karakteren som best tilsvarer elevens kunnskaper. I de tilfeller der det ikke finnes tilstrekkelig grunnlag til å gjøre en vurdering av elevens kunnskaper i faget, for eksempel på grunn av fravær, settes det ingen karakter (Skolverket 2022d).

Ved karaktersetting skal resultatene på de nasjonale prøvene tas særlig hensyn til, ved at de har større betydning for karakteren enn andre enkelte prøver eller grunnlag. Samtidig skal ikke karakteren styres helt av elevens resultat på nasjonale prøver, men heller bygge på en helhetlig vurdering av elevenes samlede prestasjoner.

Sluttvurdering i videregående

Videregående opplæring er organisert rundt kurs, og det finnes både kursplaner og fagplaner. Fagplanene beskriver faget som helhet, målet med faget og hvilke kurset som inngår i faget. Som nevnt tidligere, er det fastsatt eksamensmål for alle program i videregående utdanning. Disse angir målene med programmet (også for enkelte retninger innen programmet, om det finnes) og målene for det programets 'gymnasiearbeite'.

Karakterskalaen

Siden 2011 har Sverige hatt en sekstrinns karakterskala, fra A til F, der de fem første stegene er godkjente karakterer (A-E), mens F er underkjent, dvs. stryk.

Karakterene er kriteriebaserte og tar utgangspunkt i de langsiktige målene og innholdet i fagets kursplaner, og brukes både i grunnskolen og i videregående skole. Det finnes «*betygskriterier*» som er beskrivende kjennetegn på kunnskaper for tre av nivåene i skalaene: A (utmerket), C (god) og E (godtagbar/akseptabel). Lærerne setter karakterer ved å bruke de oppsatte kriteriene for A, C og E og se hva som best tilsvarer den enkelte elevs kunnskapsnivå. For elever som ligger mellom A og C settes karakter B, mens karakter D er for de som ligger mellom karakter C og E. For at en elev skal kunne få karakteren E må kunnskapene eleven viser tilsvare alle deler av kriteriene satt for karakteren E (Skolverket 2022c).

De siste tre tiårene har Sverige skiftet karaktersystem tre ganger. Frem til 1994 var karaktersystemet i Sverige normrelatert (relativt) og hadde fem karaktertrinn, 1 til 5, der 5 er best. Karaktersystemet bygget på at karakterfordelingen skulle være normalfordelt på aggregert nivå (med like stor andel på 1 og 5 og med 3 som gjennomsnitt), og der resultater på nasjonale prøver skulle være styrende for lærernes karaktersetning. Dette systemet hadde en høy grad av likeverdighet i karaktersetning på tvers av lærere, skoler og kommuner, og karakterene var dermed sammenlignbare på tvers av skoler i landet (Gustafsson og Erickson, 2018). Likevel ble det normbaserte karaktersystemet kritisert av flere grunner, både av lærere, skoleadministrasjon, politikere og allmennheten. Et av punktene i kritikken var at lærerne ikke brukte systemet «slik det var tänkt, utan istället applicerade prosentfordelingen inom klasser utan hänsyn till resultatens på standardprov» (Gustafsson og Erickson, 2018, s. 3). Et annet kritikkpunkt var at karaktersystemet bidro til konkurranse mellom elevene, snarere enn samarbeid, mens et tredje kritikkpunkt var at karakterene, fordi det var samme gjennomsnitt fra år til år, ikke sa noe om nivået på kunnskaper og ferdigheter (Gustafsson og Erickson, 2018).

Da læreplanen ble endret i 1994 ble den relative skalaen erstattet med målrelaterte karakterer, med tre karaktertrinn: Godkjent (G), Vel godkjent (VG) og Mye vel godkjent (MVG). Det var flere hensikter med denne endringen, dels at karakterene kunne brukes mer i resultatstyring av skolen og dels at elevene heller skulle sammenligne seg med seg selv, snarere enn med de andre elevene, som ofte ble resultatet da den relative karakterskalaen ble brukt (Lundahl et al., 2016). Det var en generell oppfatning ved innføringen av det nye målrelaterte karaktersystemet at forankringen i mål og kriterier ville gi forutsetning for likeverdighet i karaktersetning, gitt at lærerne brukte kriteriene på lik måte. Imidlertid ble det etter bare et par år oppdaget at det nye kriteriebaserte karaktersystemet hadde alvorlige problemer i forhold til likeverdighet, ved at det var store forskjeller i karakter satt for lik prestasjon, både mellom lærere og mellom skoler, og avhengig av om skolen var offentlig eller privat.

I denne sammenhengen ble det stilt spørsmål ved de nasjonale prøvenes rolle som veiledende materiale for karaktersetning. For å prøve å bøte på problemet ble

karakterskalaen endret nok en gang, til det man har i dag, med fem godkjente karakterer og F for underkjent. Ved kursstart for et kurs i videregående opplæring skal læreren gi informasjon om hva som kreves for å oppnå ulike karakterer i faget, og elever har rett til å diskutere hvordan de ligger an med læreren underveis i kurset. Ved karaktersetting skal læreren ikke sammenligne eleven med andre, men heller måle hvor godt hver elev har nådd målene for kurset. Som veiledning til karaktersetting brukes det, i beskrivelsene av hva elevene skal kunne i et enkelt fag, verdiord som «lärare skal tolka för att komma fram till en adekvat och likvärdig bedömning». Eksempler på dette vil være at man kan skille mellom de tre karakternivåene som det gis beskrivelser av på følgende måte: «eleven för E kan förstå *"det huvudsakliga innehållet och uppfatta tydliga detaljer"* och för C *"det huvudsakliga innehållet och uppfatta väsentliga detaljer"*; motsvarande krav för betygssteget A är att eleven ska *"förstå såväl helhet som detaljer"*» (Gustafsson og Erickson, 2018, s. 6).

Nivå av standardisering og sentralisering

Som i de andre nordiske landene skal sluttvurdering i Sverige ta utgangspunkt i de nasjonale læreplanene, samt kurs- og emnebeskrivelser der dette foreligger. Selv om Sverige ikke har eksamen og vurdering er forbeholdt elevenes faglærere, har nasjonale prøver hatt en standardiserende funksjon. Som beskrevet ovenfor, skal læreren ved karaktersetting ta særlig hensyn til («särskilt beaktas») elevens resultat på nasjonale prøver. Resultatet på prøven har dermed større påvirkning på sluttkarakter enn de enkelte andre prøver. Samtidig skal ikke resultatet prøven styre hele karakteren, og det skal ikke være det eneste grunnlaget som sluttkarakter settes på grunnlag av (Skolverket 2022f).

På Skolverkets sider finnes det i tillegg en mengde materiale, såkalt «bedömningsstöd» som lærere kan bruke som støtte i karaktersetting. Alt er knyttet til nøkkelordet «bedömning». Dette ordet brukes fordi materialet ikke kun skal brukes i karaktersetting, men også er ment å gi råd og veiledning både for planlegging av undervisning, for å bedømme elevens kunnskapsutvikling og for å gi tilbakemelding. Materialet er svært spesifikt, helt ned på det enkelte fag (Skolverket 2022b).

Sensur og mulighet for klage på sluttvurdering

Som nevnt gjøres sensur kun av læreren i det enkelte fag, med støtte fra nasjonale prøver i de fag der disse finnes. Elever har anledning til å klage på standpunkt-karakter, men kun i situasjoner hvor det kan påvises vesentlige formelle feil i sluttvurderingsprosessen. Med andre ord er det ikke mulig å få en ny, innholdsmessig

vurdering av standpunktkarakter, verken i grunnskolen eller i videregående opplæring.

2.5.2 Opptak til høyere utdanning

For å få tilgang på høyere utdanning i Sverige må personen ha fullført videregående opplæring. Siden 2019 gir både studieforbereidende program og yrkesfaglige program generell studiekompetanse («grundläggande behörighet»), gitt at eleven har fullført tilstrekkelig mange kurspoeng i svensk, engelsk og matematikk. For de yrkesfaglige programmene betyr dette at eleven må ta kurs utover det som vanligvis er i programmet, mens det i de studieforbereidende programmene er en integrert del av programmet. Sistnevnte gir dermed automatisk tilgang til høyere utdanning (Statsrådsberedningen 2018). I perioden 2010 til 2018 var det kun de studieforbereidende, omtalt som «högskoleförberedande», programmene som ga generell studiekompetanse, og som det var mulig å søke opptak til høyere utdanning på grunnlag av (Skollagen, 2010, kap. 15 §7, kap. 16 §4, §26 og §28). Før det igjen, i perioden 1992 til 2009, ga alle typer program i videregående opplæring, både studieforbereidende og yrkesfaglige, generell studiekompetanse og dermed tilgang til høyere utdanning (Hovdhaugen et al., 2020).

I Sverige er søkning til høyere utdanning sentralisert og opptak gjøres i flere kvoter, eller «urvalsgrupper». I gruppen «Betygsurval» konkurrerer søkerne basert på sitt vitnemål fra videregående opplæring. Minst en tredel av plassene på programmet fordeles til søkere i denne gruppen. I tillegg har man en gruppe som kalles «Högskoleprovet» som også fordeler minst en tredel av plassene. Siden 1991 har det i Sverige vært mulig for søkere som har generell studiekompetanse å gjennomføre Högskoleprovet. Dette er en form for nasjonal opptaksprøve som er ment å teste generell studieforbereidethet, og som i noen grad ligner på SAT-testen som brukes i USA. I tillegg finnes det en gruppe som heter «Alternativt urval» som er en gruppe der lærestedet selv får bestemme opptaksgrunnlaget. Dette er ofte intervju, dokumentert arbeidslivserfaring eller ulike former for tester. Maksimalt en tredel av plassene på et studium kan fordeles via denne gruppen (Antagningen.se 2022).

Generell studiekompetanse oppnås ved at eleven har fullført minst 2 250 gymnasiepoeng, som regnes ut basert på kombinasjonen av type kurs (lengre/vanskeligere kurs gir mer poeng) og omfanget på kurset. I tillegg må eleven ha oppnådd bestått-karakter i engelsk, svensk og matematikk. Imidlertid er det relativt få fag på universiteter og høyskoler som kun krever generell studiekompetanse. I stedet har de aller fleste fag har noen form for tilleggskrav, som omtales som «särskild behörighet». Disse kravene kan være alt fra krav om avansert matematikk eller

ulike realfag for opptak til noen program, til at statsvitenskap krever at eleven har tatt høyeste nivå av kurs i samfunnsfag (Hovdhaugen et al., 2020).

De som søker i «Betygskvoten» søker i stor grad opptak basert på de karakterene de kom ut av videregående med. Tidligere var det ikke anledning til å forbedre karakterer i det hele tatt. Da var det bare de som strøk og dermed ikke fikk vitnemål eller om de manglet et fag som kreves for opptak (særskild behørighet), som fikk anledning til å ta faget på Komvux (voksenopplæring) for å komplettere sitt vitnemål. I dag finnes det imidlertid to muligheter for å høyne sitt karaktersnitt fra videregående. Den ene måten er «utbyteskomplettering», der søkeren tar eksamen i et allerede bestått fag, og dersom karakteren er høyere erstatter denne den gamle og snittet øker. I slike tilfeller må dette gjøres gjennom selvstudier, da man ikke har rett til å delta i undervisning på allerede beståtte kurs. Den andre måten er «tilleggskomplettering», som innebærer å ta eksamen i nye fag, og dermed få nye karakterer på vitnemålet som kan bidra til å høyne snittet (Hovdhaugen et al., 2020).

3 Diskusjon av landssammenligningen - kvalitetssikring av sluttvurdering

Som landsammenligningen har vist, er oppbyggingen av utdanningssystemene i stor grad lik på tvers av de nordiske landene. Utdanningssystemene har lik struktur, med grunnskole til eleven er 16 år, og deretter videregående opplæring delt i studieforberedende og yrkesfaglig utdanning. Vurdering er kriteriebasert i alle land, noe som innebærer at elevene blir vurdert etter mål fastsatt i nasjonale – og i noen tilfeller lokale – læreplaner.

Det som tydeligst skiller sluttvurderingssystemene i de nordiske landene, er hvorvidt man benytter seg av eksamen som sluttvurderingsform. Per i dag er det kun Norge og Danmark som benytter seg av eksamen, mens Island og Sverige kun bruker standpunkt karakterer som sluttvurderingsform. Det samme gjelder til dels i Finland, hvor det kun er elever som ønsker å søke opptak til høyere utdanning som må gjennomføre en avsluttende eksamen. All annen sluttvurdering gjøres gjennom elevenes standpunkt karakter.

Som påpekt av Eksamensgruppa er det vesentlig at sluttvurderingen faktisk måler det den ønsker å måle (validitet), at den ikke påvirkes av tilfeldigheter (reliabilitet), og at den er rettfærdig og likeverdlig for alle elever (Blömeke, 2019a). Hvilke valg de ulike landene har tatt når det kommer til utformingen av sluttvurderingssystemet, reflekterer deres noe ulike syn på disse spørsmålene, med andre ord hvordan sluttvurdering samlet sett kan gi rettfærdig og relevant informasjon om elevenes kompetanse i de ulike fagene. Basert på beskrivelsen av sluttvurderingssystemene i de nordiske landene gitt ovenfor, vil vi i dette kapitlet derfor drøfte hvordan validitet, reliabilitet og rettfærdighet forstås og operasjonaliseres i de fem nordiske landene.

Eksamensgruppen trekker også frem håndterbarhet som et viktig fjerde element i denne sammenhengen. Håndterbarheten omhandler hva som er mulig å få til rundt de tre andre kriteriene når det gjelder praktisk organisering og gjennomførbarhet. Som Eksamensgruppa påpeker er det utfordrende å oppnå både høy validitet, reliabilitet, rettfærdighet og håndterbarhet på samme tid. Blant annet kan håndterbarheten ofte komme i konflikt med reliabiliteten, for eksempel ved at

ønsket om flere enn en sensor kan være svært ressurskrevende. En vurdering må derfor gjøres rundt hvor stor del av opplæringstiden og skolens ressurser som bør og skal brukes på å organisere og gjennomføre sluttvurdering for å sikre validitet, reliabilitet og rettferdighet.

Fra dokumentene som har ligget til grunn for denne studien har det vært utfordrende å finne informasjon om hvilke håndterbarhetsvurderinger som har blitt gjort i de ulike landene i utarbeidingen av sluttvurderingssystemet. Dette vil derfor ikke bli dekket her, til tross for at slike vurderinger sannsynligvis kan bidra til å forklare noen av valgene som har blitt tatt.

3.1 Validitet og reliabilitet

Validitet, ofte omtalt som «gyldighet», bør ifølge Eksamensgruppa regnes som det mest sentrale vurderingsteoretiske begrepet ved eksamen. Validiteten handler om hvorvidt en vurdering oppfattes som fornuftig og hva slags belegg eller argumenter som brukes for å bedømme dette. Med andre ord, å «validere en prøve eller eksamen innebærer å utvikle en argumentasjon om hva slags bevis som skal regnes som gyldige, og om hvordan tolkningen skal foregå (Blömeke et al., 2019a, s. 27; se også Markus og Borsboom, 2013). Et sentralt spørsmål er også om sluttvurderingen måler det den er utarbeidet for å måle.

Vurderingen av hvorvidt en vurdering er gyldig eller ei avhenger både av formålet med vurderingen og konteksten som vurderingen skjer innenfor. Faktorer som kan påvirke validiteten, er motstridende formål, forskjeller i hjelp og assistanse mellom lærer og elever, ulik vurdering av ulike deler av vurderingen eller eksamensoppgaven, tilbakevirkende, utilsiktede effekter (f.eks. en «washback»-effekt, hvor testingen påvirker innhold og praksis i undervisning og læring (Cheng og Curtis, 2003)), og uønskede systemiske effekter (f.eks. stress), for å nevne noen.

Reliabilitet, eller pålitelighet, handler om hvorvidt og på hvilken måte gjentatte vurderinger samsvarer (Pellegrino et al., 2001). Høy validitet avhenger av i hvilken grad man klarer å fjerne tilfeldigheter i vurderinger. Ifølge Eksamensgruppa (Blömeke et al., 2019a, s. 29) styrkes reliabiliteten ved at vurderingen er «så uavhengig som mulig av sensoren» som vurderer besvarelsen, av eksamensformen som er brukt og innholdet i denne, av tidspunkt for eksamen osv. Kravet for reliabilitet er at verken «andre sensorer eller en gjentakelse av eksamen en annen dag, med andre oppgaver eller andre eksamensformer, ville lede til et annet resultat». Reliabilitet regnes som et nødvendig, men ikke tilstrekkelig, vilkår for validitet.

I sin rapport beskrev Eksamensgruppa hvordan validitet og reliabilitet blir sikret i dagens eksamenssystem. Mens validiteten sikres gjennom at «spesielt kyn-dige fagpersoner samarbeider om oppgaver basert på nasjonale retningslinjer», avhenger reliabiliteten av at elevenes prestasjoner vurderes på en likeverdig måte,

uavhengig av hvor, når og hvordan vurderingen gjennomføres. Eksamensgruppa identifiserer en rekke faktorer som må være på plass for å sikre tilstrekkelig reliabilitet i vurderingen. For eksempel er det nødvendig med tydelige oppgaver med gode instruksjoner, slik at både elevene og sensor kjenner til hva vurderingen skal baseres på. Vurderingskriteriene, forstått som kjennetegn på måloppnåelse, må ta utgangspunkt i de konkrete målsettingene som er satt for faget. Disse må også være tydelig formulert og tilgjengelige for begge grupper. Til slutt er det nødvendig med en omfattende sensorskolering, særlig hvor to eller flere sensorer vurderer samme oppgave. Dette er med på å sikre et godt og omfattende tolkningsfellesskap, med andre ord en felles oppfatning av hvilke kriterier og refleksjoner som skal ligge til grunn for vurderingen.

Dersom vi legger disse kriteriene til grunn, kan vi se at de nordiske landenes tilnærming til validitet og reliabilitet i stor grad er sammenfallende, men at de også har noen vesentlige forskjeller. Særlig gjelder dette mellom de landene som har sentralt gitte kompetansemål og eksamen, og de som kun har lokalt bestemt sluttvurdering.

3.1.1 Validitet i lys av sluttvurderingens formål

Formålet med sluttvurdering og eksamen i de ulike landene ble beskrevet i kapittel 2. Som den beskrivelsen viser, er det relativt stort sammenfall mellom de nordiske landene, selv om det er noe variasjon i hva de ulike formålsforståelsene vektlegger. Alle systemene legger til grunn at sluttvurderingen skal gi informasjon om enkeltelevers kompetanse på slutten av et utdanningsløp i konkrete fag.

Som vi har sett, legger alle også en kriterie- eller målbasert tilnærming til grunn for vurderingen. I Norge er «grunnlaget for vurdering i fag (...) kompetansemåla i læreplanen i faget ...» (Forskrift til opplæringslova, § 3-3). Det samme gjelder i Danmark og på Island. Mens førstnevnte beskriver hvordan formålet med vurdering er å «dokumentere, i hvilken grad eksaminanden oppfyller de mål og krav, der er fastsat for faget og uddannelsen» (Almen eksamensbekendtgørelser, 2016, § 1), beskriver sistnevnte det som å «evaluate how successfully pupils meet the objectives of the National Curriculum Guide as well as individually determined learning objectives...» (Compulsory School Act, 2008, art. 27).

I det finske lovverket beskrives det hvordan «de nationella bedömningskriterierna [skal] användas» ved vurdering av elevenes teoretiske og praktiske kunnskapsnivå (Utbildningsstyrelsen, 2014, s. 50), mens sluttvurdering i Sverige skal uttrykke «i vad mån den enskilda eleven har uppnått de nationella kunskapskrav som finns för respektive ämne» (Skolverket 2022c, s. 16). Hovedformålet med sluttvurdering er dermed tilnærmet likt i de nordiske landene: å vurdere hvorvidt

og i hvilken grad elevene har oppnådd kompetansemålene fastsatt i den nasjonale læreplanen.

Hvordan man går frem for å vurdere hvorvidt dette formålet er oppnådd varierer imidlertid noe mellom de fem landene. Først og fremst finner vi variasjon i hvor 'bindende' de nasjonale kompetansemålene er. Hvis vi begynner med det norske utdanningssystemet, ser vi at det er det kompetansemålene fastsatt i den nasjonale læreplanen som ligger til grunn for både standpunkt- og eksamenskarakter. Både i vurdering av standpunkt- og eksamensprestasjon er det med andre ord felles, nasjonale mål som ligger til grunn uavhengig av hvilken skole elevene har gått på.

I noen av de andre landene er dette imidlertid ikke tilfellet. Det ser vi blant annet i Finland og på Island. Selv om de begge har nasjonale læreplaner, som fastsetter de overordnede kompetansemålene som elevene skal vurderes etter, har de også lokale læreplaner som presiserer hvordan disse målene skal forstås. Systemet åpner dermed opp for lokal tolkning og tilpasning av kompetansemålene, noe som innebærer at grunnlaget for vurdering kan – og ofte vil – variere avhengig av hvor elevene bor. Et annet eksempel finnes i Danmark. Selv om hvert enkelt fag må planlegges og gjennomføres i lys av den overordnede formålsparagrafen for den danske folkeskolen, samt fagets egne formålsbeskrivelser, ble antall bindende mål i læreplanverket redusert i 2017. De fleste målene er nå veiledende, heller enn obligatoriske. Dette betyr at den enkelte lærer og skole har større frihet når det kommer til å bestemme innhold og undervisningsform.

En slik desentralisert struktur, med mindre standardisering på tvers av skoler og deler av landet, betyr ikke nødvendigvis at validiteten er lavere. Synet på hvorvidt vurdering av måloppnåelse er rettfærdig, avhenger av hvilken elevgruppe man brukes som sammenligning (Waldow, 2014). Fordi man i Norge og Sverige har sentralt gitte mål, uten (formell) mulighet for lokal tilpasning, avhenger validiteten av hvorvidt alle elever i samme fag og på samme trinn vurderes likt. Dette er også grunntanken bak bruken av nasjonale prøver i Sverige og det norske Utdanningsdirektoratets tiltak for å bidra til et tolkningsfellesskap blant eksamenssensorer: å sørge for likeverdig vurdering for alle elevene på et gitt trinn. I land med mer lokalt tilpassede læreplaner vil nødvendigvis sammenligningsgruppen være en annen. Her brukes de lokale læreplanene til å skape en felles forståelse av innholdet og forventningene i vurderingskriteriene blant de som læreplanene retter seg mot, slik at både elever, lærere og sensorer har like forutsetninger for å forstå hva som skal til for å nå de spesifiserte målene.

En annen viktig forskjell mellom de nordiske utdanningssystemene er hvordan de går frem for å kvalitetssikre resultatene fra sluttvurderingen. I Norge og Danmark spiller sentralt gitte eksamener en viktig rolle i denne sammenhengen, da elevene her får en ekstern vurdering av sin fagkompetanse. Slike eksamener har

også identiske oppgavesett for alle elever, uavhengig av hvor de bor og hvilken skole de går på, noe som betyr at karakterene elevene får på disse, er satt med utgangspunkt i det samme vurderingsgrunnlaget. Sverige, Finland og Island har imidlertid få eller ingen sentralt gitte eksamener i sine utdanningssystem. Her baseres elevenes sluttvurderinger i all hovedsak på faglærers egen vurdering. Istedenfor en ekstern vurdering av alle elever på likt grunnlag, legger systemet her vekt på en mer allsidig og dyptgående vurdering av hver enkelt elevs prestasjon over tid. Grunntanken er at det er hver enkelt faglærer som, basert på hva elevene har lært i faget, best kan vurdere hvorvidt kompetansemålene er oppnådd. I Sverige kvalitetssikres riktignok lærerens skjønnsvurdering gjennom bruk av nasjonale prøver, både gjennom at de samme oppgavene blir gitt til alle elever, men også gjennom konkrete vurderingskriterier som hver enkelt lærer er pålagt å følge.

3.1.2 Sluttvurderingens form og innhold

Spørsmålet om sluttvurdering, enten det gjelder standpunkt eller eksamenskarakter, faktisk måler det de skal – elevenes måloppnåelse i faget – står sentralt i lovverket til alle de nordiske landene. For eksempel på Island presiseres viktigheten av at vurderingen blir «planned and carried out in such a way that it evaluates what it is supposed to evaluate [validitet] in a reliable manner» (Ministry of Education, Science and Culture, 2012a, s. 26).

En måte å styrke sluttvurderingens validitet på er å variere vurderingens form og innhold. For alle landene presiseres viktigheten av å vurdere elevenes arbeid, læring og kompetanseheving på mange og varierte måter. For eksempel beskriver det danske læreplanverket hvordan alle oppgaver og prøveformer som brukes i vurderingen skal i «al væsentlighed tilsammen dække de faglige mål, kernestoffet og det supplerende stof» som ligger til grunn for faget. Det samme ser vi på Island. Her presiserer den nasjonale læreplanguiden at, siden vurdering skal favne så mange ulike lærings- og kunnskapsaspekter som mulig, er det nødvendig å bruke en rekke ulike vurderingsformer, inkludert skriftlige, muntlige og praktiske vurderinger, samt vurderinger av ulike lengde og dybde, individuell og gruppevurdering. Dette er nødvendig for å sikre at enhver elev får en reell mulighet til å vise frem sin kompetanse, og for å forhindre at vurderingen legges opp på en måte som favoriserer visse elever eller visse typer måter å lære på. En slik variasjon kan sies å bidra til å sikre sluttvurderingens validitet, ved at man tester et større utsnitt av elevenes kompetanse og ferdigheter.

3.1.3 Tolkningsfellesskap

For å sikre god reliabilitet i sluttvurderingen, er det nødvendig å legge til rette for en felles tolkningsramme for de personene som gjennomfører vurdering og fastsetter karakterer. Dette gjelder uavhengig av om sluttvurderingen gjennomføres som standpunkt eller gjennom eksamen. I alle de nordiske landene legges det stor vekt på å skape en felles forståelse av de fastsatte kompetansemålene. Det er imidlertid en viss variasjon i hvordan landene tolker betydningen av tolkningsfellesskap og hvordan de går frem for å sikre dette.

I den norske konteksten brukes vurderingsveiledere og detaljerte kompetansebeskrivelser på hver enkelt skoletrinn for å skape et slikt tolkningsfellesskap. Når det gjelder eksamen, arrangerer Udir fellessensur i samarbeid med statsforvalterne ved alle sentralt gitte eksamener (Blömeke, 2019a). Unntaket her er sentralt gitte eksamener med lokal sensur. I den forbindelse arrangerer Udir blant annet sensorskolering, hvor sensorene med utgangspunkt i et utvalg reelle eksamensbesvarelser, diskuterer hvilken kompetanse som vises og hvilken karakter de ulike besvarelsene bør få. De utarbeider i tillegg eksamensveiledere med kjennetegn på måloppnåelse i de ulike fagene, som skal brukes som en felles referanseramme i sensurprosessen. Hensikten med disse tiltakene er å sikre et godt tolkningsfellesskap på tvers av de ulike sensorene som vurderer elevenes prestasjoner i de enkelte fagene.

Tilsvarende tiltak finner vi i de andre nordiske landene. Som beskrevet i kapittel 2, regnes de lokale læreplanene som et viktig tiltak i dette arbeidet i både Finland og på Island. Mens de nasjonale læreplanene beskriver de overordnede målene for utdanningen, skal de lokale læreplanene inneholde mer detaljerte beskrivelser av hvordan vurdering skal gjennomføres og hvilke kriterier som skal ligge til grunn. Dette er med på å gi skolen og lærerne i faget et felles grunnlag og retningslinjer for det daglige undervisnings- og vurderingsarbeidet. Dette er med på å sikre lik vurdering, om ikke mellom alle elever på samme trinn, så i hvert fall mellom alle elever på samme skole. En noe lik tilnærming har blitt benyttet i Danmark, hvor skolene og lærerne selv har en viss frihet når det kommer til hvordan de tolker og implementerer de nasjonale kompetansemålene. Samtidig har det blitt utarbeidet felles veileder (faghæfter) i alle fag, som skal bistå lærerne i deres tolkning og arbeid med kompetansemålene (Børne- og Undervisningsministeriet, 2021a). Disse skal bidra til å skape et nasjonalt tolkningsfellesskap med utgangspunkt i de overordnede kompetansemålene.

Som i Norge og Danmark, ser det svenske utdanningssystemet ut til å vektlegge et nasjonalt tolkningsfellesskap. Dette gjøres primært gjennom bruken av nasjonale prøver. Formålet med disse er å sikre «en likvärdig och rättvis bedömning och betygssättning av en elevs kunskaper» (Skolverket, 2022e). De nasjonale prøvene er bygget opp med utgangspunkt i analyser av læreplan, fag- og emneplaner. De skal

dekke så store deler av pensum som mulig, for å sikre høy validitet. For å sikre høy reliabilitet piloteres oppgavene i elevgrupper på mellom 200-1000 elever. Dette gjøres for å fjerne tilfeldige feil og mangler som kan påvirke elevenes resultater. Det presiseres også at prøvene skal gjennomføres på lik måte og på samme tid for alle elever, slik at resultatene som produseres er likeverdige i så stor grad som mulig.

3.2 Rettferdighet

Rettferdighet viser til, ifølge Eksamensgruppa, at «alle elever må ha den samme sjansen til å vise kompetansene sine under eksamen» (Blömeke et al., 2019a, s. 30). Med andre ord skal en rettferdig eksamen sikre at verken oppgaven eller resultatet påvirkes av variabler som kjønn, språk, funksjonsgrad og lignende. Det er derimot mulig å identifisere forskjeller i hva som oppfattes som rettferdig, og hva det betyr at elevene har «den samme sjansen» til å vise sin kompetanse og vurderes på et likt grunnlag. Sentralt for denne diskusjonen er såkalt institusjonell rettferdighet. Ifølge Liebig (2006, s. 23, sitert i Waldow, 2014) stiller en slik rettferdighetsforståelse spørsmål om «which rules or principles of justice are embedded in the blueprints of our social institutions, how allocative decisions (...) [are] made in these institutions and how conceptions of justice come into play in these institutions». Når det gjelder sluttvurdering, er spørsmålet dermed hvilken rettighetsforståelse som ligger til grunn for det utdanningsrammeverket beskrevet ovenfor for hvert av de nordiske landene.

Et nyttig skille i denne sammenheng er mellom innholdsmessig (substansiell) og prosessmessig (prosedural) rettferdighet. Mens førstnevnte er basert på en forståelse om hvorvidt sluttresultatet, i dette tilfellet sluttvurderingen (f.eks. en eksamenskarakter), er rettferdig, fokuserer sistnevnte på om hvorvidt sluttvurderingen var resultatet av en rettferdig prosess eller prosedyre. I et proseduralt rettferdighetsperspektiv, betyr dette at dersom sluttvurderingen ble gjennomført i tråd med gjeldende regler, vil den fastsatte karakteren per definisjon være rettferdig (uavhengig av hvilken karakter som blir satt).

Leventhal (1980) identifiserer syv komponenter som sammen er med på å sikre prosedural rettferdighet. Av disse er særlig tre av spesielt relevans her. Den første omhandler valg av «vurderingsagenter», dvs. hvilke aktører som står ansvarlige for elevenes sluttvurdering. Den andre ser på etableringen av såkalte «safeguards», dvs. instrumenter og regler som tar sikte på å sikre felles standarder for vurdering. Til slutt har vi muligheten for å klage på sluttvurderingen, som beskrives av Waldow (2014, s. 326) som «a kind of ‘safeguard ex post’, that is, a possibility for correcting breaches of the postulate of equality of opportunity after the grade or certificate has been given».

I det følgende vil vi beskrive hvordan de nordiske utdanningssystemene vurderer henholdsvis hvilke agenter, «safeguards» og muligheter for klage som sammen sikrer en rettferdig sluttvurderingsprosess. Som vi skal se er det en viss variasjon i hvilke mekanismer de ulike systemene har vektlagt når det kommer til å sikre elevenes rettferdige vurdering, noe som igjen reflekterer systemenes ulike syn på hva som teller som en rettferdig vurdering.

3.2.1 Valg av agenter og «safeguards» i sluttvurderingsprosessen

En «agent» i denne sammenhengen referer til de aktørene som står ansvarlige i sluttvurderingsprosessen. Vanlige vurderingsagenter kan være elevenes faglærer, oppnevnte sensorer, enten interne eller eksterne, eller ulike former for vurderingsnemnder. Mens Island og Sverige kun har faglærere som vurderingsagenter, har Norge, Danmark og til en viss grad Finland valgt en kombinasjon av interne og eksterne agenter.

Som vi har sett, avhenger valg av agent i stor grad av nivået av standardisering og sentralisering i sluttvurderingsprosessen. I både det finske, svenske og islandske utdanningssystemet settes alle karakterer lokalt, av elevenes faglærer. Unntaket her er den finske studenteksamen, hvor Studenteksamensnemnda tar den endelige vurderingen av eksamensprestasjonen, basert på faglærers innledende vurdering. I Danmark og Norge avhenger valg av agent av hvorvidt det gjelder standpunkt- eller eksamenskarakter, og om eksamen er lokalt eller sentralt gitt. Når det gjelder standpunktkarakter settes disse av elevenes faglærer, på lik linje med Finland, Sverige og Island. Nå det gjelder eksamen, rettes skriftlige eksamener med sentralt gitte oppgaver av to eksterne sensorer. Alle andre eksamener må ha minst en ekstern sensor, mens den andre kan være intern.

Valg av agent avhenger av hvilken rettferdighetsforståelse man legger til grunn for sluttvurderingen. Dersom systemet tar utgangspunkt i elevenes egen faglærer som vurderingsagent, vektlegger det viktigheten av en «holistisk» og dyptgående vurdering av elevenes prestasjoner. Den sterke rollen elevenes faglærer har i sluttvurderingsprosessen i alle de nordiske landene kan sies å springe ut fra et positivt syn på lærerens profesjonalitet (Waldow, 2014). Et slikt perspektiv legger til grunn at, siden lærerne kjenner elevene og har fulgt deres utvikling over tid, har de mulighet til å gi en mer helhetlig vurdering av elevenes prestasjoner enn det en ekstern sensor eller nemd kan gjøre. En risiko med valg av denne typen agenter er at vurderingen ikke gjøres på likt grunnlag på tvers av individuelle lærere og skoler. Ulike lærere kan tolke kompetansemålene på forskjellige måter, noe som kan bety at elevenes prestasjoner ikke blir sammenlignbare med hverandre. Dette kan påvirke vurderingens reliabilitet, dersom det fører til at forskjellige lærere vurderer en elevs prestasjoner ulikt. I tillegg ser man her en risiko for at faktorer som

ikke er relatert til elevenes faktiske prestasjoner, for eksempel deres personlighet eller fremtoning (såkalt «trynefaktor») kan påvirke sluttvurderingsresultatet.

Dersom systemet legger til grunn en ekstern vurderingsagent vektlegges i stedet objektivitet og nøytralitet i sluttvurderingsprosessen. Her blir rettferdighet sikret gjennom å fjerne alle faktorer, inkludert «trynefaktor», elevenes personlighet, deres forhold til lærer osv., som ikke er direkte relatert til sluttvurderingen. Elevenes prestasjoner vurderes her kun ut fra prøve- eller eksamensresultatet, som er likt for alle elevene som gjennomgår vurderingsprosessen. Et slikt system fører imidlertid også med seg visse risikoer, som kan påvirke resultatets validitet, reliabilitet og rettferdighet. Vurderingen blir mer mekanisk, og det er umulig for sensor å vite om formidlende omstendigheter ved elevenes prestasjoner. Eksterne sensorer har ofte også begrenset med kunnskap om hva elevene faktisk har gjennomgått og lært i et fag, noe som kan bety at vurderingen gjøres utenfor den konteksten som selve læringen foregikk innenfor.

For å unngå de potensielle ulempene assosiert med henholdsvis faglærer og ekstern sensor som vurderingsagenter, har de ulike utdanningssystemet introdusert en rekke «safeguards». Disse skal bidra til at elevene sikres en likeverdig vurdering, uavhengig av hvem som gjør den. I utdanningssystemer hvor faglærer spiller en viktig rolle, vektlegges faglærers profesjonalitet og evne til å gi en mangesidig og rettferdig vurdering av elevene. Faglærer kan sørge for at elevene blir vurdert på måter som tillater dem å vise frem sine ferdigheter og prestasjoner, noe som er mer krevende for standardiserte prøver som vurderer alle elever likt.

I slike systemer er det gjerne også et sterkt fokus på tolkningsfellesskap knyttet til målene fastsatt i læreplanverket (Waldow, 2014; se også (Blömeke et al., 2019a)). Dette ser vi blant annet i både Finland og Island, hvor lovverket presiserer viktigheten av å tydeliggjøre ovenfor både lærere, elever og foreldre hvilke vurderingskriterier som ligger til grunn for vurderingen og hvordan disse skal tolkes. Dette gjøres blant annet gjennom diverse læreplanveiledere¹⁶, samt de lokale læreplanene beskrevet tidligere. Tilsvarende veiledere finnes for kompetansemålene i Danmark, Norge og Sverige. I Norge har Utdanningsdirektoratet i tillegg satt i gang tiltak som skal bidra til å styrke tolkningsfellesskapet ved sensur på sentralt gitt eksamen, inkludert sensorskoler, felles sensur og eksamensveiledninger med kjennetegn på måloppnåelse (Blömeke et al., 2019a).

Sverige har, i tillegg til slikt veiledningsmateriale, en annen «safeguard»-mekanisme i sitt utdanningssystem. Dette er de nasjonale prøvene beskrevet ovenfor, som skal sikre et felles tolkningsfellesskap på tvers av individuelle lærere og skoler.

¹⁶ For Finland, se hhv. Grunderna för läroplanen för den grundläggande utbildningen (2014) og Grunderna för gymnasiets läroplan (2019). For Island, se hhv. The Icelandic National Curriculum Guide for Comprehensive Schools (2014) og The Icelandic National Curriculum Guide for Upper Secondary Schools (2012).

I retting av oppgaver må faglærer følge de detaljerte vurderingsinstruksjonene som medfølger hver enkelt prøve. Dette skal bidra til å sikre likeverdig vurdering uavhengig av hvilke lærer som retter hvilken oppgave (Skolverket 2022e). Selv om de nasjonale prøvene i Sverige skal inngå i lærerens vurderingsgrunnlag, skal de ikke styre den endelige standpunkt karakteren. Ved vurdering av standpunkt karakter kan en mer helhetlig vurdering legges til grunn, hvor «all tilgjengelig informasjon om elevens kunnskaper i forhold til kunnskapskraven» skal tas i betraktning (Skolverket 2022e). Sverige legger dermed til grunn en kombinasjon av eksterne «safeguards» og lærerens profesjonalitet, i sin tilnærming til vurdering.

Som denne gjennomgangen viser, har alle de nordiske utdanningssystemene en felles forståelse av rettferdig vurdering, som innebærer at alle elevene vurdering på likeverdig grunnlag. Det varierer imidlertid hvordan de går frem for å sikre dette. Mens noen av landene, inkludert Norge og Danmark, bruker en kombinasjon av intern og ekstern sensur, har Island og Finland vektlagt viktigheten av en holistisk vurdering basert på sterke, lokale tolkningsfellesskap. Sverige skiller seg til en viss grad ut med sitt fokus på nasjonale prøver rettet av elevenes faglærer, som tilstreber å sikre en helhetlig, men likeverdig vurdering uavhengig av sensor.

3.2.2 Mulighet for å klage på sluttvurderingen

Muligheten for å klage på sluttvurderingen ansees som en «safeguard ex post» da det åpner opp for en retrospektiv kontroll av sluttvurderingsprosessen i etterkant av satt karakter. Samtidig påpeker Waldow (2014) at muligheten for å klage kan ha både positive og negative konsekvenser for legitimiteten til sluttvurderingsprosessen og hvorvidt den oppleves som rettferdig. På den positive siden muliggjør klagerettigheter en korreksjon av en urettferdig eller ikke-valid sluttvurdering. Dette kan sees på som en styrke, da det viser systemets evne til å rette opp feil. På den negative siden kan klager være med på å fremheve systemets innebygde risiko for feil og tilfeldigheter. I tillegg kan en slik individualisert klageprosess være problematisk ut fra et rettferdighetsperspektiv, da det gir de elevene som klager, en urettferdig fordel vis-a-vis de elevene som ikke klager.

Fra gjennomgangen over ser vi at de nordiske landene, om enn på litt ulike måter, forsøker å balansere disse to perspektivene. Først og fremst: i alle de nordiske landene kan elevene klage på sin standpunkt- og/eller eksamens karakter i tilfeller hvor de kan påvise formelle feil i vurderingen. Et eksempel på dette finner vi i det danske systemet. Her har elever i de gymnasiale utdannelsene rett til å klage på standpunkt karakter ut fra såkalte «rettslige spørsmål», dvs. dersom det foreligger vesentlige formelle feil i vurderingen eller det kan påvises vesentlige prosedurale feil ved vurderingsprosessen (Bekendtgørelse om de gymnasiale uddannelser, 2017, § 74). Alle de nordiske utdanningssystemene åpner dermed opp for en svært

begrenset klagerett i tilfeller hvor det kan påvises vesentlige prosedurale feil i sluttvurderingsprosessen.

Imidlertid varierer det hvorvidt de åpner opp for en mer innholdsmessig vurdering av sluttvurderingen som foreligger. Med andre ord er det ulikt mellom landene hvorvidt elever har rett til å klage på karakteren selv i tilfeller hvor den formelle prosessen har blitt gjennomført i tråd med gjeldende regelverk. Her ser vi et skille mellom Norge, Island og Finland på den ene siden og Sverige og Danmark på den andre siden.

Dersom vi begynner med det norske systemet, ser at vi ifølge den norske forskriften til opplæringsloven § 3-17 har elever rett til å klage på det enkeltvedtaket som en sluttvurdering er. Dette gjelder både standpunkt- og eksamenskarakterer. Ved klage på standpunktkarakter kan klageinstansen imidlertid kun vurdere om læreren fulgte de gjeldende retningslinjer for vurdering i faget, mens ved eksamen er det mulig å klage på alle siden ved karakterfastsettingen. Tilsvarende paragrafer finner vi lovverket til både Finland og Island; mens elever i Finland kan be om en ny vurdering, og deretter eventuelt klage til enten det regionale forvaltningskontoret (standpunkt) eller Studenteksamensnemda (eksamen), har islandske elever mulighet til å be om en ny vurdering av sin standpunktkarakter (Compulsory School Act, 2008, art. 27). I sistnevnte tilfeller skal det opprettes en ny, ekstern sensor, som gjennomfører en ny vurdering av elevens prestasjoner, og setter en endelig standpunktkarakter i faget. I den islandske videregående opplæringen gjelder dette imidlertid kun i tilfeller hvor eleven fikk ikke bestått i faget.

I Sverige, og til dels Danmark, er det imidlertid ikke mulig å be om en ny, skjønnsmessig vurdering av avsluttende karakter. I Sverige, som kun opererer med standpunktkarakterer, er det bare vesentlige formelle feil som kan påklages. I Danmark, som vi har sett, gjelder det samme for standpunktkarakterer. Når det gjelder eksamenskarakterene åpner det danske lovverket opp for en viss skjønnsmessig vurdering, ved at elevene har rett til å klage på den faktiske vurderingen av prøven som ble gjort. Imidlertid presiseres det at skolen kun skal gi tilbud om ny vurdering/prøve dersom det foreligger «særlige omstændigheder, der giver anledning til berettiget tvivl om bedømmelsen, herunder væsentlige procedurermæssige mangler» (Almen eksamensbekendtgørelse, 2016, § 51, stk. 2). Det er med andre ord den prosedurale rettfærdighetsforståelsen som vektlegges også med henhold til eksamen.

Fra disse ulike tilnærmingene til elevenes klagerett kan vi trekke ut visse antakelser om de ulike utdanningssystemenes syn på rettfærdighet i sluttvurdering. For eksempel, rettigheten til å be om en ny, skjønnsmessig vurdering, som er grunnleggende blant annet ved skriftlig eksamen i det norske utdanningssystemet, kommer fra et ønske om å sikre elevenes rettssikkerhet. Antagelsen er at elevenes klagerett muliggjør en korreksjon av en urettferdig eller ikke-valid sluttvurdering

som ikke ville skjedd dersom eleven ikke hadde klaget. Grunntanken bak denne rettigheten stammer fra kravet om både reliabilitet og rettferdighet, nemlig at en elev skal få muligheten til å sjekke at han eller hun blir vurdert på samme måte av en annen sensor. Ved å fjerne eller begrense klagemuligheten fjerner man også muligheten til å kontrollere for tilfeldigheter eller mangler i det opprinnelige vurderingsgrunnlaget.

For utdanningssystemene uten slik klagemulighet, for eksempel Sverige, ser vi en annen forståelse av rettferdighetsbegrepet. Sentralt her er viktigheten av å behandle alle elevene «likvärdig» (Waldow, 2014). Som diskutert over spiller de nasjonale prøvene en sentral rolle i å sikre svenske elevers rettssikkerhet i sluttvurderingen. Dette er både fordi disse prøvene gjennomføres av alle elever på samme tid og på samme grunnlag, men også fordi det er med på å sikre lik vurdering av elevenes prestasjoner på tvers av individuelle lærere og skoler. Med utgangspunkt i et slikt syn er en klagerett verken nødvendig eller ønskelig, både fordi de nasjonale prøvene er tilstrekkelige til å sikre lik vurdering, og fordi det åpner opp for at noen elever får muligheten til å få sin besvarelse vurdert på nytt, mens andre elever ikke får det. Som beskrevet av Waldow (2014, s. 335): «possibly the strength of one procedural component (safeguards) compensates for the weakness of another (appeals): as national tests (...) play such an important (not least symbolic) role in safeguarding fair assessment, possibilities of appealing against assessment do not seem to be deemed necessary» i den svenske konteksten.

Vi ser dermed at klageretten i de ulike landene til en viss grad avhenger av hvor stor tiltro man har til systemenes innebygde «safeguards» som beskrevet over. I tilfeller hvor lærerens profesjonalitet legges til grunn, blir en eventuell klagerett en mulighet for noen elever til å skaffe seg en fordel sammenlignet med de elevene som ikke klager. I de landene som ikke anser lærerens profesjonalitet alene som tilstrekkelig, for eksempel Norge og Sverige, brukes andre «safeguards», henholdsvis eksamen og klagerett på den ene siden, og bruk av nasjonale prøver på den andre.

DEL 2 KUNNSKAPSKARTLEGGING

4 Systematisk kartlegging – metode

Del 2 av denne rapporten inneholder funnene fra den systematiske kunnskapskartleggingen. Formålet har vært å fremskaffe en oversikt over eksisterende forskning knyttet til organiseringen og gjennomføringen av sluttvurdering og eksamen i grunnskolen og den videregående opplæring i de nordiske landene. En slik kartlegging har også vært hensiktsmessig for å avdekke en rekke kunnskapshull på feltet. Funnene fra kunnskapsoppsummeringen blir beskrevet i kapittel 5, mens vi i dette kapitlet beskriver hvilken metode som er brukt for å fremskaffe disse.

4.1 Om metoden

Metoden systematisk kartlegging («systematic mapping review») ble utviklet for å imøtekomme behovet for å tilpasse systematiske kunnskapsoppsummeringer til bredere og mindre spesifikke problemstillinger. Systematiske kartlegginger har dermed ikke som hensikt å besvare et spesifikt spørsmål, for eksempel om effekten av et tiltak. I stedet sikter metoden mot å kartlegge, beskrive og katalogisere forliggende kunnskap om et bestemt tema, for eksempel eksamens- og sluttvurderingssystemer.

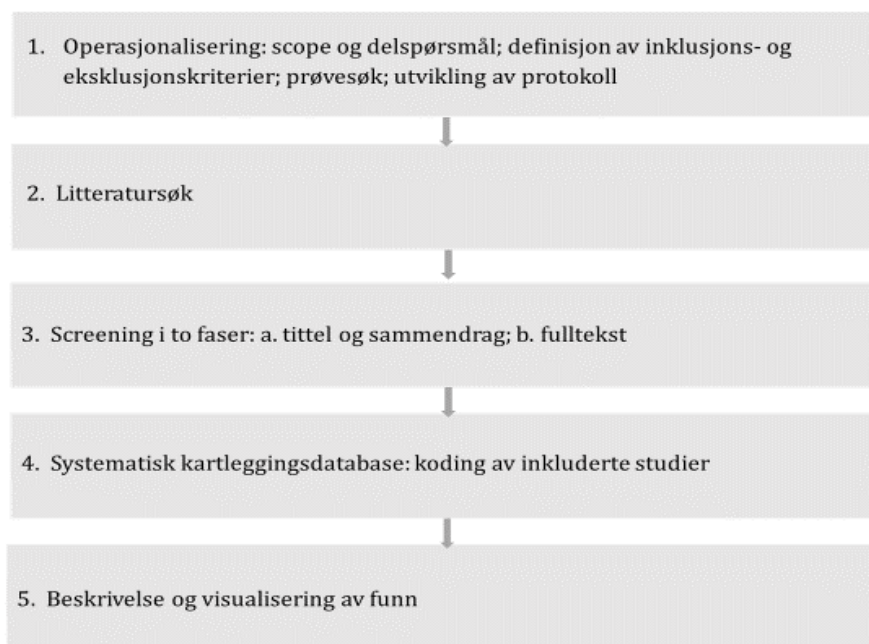
Forskningsbidrag fremskaffet gjennom den systematiske kartleggingen kan brukes for å utvikle en bedre forståelse av begreper (f.eks. eksamensordning, sluttvurdering), identifisere kunnskap for policy-relevante problemstillinger, samt identifisere beslektede kunnskapsklynger og kunnskapshull på feltet. Beslektede kunnskapsklynger kan egne seg for videre utforskning, eksempelvis i en tradisjonell systematisk kunnskapsoppsummering (James et al., 2016).

I likhet med systematiske kunnskapsoppsummeringer følger systematiske kartlegginger de samme objektive og transparente prosedyrene for å sammenstille relevant forskningsbasert kunnskap om et bestemt tema. Dermed unngås mulige fallgruver eller skjevheter som kan oppstå i tradisjonelle litteraturstudier.

Verdien av systematiske kartlegginger ligger i at de kan avdekke bredden i et forskningstema, som ofte er nødvendig for å kunne adressere politikknære spørsmål. Dermed kan slike kunnskapsoversikter avdekke både områder der det finnes

tilstrekkelig kunnskap og identifisere temaer som ikke har blitt tilstrekkelig utforsket, dvs. kunnskapshull (James et al., 2016).

Metoden anses derfor som velegnet for å besvare oppdragets problemstillinger ved å beskrive hva som finnes av forskningsbasert kunnskap om eksamen- og sluttvurderingssystemet i de nordiske landene. Figur 4.1 illustrerer de enkelte trinnene i en systematisk kartlegging, fra operasjonaliseringen av problemstilling til beskrivelse og visualisering av funnene.



Figur 4.1 Prosedyrer ved en systematisk kartlegging (ifl. James et al., 2016).

Operasjonalisering

Operasjonaliseringen av forskningsspørsmål og bredden av tematikken forutsetter en klar definisjon av inklusjons- og eksklusjonskriterier (seleksjonskriterier). I tillegg omfatter denne fasen et eksplorerende prøvesøk og en utvikling av en protokoll, dvs. en nærmere konkretisert prosjektbeskrivelse.

Inklusjons- og eksklusjonskriterier (seleksjonskriterier):

Den systematiske kartleggingen omfatter forskningsbasert kunnskap om sluttvurdering og eksamen i grunnskolen og videregående opplæring i de nordiske landene. Forskningsbasert kunnskap om systemet, organisering og gjennomføring av eksamen og sluttvurderinger, om slutteksamen i forhold til andre vurderinger og eksamensrelaterte tiltak er inkludert. Vi inkluderte forskningsbidrag fra de fem nordiske landene Norge, Danmark, Finland, Island og Sverige. Dersom forskningsbidraget var publisert på islandsk eller finsk, ble det utelatt på grunn

av språkbarrierer. Inkludert materiale ble avgrenset til de siste ti publiseringsårene, 2012 til 2022, for vitenskapelig publikasjoner, noe som er en vanlig avgrensingsperiode for denne type kunnskapsoversikter (se f.eks.: Aksnes et al., 2021; Wollscheid et al., 2020). Gitt at temaet har høy relevans for politikktutforming og praksis, inkluderte vi i tillegg offentlige publikasjoner fra de siste 20 år og andre faglige publikasjoner fra de siste fem år. Vi har valgt disse tidsavgrensningene siden de ulike publikasjonene antas å ha ulikt tidsperspektiv. I det følgende gir vi en definisjon av ulike former for publikasjoner:

- Vitenskapelige publikasjoner: fagfellevurderte tidsskriftartikler publisert på engelsk, eller på ett av de skandinaviske språkene.
- Offentlige publikasjoner: primært rapporter som har som formål å presentere og drøfte kunnskapsgrunnlaget og mulige handlingsvalg. I tillegg vil vi se på strategier for utvikling og iverksetting av offentlige tiltak for løsning av samfunnsmessige problemer og utfordringer (NOU-er og tilsvarende i de andre nordiske landene).
- Annen faglig publisering: forskningsrapporter, samt andre former for publikasjoner basert på forskning som har gjennomgått en form for kvalitetssikring, men som vanligvis er publisert utenfor kanalregisteret. I dette oppdraget vil vi begrense oss til forskningsrapporter når vi snakker om annen faglig publisering.

Følgende tabell gir en oversikt over inklusjons- og eksklusjonskriterier:

Tabell 4.1 Inklusjons- og eksklusjonskriterier

| | Inklusjonskriterier | Eksklusjonskriterier |
|-------------------|---|--|
| Fenomen | Eksamen/ Sluttvurdering: Organisering av sluttvurdering/eksamen Sluttvurdering i forhold til andre vurderinger Eksamensrelaterte tiltak Sluttvurdering sin rolle i organisering av utdanningssystemet | Undervisvurderinger som ikke ses i sammenheng med slutteksamen; Undersøkelser hvor sluttvurdering brukes kun som utfallsmål eller for å måle noe annet |
| Populasjon | Skoleelever | Studenter, spesifikke elevgrupper (f.eks., elever med en spesifikk diagnose) |
| Geografi | Nordiske land: Norge, Sverige, Danmark, Finland og Island | Land utenfor Norden |
| År | Vitenskapelige publikasjoner: 2012-2022 Offentlige publikasjoner: 2002-2022 Annen faglig publisering, forskningsrapporter: 2012-2022 | Før 2012 for vitenskapelige publikasjoner Før 2002 for offentlige publikasjoner Før 2017 for annen faglig publisering |
| Språk | Norsk, svensk, dansk, engelsk, | Finsk, islandsk |
| Publikasjonstyper | Vitenskapelige publikasjoner (primærstudier og sekundærstudier) Offentlige publikasjoner (f.eks., NOUer, meldinger) Annen faglig publisering: forskningsrapporter | Sekundærstudier (internasjonale litteraturgjennomganger, dersom de ikke er relevante for den nordiske konteksten). Bøker, kapitler, PhD-avhandlinger, konferansesammendrag, populærvitenskapelig artikler |

Litteratursøk

Som i andre forskningsprosjekter har datainnsamlingen – her bestående av litteratursøk og seleksjon av relevante publikasjoner – betydning for den videre forskningsprosessen, inkludert analyser, resultater og konklusjoner. Søkestrategien kombinerte et systematisk søk etter internasjonal forskning i databasen Web of Science (WoS) med et strategisk søk i nasjonale kilder etter offentlige dokumenter. Vi har i tillegg gjort strategiske søk i publiseringene til sentrale fagmiljøer. Søk i ulike kilder er en iterativ, heller enn en lineær prosess, noe som er vanlig for kunnskapsoppsummeringer innen samfunnsfag. Nedenfor skisserer vi vår søkestrategi og søkestrengen for Web of Science (WoS).

WoS er primærkilden for den systematiske kartleggingen og er én av de mest benyttete databasene for systematiske kunnskapsoppsummeringer og for bibliometriske formål. Den dekker primært engelskspråklige vitenskapelige tidsskrifter. Fordelen med WoS er at den dekker alle fag og land, datakvaliteten er høy, og den er basert på konsistente prinsipper for inkludering av litteratur. Den største begrensningen ved bruk av WoS er at ikke all vitenskapelig litteratur er indeksert. For eksempel er tidsskrifter på andre språk enn engelsk i liten grad indeksert og dekningsgraden av humaniora og samfunnsvitenskap er lavere enn andre fag (Aksnes og Sivertsen, 2019).

Nedenfor presenteres søkestrengen som ble brukt for databasen WoS. Søket er bygget opp av søketermer som dekker tematikken og populasjonen for prosjektet, uten å innsnevre gruppen for mye. I tillegg er det brukt avgrensinger for å spisse

søket. Disse er skissert i tabellen for inklusjonskriteriene ovenfor. Her har vi valgt å bruke avgrensning basert på tid, språk og geografi.

Vi brukte følgende søkestreng (dato: 23.mars 2022), som resulterte i 3765 treff.

- "exam*" OR "final grad*" OR "final assessment*" (Topic) and "school*" OR "high school*" OR "upper secondary school*" OR "pupil*" (Topic) and "Norw*" OR "Swe*" OR "Dan*" OR "Fin*" OR "Ice*" (Topic) and 2012/2022 (Publication Date) and English OR Norwegian OR Swedish OR Danish OR Finnish OR Icelandic (Language) and NORWAY or ICELAND or SWEDEN or DENMARK or FINLAND (Countries/Regions)

I første omgang begynte vi med en pilotscreening hvor alle i prosjektet vurderte 50 publikasjoner for sammen å oppnå en fellesforståelse om fenomenet og seleksjon av relevante studier. I denne screeningen ble titler og sammendrag av referansene fra det systematiske søket i WoS vurdert basert på inklusjonskriteriene skissert ovenfor. Søket resulterte i over 3700 treff (tittel og sammendrag).

Gjennom det systematiske søket antok vi å kunne identifisere de sentrale fagmiljøene i de nordiske landene. Resultatet fra det systematiske søket kunne dermed bidra til å informere det strategiske søket i nasjonale kilder og supplerende søk i referanselister til relevante publikasjoner.

Det ble også foretatt strategisk søk i nasjonale kilder. For å identifisere relevante publikasjoner utenfor de fra WoS, kombinerte vi ulike strategiske metoder som søk i nasjonale kilder, søk i referanselister fra allerede inkluderte studier og søk på nettsider fra utvalgte fagmiljøer. På NIFU har vi en egen database over ikkevitenskapelige publikasjoner som er registret i Cristin/NVI som dekker tidsavgrensningen for annen faglig publisering, som også ble benyttet.

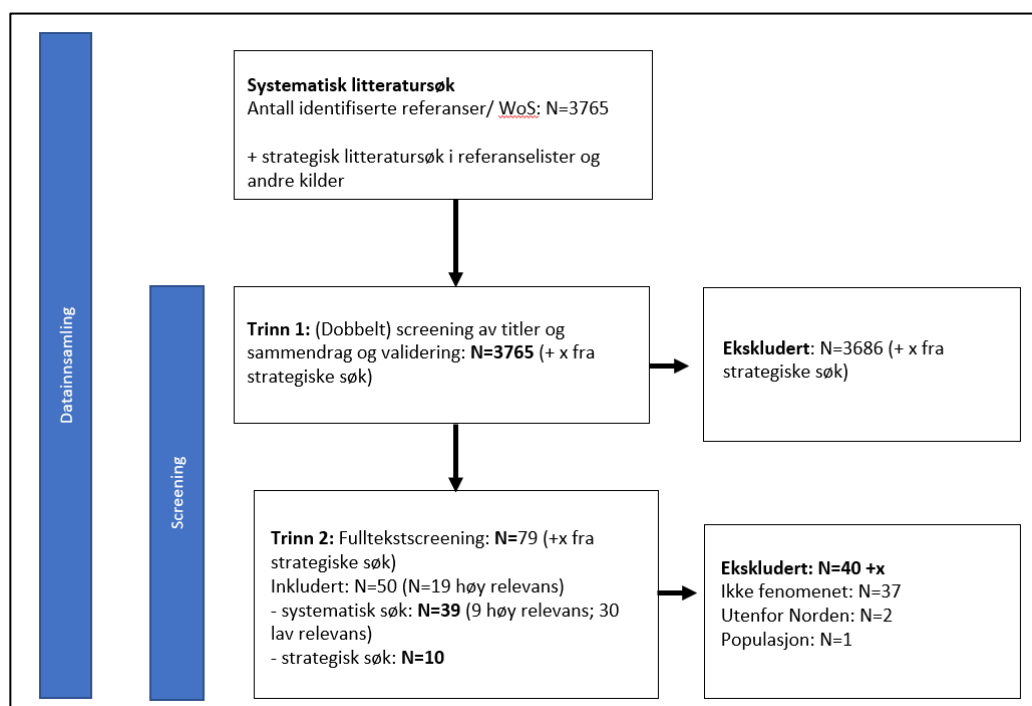
Screening av tittel og sammendrag, og av fulltekst

Screening av resultatene fra søket mot seleksjonskriteriene i den systematiske kartleggingen foregikk i en tottrinnsprosess. I første omgang ble resultatene fra litteratursøket importert for en innledende screening. Her ble tittel og sammendraget til artiklene lest av to forskere som uavhengig av hverandre vurderte relevans for problemstillingen opp mot inklusjons- og eksklusjonskriteriene. Hensikten med dette var å ekskludere publikasjoner som ikke passer oppdragets fokus og prosjektets problemstillinger (se tabell 4.1). Usikre vurderinger med tanke på inklusjon eller eksklusjon ble gjennomgått og validert på nytt av en tredje forsker. På grunn av et høyt antall treff var denne typen screening relativt tidkrevende.

Dernest ble relevante bidrag innhentet i fulltekst. Fulltekstartiklene ble lest av minst én forsker med tanke på inklusjon og eksklusjon. Ved usikkerhet ble bidraget diskutert med en annen forsker/prosjektleder. Vi endte opp med 39 inkluderte studier fra det systematiske søket i WoS i tillegg til 11 studier fra det strategiske søket i andre kilder.

Dokumentasjon av screeningen, med antall ekskluderte artikler på hvert steg og grunner for eksklusjon for fulltekstbidrag, er en del av metoden i en systematisk kartlegging. Dette gjøres for å vise transparens i prosessen, for eksempel illustrert av et PRISMA-flytdiagram (Tricco et al., 2018). Et slik diagram som gir en oversikt over screeningen er vist i figur 4.2.

I gjennomføringen av screeningen av sammendrag og tittel (Trinn 1) og fulltekstartikler (Trinn 2) fra det systematiske søket brukte vi det digitale verktøyet Covidence¹⁷. Screening av referanser som vi identifiserte gjennom det strategiske søket ble gjort manuelt.



Figur 4.2 Utvalg av relevante publikasjoner (PRISMA-diagram)

Koding, analyse og kartlegging

I systematiske kunnskapsoversikter omfatter koding og dataekstraksjon både metadata (informasjon som beskriver studien og dens metoder) og funn fra de inkluderte bidragene. I systematiske kartlegginger kan derimot kodingen begrenses til

¹⁷ <https://www.covidence.org/>

metadata og/eller mindre omfattende informasjon om resultatene fra studien. Dersom et forskningsbidrag inneholder et fyldig sammendrag, kan det være nok å hente ut data fra selve sammendraget. I dette prosjektet har vi inkludert forskningsbidrag knyttet til eksamen- og sluttvurderingssystemene i de nordiske landene, og de ulike bidragene belyser tematikken i varierende grad. Følgelig har vi i første omgang kategorisert forskningsbidragene ut ifra om de har høy eller lav relevans for kartleggingens problemstilling.

Forskningsbidrag vurdert til å ha høy relevans, omtaler systemet for og organiseringen av eksamen og sluttvurderingen i de nordiske landene. Ettersom det dukket opp relativt få studier med høy relevans for kartleggingens problemstilling, har vi valgt å gi en mer fyldig beskrivelse av de inkluderte studiene (se delkapittel 5.2).

Inkluderte forskningsbidrag vurdert til lav relevans, tar ikke direkte opp eksamens- eller sluttvurderingssystemene i de nordiske landene. Likevel tar bidragene opp relatert tematikk, og de kan dermed belyse sentrale relevante kunnskapsklynger som er relatert til tematikken sluttvurdering (se delkapittel 5.3).

I tillegg til å identifisere hvilke områder det eksisterer forskning på, kan vi på bakgrunn av vår systematiske kartlegging også identifisere noen kunnskapshull som kan indikerer behov for videre forskning (se delkapittel 6.1.2). Se også kapittel 5 der de aktuelle kunnskapsklyngene blir presentert.

4.2 Metodiske styrker og begrensninger

Gjennom metoden systematisk kartlegging er det mulig å både avdekke områder der det finnes kunnskap og identifisere kunnskapshull (James et al., 2016). Rapporten presenterer kartlagt kunnskap eller temaområder av høy relevans for problemstillingen, samt temaområder av lavere relevans (kunnskapsklynger). Sistnevnte vurderes å være tematisk i gråsonen, men allikevel delvis relevant for problemstillingene.

Den metodiske tilnærmingen er basert på transparente og systematiske prosedyrer. Dens styrke ligger dermed i at den muliggjør en innhenting og beskrivelse av den foreliggende forskningsbaserte kunnskapen om eksamen og avsluttende prøvers rolle i sluttvurderingssystemet, mulige kunnskapshull og implikasjoner for videre forskning på en systematisk og dermed etterprøvable måte.

Begrensningen i fremgangsmåten er både knyttet til tiden og ressursene vi hadde til rådighet, samt begrensninger i språkferdigheter når det gjelder inklusjon av mulig relevante forskningsbidrag fra Finland og Island. En annen begrensning var knyttet til antall tilgjengelige og relevante datakilder og -baser.

For å motvirke større geografiske skjevheter, har vi konsultert fageksperter fra de ulike landene, inkludert Finland og Island, for å validere datamaterialet. Likevel er datamaterialet karakterisert ved en utvalgsskjevhet i favør av Norge: vi har

inkluderte 13 av 20 forskningsbidrag fra Norge, som ble vurdert til å ha høy relevans for problemstillingen. Vi kan dermed ikke garantere at datamaterialet er like representativt for alle av de fem nordiske landene, da vi kan anta at det foreligger noen studier på finsk og islandsk som ikke ble fanget opp i kartleggingen.

5 Funn fra kartleggingen

I dette kapittelet presenteres sentral forskning på sluttvurdering i Norden. Studiene som presenteres er de inkluderte bidragene fra de systematiske og strategiske søkene fra kartleggingen. Da det finnes relativt lite forskning på selve organiseringen av eksamens- og sluttvurderingssystemet i de nordiske landene, vil vi i vår kartlegging belyse sentrale funn på tvers av landene. Disse funnene er organisert/relatert etter hvilket nivå forskningen på sluttvurdering forholder seg til: makronivå (systemet), mesonivå (institusjonelle faktorer og organisering) eller mikronivå (gjennomføring av og erfaringer med sluttvurdering). I tillegg presenterer vi sentrale kunnskapsklynger (temaklynger), som viser studier som er relatert til tematikken sluttvurdering, men som ikke anses å ha høy relevans for problemstillingen.

5.1 Oversikt over de inkluderte studiene

Den systematiske kartleggingen omfatter forskningsbasert kunnskap om eksamen- og sluttvurderingssystemet i grunnskolen og videregående opplæring i de nordiske landene. Til sammen 19 av de inkluderte forskningsbidragene ble vurdert til å ha høy relevans. Selv om litteratursøket omfattet alle de nordiske landene, er det kun forskningsbidrag fra Norge, Sverige og Danmark som ble vurdert til å ha høy relevans for problemstillingen. Tabell 5.1 gir en oversikt over de inkluderte studiene. Tabellen er organisert med utgangspunkt i studienes referanse, publiseringskanal, språk, metode, tema og populasjon og er gruppert etter tre hovedkategorier eller nivåer (makro, meso og mikro) som ble utviklet gjennom en induktiv-deduktiv tilnærming.

Tabell 5.1 Kartlegging av de inkluderte studiene fra systematisk og strategisk søk

| | Førsteforfatter, år, land | Publiseringstype/Språk | Navn av Tidsskrift/ Utgiver | Metode | Hovedområde/ Nivå | Tema | Populasjon |
|-----|----------------------------------|-------------------------------|---|--|--------------------------|---|---|
| 1. | Bensnes, 2019 Norge | Tidsskrift/ Engelsk | The Scandinavian Journal of Economics | Kvantitativ | Mikro | Organisering av eksamen/ sluttvurderingssystemet | Elever i VGS |
| 2. | Bjørnset*, 2020 Norge | Rapport/ Norsk | Fafo | Kvantitativ og kvalitativ | Mikro | Evaluering av matematikkeksamen ved 10. trinn | Lærere |
| 3. | Bukh, 2022 Danmark | Tidsskrift/ Engelsk | Sage Open | Kvalitativt (intervju med skoleledere) | Meso | Sluttvurdering/ eksamensrelaterte tiltak | Rektorer i Folkeskolen (grunnskolen) |
| 4. | Bøhn*, 2018 Norge | Tidsskrift/ Engelsk | Journal of Language Teaching and Research | Verbale protokoller og halvstrukturerter intervjuer med lærere | Mikro | Eksamensordning knyttet til muntlig eksamen i engelsk | Sensorer og lærere som er involvert i eksamen |
| 5. | Falch, 2014 Norge | Tidsskrift/ Engelsk | Labour Economics | Kvantitativ (registerdata SSB) | Meso | Organisering av eksamen/avsluttende eksamen vgs. | Elever i VGS (siste året) |
| 6. | Gustafsson*, 2018, Sverige | Tidsskrift/ Svensk | Acta Didactica | Oversikt over endringer ved nasjonale prøver over tid | Makro | Ordning ved nasjonale prøver og endringer over tid | Nasjonale prøver og lærernes rolle |
| 7. | Hovdhagen*, 2018 Norge | Tidsskrift/ Norsk | Acta Didactica | Kvantitativ (karakterstatistikk) og kvalitativ (intervjuer) | Mikro | Funksjon av den norske ordningen med henholdsvis eksamens- og standpunktkarakterer for lærernes vurderingspraksis | Lærere |
| 8. | Hvidman, 2021 Danmark | Tidsskrift/ Engelsk | Journal of Human Resources | Kvantitativ (registerdata, statistics Denmark) | Meso | Sluttkarakterer (GPA) og ny sluttvurderingsreform (2007) i Danmark | Elever i VGS |
| 9. | Kupiainen*, 2016 Finland, | Tidsskrift/ Engelsk | Journal of Educational Research Online | Eksplorativ: analyse av eksamensresultater | Meso | Sammenlignbarhet av eksamenskarakterer i ulike fag og endringer over tid i elevenes valg av eksamen | Elever i VGS |
| 10. | Modell, 2021 Sverige | Tidsskrift/ Engelsk | Sport, education and society | Kvalitativ (intervju, fokusgrupper) | Mikro | Sluttevaluering | Elever i VGS |
| 11. | Prøitz*, 2013 Norge | Tidsskrift/ Norsk | Education Inquiry | Kvalitativ: lærerintervjuer | Meso | Betydningen av skolefaget for karaktersetting | Lærere |

| | | | | | | | |
|-----|---------------------------------|------------------------|---------------------------------|--|-------|---|---|
| 12. | Sandvik, 2021 Norge | Tidsskrift/ Engelsk | Teacher and Teaching | Kvalitativ (intervju, fokusgrupper) | Mikro | Evalueringsformer og sluttvurdering | Elever i VGS (siste året) |
| 13. | Sandvik & Sommer, 2021 Norge | Tidsskrift/ Engelsk | Languages | Mixed-method (kvantitativ survey og kvalitative intervju) | Mikro | Studentperspektiv på elevmedvirkning i vurderinger | Elever i VGS (første året) |
| 14. | Skovholt, 2021 Norge | Tidsskrift/ Engelsk | Journal of Pragmatics | Kvalitativt (videoopptak av muntlig eksamen) ved 2 ungdoms- og 2 VGS | | Organisering av (muntlig) eksamen | Eksamen til elever i ungdomsskole/ VGS (secondary school) |
| 15. | Waldow, 2014 Sverige | Tidsskrift/ Engelsk | Comparative Education Review | Komparativ | Makro | Organisering av eksamen/sluttvurderingssystemet i ulike velferdsregimer | Aktører som er involvert i eksamen og sluttvurdering |
| 16. | Björnsson*, 2021 Norge | Rapport/ Norsk | ILS/ Universitetet i Oslo (UiO) | Klassiske metoder: beregninger av kappa, vektet kappa og intraklasse-koeffisienter | Meso | Evaluering av sensorreliabilitet | Sensorer som er involvert i eksamen |
| 17. | Hovdhagen*, 2014 Norge | Rapport/ Norsk | NIFU | Kvantitativ og kvalitativ | Meso | Avvik mellom standpunkt karakterer og eksamens karakterer | Private og offentlige skoler |
| 18. | Sjaastad*, 2016, Norge | Rapport/ Norsk | NIFU | Kvantitativ og kvalitativ | Meso | Organisering av eksamen | Elever i VGS |
| 19. | Tveit*, 2018 Norge | Tidsskrift/ Norsk | Acta Didactica | Dokumentanalyse av Stortingsmeldinger, Utdanningsdirektoratets årlige rapporter, Utdanningsspeilet samt intervjuer | Makro | Eksamens rolle i norsk grunnsopplæring | Eksamens |

*identifisert gjennom strategisk søk.

5.2 Oppsummering av de inkluderte studiene

I dette delkapitlet oppsummeres de inkluderte studiene fra vår systematiske kartlegging. For å belyse noen sentrale sider ved sluttvurderingssystemet på tvers av de nordiske landene, er studiene kategorisert etter tre hovedområder:

1. Overordnet systemperspektiv på eksamen og sluttvurderinger (*makro-nivå*).
2. Institusjonelle faktorer og organiseringen av eksamen og sluttvurderinger (*mesonivå*).
3. Gjennomføring, praksiser og erfaringer knyttet til eksamen og vurderinger (*mikronivå*).

De tre grupperingene av bidragene kom frem gjennom kodingen av de inkluderte forskningsbidragene¹⁸. Videre brukes grupperingene i presentasjonen av forskningsbidragene, ettersom de får frem ulike nivåer og perspektiver på eksamen- og sluttvurderingssystemet på tvers av de nordiske landene.

5.2.1 Studier knyttet til et overordnet systemperspektiv på eksamen og sluttvurderinger

Den første grupperingen inkluderer forskningsbidrag som tar for seg det overordnede systemperspektivet på eksamen og sluttvurdering i de nordiske landene. Denne grupperingen dekker de mer overordnede formålene, strukturene og perspektivene knyttet til de ulike sluttvurderingssystemene. Et slikt makroperspektiv kan for eksempel dreie seg om eksamens rolle i sertifisering og styring av utdanningen, eller rettferdighetsaspektet i eksamenssystemet. Følgende tre studier er presentert her: Gustafsson og Erickson (2018), Tveit og Olsen (2018) og Waldow (2014).

Gustavsson, J-E og Ericksson, G. (2018). Nationella prov i Sverige – tradition, utmaning, förändring. *Acta Didactica*.

Land: Sverige

Tema for artikkelen er nasjonale prøver i Sverige. Formålet med artikkelen er å gi en oversikt over hvilke endringer prøvene har gjennomgått de siste tiårene, sett utfra de ulike styringsdokumentene og karaktersystemene de har vært knyttet til. Metodisk sett er artikkelen hovedsakelig deskriptiv. Den beskriver på hvilken måte og i hvilken grad de nasjonale prøvene har vært i stand til å møte de

¹⁸ Det vil være en glidende overgang mellom kategoriene, og noen studier vil kunne plasseres i skjæringsfeltet mellom kategoriene.

utfordringer som endringene i det svenske karaktersystemet over tid har krevd. I tillegg diskuterer artikkelen generelle metodiske utfordringer, både ved karaktersetting og ved hvordan nasjonale prøver er utformet i Sverige.

Forfatterne åpner med å påpeke at diskusjonen om karakterer, karaktersetting og nasjonale prøver i Sverige i stor grad har omhandlet likeverdighet i bedømming og i karaktersetting. Bakgrunnen for innføring av en form for standardiserte prøver, var at det tidlig på 1940-tallet ble oppdaget at det var variasjon i lærernes bedømming mellom klasser og kommuner. Hensikten med nasjonale prøver var dermed å bistå lærerne med informasjon om egne elevers prestasjonsnivå, sammenlignet med andre elever. Fra starten av 1960-tallet til 1994 var karaktersystemet i Sverige normrelatert (relativt). I 1994 ble det innført en ny læreplan, og man gikk over til et målrelatert karaktersystem. De nasjonale prøvene ble beholdt, men hensikten med dem ble endret fra styrende til veiledende.

Imidlertid ble det etter et par år oppdaget at det var store forskjeller i de fastsatte karakterene både mellom lærere, mellom skoler og etter om skolen var offentlig eller privat. Det ble også påpekt at det var karakterinflasjon, og det ble stilt spørsmål ved de nasjonale prøvenes rolle som veiledende materiale for karaktersetting. Disse utfordringene ledet til en utredning i 2007. Denne kom frem til at nasjonale prøver kun burde brukes til likeverdig og rettferdig karaktersetting og som grunnlag for analyse av om kunnskapskravene på skolenivå, kommunenivå og nasjonalt nivå var nådd. I likhet med tidligere karakterer skulle de målrelaterte karakterene gi informasjon om den enkelte elevs studieresultat, samt brukes i opptak til høyere utdanning. Karakterene skulle også gi informasjon om utvikling i resultater på skole-, kommune- og nasjonalt nivå. Dermed ble mandatet for de målrelaterte karakterene mye bredere enn for de normrelaterte karakterene.

Forfatterne påpeker videre at det er mangler i de nasjonale prøvene når det gjelder å gi lærerne tilstrekkelig støtte for likeverdig karaktersetting, og at dette også gjelder etter innføringen av målrelaterte karakterer. Blant annet blir det pekt på at det er relativt stor variasjon i prøvekarakterer fra år til år, til tross for at det kun er moderate prestasjonsforskjeller på nasjonalt nivå i et fag mellom ulike år. Dette kobler forfatterne til typen prøver brukt. I Sverige handler det ofte om å skrive en tekst snarere enn å bruke mer standardiserte prøvetyper, hvor forfatterne argumenterer for at førstnevnte kommer med en rekke validitets- og reliabilitetsproblemer. Denne utfordringen er også påpekt av OECD, som heller foreslår å bruke mer flervalgsoppgaver, da bruk av disse vil kunne øke reliabiliteten til prøvene. Videre viser analyser at lærernes karaktersetting blir påvirket av prøveresultatene, ved at gjennomsnittet for slutt karakter følger gjennomsnittet for karakter på nasjonale prøver i faget. Samtidig er det konsekvente mønstre over tid som viser at 30-40 prosent av elevene i siste år på grunnskolen (9. trinn), får bedre slutt karakter enn hva de nasjonale prøvene tilsier at de skulle ha fått.

Hovedkonklusjonene i artikkelen er relatert til prøvenes to hovedformål: som støtte for likeverdig karaktersetting og som støtte for oppfølging og analyse av resultater på skolenivå, kommunenivå og nasjonalt nivå. Konklusjonen i forhold til det første hovedmålet, er at nasjonale prøver fungerte godt for å styre karaktersetting, og for å sikre likeverdighet og rettferdige bedømmelser i det normbaserte karaktersystemet. Da var også prøvene utformet for å løse nettopp denne oppgaven. Ved innføringen av det målbaserte karaktersystemet, ble prøvenes støtte til likeverdig karaktersetting et mer underordnet mål, siden man antok at dette skulle garanteres gjennom målene og kriteriene fastsatt i læreplanverket. Dermed var prøvene heller ikke utformet for å styre eller støtte karaktersetting, fordi det ikke fantes noen regler om hvordan resultater på nasjonale prøver skal kombineres med annen informasjon som læreren har om elevens resultater.

Konklusjonen relatert til det andre hovedmålet, oppfølging, er at en forutsetning for å kunne gjøre slike sammenligninger er at prøvene er sammenlignbare over tid. Siden det er validitets- og reliabilitetsproblemer i de brukte prøvetypene, er de ikke brukbare for dette formålet. I tillegg dekker ikke prøvene hele fagdomenet, da de har for begrenset omfang av oppgaver. Dermed er konklusjonen at systemet med nasjonale prøver i Sverige ikke gir tilstrekkelig grunnlag for oppfølging og evaluering, verken på lokalt eller nasjonalt nivå.

Tveit, S. og Olsen, R. V. (2018). Eksamens mange roller i sertifisering, styring og støtte av læring og undervisning i norsk grunnopplæring. *Acta Didactica*.

Land: Norge

Artikkelen tar opp hvilke roller eksamen har i norsk grunnopplæring. I et samspill mellom teori og empiri viser artikkelforfatterne hvordan vurdering defineres som en prosess for å fastsette måloppnåelse (eller standardoppnåelse). Videre skilles det mellom tre roller for vurdering: vurdering brukt til å *sertifisere*, *styre* og *støtte* læring og undervisning. Disse aspektene kommer til uttrykk i studiens datamateriale på ulikt vis:

Eksamens rolle i å *sertifisere* læring og undervisning:

- Eksamen som sluttvurdering (karaktersetting).
- Eksamensresultater som grunnlag for sertifisering og seleksjon.

Eksamens rolle i å *styre* læring og undervisning:

- Eksamensdata som grunnlag for beslutningstakeres kvalitetsutviklingstiltak.
- Eksamen som kontroll av praktiseringen av læreplaner, forskrifter og retningslinjer.

Eksamens rolle i å *støtte* læring og undervisning:

- Eksamens rolle i å støtte elevenes læring.
- Eksamens rolle i å støtte lærernes undervisning.

Artikkelforfatterne påpeker at forståelsen av eksamen som kvalitetssikring også bør ses i sammenheng med eksamens «kalibrerende rolle». Av studiens datamateriale framgår det implisitt at denne kalibreringen oppfattes å bidra til kvalitetssikring, altså at eksamenssystemet representerer en kvalitetssikring av standpunkt-karaktersettingen. Artikkelforfatterne mener det er grunn til å reise spørsmål ved validitetsgrunnlaget for sammenligning av to vidt forskjellige typer vurderingsformer med delvis ulike kriterier. Når politikere og andre policyutviklere omtaler eksamen som en «kvalitetssikring» er det derfor grunn for å spørre: *Hvordan* bidrar eksamen til kvalitetssikring? Og hva er det som kvalitetssikres?

Dersom man med kvalitetssikring mener å undersøke *kvaliteten til eksamensprøvene* per se, bør man definere hva slags prinsipper, prosedyrer og rammer for kvalitetssikring som bør anvendes. Er derimot svaret at eksamen bidrar til *kvalitetssikring av lærernes standpunkt-vurdering*, påpeker artikkelforfatterne utfordringene med en slik praksis på enkeltlærer- eller skolenivå. Dette *kan* gjøres på forsvarlige og hensiktsmessige måter om man har tilstrekkelig innsikt i de statistiske mulighetene og begrensningene ved aggregerte dataserier for små utvalg av elever. Imidlertid kan slike sammenlikninger være mer meningsfulle for data aggregert til høyere nivåer (større kommuner, fylker og nasjonalt nivå). I så fall kan man eksempelvis undersøke om man over tid observerer betydelige skjevheter mellom standpunkt- og eksamenskarakterer på tvers av fag, mellom skoler med ulike karakteristikk, mellom kjønn etc. Dersom slike skjevheter observeres, bør man drøfte om disse forskjellene er akseptable og rimelige.

Videre påpekes det at bruk av eksamen som grunnlag for kvalitetsutvikling i sektoren er i ferd med å bli utbredt, og det kan se ut som at en slik rolle tas for gitt av aktørene i systemet. Gjennom å definere dette tydeligere som et av formålene med eksamen, vil man kunne få økt bevissthet knyttet til muligheter og begrensninger ved en slik bruk.

Studien påpeker også at trekkordningen i eksamenssystemet innebærer at elever blir fordelt på ulike eksamener. I all hovedsak skal dette være basert på

tilfeldige utvalg, men lokalt må man også ta hensyn til at lærere har bestemte elevgrupper i flere fag, at de har flere grupper med eksamen på samme dag og at de kan være oppnevnt som ekstern sensor ved andre skoler de samme dagene. Dette fører til store variasjoner, hvilket elevene opplever som urettferdig.

Videre hviler trekkordningen på et prinsipp som er lite omtalt i policydokumentene, men som artikkelforfatterne mener er så institusjonalisert i vårt utdannings-system at vi tar det for gitt. Prinsippet går fram av Stortingsmelding nr. 20 (2012-2013): «Trekkordningen innebærer at elevene ikke skal ha eksamen i alle fag, men at de skal være eksamensforberedt i de fagene der eksamen er en mulig sluttvurdering ved siden av standpunktkarakteren» (s. 65-66). Gjennom dette får eksamen også en ikke-formalisert, men likefullt reell rolle: Vissheten om at det kan komme en eksamen i faget skal bidra til å holde elevene på «tå hev». Eksamen har dermed også en rolle som et ytre press som skal bidra til at elevene arbeider godt i alle fagene i skolen. Denne rollen hevder artikkelforfatterne bør undersøkes nærmere, da den er svært lite belyst i policydokumenter og forskning. For å sikre en mer rettferdig trekkordning burde det vurderes å balansere de praktiske hensynene mot noen prinsipper for trekking fastsatt i rammeverket for eksamen.

Artikkelforfatternes overordnede konklusjon er at eksamen i grunnopplæringen har en rekke ulike funksjoner som burde tydeliggjøres og defineres som formelle formål ved eksamenssystemet. Manglende avklaringer av eksamensformål kan føre til at politikere og andre policyutviklere forvalter og videreutvikler eksamenssystemet på sviktende grunnlag. Videre kan det føre til at Utdanningsdirektoratet, Fylkesmannen og andre tilsynsmyndigheter, skoleeiere, skoleledere, lærere, foreldre og elever bruker eksamensdata på uheldige måter.

Artikkelen konkluderer med at lovverk, forskrift og rammeverk for eksamen bør gjennomgås og endres til eksplisitt å uttrykke de rollene eksamen skal fylle og ikke. Slik kan man få et bedre utgangspunkt for å utvikle eksamensprøver og sikre hensiktsmessig bruk av eksamensresultater både nasjonalt og lokalt. Alternativt kan en slik gjennomgang føre til konklusjonen om at det er vanskelig, og kanskje umulig, for eksamen å skulle ha så mange roller. I så fall vil det å gjøre disse funksjonene til eksplisitte definerte formål kunne bidra til en nødvendig diskusjon om hvilke endringer som bør gjøres i eksamenssystemet for å ivareta hensyn knyttet til sertifisering, styring og støtte av læring og undervisning i norsk grunnopplæring.

Waldow (2014) Conceptions of Justice in the Examination Systems of England, Germany, and Sweden: A Look at Safeguards of Fair Procedure and Possibilities of Appeal. *Comparative Education Review*.

Land: Sverige (sammenlignet med Tyskland og England)

Artikkelen tar for seg eksamens- og sluttvurderingssystemene i videregående opplæring i Sverige, Tyskland og England, og undersøker hvordan disse systemene er lagt opp for å sikre en rettferdig sluttvurderingsprosess. Med andre ord fokuserer artikkelen på hvilke prinsipper for prosedural rettferdighet som er bygd inn i sluttvurderingssystemet i disse tre landene, med særlig fokus på hvilke aktører som er involvert i prosessen (f.eks. elevenes faglærer versus en ekstern sensor) og muligheten for ytre kvalitetssikring (f.eks. gjennom nasjonale prøver eller elevenes klagemuligheter).

Studien viser hvordan tre land som i utgangspunktet oppfatter seg selv som meritokratiske, har ulike forståelser av hva som gjør sluttvurderingen rettferdig. Dette påvirker igjen hvilke mekanismer de har bygd inn i sluttvurderingsprosessen for å oppnå dette. Forskjellene kan, ifølge forfatteren, kun forstås dersom man ser sluttvurderingssystemene i sammenheng med den øvrige sosiale og historiske konteksten i de tre landene.

Når det gjelder hvilke aktører som ansees som sentrale for å sikre en rettferdig sluttvurderingsprosess, inntar læreren en sentral (men ikke eksklusiv) rolle både i Sverige og Tyskland. I Tyskland anses lærerens profesjonalitet som den primære faktoren for å sikre en likeverdig vurdering, selv om denne i økende grad suppleres med sentraliserte instrumenter for vurdering. Lærerens profesjonalitet står også sentralt i Sverige, men her kvalitetssikres sluttvurderingsprosessen gjennom de nasjonale prøvene.

Mens det i Tyskland og Sverige dermed er et fokus på å sikre kvaliteten på sluttvurderingsprosessen, har man i England et større fokus på å sikre kvaliteten på sluttvurderingsresultatet. Dette gjøres nesten utelukkende av eksterne eksamenstyver ('examination boards'). En slik ekstern vurdering av elevenes prestasjoner ansees her som sentralt for å sikre en rettferdig vurdering av elevene på likt grunnlag. Dette står i sterk kontrast til systemet i både Sverige og Tyskland, hvor en rettferdig vurdering sikres gjennom at elevenes faglærer kjenner hver enkelt elev og dermed kan gjennomføre en holistisk vurdering av deres prestasjoner og læring.

Ifølge studien vil de ulike landenes forståelse av hva som regnes som en rettferdig vurdering også påvirke hvilke mekanismer for kvalitetssikring de anser som sentrale. I Sverige legges det til grunn at de nasjonale prøvene er en tilstrekkelig «safe guard» for å sikre et likeverdig og rettferdig sluttvurderingsresultat. Det svenske utdanningssystemet åpner derfor ikke opp for muligheten til å klage på den innholdsmessige vurderingen som er gjort. I England har bruk av eksternt

rettet eksamen den samme «safe guard»-funksjonen, mens Tyskland legger vekt på muligheten til å klage.

Oppsummering

Vi har inkludert tre studier av høy relevans – én fra Norge og to rettet mot den svenske konteksten – som tar for seg det overordnede systemperspektivet på eksamen og sluttvurderinger i de nordiske landene.

Gustavsson og Ericksson (2018) studerer hvilke endringer som er gjort med de nasjonale prøvene i Sverige de siste tiårene. De konkluderer med at nasjonale prøver fungerte godt for å styre karaktersetting og for å sikre likeverdighet og rettferdighet i bedømmelser da karaktersystemet var normbasert. Etter innføring av et kriteriebasert karaktersystem i 1996, samt validitets- og reliabilitetsproblemer assosiert med dagens prøver, argumenterer forfatterne for at det nåværende systemet med nasjonale prøver i Sverige ikke lenger gir tilstrekkelig grunnlag for oppfølging og evaluering, verken lokalt eller nasjonalt.

Tveit og Olsen (2018) undersøker eksamens rolle i norsk grunnopplæring. De viser til at eksamen har en rekke funksjoner, som burde tydeliggjøres og defineres som formelle formål ved eksamenssystemet i Norge. Manglende klargjøring av eksamens formål kan føre til at forvaltningen og videreutviklingen av eksamenssystemet gjøres på sviktende grunnlag, eller at eksamensdata brukes på en uheldig måte. Artikkelforfatterne argumenterer for en gjennomgang av lovverk, forskrift og rammeverk for eksamen. Videre peker de på at eventuelle endringer eksplisitt bør uttrykke de funksjonene eksamen skal fylle og ikke. En alternativ konklusjon er at det er vanskelig, og kanskje umulig, for eksamen å skulle ha så mange roller. I så fall vil det å gjøre eksamens roller til eksplisitte definerte formål kunne bidra til en diskusjon om hvilke endringer som bør gjøres i eksamenssystemet for å ivareta ulike hensyn knyttet til sertifisering, styring og støtte av læring og undervisning i norsk grunnopplæring.

Waldow (2014) undersøker ulike forståelser av rettferdighet i sluttvurderings-systemene i Sverige, Tyskland og England. Artikkelen viser hvordan lærerens profesjonalitet, koblet med visse eksterne kvalitetssikringsmekanismer (hhv. nasjonale prøver, og muligheten til å klage på karakter) ansees som sentralt for å sikre rettferdig vurdering i henholdsvis Sverige og Tyskland. I England vektlegges i stedet viktigheten av likeverdig vurdering på tvers av elever og klasser, noe som sikres gjennom bruk av eksterne eksamensstyrer.

5.2.2 Studier knyttet til institusjonelle faktorer og organisering av eksamen og sluttvurdering

Den andre gruppen av inkluderte forskningsbidrag i vår kartlegging omfatter bidrag som tar for seg institusjonelle faktorer og organisering av eksamen og sluttvurderinger i de nordiske landene. Dette meso-perspektivet kan for eksempel dreie seg om trekkordningen for slutteksamen og hvordan karaktersetting er ulik mellom fag. Følgende ni studier er presentert her: Bensnes (2019), Björnsson og Skar (2021), Bukh og kollegaer (2022), Falch, Nyhus og Strøm (2014), Hvidman og Sievertsen (2021), Hovdhaugen og kolleger (2014), Kupiainen, Marjanen og Hautamäki (2016), Prøitz (2013) og Sjaastad, Carlsten og Wollscheid (2016).

Bensnes, S. S. (2019). Scheduled to gain: Short- and longer-run educational effects of examination scheduling. *The Scandinavian Journal of Economics*.

Land: Norge

Denne studien tar for seg eksamenssystemet i Norge, og den underliggende antagelsen om at eksamenskarakterer er en pålitelig indikasjon på elevers ferdigheter. Artikkelforfatteren påpeker at flere nyere studier har sett på hvordan ulike tilfeldige og forstyrrende faktorer på eksamen (slik som luftforurensning og pollen-spredning på eksamensdagen), kan ha relativt store effekter på elevers eksamensresultater. Også institusjonelle faktorer uten sammenheng med elevers kognitive ferdigheter kan påvirke eksamenssituasjonen og elevers eksamensresultater. Studien tar for seg en slik institusjonell faktor: eksamenstimeplanen. Eksamenstimeplanen kan anses som et institusjonelt rammeverk rundt eksamen, der elever på en gitt dato hvert år får vite når deres skriftlige slutteksamen blir holdt og hvilke fag de vil bli eksaminert i. Hvilke fag elevene blir trukket ut til er tilfeldig og vil variere fra elev til elev ettersom de kan ha ulike sammensetninger av fag og fagfordypninger. Denne tilfeldige uttrekkingen fører dermed også til variasjoner i forberedelsestiden elevene får til eksamen.

Formålet med studien er å estimere hvilken effekt tilfeldige variasjoner i eksamenstimeplanen har på elevers eksamensresultater, for deretter å følge denne effekten til langsiktige resultater. Studien benytter kvantitativ metode og statistiske analyser knyttet til eksamenskarakterer for videregående elever i Norge og dekker eksamener gjennomført i perioden 2008-2012. Eksamensdataene er koblet til de årlige listene for eksamensdatoer fra Utdanningsdirektoratet.

Studien viser at videregåendeelevers eksamensprestasjon påvirkes av eksamenstimeplanen. Når elever blir tilfeldig tildelt en eksamensperiode som gir dem lengre forberedelsestid før eksamen, har dette en positiv effekt på eksamensresultatene. Resultatene fra studien antyder at en økning i forberedelsestid fra 5-8

dager til 9-12 dager kan bedre karakterresultatene med 5,3 prosent (av et standardavvik). Effekten er sterkere i naturfag og matematikk, samt sterkere for jenter enn for gutter. Resultatene indikerer at elever kun bruker et begrenset antall dager – mellom én og to uker – til å forberede seg på eksamen, uavhengig av hvor mye forberedelsestid elevene er gitt. Dersom elevene er tildelt mindre forberedelsestid enn dette vil de skåre dårligere på eksamen.

Studien ser også på mer langsiktige konsekvenser av forberedelsestid til eksamen i videregående skole ved å se på resultater i høyere utdanning. Her viser studien at variasjoner i eksamensresultat som følge av eksamenstidplanen er store nok til å også påvirke elevers senere utdanningsmuligheter og prestasjoner.

Samlet sett viser studien hvordan enkelte eksamensresultater blir påvirket av en tilfeldig institusjonell faktor som ikke er relatert til elevers dyktighet eller kognitive ferdigheter. Videre peker artikkelforfatteren på to momenter som følger av disse funnene: 1) Hvor nøyaktig eksamen reflekterer en elevs ferdigheter vil delvis avhenge av hvor tilbøyelige eksamenskarakterene er til å bli påvirket av tilfeldige variasjoner i eksamenssituasjonen; og 2) Jo mer tilbøyelige for påvirkning de er, desto mindre nyttig er eksamen som en indikasjon på egnethet for høyere utdanning. Funnene i artikkelen kan bidra til å skape bevissthet om hvordan slutteksamen bør gjennomføres og hvor mye elevers eksamensprestasjoner bør vektlegges sammenlignet med andre prestasjonsmål.

Björnsson, J.K., Skar, G. B. (2021). Sensorreliabilitet på skriftlig eksamen i videregående opplæring. Estimer av sensorreliabilitet basert på analyser av karakterforslag fra to uavhengige sensorer før fellessensur - Resultater fra 40 utvalgte fag. *Rapport, ILS/ Universitetet i Oslo (UiO)*.

Land: Norge

Denne rapporten undersøker sensorreliabilitet, ved å sammenligne karakterforslagene til de to sensorene. Studien baserer seg på vurderinger av over 700 000 elevbesvarelser i perioden 2015-2019, alle gjennomført av to eksterne sensorer.

Sensorreliabilitet (interrater reliability) innebærer at vurderingene fra to eller flere sensorer samsvarer. Forfatterne påpeker at det i dag ikke er mulig å analysere sensorreliabilitet på endelige eksamenskarakterer, fordi den endelige karakteren må være lik fra begge sensorer. De brukte derfor foreløpige karakterer til å foreta disse analysene, da disse viser sensorenes initiale karaktersetting, før diskusjon med den andre sensoren. Rapporten bruker klassiske statistiske metoder som beregninger av kappa, vektet kappa og intraklasse-koeffisient for å evaluere sensorreliabiliteten.

Funnene viser at vurderingene som er gjenspeilet i karakterene i noen fag preges av veldig god sensorreliabilitet, mens at dette ikke er tilfelle i andre fag. I

fagene norsk-, fremmedspråk og samfunnsfag ble det påvist en stabilt lav sensorreliabilitet over tid, mens for matematikk og realfagene var sensorreliabiliteten høy. For andre fag ble det påvist noen svingninger over tid. I sistnevnte kan eksamenskarakteren påvirkes av hvilke sensorer som har vurdert besvarelsen, noe som betyr at den satte karakteren ikke kun er et uttrykk for den kompetansen kandidaten har. Her ble det også påvist forskjeller innad i de ulike fagene.

For å undersøke dette videre gjennomførte forfatterne en såkalt MFRM-analyse («Many-Facet Rasch Model»). Denne metoden innebar at elever og sensorer ble koblet til hverandre på et utvalgt antall eksamener i fagene der det var påvist lav reliabilitet. Analysen viste at eksamen generelt er bedre på å skille mellom sensorers streghet enn mellom kandidatenes kompetanse. Videre viste analysene også at det ikke er statistisk grunnlag for å skille mellom seks nivåer av elevene kompetanse innenfor det gitte faget, slik dagen karaktersystem fungerer.

Generelt konkluderer rapporten at det foreligger store variasjoner i sensorreliabilitet i eksamener i norsk videregående opplæring. Disse forskjellene har ulike årsaker i ulike fag. Rapporten anbefaler derfor å utforske dette nærmere for hvert fag, slik at en kan iverksette passende tiltak for å styrke reliabiliteten.

Bukh, P. N., Christensen, K. S. og Poulsen, M. L. (2022). Performance funding: Exam results, stakes, and washback in Danish schools. *Sage Open*.

Land: Danmark

Denne studien ser på prestasjonsbasert skolefinansiering som tar utgangspunkt i elevers eksamensresultater i dansk grunnskole, og knytter dette opp mot såkalte washback-effekter og utilsiktede konsekvenser. Tidligere forskning har vist at viktige prøver og eksamen (såkalte high-stakes prøver) har washback-effekter som påvirker lærere og læring. Det er også vist at slike prøver kan ha utilsiktede konsekvenser, i form av negative washback-effekter. Når skolefinansiering knyttes til prøveresultater, blir nettopp eksamenskarakterene særlig viktige, ettersom eksamensresultatene får betydelige konsekvenser for skolens økonomiske situasjon.

Formålet til studien er å *studere prestasjonsbasert skolefinansiering som tar utgangspunkt i elevers eksamensresultater*. Studien har en kvalitativ metodetilnærming, og ser på rektorers oppfattelse og tolkning av hvordan prestasjonsbasert finansiering påvirker innsatsen og washback-effekten knyttet til viktige eksamener, gjennom bruk av intervju og dokumentstudier.

Studien bidrar til den eksisterende forskningslitteraturen ved å demonstrere at washback-effekter og utilsiktede konsekvenser kan oppstå innen grunnskoleutdanningen når prestasjonsfinansiering innføres. Blant annet indikerer studien at implementeringen av prestasjonsfinansiering kan øke innsatsen og ha washback-effekt, men at omfanget av dette delvis avhenger av valg rektoren tar. Rektorene

som ble intervjuet hadde forutinntatte og negative meninger om prestasjonsfinansiering, og de så på prestasjonsbasert finansing med skepsis. Likevel fokuserte de på mulighetene og endret sin praksis for å kunne nå målene.

Studien illustrerer også flere initiativer som kunne tolkes som «teaching to the test» (dvs. å lære med mål å gjøre det best mulig på prøven), innsnevring av pensum og vektlegging av bestemte og marginaliserte studentgrupper. Selv om dette er mulige negative washback-effekter, ble disse utilsiktede konsekvensene rasjonalisert som positive av rektorene og ofte ansett som generelle initiativ som var positive for elevene. På bakgrunn av dette argumenterer artikkelforfatterne for at mulige utilsiktede konsekvenser bør vurderes nøye før nye finansieringsmekanismer introduseres.

Falch, T., Nyhus, O. H. og Strøm, B. (2014). Causal effects of mathematics. *Labour Economics*.

Land: Norge

Studien tar utgangspunkt i hvordan 16-åringer i grunnskolen i Norge blir tilfeldig fordelt til en obligatorisk avsluttende eksamen i enten matematikk eller språk (engelsk eller norsk). Ettersom elevene blir varslet om eksamensfag og dato kort tid i forveien, har de en intensiv forberedelsestid i forkant av den avsluttende eksamen. På bakgrunn av dette er formålet med studien å *estimere den kausale effekten av øving og forberedelse umiddelbart etterfulgt av en avsluttende eksamen i matematikk, sett i forhold til eksamen i språk (engelsk eller norsk)*.

Studien tar i bruk kvantitativ metode og statistisk analyse, med utgangspunkt i registerdata fra Statistisk sentralbyrå (SSB). Dataene dekker alle elever som fullførte obligatorisk utdanning i perioden 2002-2004.

Funn fra studien viser at en intensiv forberedelsesperiode på 2-5 dager i matematikk sammenlignet med språk har spesielt tre implikasjoner: i) det øker sannsynligheten for å bestå videregående skole, ii) det øker sannsynligheten for opptak til høyere utdanning, og iii) det øker sjansen for å ta naturvitenskapelige og teknologiske utdanningsprogram i høyere utdanning. Fem dager med intensiv øving er estimert til å øke disse sannsynlighetene med henholdsvis ca. 1.0, 0.8 og 0.5 prosentpoeng. Samlet sett ser den kausale effekten ut til å være noe sterkere for menn enn for kvinner, men analysen indikerer at kjønnsforskjeller har en kompleks påvirkning med hensyn til elevenes tidligere matematikkferdigheter.

Videre tyder studien også på at matematikkferdigheter predikerer utdannings-suksess bedre enn språkferdigheter. Artikkelforfatterne påpeker at en implikasjon for politikktutforming ut ifra disse funnene er at alle elever bør ha slutteksamen i matematikk. Argumentet er at hvis alle elever på forhånd visste at de kom til å ha slutteksamen i matematikk, ville de mest sannsynlig forberede seg til denne eksamen gjennom hele skoleåret. Effekten av en slik institusjonell endring er derimot

ikke mulig å forutsi på bakgrunn av denne studien, ettersom elevenes innsats over en lengre periode mest sannsynlig vil være annerledes enn situasjonen med en kort og intensiv forberedelsesperiode rett før slutteksamen. Likevel indikerer denne studien at en intensiv forberedelsesperiode i forkant av en matematikk-eksamen er fordelaktig.

Hvidman, U. og Sievertsen, H. H. (2021). High-stakes grades and student behaviour. *Journal of Human Resources*.

Land: Danmark

Studien tar for seg slutteksamen og sluttkarakterer for elever i videregående skole i Danmark. Karakterer gir viktige indikasjoner på elevers sannsynlighet for å uteksamineres, det påvirker individets utdanningsløp (slik som muligheten for å fortsette med høyere utdanning), samt elevenes langsiktige arbeidsmarkedsutfall. På bakgrunn av dette ønsker artikkelforfatterne å undersøke hvordan elever responderer på forandringer i sitt karaktersnitt. Elever bruker karaktersnittet som et signal om sine egne evner, men elever blir også påvirket av at snittet brukes av universiteter i opptaksprosedyren. Formålet med studien er å *undersøke om insentivene som er forbundet med eksamenskarakterer, påvirker elevers investering i utdanningen*.

Studien identifiserer effekten av sluttkarakterer på elevers adferd, ved å utnytte registrering av danske videregåendeelevers karaktersnitt etter en ny reform i 2007. Reformen innebar implementering av et nytt karaktersystem i Danmark, hvilket førte til at mange elever fikk endret sitt karaktergjennomsnitt (i artikkelen omtalt som «*karaktersjokket*»). Studien tar i bruk kvantitativ metode og statistisk analyse, der det er tatt i bruk administrative data fra Danmarks Statistikk. Data-grunnlaget inkluderer alle elever som fullførte treårig videregående utdanning i 2008.

Reformen innebar at eksisterende karakterer ble omregnet i tråd med den nye karakterskalaen. Siden den nye skalaen hadde færre karakterkategorier kunne elever bli satt både opp og ned i forhold til tidligere karakter. Funn fra studien viser hvordan danske videregåendeelever responderte på endringer i sitt karaktersnitt som følge av implementeringen av et nytt karaktersystem. Elever som ble satt ned i karakter i førsteåret på grunn av den nye reformen, hadde en bedre prestasjon i deres andre og tredje år i videregående skole. For jenter i førsteåret på videregående, førte karaktersjokket til karakterforbedring i deres andre og tredje skoleår. For gutter derimot økte ikke karakterene gjennom videregående nok til å veie opp for karaktersjokket. Karakterforbedring i videregående skole gjenspeiles videre i økt sannsynlighet for å begynne på universitetet. Funnene indikerer at elever justerer sin innsats og sine utdanningsvalg i lys av endringer i deres

karactersnitt, selv i tilfeller der endringen ikke kan relateres til deres egne prestasjoner. Dette underbygger argumentet om at insentiver er viktige for elevers investering i utdanning.

Artikkelforfatterne påpeker videre at resultatene ser ut til å være relevante både i en dansk kontekst, men også for andre land. Selv om det danske utdanningssystemet i noen grad er forskjellig fra utdanningssystemer i andre europeiske land og USA, så har alle til felles at sluttvurderingen i videregående skole er viktig for opptak til høyere utdanning. Funnene i studien peker på hvordan opptak til universitetet kan utgjøre et insentiv for elever for å arbeide hardere. Mens tidligere litteratur hovedsakelig har vektlagt økonomiske insentiver for å motivere elever til innsats, indikerer denne studien at insentiver relatert til opptak ved universitet også er kan være viktig.

Hovdhaugen, E., Seland, I., Lødding, B., Prøitz, T. og 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. *Rapport, NIFU.*

Land: Norge

Denne rapporten dreier seg om avvik mellom eksamens- og standpunkt karakterer og omtaler også rutiner og praksis for standpunkt karakter setting. En sentral problemstilling er om grad av avvik mellom eksamens- og standpunkt karakter kan være relatert til om skolen er offentlig eller privat. Rapporten bygger på kvantitative data og kvalitative intervjuer med lærere og elever. Oppsummert påvises det et litt større avvik mellom standpunkt- og eksamens karakterer ved noen av de private skolene, også etter kontroll for andre forhold med betydning for karakterer. Videre viser rapporten at elevenes standpunkt karakterer har størst påvirkning på hvor stort avviket mellom de to karakterene er. Elever med gode karakterer har større sannsynlighet for å gå ned i karakter på eksamen sammenlignet med elever med svakere karakterer. Dette henger sammen med at standpunkt karakterer i gjennomsnittet ligger over eksamens karakterer uansett skoletype.

Når det gjelder retningslinjer og rutiner som grunnlag for lærernes standpunkt vurdering legges det et mangfold til grunn når lærerne arbeider med vurdering og karakter setting. Eksempler på formelle retningslinjer er lærernes fortolkning og bruk av vurderingsforskriften, eksamensforberedende prøver (heldagsprøver) som inkluderer eksamensveiledningene, erfaringer med elevsamtalene som pålagt av loven og erfaringer med sensorarbeid i tillegg til skoleinterne prosjekter rettet mot praksis i vurdering. Funnene viser at læreren på de skolene som ble

besøkt forhold seg likt til de formelle rutinene og praksisene for vurdering og karaktersetting.

Et viktig spørsmål – også med tanke på det nåværende prosjektet – er i hvilken grad eksamenskarakteren faktisk kan og skal brukes for å kalibrere standpunkt-karakteren. Først og fremst forutsetter en slik kalibrering at eksamenskarakterene er stabile over tid, noe de ikke er. Eksamen oppleves som en smalere testing av elevenes kompetanse enn standpunktkarakteren, selv om begge former skal vurdere eleven ut fra kompetansemålene.

Lærere med sensorerfaring opplever de to vurderingssituasjonene som ulike. Når de setter stand-punktkarakterer påvirkes lærernes vurdering av kjennskap til eleven og dens faglige utvikling med positivt bias, mens i en eksamenssituasjon vil de vurderer den anonyme elevbesvarelsen kun utfra vurderingskriteriene.

Kupiainen, S., Marjanen, J. og Hautamäki, J. (2016). The problem posed by exam choice on the comparability of results in the Finnish matriculation examination. *Journal of Educational Research Online*.

Land: Finland

Artikkelen diskuterer reformer i finsk videregående utdanning, som i dag gjør at elever har større valgmuligheter i hvilke eksamener de skal inkludere i sin finske avgangseksamen fra studieforberevende program i videregående skole, den såkalte finske studenteksamen (matrikulasjonseksamen). Det er denne eksamen som gir de som har fullført studieforberevende videregående opplæring rett til å søke opptak i høyere utdanning.

Hovedmodellen for opptak i høyere utdanning er at avgangselever søker basert på gjennomsnittskarakteren de oppnår på studenteksamen, ikke karaktersnittet de har oppnådd i skolen. I det finske systemet må avgangselevne bestå minst fire eksamener, der morsmåls eksamen (finsk/svensk) er obligatorisk. I tillegg kan elevene velge blant følgende fagområder for de andre tre eksamener de skal ta: matematikk, det andre nasjonale språket (finsk for svensktalende, svensk for finsktalende), fremmedspråk og generelle fag (realfag/humaniora/ samfunnsfag). I den siste gruppen er det nå 10 ulike fag å velge mellom. For å sikre at studenteksamen er sammenlignbart over tid settes karakterer utfra en normalfordelingsdistribusjon, gauss-kurven, som anvendes på alle som tar eksamen ved det eksamenstilfellet.

Reformer i løpet av 2000-tallet har endret sammensetningen av eksamener elevene tar til sin studenteksamen. Først økte antallet valgfrie eksamener ved at nasjonalt andrespråk (svensk for finsktalende og finsk for svensktalende) ikke lenger var en obligatorisk eksamen. Deretter økte antallet valgfrie eksamener ved at en felles eksamen i generelle fag ble oppdelt i ti ulike eksamener, avhengig av om elevene hadde tatt kurs i det enkelte faget.

Analysen er kvantitativ og tar utgangspunkt i data fra studenteksamen i perioden 2006-2009, samt våren 2012. I tillegg vurderes data fra eksamen sammenholdt med karakterer for nærmere 2000 elever ved 35 skoler innsamlet våren 2012. Artikkelen analyserer tre forskningsspørsmål:

Får studenter som tar ulike eksamener ulikt karakternivå i gjennomsnitt, noe som kan sette sammenlignbarheten i fare?

Er det systematiske forskjeller i valg av hvilke eksamener elevene skal gå opp til i perioden 2006-2012, noe som indikerer at elever velger strategisk hvilke eksamener de tar?

Er det systematiske forskjeller mellom skoler i relasjonen mellom karakter i et fag og karakter på tilsvarende studenteksamen i faget?

Analysene av forskjeller i karakternivå mellom ulike eksamener peker på at det er dårligst gjennomsnittskarakter i fag som anses som vanskelige, slik som avansert matematikk, fysikk, kjemi og det nasjonale andrespråket. En sammenligning av gjennomsnittskarakter på de andre fagene i studenteksamen for elever som har tatt grunnleggende og avansert matematikk viser at elever som har tatt avansert matematikk gjennomgående har høyere snittkarakter. Begge disse funnene tyder på at hvor vanskelig det er å få god karakter avhenger av hvor gode de andre elevene som tar samme eksamen er.

Analysen av hvilke eksamener elevene velger å ta, peker på at det over tid har vært endring i hvilke eksamener elever velger seg, og at eksamener som anses som «enklere» å få gode karakterer på blir mer populære over tid. Korrelasjonen mellom gjennomsnittskarakter på studenteksamen i et fag og lærersatt karakter i det enkelte fag er relativt høy, 0,79, men det er variasjon mellom fag. Derimot ser det ikke ut til å være systematisk variasjon mellom skoler.

Artikkelen konkluderer med at mens det tidligere var en ganske homogen eksamen, der to eksamener var valgfrie og de andre to ble valgt blant et avgrenset antall fag, er det nå mange valg involvert i sammensetning av finsk studenteksamen. Dette gjør at gjennomsnittet ikke alltid er sammenlignbart på tvers av elever som har tatt ulike kurs. At elevene som tar den enkelte eksamen kun sammenlignes med andre elever som har tatt samme eksamen, gjør at sammenligningsgrunnlaget blir forskjellig for hver eksamen. Dermed bidrar bruken av gauss-kurven i karaktersetningen for den enkelte eksamen til at sammenlignbarheten mellom fag blir dårligere. Grunnen til dette er at elever med ulike akademiske evner velger ulike kurs, og dermed kan ikke disse sammenlignes med hverandre. Siden dette også er informasjon som elevene selv besitter, ser man en økning over tid i populariteten til eksamener hvor elevene oppfatter at det er ganske enkelt å få en god karakter. Med andre ord leder dette til en strategisk tilpasning blant elevene.

Prøitz, T. S. (2013). Variations in grading practice – subjects matter. *Education Inquiry*.

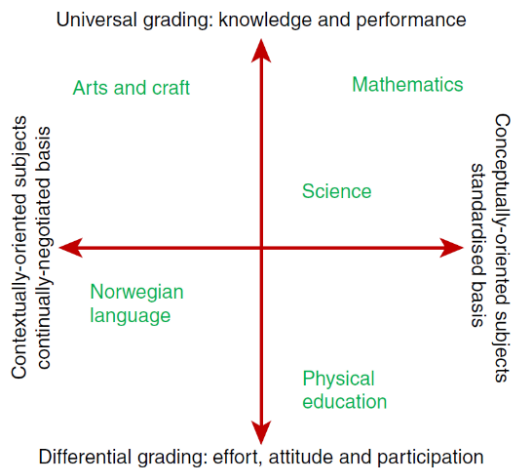
Land: Norge

Denne studien ser på *relevansen og viktigheten av faget i læreres karaktersetting, slik lærere selv oppfatter det*. Studien er basert på kvalitativt materiale fra 41 semi-strukturerte intervjuer med lærere i fem fag, fra fire ungdomskoler og to videregående skoler i Norge.

Studien viser at de særegne kjennetegnene til ulike skolefag betyr noe for hva som brukes som evidens og hvilke elementer som vektlegges av lærere ved karaktersetting. Dette innebærer at ulike fag har ulike utfordringer med å oppfylle anbefalingene for karaktersetting.

Eksamenssystemet, med nasjonalt utviklede eksamen og eksterne sensorer, er ment å sørge for en kalibrering av læreres karaktersetting. Imidlertid er det ikke alle fag som har eksamen. Systemet dekker kun noen av de «viktige» og mer teoretiske kjernefagene (som norsk og matematikk), og ikke mer praktiske «tilleggsfag» (som kunst og håndverk og kroppsøving). Artikkelforfatterne påpeker også at de fleste karakterene på vitnemålet fra ungdomsskolen og videregående skole, er satt av elevenes egne lærere. Utdanningssystemet er dermed avhengig av at lærere er i stand til å gi rettfærdige vurderinger av elevenes prestasjoner. Ettersom karakterer har viktige konsekvenser for elevens videre utdanningsmuligheter (blant annet opptak til høyere utdanningsnivåer), er det et sentralt prinsipp at karakterer skal være sammenlignbare, pålitelige og valide.

Blant annet viser studien at lærere i mer konseptuelle fag (som matematikk, naturfag og kroppsøving), ser ut til å behandle karaktergrunnlaget på en mer standardisert måte. Lærere i fag som er mer kontekstuelle (som norsk og kunst og håndverk), behandler karaktergrunnlaget som mer åpent for kontinuerlig forhandling. Dette reflekterer to ulike måter å forstå rettfærdig karaktersetting på: universell karaktersetting (som brukt i matematikk og naturvitenskap), og karaktersetting tilpasset den enkelte elev (som brukt i norskfaget og kunst og håndverk). I studien vises det hvordan de to grupperingene danner to kontinuum i en modell, som fanger variasjonen i både behandlingen av karaktergrunnlaget og tilnærmingen til rettfærdighet i karaktersetting.



Figur 5.1 Behandlingen av vurderingsmaterialet og rettferdighet i de fem fagene

Studien viser at lærerne i fagene tar hensyn til elementer ved elevenes deltakelse og innsats når de setter karakterer, men at dette gjøres i ulik grad mellom enkeltlærere og mellom fag. Lærere anser deltakelse og innsats som et viktig element, men et som ikke tydelig nok anerkjennes i de satte vurderingskriteriene og forskriften. Dette illustrerer en uløselig konflikt som læreren står i, mellom undervisnings- og læringsprosjektet som skjer i timen, og forventningen om at elevens kompetanse skal anses som et isolert fenomen.

Funnene indikerer også at ulike epistemologiske kulturer i ulike fag, kan føre til ulikheter i validiteten og reliabiliteten til lærernes vurderinger. Funnene indikerer også generelle svakheter i validiteten (og reliabiliteten) ved karakterer, i tilfeller der de blir brukt som nasjonale og internasjonale indikatorer for læringsutbytte. For eksempel anså informantene det begrensede rammeverket for matematikk og naturvitenskap som viktig for å sørge for rettferdig karaktersetting.

Spesielt i matematikk kan det begrensede karaktergrunnlaget, som består av resultater fra lærerlagde prøver og resultatene fra tentamen, føre til underrepresentasjon og risiko for å sette karakterer på et snevert grunnlag. En mer åpen karaktersetting (som for eksempel i kroppsøvningsfaget) kan fange opp aspekter ved elevene som styrker validiteten til lærerens slutninger, men som på samme tid kan introdusere en risiko for irrelevans da lærere kan ta irrelevante prestasjons- eller adferdsaspekter med i beregningen når de setter karakterer.

Sjaastad, J.; Carlsten, T.C.; Wollscheid, S. (2016): Får elevene den opplæringen de har krav på? Kartlegging av undervisningstimer med kvalifiserte lærere i videregående opplæring. *Rapport*, NIFU.

Land: Norge

Problemstillingen for denne rapporten er om elevene i videregående opplæring i Norge får den opplæring de har krav på, målt i årstimetall og undervisning med kvalifisert lærer. For å besvare problemstillingen kombinerte studien data fra spørreundersøkelser til skoleeiere, skoleledere og lærere med data fra skolebesøk der elever, lærere, tillitsvalgte og skoleledere ble intervjuet.

Når det gjelder organisering av eksamensperioden viser studien at den enkelte skole planlegger å nå kompetansemålene før eksamen, gjennom å øke timetallet per uke, gjerne med fleksible økter til ulike fag. Samtidig vises det at eksamensperioden er den enkeltårsaken som gjør at flest undervisningstimer forsvinner. Lærere og skoleledere peker på utfordringen som tilfeldig trekking til eksamen gir for planlegging av opplæring i eksamensperioden. Mens noen klasser og utvalgte elever blir trukket ut til eksamen, mangler andre klasser og elever lærer fordi vedkommende er opptatt med eksamensrelaterte oppgaver. Dessuten benyttes klasserom til gjennomføring av eksamen. Samtidig pekes det på utfordringen at undervisningen skal fortsette også etter eksamensperioden, til tross for at kompetansemålene er nådd, eksamen er gjennomført og mange elever derfor har lav motivasjon.

Generelt viser funnene at omtrent halvparten av lærere og skoleledere var «enig» eller «svært enig» i at eksamensperioden burde endres. Mens lærere i matematikk, naturfag og engelsk var spesielt kritiske, viste lærere i ulike yrkesfag, som helsefag, byggfag og elektrofag et noe mindre endringsbehov.

Oppsummering

I vår systematiske kartlegging har vi inkludert ni forskningsbidrag som er knyttet til institusjonelle faktorer, herunder seks bidrag fra Norge, to fra Danmark og ett bidrag fra Finland.

Bensnes (2019) utforsker den underliggende antagelsen om at eksamens karakterer er en pålitelig indikasjon for elevers ferdigheter. Konkret ønsker forfatteren å estimere hvilken effekt tilfeldige variasjoner i eksamenstimeplanen har på elevers eksamensresultater på kort og lengre sikt. Eksamensprestasjonen til videregående elever viser seg til å være tilbøyelig til å bli påvirket av eksamenstimeplanen. Når elever blir tilfeldig tildelt en eksamensperiode som gir dem lengre forberedelsestid før eksamen, har dette en positiv effekt på eksamensresultatene. To potensielle slutninger følger av dette funnet: For det første, hvor nøyaktig

eksamen reflekterer elevers ferdigheter, vil delvis avhenge av hvor tilbøyelige eksamenskarakterene er til å bli påvirket av tilfeldige variasjoner i eksamenssituasjonen. For det andre, jo mer tilbøyelige til å bli påvirket de er, desto mindre nyttig er eksamen som et rangeringsverktøy. Artikkelforfatteren påpeker at funnene kan bidra til å informere hvordan slutteksamen bør utføres og hvor mye man skal vektlegge elevers eksamensprestasjoner sett i forhold til andre prestasjonsmål.

En studie av Björnsson og Skar (2021) analyserte sensorreliabilitet basert på de foreløpige karakterene på skriftlig eksamen i videregående opplæring satt av de to sensorene som foretok den eksterne sensureringen. Konklusjonen er at det er store variasjoner i sensorreliabilitet ved norske eksamener. Disse forskjellene har ulike årsaker i ulike fag, og det anbefales å utforske dette nærmere for hvert fag, slik at en kan iverksette passende tiltak.

Bukh og kollegaer (2022) studerte prestasjonsbasert skolefinansiering med utgangspunkt i elevers eksamensresultater. De demonstrerte at washback-effekter og utilsiktede konsekvenser kan oppstå innen grunnskoleutdanningen når prestasjonsfinansiering innføres. Blant annet indikerer studien at implementeringen av prestasjonsfinansiering har potensial til å øke innsatsen og fremme washback-effekter, men at dette delvis avhenger av rektors valg. Studien illustrerer også flere initiativer som kunne tolkes som «teaching to the test», innsnevring av pensum og vektlegging av bestemte studentgrupper. Selv om dette er mulige negative washback-effekter, ble disse utilsiktede konsekvensene rasjonalisert som positive av rektorene og ofte ansett som generelle initiativ som var til fordel for elevene.

Studien til Falch og kollegaer (2014) tar utgangspunkt i hvordan 16-åringer i grunnskolen i Norge blir tilfeldig fordelt til en obligatorisk avsluttende eksamen i enten matematikk eller språk (engelsk eller norsk). Det påvises at en intensiv forberedelsesperiode på 2-5 dager i matematikk i stedet for språk har viktige implikasjoner: det øker sannsynligheten for å fullføre videregående skole og det gjør det mer sannsynlig at elevene blir tatt opp til høyere utdanning, særlig innen naturvitenskapelige og teknologiske utdanningsprogram. Studien indikerer også at matematikkferdigheter er et viktigere mål på utdanningssuksess enn språkferdigheter. Artikkelforfatterne påpeker at en implikasjon for politikktutforming ut ifra disse funnene, er at alle elever bør gjennomføre en slutteksamen i matematikk.

Hvidman og Sievertsen (2021) undersøkte i en dansk kontekst om insentivene forbundet med eksamenskarakterer, påvirker elevers investering i utdanningen. Studien viser at elever tilpasser sin innsats og sine utdanningsvalg i tråd med endringer i deres karaktersnitt, selv i tilfeller hvor endringen ikke kan relateres til deres tidligere prestasjoner. Dette underbygger argumentet om at insentiver er viktige for å forstå elevers investering i sin utdanning. Artikkelforfatterne påpeker videre at resultatene ser ut til å være relevant både for en dansk kontekst, men også for utdanningssystemer i andre land.

Hovdhaugen og kollegaer (2014) studerer avvik mellom eksamens- og standpunkt karakterer i private og offentlige videregående skoler. Studien omfatter også læreres rutiner og praksis for standpunkt karaktersetting. Oppsummert finner de et litt større avvik mellom standpunkt- og eksamens karakterer ved noen av de private skolene, også etter kontroll for andre forhold med betydning for karakterer. Det vises at elevenes standpunkt karakterer har størst påvirkning på hvor stort avviket mellom de to karakterene er. Elever med gode karakterer har med andre ord større sannsynlighet for å gå ned i karakter på eksamen sammenlignet med elever med svakere karakterer. Dette henger sammen med at standpunkt karakterer i gjennomsnitt ligger over eksamens karakterer uansett skoletype.

Kupiainen og kollegaer (2016) diskuterer en reform i finsk videregående utdanning, som i dag gjør at elever har større valgmuligheter i hvilke eksamener de skal inkludere i sin finske avgangseksamen fra studieforberedende program. De konkluderer med at det som tidligere var en ganske homogen eksamen, der to eksamener var valgfrie og de andre to ble valgt blant et avgrenset antall fag, i dag innebærer mange valg knyttet til hvilken sammensetning elevene ønsker. At elevene som tar den enkelte eksamen kun sammenlignes med andre elever som tar samme eksamen, gjør at sammenligningsgrunnlaget blir forskjellig for hver eksamen. Grunnen til dette er at elever med ulike akademiske evner velger ulike kurs, noe som betyr at de ikke blir sammenlignbare. Dette er informasjon som også elevene besitter. Dette leder til en strategisk tilpasning blant elevene, noe som gjør at vi ser en økning i populariteten til eksamener der det oppfattes som relativt enkelt å få en god karakter.

I en norsk rapport stiller Prøitz (2013) spørsmål ved introduksjonen av et felles rammeverk for anbefalinger og regelverk for karaktersetting: Kan ulikheter i læreres karaktersetting bli bedre forstått ved at karaktersettingen sees i lys av faget de underviser i, og er det mulig eller passende å bruke et felles rammeverk for alle fag? Det påvises at faget har betydning når lærere setter slutt karakter, og at ulike fag har ulike utfordringer eller hindre når det gjelder å imøtekomme statlige anbefalinger eller regelverk for karaktersetting. Det vises at lærere i mer konseptuelle fag (som matematikk, naturfag og kroppsøving), ser ut til å behandle karaktergrunnlaget på en mer standardisert måte. Lærere i mer kontekstuelle fag (som norsk og kunst og håndverk), ser karaktergrunnlaget som mer åpent for kontinuerlig forhandling.

Sjaastad og kollegaer (2016) spurte om elevene i videregående skoler i Norge får opplæring i forhold til det årstimetallet og med de kvalifiserte lærere som de har krav på. De finner at enkelte skoler planlegger å nå kompetansemålene før eksamen gjennomføres, gjennom å øke timetallet per uke, gjerne med fleksible økter til ulike fag. Samtidig synes eksamensperioden å være den enkeltårsaken som bidrar til at flest undervisningstimer forsvinner. Det vises til utfordringen at

undervisningen skal fortsette etter at eksamensperioden er over, til tross for at kompetansemålene er nådd, eksamen gjennomført og mange elever har lav motivasjon. Omtrent halvparten av lærere og skoleledere var «enig» eller «svært enig» i at eksamensperioden burde endres. Mens lærere i matematikk, naturfag og engelsk var spesielt kritiske, ga lærere i ulike yrkesfag (som helsefag, byggfag og elektrofag) i mindre grad uttrykk for stort endringsbehov.

5.2.3 Studier knyttet til gjennomføring, praksiser og erfaringer knyttet til eksamen og sluttvurdering

Den siste gruppen av inkluderte forskningsbidrag i kartleggingen tar for seg gjennomføring, praksis og erfaringer knyttet til eksamen og sluttvurderinger i grunnskolen og videregående opplæring i de nordiske landene. Grupperingen dekker i noen grad også lærer- og elevperspektivet på eksamen og sluttvurderinger. Et slikt mikroperspektiv kan for eksempel dreie seg om hvordan elevene opplever vurdering og karaktersetning, eller gjennomføring av muntlig eksamen. Følgende syv studier er presentert her: Bjørnset og kollegaer (2020; Fafø-rapport), Bøhn (2018), Hovdhaugen, Prøitz og Seland (2018), Modell og Gerdin (2021), Sandvik og kollegaer (2021), Sandvik og Sommervold (2021) og Skovholt og kollegaer (2021).

Bjørnset, M.; Fossum, A.; Rogstad, J. og Smestad, B. (2020). På like vilkår? Evaluering av matematikkeksamen på 10. trinn 2017-2019. *Rapport, Fafo*.

Land: Norge

Fafo har utgitt tre rapporter om eksamen, der hver rapport har prioritert å trekke frem et særskilt tema. Felles for de tre rapportene er at de ønsker å besvare spørsmålet om eksamen både er og oppleves å være rettferdig, i den forstand at kandidatene vurderes på likt grunnlag. I den siste rapporten (Bjørnset et al., 2020) oppsummeres hovedfunnene som gjelder hele prosjektet.

1. *Bjørnset, M.; Fossum, A.; Rogstad, J. og Smestad, B. (2020). På like vilkår? Evaluering av matematikkeksamen på 10. trinn 2017-2019. Sluttrapport. Fafo.* Denne rapporten har helhetlig vurdert matematikkeksamen i perioden 2017-19.

2. *Bjørnset, M., Fossum, A., Rogstad, J., Smestad, B. og Talberg, N. (2018). Digitale skillelinjer. Evaluering av matematikkeksamen på 10. trinn våren 2018. Fafo-rapport 2018: 36.*

3. *Andresen, S., Fossum, A., Rogstad, J. og Smestad, B. (2017). På prøve. Evaluering av matematikkeksamen på 10. trinn våren 2017. Fafo-rapport 2017: 36.*

Følgende spørsmål i prosjektet er relevant for den nåværende kartleggingen:

1. Foreligger det en god sammenheng mellom eksamenen, læreplanen og opplæringen som er gitt?
2. I hvilken grad finnes det samsvar mellom ulike sensorers vurdering?
3. Er det oppgaver av ulike vanskegrad i eksamen, som kan måle alle kompetansenivåer?
4. Hvordan vurderer elevene eksamens arbeidsmengde sammenlignet for eksempel til den tiden de har fått til rådighet totalt?
5. Er oppgavene utformet på en måte at de måler elevens matematikkompetanse?
6. Hvordan vurderes eksamens- og sensurenes utvikling i tidsperioden?
7. Er eksamensoppgavene utarbeidet på en forståelig måte med hensikt til tekst og illustrasjon?
8. Hvordan har elevene har vært forberedt på å bruke digitale hjelpemidler på eksamen.

Prosjektet bruker både kvantitative og kvalitative data. De kvantitative dataene er en spørreskjemaundersøkelse som ble sendt til matematikklærere på 10. trinn som hadde hatt elever som gikk opp til eksamen. Kvalitative intervjuer ble gjennomført med lærere og elever ved fire skoler og det ble gjort klasseromsobservasjon. I tillegg fikk sensorer spørsmål gjennom Utdanningsdirektoratets spørreundersøkelse til sensorer, og eksamensbesvarelser ble analysert.

Bjørnset og kollegaer (2020) konkluderer med at eksamen i den gitte evalueringsperioden fremstår som rettferdig og av høy kvalitet. Dette betyr ifølge

forfatterne at de ulike eksamensoppgavene relaterte godt til undervisningen, og de nedfelte kompetansemålene. De peker på at matematikkeksamen har høy legitimitet blant elevene og lærere, noe som betyr at nesten alle deler av læreplanen har blitt gjenstand for prøve i de tre årene. Samtidig nevnes det at en økende andel av sensorene mener at det er deler av læreplanen som aldri blir testet. Innholdet i eksamensoppgavene oppleves i å være i tråd med opplæringen som er gitt, både av elever og lærere.

For bruken av digitale verktøy fant de derimot at elevene har fått ulik opplæring, noe som ga dem ulike muligheter for å bruke digitale hjelpemidler på eksamen. Ifølge forfatterne betyr denne typen forskjeller at prinsippet om at kandidatene skal ha like betingelser til å besvare de ulike oppgavene blir utfordret.

Samlet sett synes Bjørnset og kollegaer (2020) at eksamen holdt god kvalitet. Dette støttes av en analyse av arbeidsmengde og vanskelighetsgrad på eksamen med utgangspunkt i vurderingsskjemaene. Generelt sett var vanskelighetsgraden tilstrekkelig variert slik at alle elever uansett prestasjonsnivå fikk muligheten til å vise sin kompetanse. Det var imidlertid ett unntak: svakt presterende elever. Analysene viste at det er forholdsvis få elever som opplever tidspress.

For språk og bruk av illustrasjoner, synes forfatterne å være bekymret over at flertallet av oppgaver stiller sterke språklige krav. De synes at det er ikke er nødvendig å teste leseferdighet i alle eksamensoppgavene, selv om dette er en grunnleggende ferdighet også i matematikk. Begrepskjennskap er også knyttet til andre forhold enn matematisk kompetanse, for eksempel om man er født utenfor Norge, noe som kan utfordre rettferdighetsprinsippet. Selv om hovedkonklusjonen er at eksamen oppfattes som rettferdig, stiller forfatterne spørsmål ved den språklige kompleksiteten i noen oppgaver og om dette har negative konsekvenser for noen elevgrupper.

Det blir også trukket frem i sensorenes vurderinger at svakt presterende elever i mindre grad gis mulighet til å vise sin helhetlige kompetanse, fordi mange av oppgavene er flervalgsoppgaver eller oppgaver der det bare oppgis et svar. Ut over dette ble det påvist stor ulikhet i vurderingene mellom sensorene på de oppgavene med krav til bruk av digitale hjelpemidler, oppgaver med høyere krav til kommunikasjon og oppgaver hvor elevene selv måtte velge en hensiktsmessig metode.

Når det gjelder utviklingen i perioden 2017-2019 er konklusjonen at eksamenene har vært tilstrekkelig gode. I denne sammenhengen trekkes det frem at oppgavens utforming og til dels selve spørsmålsformuleringen har vært noenlunde likt, noe som ikke er en selvfølge. Når det gjelder endringer har det skjedd en endring i vektingen mellom del 1 og del 2 som har ført til forbedrende eksamensresultater, og andelen oppgaver med flervalg har økt radikalt.

Bøhn, H. (2018). Assessing content in a curriculum-based EFL oral exam: The importance of higher-order thinking skills. *Journal of Language Teaching and Research*.

Land: Norge

Konteksten for denne studien er undervisning i engelsk som fremmedspråk i norsk videregående skole. Et definerende trekk ved denne konteksten er mangelen på en felles sensureringsskala eller vurderingsmatrise for å veilede lærere i deres vurdering av muntlig prestasjoner i engelsk. Det finnes også få empiriske studier som beskriver hvordan innhold blir vurdert i slike settinger. Kunnskap om hvilke aspekter ved elevenes presentasjon som sensor vektlegger i sin vurdering er viktig for å vurdere resultatenes validitet. I tillegg er slik kunnskap viktig for forståelse av mulige washback-effekter knyttet til hva lærere prioriterer i språkfag. Formålet med denne studien er derfor å se *hva engelsklærere i norske videregående skoler anser som relevant faginnhold for vurdering i en muntlig engelskeksamen*.

Studien analyserer kvalitative data fra muntlige protokoller og semistrukturerte intervjuer, som involverte ti engelsklærere ved videregående skole i Norge. Læreres orientering blir også sammenlignet med fagaspekter i emnets læreplan.

Engelskarakterer i videregående skole er hovedsakelig gitt i form av generelle prestasjonskarakterer, fastsatt av hver enkelt faglærer på bakgrunn av ulike former for klasseromsvurdering. I tillegg blir ca. 20 prosent av elevene tilfeldig trukket ut til en skriftlig eksamen, og 5 prosent til en muntlig eksamen. Når en elevgruppe blir trukket ut til muntlig engelskeksamen, kreves det at elevenes engelsklærer er sensor. I tillegg blir en engelsklærer utenfor elevenes skole tildelt rollen som ekstern sensor. Når det gjelder operasjonaliseringen av hva som skal vurderes i faget, er det en tydelig forskjell mellom muntlig og skriftlig eksamen. Skriftlige eksamen blir administrert nasjonalt av Utdanningsdirektoratet, mens muntlig eksamen blir kontrollert lokalt av kommunene. Dette betyr at det er nasjonalt utviklede vurderingsskalaer og nasjonalt designede eksamensoppgaver for skriftlig eksamen, mens det for muntlig eksamen eksisterer ulike typer av lokalt utviklede skalaer og oppgaver. Dermed mangles en felles vurderingsskala for lærerne og sensorer i muntlige engelskeksamen i norske videregående skoler.

Når det gjelder oppfattelse av vurderingsaspekter i engelsk muntlig eksamen, indikerer resultatene fra studien at lærere forstår emnet i veldig generelle termer. Engelsklærerne bekrefter at aspektene som er listet opp i læreplanen, bortsett fra metakognitive strategier, er relevante aspekter å vurdere. Ettersom en rekke av disse aspektene er veldig omfattende, understreket de på samme tid at det er urealistisk å forvente at elevene husker detaljer fra alle de mulige temaene i faget. Som en konsekvens, ser det ut til at lærerne adopterer en vurderingsstrategi der testingen av elevenes evner og prosesser (beskrivelse, analyse, evaluering etc.) blir

viktigere enn vurderingen av et klart definert faginnhold. Dette betyr med andre ord at detaljene i faginnholdet blir betydelig nedprioritert.

Funn fra studien indikerer altså at lærere generelt sett er mer opptatt av elevenes evne til å presentere og reflektere (altså evner- og prosessaspekter), enn det spesifikke faginnholdet i hva som ble presentert. Lærerne virker spesielt orientert mot høyere kognitive resoneringsevner («higher order thinking skills»), slik som evnen til å analysere, evaluere og reflektere over et gitt tema. Artikkelforfatteren påpeker at dette funnet støtter funn fra tidligere studier, som viste at læreres operasjonalisering av innholdskonstruksjoner i all hovedsak involverer evner og ferdigheter.

Artikkelforfatteren peker på at studien har to sentrale implikasjoner: Den første relateres til betydningen av å varsle elever om viktigheten av høyere kognitive resoneringsevner i engelskfag på videregående nivå. Ettersom sensorer ser ut til å vektlegge slike evner i denne konteksten, bør også lærere sørge for undervisningsoppgaver og undervisningsmaterieell som lar elever utvikle sin evne til å analysere, evaluere og reflektere over faginnhold. Den andre implikasjonen er at validitetsutfordringene knyttet til rollen metakognitive strategier har i det målte konstruktet bør adresseres. Det påpekes også at ulikhetene i oppfatningen knyttet til hva slags adferd som indikerer prestasjoner på de ulike nivåene bør adresseres.

Artikkelforfatteren konkluderer med at en mulig løsning til disse problemene er introduksjonen av en felles vurderingsskala og retningslinjer for sensorer i vurdering av innhold ved muntlig engelskeksamen. I tillegg ser det ut til at mer sensurtrening kan være fordelaktig, ettersom dette er rapportert til å ha positiv effekt på reliabiliteten.

Hovdhaugen, E., Prøitz, T. S. og Seland, I. (2018). Eksamens- og standpunktkarakterer – to sider av samme sak? *Acta Didactica*.

Land: Norge

Studien tar utgangspunkt i standpunktkarakterer og eksamenskarakterer i Norge, og stiller følgende spørsmål: *Hvilken funksjon har eksamenskarakterer og standpunktkarakterer for læreres vurderingspraksis?* Analysen bygger på nasjonal karakterstatistikk og intervjuer med lærere som underviser i norsk eller matematikk i åtte videregående skoler i Norge, og som har erfaring med å vurdere elevenes sluttkompetanse i disse fagene ved hjelp av begge vurderingsordningene.

Studien peker på at standpunkt- og eksamenskarakterer blir satt på ulike premisser og springer ut av ulike kontekster. Mens eksamen viser elevens samlede kompetanse slik den kommer til uttrykk på eksamensdagen, skal standpunktvurdering gi informasjon om kompetansen eleven har oppnådd i faget ved avslutningen av opplæringen.

Studien viser også at lærere ser på det å sette standpunkt og det å sette eksamenskarakter 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 standpunktkarakteren et resultat av lærerens kjennskap til eleven og elevens progresjon over tid. Sensor vil ifølge lærerutsagnene 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 er faglærers oppgave og skjønn å vurdere standpunktkarakteren som en gjenspeiling av elevens progresjon gjennom opplæringen.

Funn fra studien viser videre at faglærer ofte vil bestrebe seg på å sikre at eleven ikke får strykkarakter ved standpunktvurdering. Sensor har på den andre siden ingen tilsvarende bindinger til eleven og elevens fremtidsutsikter i videre utdanning eller jobb. De vil derfor, ifølge lærerne, sette strykkarakter hvis vurderingskriteriene tilsier dette. På den annen side vil sensor ofte, med grunnlag i vurderingskriteriene, være tilbakeholden med å gi toppkarakter på eksamen.

Artikkelforfatterne konkluderer med å påpeke at 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. Studiens funn 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.

Fastsetting av standpunktkarakter skjer på grunnlag av en lærer-elev-relasjon, preget av en form for dialog og lærerens inngående kjennskap til eleven som person. Eksamenskaraktersetting gjennomføres med gjensidig anonymitet og et såkalt sensorblikk. Artikkelforfatterne poengterer at 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. Studien 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 dermed et spørsmål om man isteden bør finne alternative måter å kvalitetssikre standpunktvurderingen og eksamen på – hver for seg og basert på de to vurderingsordningenes ulike premisser.

Modell, N. og Gerdin, G. (2021). 'But in PEH it still feels extra unfair': students' experiences of equitable assessment and grading practices in physical education and health (PEH). *Sport, education and society*.

Land: Sverige

Bakgrunnen for studien er utfordringene knyttet til vurdering i kroppsøvningsfaget. Studien ser på elevers erfaringer med vurdering i det svenske skolefaget kroppsøving og helse, ettersom elevers oppfatning av vurdering ofte er neglisjert i forskningen. Målet til studien er *å studere elevers erfaringer med å ha like muligheter til å demonstrere deres kunnskap og evner når de blir vurdert i kroppsøving og helse*. Denne kvalitative studien tar utgangspunkt i data som er samlet inn ved fokusgruppeintervju med 38 elever fra fire forskjellige videregående skoler i Sverige.

Studien viser at elever oppfatter kroppsøving og helse som et urettferdig fag når det kommer til vurdering og karaktersetting, ettersom alle har ulike fysiske forutsetninger. Elevene oppfattet det også som urettferdig at noen elever blir gitt muligheter til å vise sin kunnskap og sine evner i kroppsøving og helse, mens andre ikke får denne muligheten. Både elever som anser seg som lite sportslige, og elever som anser seg som veldig sportslige, følte at de ble urettferdig vurdert.

Studien peker blant annet på at det er en frustrasjon blant elevene over at det ikke er rom for å bli bedre i for eksempel fotball. I stedet bør elever vise at de allerede kan spille fotball. Elevenes tidligere erfaringer med sport blir dermed en forutsetning for å kunne få høy karakter, hvilket blir oppfattet som urettferdig. I tillegg oppfatter elever det som urettferdig at det gjelder å vise kunnskap og evner innenfor «de riktige» sportsgrenene i kroppsøving og helse. Dette understreker at noen idrettsgrener har en høyere status i karaktersettingen enn andre, hvilket betyr at de elevene som får mulighet til å vise sine evner i faget er privilegert. For eksempel prioriteres fotball, men elever som er gode i hesteridning ikke gis samme mulighet til å vise sine ferdigheter innenfor denne idrettsgrenen.

Studien peker videre på at de urettferdige vurderingspraksisene hovedsakelig ser ut til å stamme fra manglende like muligheter til å lære og demonstrere kunnskap og evner i kroppsøving og helse. Dette anses igjen som et resultat av kulturell-kognitive forhold. Som funnene viser er både lærere og læreres praksis med vurdering og karaktersetting i kroppsøving og helse, formet av kulturelle-kognitive forhold assosiert med de normer og verdier vi finner i konkurransedrevne sporter. Dette går ofte på bekostning av vurdering gjort med utgangspunkt i mer regulative og normative forhold, slik det er beskrevet i den svenske utdanningsloven og pensumet i kroppsøving og helse. Funn fra studien demonstrerer også hvordan både lærere og elever er aktivt involvert i å reprodusere slik undervisnings- og vurderingspraksis. Dermed korresponderer funn fra denne studien med tidligere

forskning, som viser hvordan vurdering i kroppsøving ikke er basert på kriterier fra læreplanen, men heller det «å være god i riktig type idrett».

Artikkelforfatterne trekker også frem det faktum at karaktersetning ikke blir ansett som rettferdig. Ifølge den svenske riksrevisjonen (the Swedish National Audit Office) kan dette innebære at karaktersetningssystemet risikerer å miste sin troverdighet. Artikkelforfatterne konkluderer med å etterspørre mer arbeid og forskning knyttet til elever og lærere i kroppsøving og helsefag. Dette er nødvendig for å rette større oppmerksomhet på og få en bedre implementering av læring og måloppnåelse av pensum. Ved å adressere den pågående utfordringen med å gjennomføre vurdering i kroppsøving og helse, argumenterer artikkelforfatterne for at man må bli flinkere til å sammenstille vurderingsprosesser med intensjonene til læreplanen knyttet til lik utdanningskvalitet og læring for likestilling.

Sandvik, L. V., Smith, K., Strømme, A. et al. (2021). Students' perceptions of assessment practices in upper secondary school during COVID-19. *Teacher and Teaching*.

Land: Norge

Bakgrunnen for studien er vurderingspraksisen i videregående skoler under Covid-19 pandemien. Under nedstengingen våren 2020, endret forholdene for undervisning, læring og vurdering seg. I Norge ble skolene stengt og alle undervisningsaktiviteter foregitt digitalt. Også avsluttende eksamener ble avlyst. Både nedstengingen og kanseleringen av eksamen påvirket undervisningen og vurderingspraksisen, og hadde stor påvirkning på elevenes erfaringer med vurderings- og læringskvaliteten. På bakgrunn av dette er målet med studien å undersøke hvordan elever i videregående skole i Norge opplevde vurderingspraksisen under den første nedstengingen under Covid-19 pandemien. Studien ser også på hvordan elevenes motivasjon og mestringfølelse («self-efficacy») ble påvirket av undervisnings- og vurderingspraksisene.

Artikkelen presenterer kvalitative funn fra en studie som undersøkte elevers oppfattelse av undervisning og vurderingspraksis i videregående skole i Norge våren 2020. Data ble samlet fra fokusgruppeintervju med til sammen 24 elever i sisteåret fra fire videregående skoler i Norge.

Funn fra studien viser at elevene verdsetter læreres tilbakemelding som støtte i læringen. Elevene erfarte fleksibilitet i undervisningspraksisen, og de ble også vist tillit til at de gjennomførte det arbeidet de ble gitt. Når det gjaldt vurdering, følte de likevel at dette var urettferdig og at elevmedvirkning i vurderingen manglet. Elevene rapporterte også at det var lite variasjon i oppgavene, som hovedsakelig var individuelle skriftlige oppgaver. Studien poengterer også at en viktig beskjed fra informantene, var at elevene savnet samarbeid med andre elever og at mange av dem ble sittende alene med sitt skolearbeid.

Funn fra studien viser også at når elevene visste at deres sluttvurdering ville gjenspeile arbeidet og utviklingen i et fag gjennom hele året, så rapporterte de at motivasjonen økte og at deres stress knyttet til prøver minket. Dette så ut til å være viktig for deres investering i ulike oppgaver, deres vurdering av kompetanse for å utføre en oppgave, deres tro på viktigheten, nytten og relevansen til ulike oppgaver de ble gitt i et fag, samt deres personlige motivasjon for en vurderingsoppgave.

På bakgrunn av funn fra studien, påpeker artikkelforfatterne at elevers involvering i undervisning og vurderingspraksis bør bli gitt mer oppmerksomhet når disse praksisene utvikles. Dette gjelder ikke kun ved digital undervisning, men også under normale undervisnings- og lærings situasjoner. Ved å forstå vurdering som læring, understreker man en langvarig eller bærekraftig forståelse av vurdering. En slik forståelse av vurdering kan også være en måte for å oppnå mer likestilling i vurderinger. Studien viste at elever oppfatter struktur og organisering, elevmedvirkning og forutsigbarhet i vurdering som viktige elementer for å skape god vurdering.

Studien tar utgangspunkt i en periode der skolesituasjonen var kaotisk og vanskelig for alle som var involvert, og artikkelforfatterne understreker noen av implikasjonene for videre undervisning og vurderingspraksis. En av implikasjonene som trekkes frem, er at det kan se ut til at utdanningssystemet i Norge vil ha færre og færre eksamener, og derfor må lærere være forberedt på å sikre kvaliteten på sluttvurderingen i større grad. Studien indikerer at det er store variasjoner i læreres vurderingspraksis, hvilket fører til usikkerhet for elevene når det kommer til vurdering. Artikkelforfatterne peker derfor på at det er nødvendig å forbedre vurderingsferdighetene i kontekster utenom eksamen.

Sandvik, L. V. og Sommervold, O. A. (2021). Students' perceptions of involvement in the assessment of oral competence in English as a second language. *Languages*.

Land: Norge

Bakgrunnen for studien er elevmedvirkning i vurdering og hvordan dette kan bli ansett som en nøkkelfaktor når det kommer til å øke elevers læringsutbytte og motivasjon. Tidligere forskning har vist at studentmedvirkning og vurderingen av muntlig kompetanse i faget engelsk som andrespråk har vært utfordrende. På bakgrunn av dette er formålet med studien å undersøke hvordan elever i videregående skoler i Norge oppfatter deres involvering i vurderingspraksisen av engelsk muntlig kompetanse.

Studien benytter en mixed-method-tilnærming, der det ble brukt en survey (N=116) og to fokusgruppeintervjuer (N=8). Informantene var førsteårselever fra en videregående skole i Norge.

Artikkelforfatterne påpeker at formativ vurdering i løpet av de siste tiårene har blitt et sentralt aspekt i utdanningen for å øke elevers læringsutbytte. I tillegg har det blitt et fokusområde i det norske skolesystemet gjennom 2000-tallet. Det har derimot blitt vist at læringsprosessen i engelsk begrenses av nasjonale tester og eksamen, noe som kan føre til en washback-effekt der engelskfagets læreplanmål blir nedspilt på bekostning av de nasjonale vurderingskriteriene. Dette kan ifølge artikkelforfatterne føre til at mange lærere blir mer opptatt av å forberede elevene på de ulike testene og eksamenene, fremfor å sørge for gode læringsprosesser som bidrar i utviklingen av elevers interkulturelle forståelse, kommunikasjon og identitetsutvikling i faget.

Funn fra studien viser at elevene ønsket å være mer involvert i vurderingspraksisen. Videre anså elevene denne økte medvirkningen som en måte å øke deres muntlige kompetanse i engelsk. Studien antyder også at det er uoverensstemmelse mellom hva elever tror målet er og hva læreren faktisk vurderer dem etter. Denne uoverensstemmelsen betyr ikke at elever ikke vet hva vurderingskriteriene er, men det kan indikere at elevene til en viss grad ikke er klar over hvordan kriteriene blir vurdert og hvilke kriterier som blir vektlagt. Videre viser studien at elevene erfarte en usikkerhet knyttet til hva de faktisk blir vurdert etter når de fikk tilbakemelding fra lærere, og dette påvirket også om elevene oppfattet tilbakemeldingen som nyttig eller ikke.

Ifølge artikkelforfatterne antyder studien at økt involvering i utviklingen av mål og kriterier, samt mer dialogbasert tilbakemelding, er fordelaktige tiltak for å styrke elevers læringsutbytte.

Skovholt, K., Solem, M. S., Vonen, M. N., Sikveland, R. O., og Stokoe, E. (2021). Asking more than one question in one turn in oral examinations and its impact on examination quality. *Journal of Pragmatics*.

Land: Norge

Formålet med denne studien er å analysere hvordan sensorer stiller spørsmål til eksamenskandidater ved muntlig eksamen i skolen, både ved slutten av ungdomsskolen og ved slutten av videregående skole. Mer spesifikt ønsket artikkelforfatterne å undersøke hvorvidt det å stille mer enn et spørsmål samtidig påvirket elevenes resultater, og hvordan spørsmålsformatet åpnet for eller begrenset kandidatens muligheter for å produsere det som senere vurderes som svar av høy kvalitet.

Datamaterialet omfattet 36 (23+13) videoopptak av muntlig eksamen med seks sensorer på tvers av fire skoler i Norge: to ungdomsskoler og to videregående skoler med totalt 18 timer opptak. Data ble analysert gjennom diskursanalyse. For å måle muntlige ferdigheter må norske elever ta muntlig eksamen i et fag ved slutten

av ungdomsskolen, og ved slutten av tredje klasse i videregående skole. Norske muntlige eksamener vurderer både kompetanse knyttet til faginnhold og elevens muntlige ferdigheter. Muntlige ferdigheter regnes som én av fem nøkkelkompetanser som vurderes i den norske læreplanen.

Generelt ble sensorer advart mot å stille mer enn et spørsmål av gangen ved muntlig eksamen. Likevel viste denne studien at bredere spørsmål, som inneholder flere spørsmål samtidig, er en utbredt praksis. Denne studien fokuserte på de to mest hyppige formatene – likestilte («side-ordered») spørsmål og spesifiserende («sub-ordered») spørsmål. Studien analyserte utformingen av «multi-unit questions» (MUQs) og demonstrerte måter hvordan MUQs begrenset eller fremmet eksamenskandidatens muligheter for å gi utdypende og uavhengige svar av høy kvalitet.

Et hovedfunn er at to eller flere spørsmål spurt i samme omgang fungerer primært som et verktøy for å understøtte og sikre gode svar og forebygge misforståelser. Samtidig kunne samme praksis også hindre kandidaten i å vise fagspesifikk kompetanse. Når det gjelder likestilte spørsmål, bidro dette formatet positivt til å fremme kandidatens svar. Når det gjelder underordnede spørsmål, kunne dette formatet virke begrensende, fordi de spesifiserende spørsmålene kunne hindre kandidaten i å besvare de mer generelle spørsmålene på en uavhengig måte. Ved å spørre to spørsmål i samme omgang får kandidaten begrensede muligheter til å besvare det første spørsmålet før det andre blir stilt. Følgelig kan dette betyr færre muligheter for å besvare det opprinnelige spørsmålet i en muntlig eksamen.

Artikkelforfatterne argumenterer for at denne studien kan bidra til en større diskusjon om hvilke test-format som er mest hensiktsmessige og hvordan flere spørsmål kan påvirke validitet, reliabilitet og rettferdighet i muntlig eksamen. Som det ble påvist kan likestilte og spesifiserende MUQs fungerer som et understøttende verktøy som styrer kandidaten i retning av det foretrukne svaret.

Artikkelforfatterne konkluderer med at det er viktig å være oppmerksom på den rollen spørsmålsformatet i en eksamen spiller, og øke oppmerksomheten rundt hvordan sensorenes måte å spørre på kan påvirke kandidatens resultater. Dette kan bidra til å øke kvaliteten til muntlig eksamen, sikre sensorenes kontroll over hvilke spørsmål som stilles og dermed redusere risikoen for tilfeldigheter i spørsmålsstillingen.

Oppsummering

Vi har inkludert syv forskningsbidrag om gjennomføring, praksiser og/eller erfaringer knyttet til sluttvurdering og eksamen. Av disse er seks studier fra Norge, og én studie er fra Sverige.

I en av tre Fafo-rapporten ønsker Bjørnseth og kollegaer (2020) å besvare spørsmålet om eksamen både er og oppleves som rettferdig, slik at kandidatene

vurderes på likt grunnlag. Overordnet fremstår eksamen i den gitte evalueringsperioden (2017-2020) som rettfærdig og av høy kvalitet: de ulike eksamensoppgavene var knyttet til den faktisk gjennomførte undervisningen og de nedfelte kompetansemålene. Forfatterne peker på matematikkksamens høye legitimitet, noe som kommer av at nesten alle deler av læreplanen har blitt utprøvd i de tre årene. Samtidig nevnes at en økende andel av sensorene mener at det er deler av læreplanen som aldri ble testet. Generelt sett var vanskegraden tilstrekkelig variert slik at alle elever uansett prestasjonsnivå fikk muligheten til å vise sin kompetanse. Unntaket her var de svakt presterende elevene. For språk og bruk av illustrasjoner, synes forfatterne å være bekymret flertallet av oppgaver stiller språklige krav. De synes at det er ikke nødvendig å teste leseferdigheten i alle eksamensoppgavene, selv om dette er en grunnleggende ferdighet også i matematikk.

For engelskfaget ser Bøhn (2018) på hva engelsklærere i norske videregående skoler anser som relevant faginnhold for vurdering i en muntlig engelskksamen. Funn indikerer at lærere generelt sett er mer opptatt av elevenes evne til å presentere og reflektere (altså evner- og prosessaspekter), enn det spesifikke faginnholdet i hva som ble presentert. Lærerne virker spesielt orientert mot høyere kognitive resoneringsevner, slik som evnen til å analysere, evaluere og reflektere over et gitt tema.

Hovdhaugen og kollegaer (2018) ønsket å besvare følgende spørsmål: Hvilken funksjon har eksamenskarakterer og standpunkt karakterer for læreres vurderingspraksis? De konkluderer med å påpeke at 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. De peker på flere svakheter ved ideen om at eksamen kan fungere som kvalitetssikring av standpunkt karakterene. 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. Artikkelforfatterne poengterer at de to vurderingsordningene har styrker hver for seg som det er flere grunner til å forsvare, men også svakheter som kan kritiseres.

Målet til studien av Modell og Gerdin (2021) er å studere svenske elevers erfaringer med å ha like muligheter til å demonstrere deres kunnskap og evner når de blir vurdert i kroppsøving og helse. Studien konkluderer med å etterspørre mer arbeid og forskning knyttet til elever og lærere i kroppsøving og helsefaget, dette for å rette større oppmerksomhet på og få en bedre implementering av læring og måloppnåelse av pensum. Ved å adressere den pågående utfordringen knyttet til vurdering i kroppsøving og helse, argumenterer artikkelforfatterne for at man må

bli flinkere til å sammenstille vurderingsprosesser med intensjonene til læreplanen knyttet til lik utdanningskvalitet og læring for likestilling.

Sandvik og kollegaer (2021) undersøkte hvordan elever i videregående skole i Norge opplevde vurderingspraksisen under den første nedstengningen under Covid-19 pandemien. De påpekte at elevmedvirkning i vurderingen manglet i denne perioden. Elevene rapporterte også at det var lite variasjon i oppgavene, som hovedsakelig var individuelle skriftlige oppgaver. De fant også at når elevene visste at deres sluttvurdering ville gjenspeile arbeidet og utviklingen i et fag gjennom hele året, så rapporterte de at motivasjonen økte og at deres stress knyttet til prøver minket. Studien viste at elever oppfatter struktur og organisering, elevmedvirkning og forutsigbarhet i vurdering som viktige elementer for å skape bærekraftig vurdering. Studien tar utgangspunkt i en vanskelig periode og understreker noen av implikasjonene for videre vurderingspraksis, for eksempel at det kan se ut til at utdanningssystemet i Norge vil ha færre og færre eksamener, og derfor må lærere være forberedt på å sikre kvaliteten på sluttvurderingen i større grad.

Sandvik og Sommervold (2019) undersøkte hvordan elever i videregående skoler i Norge oppfatter deres involvering i vurderingspraksisen av engelsk muntlig kompetanse. De finner at elevene ønsket å være mer involvert i vurderingspraksisen. Videre anså elevene denne økte medvirkningen som en måte å øke deres muntlige kompetanse i engelsk på.

Skovholdt og kollegaer (2021) analyserte hvordan sensorer stiller spørsmål til eksamenskandidater ved muntlig eksamen i skolen, både ved slutten av ungdomsskolen og videregående skole. Mer spesifikt ønsket artikkelforfatterne å undersøke hvorvidt det å stille mer enn et spørsmål samtidig påvirket elevenes resultater, og hvordan spørsmålsformatet åpnet for eller begrenset kandidatenes muligheter for å produsere det som senere vurderes som svar av høy kvalitet. De konkluderte med at det er viktig å være oppmerksom på rollen spørsmålsformatet i ulike eksamener spiller, og øke oppmerksomheten rundt hvordan sensorers måte å spørre på kan påvirke kandidatens svar.

5.3 Kunnskapsklynger

Et sentralt formål med vår systematiske kartlegging har vært å identifisere forskningsbidrag knyttet til eksamen- og sluttvurderingssystemet i de nordiske landene. I gjennomgangen av forskningsmaterialet identifiserte vi også noen sentrale kunnskapsklynger som vi anså å være tematisk beslektet/overlappende med hovedproblemstillingen. Funnene fra kunnskapsklyngene er det verdt å se nærmere på, ettersom de bidrar til å nærmere belyse kunnskapslandskapet knyttet til sluttvurderinger og eksamen i Norden.

I vår systematiske kartlegging dukket det opp mange forskningsbidrag som ikke direkte omhandler eksamens- og sluttvurderingssystemene i Norden, men som likevel er relatert til tematikken (se vedlegg 2). Disse bidragene ble vurdert til å ha lav relevans med tanke på kartleggingens problemstilling. Likevel anses de som relevante ettersom de kan sette hovedfunnene i kontekst og gi et bredere bilde av kunnskapslandskapet.

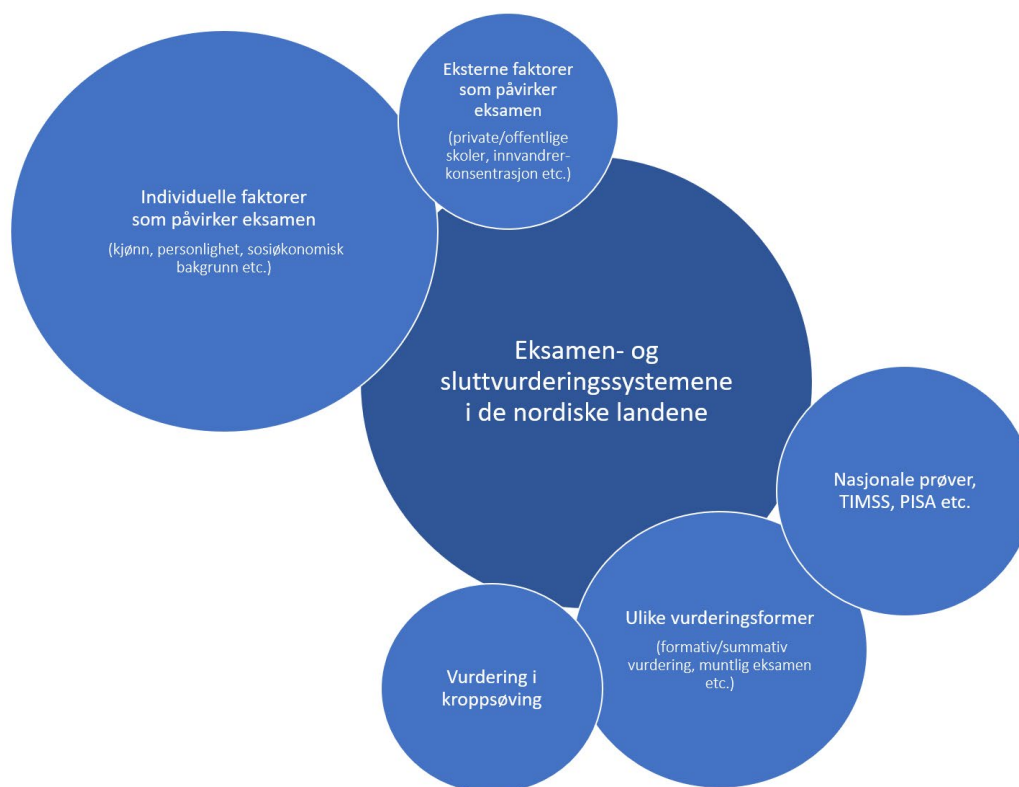
På bakgrunn av kodingen av forskningsbidragene kunne vi identifiserte tre overordnede kunnskapsklynger (eller temaklynger) knyttet til eksamen- og sluttvurderingssystemene i de nordiske landene. Dette er forskning knyttet til:

- Individuelle faktorer ved eleven som påvirker eksamen.
- Eksterne faktorer som påvirker eksamen.
- Andre vurderingsformer, herunder fagspesifikke vurderingsformer og standardiserte prøver.

Figur 5.2 viser kunnskapsklyngene identifisert gjennom den systematiske kartleggingen. Sirklene i modellen er omtrentlig skalert ut ifra antall forskningsbidrag som ble inkludert. 19 bidrag er knyttet til systemet for og organiseringen og gjennomføringen av sluttvurdering og eksamen i de nordiske landene (inkludert med høy relevans). Dette er hovedfokuset i vår kartlegging, og det ble foretatt både systematiske og strategiske litteratursøk for tematikken.

Forskningsbidragene i de andre kunnskapsklyngene er inkludert med lav relevans, og ble identifisert på bakgrunn av det systematiske litteratursøket. Disse bidragene er kun eksempler på studier fra andre og relaterte kunnskapsområder¹⁹. I vår kartlegging identifiserte vi 15 studier knyttet til individuelle faktorer som påvirker eksamen, 4 bidrag om eksterne faktorer som påvirker eksamen, 9 studier om ulike vurderingsformer, 7 studier knyttet til nasjonale prøver, TIMSS og PISA, samt 4 studier om fagspesifikk vurdering i kroppsøvingfaget. Kunnskapsklyngene er ikke sterkt avgrenset, og det vil være en glidende overgang mellom noen av klyngene. For en detaljert oversikt over forskningsbidragene i de ulike kunnskapsklyngene, se tabell 5.1 og vedlegg 2.

¹⁹ Hadde man kartlagt disse spesifikke kunnskapsklyngene i mer detalj, for eksempel gjennom ytterligere strategiske litteratursøk, ville man fått et større antall treff også på disse kunnskapsområdene. Imidlertid vil et slikt arbeid gå utover dette prosjektets forskningsspørsmål og rammer.



Figur 5.2 Kunnskapsklynger identifisert gjennom den systematiske kartleggingen

I de følgende avsnittene sammenfattes noen fellestrekk ved de ulike kunnskapsklyngene relatert til eksamen- og sluttvurderingssystemene i de nordiske landene.

5.3.1 Individuelle faktorer som påvirker eksamen

I vår systematiske kartlegging finner vi en rekke forskningsbidrag som ser på hvordan ulike individuelle faktorer ved elever påvirker eksamen og/eller sluttkarakterer. Av de inkludert bidragene med lav relevans, kan 15 studier grupperes i denne kunnskapsklyngen (se vedlegg 2). Forskningsbidragene ser blant annet på hvordan individuelle faktorer som kjønn, alder, personlighet, fysisk helse, intelligens, evne til stresshåndtering, sosioøkonomisk bakgrunn eller lære- lese- og regnevansker påvirker elevens eksamens- eller sluttkarakterer i grunnskolen og videregående opplæring. For eksempel ser Evensen (2019) på sammenhengen mellom elevers mentale helse og eksamenskarakterer, Sæle og kollegaer (2016) ser på hvordan elevers kognisjon, problemløsningsevner og fremtidsplaner spiller inn på deres sluttkarakterer, mens Bensnes (2016) studerer hvordan pollensesongen og pollenallergi påvirker elevers prestasjoner på slutteksamen.

Studiene i denne kunnskapsklyngen sier dermed ikke noe om selve organiseringen eller systemet rundt eksamen og sluttvurdering i de nordiske landene.

Derimot er eksamen og slutt karakterer ofte utfall i studiene, og forskningsbidragene kan dermed si noe om faktorer som har betydning for elevers eksamensprestasjoner og sluttresultater. Som en forlengelse av dette, problematiserer mange av studiene i denne klyngen rettferdighetsaspektet ved eksamen og sluttvurderinger. Ettersom individuelle faktorer *utenom* skolefaglige evner (slik som kjønn, allergi eller sosioøkonomisk bakgrunn) kan påvirke elevers karakterer, aktualiseres blant annet spørsmålet om eksamen sikrer elever en lik og rettferdig sjanse til å vise sin faglige kompetanse.

5.3.2 Eksterne og institusjonelle faktorer som påvirker eksamen

En annen kunnskapsklynge som kan identifiseres i dette kunnskapslandskapet, er forskningsbidrag som ser på hvordan ulike eksterne og institusjonelle faktorer påvirker elevers eksamen- og slutt karakterer. Av de inkluderte bidragene med lav relevans, kan fire studier grupperes i denne kunnskapsklyngen. De fire studiene er eksempler fra et større kunnskapsområde, som også har berøringspunkter med studier som mer konkret omhandler eksamen- og sluttvurderingssystemet.

I denne kunnskapsklyngen finner vi studien fra Hardoy og kollegaer (2018), som undersøker om innvandrerkonsentrasjon ved en skole påvirker elevers karakterer. Videre ser Kortelainen og Manninen (2019) om elevers sluttresultater påvirkes av om de går på en privat eller offentlig skole, Karlson (2015) undersøker om elevers plassering i ulike nivåinndelinger (high/low-track class) påvirker deres resultater og Bogetoft og kollegaer (2015) sammenligner de nordiske landene for å se om investeringer i de ulike utdanningssystemene påvirker elevenes karakternivå.

Forskningsbidragene i denne kunnskapsklyngen sier ikke noe om selve organiseringen eller systemet rundt eksamen og sluttvurdering i de nordiske landene. Derimot er eksamen og slutt karakterer også her det sentrale utfallet i studiene. Forskningsbidragene kan dermed si noe om eksterne og institusjonelle faktorer som kan ha betydning for elevers eksamen- og sluttresultater.

5.3.3 Omtale av andre vurderingsformer

Den siste temaklyngen som vi identifiserte i vår systematiske kartlegging, er forskningsbidrag som omhandler andre vurderingsformer utenom tellende eksamen eller sluttvurderinger. Dette er for eksempel studier som tar for seg læreres vurderings- og tilbakemeldingspraksis, formative og summative vurderingsformer, eller andre prøver og tester i grunnskolen og videregående opplæring i Norden. Til sammen ni studier er kategorisert i denne kunnskapsklyngen (se vedlegg 2). For eksempel ser Granbom (2016) på fordelene ved formativ vurdering og økt

elevmedvirkning i biologifag, Mäkipää (2021) studerer hvilke typer tilbakemeldinger elever oppfatter som oppmuntrende, og Jalava og kollegaer (2014) ser på hvordan ulike vurderingspraksiser påvirker elevers motivasjon til å gjøre det bra på prøver.

I denne temaklyngen kan vi finne en undergruppe av forskningsbidrag som omhandler studier på standardiserte nasjonale eller internasjonale prøver. Her finner vi forskning knyttet til blant annet PISA, TIMSS og de ulike nasjonale prøvene og deres betydning i de nordiske landene. Syv studier fra vår kartlegging kan kategoriseres i denne undergruppen (se vedlegg 2).

Vi kan også skille ut en annen undergruppe i denne kunnskapsklyngen, og det er studier knyttet til tester og vurderinger i kroppsøvningsfaget. Selv om vi finner forskningsbidrag knyttet til andre fagspesifikke vurderinger (slik som studier på muntlig eksamen i engelsk), viser vår kartlegging at det er spesielt mange studier knyttet til vurdering i kroppsøvningsfaget. Til sammen fire forskningsbidrag fra kartleggingen er kategorisert i denne undergruppen (se vedlegg 2), men funn fra litteratursøket viser tydelig at dette er et større forskningsfelt. At det er foretatt mye forskning på vurderingspraksisen i akkurat kroppsøvningsfaget, kan skyldes en relativt fremtredende diskusjon knyttet til utfordrende vurderinger i faget – både når det gjelder beskrivelser av fysiske krav i kroppsøving og rettferdig vurdering av elever med ulike fysiske forutsetninger.

6 Diskusjon av kunnskapskartleggingen

6.1 Kunnskapslandskapet

Formålet med vår systematiske kartlegging er å identifisere hva som finnes forskning knyttet til systemet for og organiseringen og gjennomføringen av sluttvurdering og eksamen i grunnskolen og videregående opplæring i Norden. I beskrivelsen av kunnskapslandskapet er det også sentralt å identifisere temaområder det er forsket lite på – såkalte *kunnskapshull* i forskningslitteraturen. Vår systematiske kartlegging viser at det er gjort en del forskning knyttet til eksamen og sluttvurdering i de nordiske landene. Samtidig viser en nærmere gjennomgang av forskningslitteraturen at kunnskapslandskapet er noe fragmentert, med både tydelige kunnskapsklynger og kunnskapshull.

6.1.1 Distinkte forskningsområder

Det er relativt mye forskning på individuelle og eksterne (institusjonelle) faktorer som kan påvirke elevers eksamen- eller sluttkarakterer. Studiene som tar for seg individuelle faktorer, ser blant annet på hvordan elevers kjønn, alder, fysiske og mentale helse, personlighet, intelligens, evne til stresshåndtering, fremtidsplaner, sosioøkonomiske bakgrunn eller lære-, lese- og regnevansker kan påvirke deres eksamens- eller sluttkarakterer.

Studiene som tar for seg eksterne faktorer, ser ofte om institusjonelle faktorer påvirker elevers eksamen- eller sluttkarakterer. Eksterne faktorer er for eksempel innvandrerkonsentrasjon ved en skole, om skolen er privat eller offentlig, om elever blir plassert i en nivådelt klasse eller om landets investering i utdanningssystemet kan påvirke elevers eksamens- eller sluttkarakterer.

Fellesnevneren for denne forskningslitteraturen, er at eksamens- eller sluttkarakterer er satt som utfall i studiene. Mye av forskningen anvender en kvantitativ tilnærming, med bruk av registerdata om elevers sluttresultater. Forsknings-

bidragene studerer dermed ikke selve organiseringen eller systemet rundt eksamen og sluttvurdering i de nordiske landene, men de bidrar til å belyse aspekter som kan ha betydning for elevenes resultater og karakterer. Denne forskningslitteraturen bidrar også til å problematisere rettferdighetsaspektet ved eksamens- og sluttvurderingssystemene i de nordiske landene, ettersom mange av studiene peker på individuelle eller eksterne faktorer utenom elevers skolefaglige kunnskapsnivå som kan påvirke deres resultater.

Det er også foretatt en del studier av ulike vurderingsformer og vurderingspraksiser i grunnskolen og videregående opplæring i de nordiske landene. Blant annet er det gjort en del kvalitative studier om læreres oppfatning av ulike vurderingspraksiser i skolen, samt hvordan elever oppfatter eller blir påvirket av ulike vurderings- og tilbakemeldingspraksiser. I tillegg er det foretatt en del studier knyttet til standardiserte prøver (som nasjonale prøver, TIMSS og PISA). Vi finner også spesielt mange studier knyttet til de fagspesifikke vurderingene i kroppsøvingfaget. Fellesnevneren for mye av denne forskningslitteraturen, er at de ikke hovedsakelig tar for seg eksamen eller sluttvurderinger. I stedet vektlegges den mer generelle vurderings- og tilbakemeldingspraksisen, samt andre prøver og tester i grunnskolen og videregående opplæring i Norden.

6.1.2 Områder med lite forskning

Gjennom vår systematiske kartlegging kommer det også frem noen kunnskapshull i forskningen på sluttvurdering og eksamen i Norden. Generelt sett er det relativt lite forskning på selve eksamen- og sluttvurderingssystemene i grunnskolen og videregående opplæring i de nordiske landene – hvilket nettopp var fokuset for vår systematiske kartlegging. Dette innebærer både forskning på selve systemet for eksamen og sluttvurderinger, men også organiseringen og gjennomføringen av eksamen og sluttvurderinger. Det er følgelig også lite forskning som ser på den helhetlige strukturen og prosessen ved eksamen og sluttvurderinger, fra systemnivå til det mer praksisnære nivået i grunnskolen og videregående skole.

De få forskningsbidragene vi har funnet, tar opp aspekter ved organiseringen, gjennomføringen eller praksis knyttet til eksamen og sluttvurderinger. Her finner vi også et par kvalitative studier som tar opp lærere og elevers perspektiver på og erfaringer med eksamen og sluttvurderinger. Kun et fåtall av forskningsbidragene i denne gruppen omhandler det overordnede systemet for eksamen og/eller sluttvurderinger i grunnskolen eller videregående opplæring.

Vi finner også få studier fra Danmark, Finland og Island. Som tidligere nevnt i metodebegrensningene, kan dette skyldes språkbegrensninger (for Finland og Island) og begrensninger i søkekildene (for Danmark, Finland og Island). Videre er

det få studier som sammenligner selve organiseringen eller systemet rundt eksamen og sluttvurdering på tvers av de nordiske landene.

Det er også få studier som konkret omhandler rettferdighetsperspektivet ved eksamen- og sluttvurderingssystemet i Norden. Riktignok har en del studier blitt gjort på dette området i Sverige, da hovedsakelig knyttet til de svenske nasjonale prøvene. Siden Sverige har et noe annet eksamens- og sluttvurderingssystem, samt har et annet formål med sine nasjonale prøver, er ikke alle disse studiene relevante for en norsk kontekst. Som tidligere nevnt, er det også gjort en del studier knyttet til individuelle og eksterne faktorer som kan påvirke elevers eksamens- og sluttkarakterer. Hovedsakelig studeres faktorer utenfor elevers faglige kunnskapsnivå (som kjønn, fødselsmåned, helse og familiebakgrunn) og hvorledes disse faktorene kan ha negativ effekt på elevers sluttvurdering og eksamensprestasjon. Selv om mange av studiene (ofte indirekte) problematiserer rettferdighetsaspektet ved eksamens- og sluttkarakterer, tar de ikke direkte opp det overordnede rettferdighetsperspektivet.

Mange av kunnskapshullene vi har funnet i vår systematiske kartlegging, samsvarer med kunnskapshullene i Eksamensgruppas kunnskapsgrunnlag for evaluering av eksamensordningen i Norge (Blömeke et al., 2019a). Vi må også understreke at kunnskapshull i forskningslandskapet ikke nødvendigvis betyr komplett fravær av metodisk godt arbeid av høy kvalitet på eksamen i de nordiske landene. Likevel mangler et systematisk forsknings- og kunnskapsgrunnlag knyttet til det overordnede systemet for og organiseringen av eksamen og sluttvurderinger i landene.

6.1.3 Diskusjon

Eksamen og sluttvurderinger har en relativ sentral plass i de ulike utdanningssystemene i de nordiske landene. I tillegg har eksamen og sluttvurderinger stor betydning for den enkelte elevs mulighet for videre skoleopptak, utdanningsløp og yrkesmuligheter. På tross av dette, finner vi store kunnskapshull på området. Blant annet finner vi et begrenset forskningsgrunnlag knyttet til det overordnede systemet, organiseringen og gjennomføringen av eksamen og sluttvurderinger i de nordiske landene. Dette inkluderer mangelfull forskning på den helhetlige prosessen fra systemnivå til det mer praksisnære nivået i grunnskolen og videregående skole. I tillegg mangler også forskning som knytter eksamen- og sluttvurderingssystemet opp mot kvalitet- og rettferdighetsperspektivet. At dette er områder med lite robust og systematisk kunnskap, kan skyldes at det vil være relativt arbeidskrevende å gjennomføre denne typen forskning. Blant annet vil forskning som er relevant for formålet kunne omfatte større forskningsprosjekter og designforsøk i

utvalgte skoler, samt studiedesign som tar hensyn til både kvantitative og kvalitative sider ved ulike nivåer i eksamen- og sluttvurderingssystemet.

Til sammenligning er det mindre krevende å gjennomføre studier som tar for seg enkelte aspekter knyttet til eksamen, prøver, karakterer eller sluttvurderinger i skolen. Dette gjenspeiles i andelen forskningsbidrag knyttet til individuelle og eksterne faktorer som påvirker elevers eksamenskarakterer, ettersom de studerer et avgrenset fenomen med eksamensresultater som utfall. Det samme gjelder studier som sammenligner resultater fra for eksempel nasjonale prøver med elevers standpunktkarakter. Selv om disse forskningsbidragene gir nyttig informasjon om spesifikke deler ved eksamen og vurderinger i skolen, sier de lite om det helhetlige og overordnede systemet for eksamen og sluttvurderinger.

Ettersom det mangler forskning på det overordnede eksamens- og sluttvurderingssystemet i de nordiske landene, er det ikke overraskende at det også mangler gode sammenligninger mellom landene. Selv om det er mange likheter ved utdanningssystemene i de nordiske landene, har landene ulike eksamens- og sluttvurderingssystemer. Ikke all informasjon vil derfor være overførbar eller relevant landene imellom. Likevel kan en sammenligning av de overordnede utdanningssystemene i Norden bidra til å belyse alternative måter for å organisere og gjennomføre eksamen og sluttvurderinger i grunnskolen og videregående opplæring.

7 Konklusjon og mulige forskningsimplikasjoner

I dette avslutningskapitlet vil vi oppsummere konklusjonene fra landsammenligningen og kartleggingen, samt komme med forslag til temaer for videre forskning på feltet sluttvurdering.

Konklusjon fra landsammenligningen

Som kartleggingen av de nordiske utdanningssystemene viser, er oppbyggingen av systemene i stor grad lik på tvers av landene. Med unntak av Finland, hvor obligatorisk grunnskole starter når elevene er syv år, er skolestart i alle landene det året elevene fyller seks, og grunnskolen varer til eleven er 16 år. Det er noe variasjon i fra hvilke skoletrinn karakterer settes. I Sverige får elevene karakterer fra 6. trinn, i Norge, Danmark og Finland fra 8. trinn, mens Island har karakterer fra 10. trinn. Felles for alle er imidlertid at første gang sluttvurdering gjøres er siste skoleår i grunnskolen.

Etter fullført grunnskole tilbyr alle landene videregående opplæring, som elevene søker seg til etter eget ønske. Alle landene skiller mellom studieforberevende program og yrkesfaglige program, som gir ulike former for sluttkompetanse og videre muligheter. I Danmark, Finland og Sverige kreves et vitnemål med bestått i alle fag fra grunnskolen for å kunne få opptak i ordinær videregående opplæring, mens elever i Norge og Island automatisk har rett til opptak i videregående opplæring etter gjennomført grunnskole. I Danmark og Finland tilbys elever som ikke har bestått alle fag i grunnskolen, et 11. skoleår.

Sluttvurderingen som elevene får i videregående opplæring, er hovedgrunnlag for opptak til høyere utdanning i alle land unntatt Finland. I Finland er det en kombinasjon av resultater på Studenteksamen, karakterer på vitnemålet og/eller lærestedsspesifikke prøver som bestemmer opptak til høyere utdanning. Studenteksamen er et sett med prøver som eleven må bestå for å være kvalifisert til å søke høyere utdanning i Finland. Dette kommer i tillegg til at eleven må bestå all fagene i videregående (med karakter satt av lærer).

Alle landene legger til grunn en kriteriebasert vurderingsform, med ferdighets- og kompetansemål fastsatt i en nasjonal læreplan. Sammenhengen mellom vurdering og læreplan står dermed sterkt i alle de nordiske landene. Imidlertid er det noe variasjon knyttet til hvor bindende disse målene er, samt muligheten for lokal tilpasning. Mens både Norge, Danmark og Sverige kun opererer med nasjonale læreplaner, har både Finland og Island innslag av lokale læreplaner som i stor grad styrer undervisningen og vurderingen på hver enkelt skole. De to siste systemene åpner dermed opp for en mer desentralisert og mindre standardisert tilnærming til sluttvurdering enn det vi ser i de øvrige landene.

Det som tydeligst skiller sluttvurderingssystemene i de nordiske landene, er hvorvidt man benytter seg av eksamen som sluttvurderingsform. Per i dag er det kun Norge og Danmark – og til dels Finland – som benytter seg av eksamen, mens Island og Sverige kun har standpunkt karakterer som sluttvurderingsform. Med andre ord innebærer det at i Sverige og Island, og egentlig også i Finland, er det læreren som har faget som også setter karakter på elevene. Studenteksamen i Finland vurderes først av lærer, men blir siden vurdert av eksamensnemnden i tillegg, som setter endelig karakter.

Ulikhet i hvordan sluttvurderingen gjennomføres påvirker også hvordan de ulike systemene går frem for å sikre sluttvurderingens validitet, reliabilitet og rettferdighet. I alle landene settes standpunkt karakter av elevenes faglærer. Dette skal gjøres med utgangspunkt i en rekke ulike vurderingsformer og -metoder, slik at elevene får mulighet til å illustrere hva de kan og hva de har lært på en variert måte. Dette, samt at faglæreren kjenner elevene og har mulighet til å følge dem over tid, er med på å sikre at elevenes reelle kompetanse faktisk fanges opp.

I alle landene vektlegges også viktigheten av tolkningsfelleskap når det gjelder å sikre en likeverdig vurdering av elevene på tvers av klasser og skoler. Mens Norge, Sverige og Danmark vektlegger et nasjonalt tolkningsfelleskap, med sensur- og veiledningsmateriell til å bistå den enkelte faglærer, brukes de lokale læreplanene i Finland og på Island til å sikre en felles forståelse av de fastsatte kompetansemålene.

I tillegg har noen av landene, herunder Norge, Danmark og Sverige, introdusert ytterligere «safe guards» for å sikre at sluttvurderingen gjøres på riktig måte. I Norge og Danmark brukes eksamen som en form for ekstern kvalitetssikring, der elevene fullfører samme prøve og vurderes etter de samme kriteriene. I Sverige fyller de nasjonale prøvene en tilsvarende funksjon, ved at de bidrar til en likeverdig vurdering av elevenes måloppnåelse på tvers av de ulike faglærerne. I tillegg åpner alle de nordiske landene opp for muligheten for å klage på sluttvurderingen, selv om denne muligheten begrenser seg til formelle og prosessuelle feil når det gjelder standpunkt karakter. Mens man i Norge kan be om en ny innholdsmessig

vurdering av eksamenskarakteren, begrenses muligheten til å klage på eksamen i Danmark seg til såkalte «rettslige feil», dvs. formelle feil ved vurderingsprosessen.

Konklusjon fra kartleggingen

I vår systematiske kartlegging har vi hovedsakelig rettet oppmerksomheten mot forskning som tar for seg selve systemet for og organiseringen og gjennomføringen av sluttvurdering og eksamen i de nordiske landene. Vår kartlegging viser at det er gjort en del forskning knyttet til dette området i Norden. En nærmere gjennomgang av litteraturen viser derimot at kunnskapslandskapet har tydelige kunnskapsklynger og kunnskapshull.

Når det gjelder kunnskapshull, finner vi et begrenset forskningsgrunnlag knyttet til det overordnede systemet, organiseringen og gjennomføringen av eksamen og sluttvurderinger i de nordiske landene. Dette inkluderer manglende forskning på den helhetlige prosessen fra systemnivå til det mer praksisnære nivået i grunnskolen og videregående opplæring. Ettersom det mangler forskning på selve organiseringen og systemet rundt eksamen og sluttvurderinger i de nordiske landene, er det heller ikke overraskende at det også er begrenset med gode sammenligninger mellom landene. Det er også relativt lite forskning som knytter eksamen og sluttvurderingssystemet i Norden opp mot kvalitet- og rettferdighetsperspektivet. Ettersom denne tematikken var et sentralt utgangspunkt for vår systematiske kartlegging – noe som også gjenspeiles i vårt litteratursøk og seleksjonskriterier for kartleggingen – var det spesielt påfallende hvor få nordiske forskningsbidrag vi fant på akkurat dette området.

Til sammen identifiserte vi 19 forskningsbidrag av høy relevans som omhandler selve organiseringen av eller systemet rundt eksamen og sluttvurderinger i de nordiske landene. Studiene kategoriserte vi i tre hovedområder. Denne grupperingen ble valgt ettersom den får frem ulike nivåer og perspektiver på eksamen og sluttvurderingssystemet på tvers av de nordiske landene:

- Studier som tar for seg et overordnet systemperspektiv på eksamen og sluttvurdering (tre studier).
- Studier som tar for seg institusjonelle faktorer og organiseringen av eksamen og sluttvurderinger.
- Studier som tar for seg gjennomføringen, praksiser og erfaringer knyttet til eksamen og vurderinger (syv studier).

Selv om vi fant relativt få forskningsbidrag knyttet til selve organiseringen og systemet for eksamen og sluttvurdering i de nordiske landene, kunne vi identifisere tre relaterte kunnskapsklynger:

- Forskning på individuelle faktorer ved eleven som påvirker eksamen.
- Forskning på eksterne faktorer som påvirker eksamen.
- Forskning på andre vurderingsformer (herunder fagspesifikke vurderingsformer og standardiserte prøver).

Forskningsbidragene fra disse kunnskapsklyngene tok ikke opp selve organiseringen eller systemet rundt eksamen og sluttvurdering. Likevel bidro forskningsbidragene fra kunnskapsklyngene til å belyse ulike variabler som kan påvirke elevens eksamen- og sluttkarakterer, samt belyse spesifikke sider ved eksamen, prøver eller vurderingspraksiser.

Mulige forskningsimplikasjoner

På bakgrunn av funnene fra del 1 og del 2 kan vi trekke frem noen forslag som kan være relevante for videre forskning på feltet. Forslagene reflekterer særlig kunnskapshullene og kunnskapsklyngene som ble identifisert:

- Studier som tar for seg elevperspektivet på sluttvurdering.
- En mer helhetlig tilnærming i forskningen på eksamen og sluttvurderinger i skolen, dvs. studier som ser vurdering og sluttvurdering på ulike nivåer i forhold til hverandre.
- Studier som tar for seg rettferdighetsperspektivet i organisering av sluttvurdering, og hvordan individuelle faktorer påvirker sluttvurderingens reliabilitet.
- Internasjonale sammenligninger av sluttvurdering i ulike land; organisering, ressursbruk, involvering av ulike aktører, prinsipper for utforming av systemer og ordninger.
- Mer inngående studier og sammenligninger av de ulike sluttvurderingssystemene i Norden, for eksempel hvorfor noen land har eksamen og andre ikke, og hva forskjellene betyr for organisering, ressursbruk, elever og andre aktører.

I tillegg vil mer kunnskap kunne frembringes ved supplerende kunnskapskartleggingen, blant annet ved å utvide den geografiske rekkevidden til europeiske land utenfor Norden (slik som Sentral-Europa), eller initiere egne systematiske kartlegginger for de kunnskapsklyngene som er skissert i denne rapporten.

Referanser

Referanser merket med * er inkludert i kunnskapskartleggingen

- Aksnes, D. W., & Sivertsen, G. (2019). A Criteria-based Assessment of the Coverage of Scopus and Web of Science. *Journal of Data and Information Science*, 4(1), 1-21. <https://doi.org/10.2478/jdis-2019-0001>
- Aksnes, D. W., Bergene, A. C., Fossum, L. W. og Wollscheid, S. (2021). *Nordisk forskning om diskriminering, trakassering og likestilling: En forenklet kunnskapsoversikt*. NIFU-rapport 2021:14. <https://hdl.handle.net/11250/2830908>
- Andresen, S., Fossum, A., Rogstad, J. og Smestad, B. (2017). På prøve. Evaluering av matematikkeksamen på 10. trinn våren 2017. Fafo-rapport 2017: 36.
- Almen eksamensbekendtgørelsen. (2016). *Bekendtgørelse om prøver og eksamen i de almene og studieforberedende ungdoms- og voksenuddannelser*. (BEK nr 343 af 08/04/2016). Retsinformation. <https://www.retsinformation.dk/eli/lta/2016/343>
- Bekendtgørelse om de gymnasiale uddannelser. (2017). *Bekendtgørelse om de gymnasiale uddannelser (BEK nr 497 af 18/05/2017)*. Retsinformation. <https://www.retsinformation.dk/eli/lta/2017/497>
- *Bensnes, S. S. (2019). Scheduled to gain: Short- and longer-run educational effects of examination scheduling. *The Scandinavian Journal of Economics*, 122(3). <https://doi.org/10.1111/sjoe.12363>
- *Bjørnset, M., Fossum, A., Rogstad, J. og Smestad, B. (2020). *På like vilkår? Evaluering av matematikkeksamen på 10. trinn 2017-2019*. Fafo-rapport 2020:01. <https://www.fafo.no/images/pub/2020/20736.pdf>
- Bjørnset, M., Fossum, A., Rogstad, J., Smestad, B. og Talberg, N. (2018). Digitale skillelinjer. Evaluering av matematikkeksamen på 10. trinn våren 2018. Fafo-rapport 2018: 36.
- *Björnsson, J.K. og Skar, G. B. (2021). *Sensorreliabilitet på skriftlig eksamen i videregående opplæring: Estimer av sensorreliabilitet basert på analyser av karakterforslag fra to uavhengige sensorer før felles-sensur - Resultater fra 40 utvalgte fag*. Universitetet i Oslo. <https://www.duo.uio.no/bitstream/handle/10852/85597/Bj%25C3%25B6r>

[nsson-Skar-2021-Sensorreliabilitet-pa-skriftlig-eksamen-i-vgo.pdf?sequence=1&isAllowed=y](#)

Blömeke, S., Skillinghaug, S., Blikstad-Balas, M., Eggen, P. O., Fjørtoft, H., Gamlem, S. T. M., Prøitz, T., Tveit, S., Helgesen, R., Johannessen, S., Minken, M., Waage, A. og Walker, M. J. (2019a). *Kunnskapsgrunnlag for evaluering av eksamensordningen*. Utdanningsdirektoratet. <https://www.udir.no/tall-og-forskning/finn-forskning/rapporter/Kunnskapsgrunnlag-for-evaluering-av-eksamensordningen/>

Blömeke, S., Skillinghaug, S., Blikstad-Balas, M., Eggen, P. O., Fjørtoft, H., Gamlem, S. T. M., Tveit, S., Gilje, Ø., Eira, K. I., Helgesen, R., Johannessen, S., Minken, M., Waage, A. og Walker, M. J. (2019b). *Vurderinger og foreløpige anbefalinger fra Eksamensgruppa*. Utdanningsdirektoratet. <https://www.udir.no/tall-og-forskning/finn-forskning/rapporter/vurderinger-og-forelopige-anbefalinger-fra-eksamensgruppa/>

*Bukh, P. N., Christensen, K. S. & Poulsen, M. L. (2022). Performance funding: Exam results, stakes, and washback in Danish schools. *SAGE Open*. <https://doi.org/10.1177%2F21582440221082100>

*Bøhn, H. (2018). Assessing content in a curriculum-based EFL oral exam: The importance of higher-order thinking skills. *Journal of Language Teaching and Research*, 9(1). <http://dx.doi.org/10.17507/jltr.0901.03>

Børne- og Undervisningsministeriet (2017). Læreplan Dansk A – toårigt hf, august 2017 [bilag 4] i *Hf – læreplaner 2017*. Børne- og Undervisningsministeriet. <https://www.uvm.dk/gymnasiale-uddannelser/fag-og-laereplaner/laereplaner-2017/hf-laereplaner-2017>

Børne- og Undervisningsministeriet. (2021a, 4. november). *Faghæfte - Fælles Mål, læseplan og vejledning*. <https://emu.dk/grundskole/engelsk/faghaefte-faelles-maal-laeseplan-og-vejledning>

Børne- og Undervisningsministeriet. (2021b). *Prøve- og eksamensinformation for gymnasiale uddannelser og eux-forløb: Gældende for tilrettelæggelse af uddannelser, der er påbegyndt 1. august 2017 eller senere*. Børne- og Undervisningsministeriet.

Børne- og undervisningsministeriet. (2022a). *Bonusordning. Mulighed for justering af karaktergjennemsnit*. UddannelsesGuiden. <https://www.ug.dk/videregaendeuddannelse/altomoptagelse/bonusordning>

Børne- og undervisningsministeriet. (2022b). *Gymnasiale uddannelser*. Børne- og undervisningsministeriet. <https://www.uvm.dk/gymnasiale-uddannelser>

Børne- og Undervisningsministeriet. (2022c, 21. juli). *Lempelse af bindinger i regelsættet om Fælles Mål*. <https://www.uvm.dk/folkeskolen/fag-timetal-og->

[overgange/faelles-maal/lempelse-af-faelles-maal/lempelse-af-bindinger-i-regelsaettet-om-faelles-maal](#)

Børne- og Undervisningsministeriet. (2022d). *Uddannelses Guiden*.

Uddannelsesguiden. <https://www.ug.dk/>

Cheng, L. & Curtis, A. (2003). Washback or Backwash: A Review of the Impact of Testing on Teaching and Learning. I L. Cheng, Y. Watanabe (Red.), *Washback in Language Testing*. New York: Routledge.

Compulsory School Act. (2008). *Icelandic Compulsory School Act, Edt. 145b*.

Ministry of Education, Science and Culture.

https://www.government.is/media/menntamalaraduneyti-media/media/frettatengt2016/91_2008-Compulsory-School-Act-ENGLISH-Uppfaert-Jan-2017.pdf

Danmarks evalueringsinstitut. (2016). *Årskarakterer og prøvekarakterer i det almene gymnasium (stx)*. Danmarks evalueringsinstitut.

https://www.eva.dk/sites/eva/files/2017-07/Notat%20om%20arskarakterer%20og%20provekarakterer%20pa%20stx%2021.11.2016%20Endelig_www.pdf

European Commission. (2022a, 20. juli). *Eurydice: Denmark*. Eurydice.

<https://eurydice.eacea.ec.europa.eu/national-education-systems/denmark/overview>

European Commission. (2022b, 16. juni). *Eurydice: Finland*. Eurydice.

<https://eurydice.eacea.ec.europa.eu/national-education-systems/finland/finland>

European Commission. (2022c, 10. juni). *Eurydice: Iceland*. Eurydice.

<https://eurydice.eacea.ec.europa.eu/national-education-systems/iceland/iceland>

European Commission. (2022d, 15. juli). *Eurydice: Sweden*. Eurydice.

<https://eurydice.eacea.ec.europa.eu/national-education-systems/sweden/sweden>

*Falch, T., Nyhus, O. H. & Strøm, B. (2014). Causal effects of mathematics. *Labour Economics*, 31. <http://dx.doi.org/10.1016/j.labeco.2014.07.016>

Folkeskoleloven. (2021). *Bekendtgørelse af lov om folkeskolen*. (LBK nr 1887 af 01/10/2021). Retsinformation.

<https://www.retsinformation.dk/eli/lta/2021/1887>

Forskrift til opplæringslova. (2006). *Forskrift til opplæringslova*. (FOR-2006-06-23-724). Lovdata. <https://lovdata.no/LTI/forskrift/2006-06-23-724>

Förordning om grundläggande utbildning (1998). *Förordning om grundläggande utbildning* (20.11.1998/852). Finlex.

<https://www.finlex.fi/sv/laki/ajantasa/1998/19980852>

- *Gustavsson, J-E & Ericksson, G. (2018). Nationella prov i Sverige – tradition, utmaning, förändring. *Acta Didactica Norge*, 12(4), <https://doi.org/10.5617/adno.6434>
- Gymnasielag (2018). *Gymnasielag*. (714/2018). Finlex. <https://www.finlex.fi/sv/laki/alkup/2018/20180714>
- *Hovdhaugen, E., Prøitz, T. S. og Seland, I. (2018). Eksamens- og standpunktkarakterer – to sider av samme sak?. *Acta Didactica Norge*, 12(4), <https://doi.org/10.5617/adno.6276>
- Hovdhaugen, E., Sandsør, A. M. J., Rønsen, E. og Carlsten, T.C. (2020). Opptak til høyere utdanning: En studie av konsekvensene av kvote og poengberegning med fokus på psykologi, medisin, jus og grunnskolelærerutdanning. NIFU-rapport 2020:4. NIFU. <https://hdl.handle.net/11250/2648860>
- *Hovdhaugen, E., Seland, I., Lødding, B., Prøitz, T. og Vibe, N. (2014). *Karakterer i offentlige og private videregående skoler: En analyse av eksamens- og standpunktkarakter i norsk og matematikk og rutiner for standpunktvurdering i offentlige og private videregående skoler*. NIFU-rapport:2014-24. NIFU. <http://hdl.handle.net/11250/280144>
- *Hvidman, U. & Sievertsen, H. H. (2021). High-stakes grades and student behaviour. *Journal of Human Resources*, 56(3). doi: 10.3368/jhr.56.3.0718-9620R2
- James, K. L., Randall, N. P., & Haddaway, N. R. (2016). A methodology for systematic mapping in environmental sciences. *Environmental evidence*, 5(1), 1-13. <https://doi.org/10.1186/s13750-016-0059-6>
- Jónasson, J. T. (2004). Higher education reforms in Iceland at the transition into the twenty-first century, in Fägerlind, Ingemar & Strömqvist, Görel (eds). *Reforming higher education in the Nordic countries – studies of change in Denmark, Finland, Iceland, Norway and Sweden*, 137-188, UNESCO – International Institute for Educational Planning.
- Jónasson, J. T. (1997). Students Passing the Icelandic University Entrance Examination (UEE) 1911-94, *European Journal of Education*, 32(2): 209-220
- *Kupiainen, S., Marjanen, J. & Hautamäki, J. (2016). The problem posed by exam choice on the comparability of results in the Finnish matriculation examination. *Journal of Educational Research Online*, 8(2). <http://hdl.handle.net/10138/232733>
- Lag om grundläggande utbildning. (1998). *Lag om grundläggande utbildning* (628/1998). Finlex. <https://www.finlex.fi/sv/laki/smur/1998/19980628>
- Leventhal, G.S. (1980). What Should Be Done with Equity Theory? New Approaches to the Study of Fairness in Social Relationships. I K. J. Gergen, M. S. Greenberg, & R. H. Willis (Red.), *Social Exchange: Advances in Theory and Research*. Plenum. doi:10.1007/978-1-4613-3087-5_2

- Liebig, S. (2006). Soziale Gerechtigkeit—Modelle und Befunde der soziologischen Gerechtigkeitsforschung. I O. Höffe, S. Liebig, and B. von Maydell (Red.), *Fachgespräch Gerechtigkeit*. Berlin-Brandenburgische Akademie der Wissenschaften.
- Lilja, P. (2011). Lärarlegitimation-professionalisering med förhinder? *Arbetsmarknad & Arbetsliv*, 17(4), 29-42.
- Lov om de gymnasiale uddannelser. (2020). *Bekendtgørelse af lov om de gymnasiale uddannelser*. (LBK nr 957 af 22/06/2022). Retsinformation. <https://www.retsinformation.dk/eli/lt/2022/957>
- Lundahl, C., Hultén, M. & Tveit, S. (2016). *Betygssystem i internationell belysning*. Stockholm: Skolverket.
- Markus, K. A. & Borsboom, D. (2013). *Frontiers of Test Validity Theory: Measurement, Causation, and Meaning*. New York: Routledge.
- Ministry of Education, Science and Culture (2012). *The Icelandic National Curriculum Guide for Compulsory Schools – General Section*. https://www.government.is/library/01-Ministries/Ministry-of-Education/Curriculum/adalnorsk_greinask_ens_2014.pdf
- *Modell, N. & Gerdin, G. (2021). 'But in PEH it still feels extra unfair': students' experiences of equitable assessment and grading practices in physical education and health (PEH). *Sport, Education and Society*. <https://doi.org/10.1080/13573322.2021.1965565>
- Opplæringslova. (1998). *Lov om grunnskolen og den vidaregåande opplæringa*. (LOV-1998-07-17-61). Lovdata. <https://lovdata.no/dokument/NL/lov/1998-07-17-61>
- Pellegrino, J.W., Chudowsky, N., Glaser, R., & National Research Council (U.S.) (Red.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Prøvebekendtgørelsen. (2021). *Bekendtgørelse om folkeskolens prøver*. (BEK nr 1858 af 27/09/2021). Retsinformation. <https://www.retsinformation.dk/eli/lt/2021/1858>
- *Prøitz, T. S. (2013). Variations in grading practice – subjects matter. *Education Inquiry*, 4(3). <https://doi.org/10.3402/edui.v4i3.22629>
- *Sandvik, L. V. & Sommervold, O. A. (2021). Students' perceptions of involvement in the assessment of oral competence in English as a second language. *Languages*, 6. <https://doi.org/10.3390/languages6040000>
- *Sandvik, L. V., Smith, K., Strømme, A. et al. (2021). Students' perceptions of assessment practices in upper secondary school during COVID-19. *Teacher and Teaching*. <https://doi.org/10.1080/13540602.2021.1982692>
- *Sjaastad, J., Carlsten, T.C., Wollscheid, S. (2016). *Får elevene den opp-læringen de har krav på? Kartlegging av undervisningstimer med kvalifiserte lærere i*

- videregående opplæring. NIFU-Rapport 2016:26.
<http://hdl.handle.net/11250/2425531>
- Skolverket. (2022a, 29. juli). *Antagning till gymnasieskolan*. Skolverket.
<https://www.skolverket.se/regler-och-ansvar/ansvar-i-skolfragor/antagning-till-gymnasieskolan>
- Skolverket (2022b). *Bedömning i grundskolan*. Skolverket.
<https://www.skolverket.se/undervisning/grundskolan/bedomning-i-grundskolan>
- Skolverket. (2022c). *Läroplan för grundskolan samt för förskoleklassen och fritidshemmet (Lgr22)*. Skolverket.
<https://www.skolverket.se/undervisning/grundskolan/laroplan-och-kursplaner-for-grundskolan>
- Skolverket (2022d). *Läroplan, program och ämnen i gymnasieskolan*. Skolverket.
<https://www.skolverket.se/undervisning/gymnasieskolan/laroplan-program-och-amnen-i-gymnasieskolan>
- Skolverket. (2022e, 7. april). *Nationella prov*. Skolverket.
<https://www.skolverket.se/innehall-a-o/landningssidor-a-o/nationella-prov>
- Skolverket. (2022f, 19. juli). *Sätta betyg i gymnasieskolan*. Skolverket.
<https://www.skolverket.se/undervisning/gymnasieskolan/betyg-i-gymnasieskolan/satta-betyg-i-gymnasieskolan>
- *Skovholt, K., Solem, M. S., Vonen, M. N., Sikveland, R. O., & Stokoe, E. (2021). Asking more than one question in one turn in oral examinations and its impact on examination quality. *Journal of Pragmatics*, 181.
<https://doi.org/10.1016/j.pragma.2021.05.020>
- Statsrådsberedningen. (2018, 19. februar). *Grundläggande behörighet på yrkesprogram*. Regeringskansliet.
<https://www.regeringen.se/pressmeddelanden/2018/02/grundlaggande-behorighet-pa-yrkesprogram/>
- Stefansson, Thorarinn & Karlsdottir, Raghheidur (2008). Iceland, I W. Hörne, H. Döbert, B. von Kopp & W. Mitter (Red.), *The Education Systems of Europe* (s. 356-377). Springer.
- Studentexamensnämnden. (2022, 12. august). *Begäran om omprövning*. Ylioppilastutkintolautakunta Studentexamensnämnden.
<https://www.ylioppilastutkinto.fi/sv/studentexamen/omprovning>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garrity, C., ... Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, 169(7), 467-473. doi: 10.7326/M18-0850

- *Tveit, S. og Olsen, R. V. (2018). Eksamens mange roller I sertifisering, styring og støtte av læring og undervisning i norsk grunnopplæring. *Acta Didactica Norge*, 12(4). <https://doi.org/10.5617/adno.6381>
- Uddannelses- og Forskningsministeriet. (2021, 20. oktober). Karakterskalaer. Uddannelses- og Forskningsministeriet. <https://ufm.dk/uddannelse/anerkendelse-og-dokumentation/dokumentation/karakterskalaer>
- Universitets- och högskolerådet. (2022, 22. februar). *Plastfördelning och urval. Antagning.se*. <https://www.antagning.se/sv/efter-anmalan/plastfordelning-och-urval/>
- Universitets- og høyskoleloven. (2005). *Lov om universiteter og høyskoler* (LOV-2005-04-01-15). Lovdata. <https://lovdata.no/lov/2005-04-01-15>
- University of Iceland. (u. å.). *Admission requirements in undergraduate studies*. University of Iceland. https://english.hi.is/university/admission_requirements
- Upper Secondary Education Act. (2008). *Icelandic Upper Secondary Education Act*. Ministry of Education, Science and Culture. <https://www.government.is/media/menntamalaraduneyti-media/media/law-and-regulations/Upper-Secondary-Education-Act-No.-92-2008.pdf>
- Utbildningsdepartementet. (2010).. *Skollag (2010:800)*. Sveriges Riksdag. https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/skollag-2010800_sfs-2010-800
- Utbildningsstyrelsen (2014). *Grunderna för läroplanen för den grundläggande utbildningen 2014*. https://maailmansuurinvanhempainilta.fi/sites/default/files/documents/s/grunderna_for_laroplanen_for_den_grundlaggande_utbildningen_2014.pdf
- Utbildningsstyrelsen (2018). *Finnish education in a nutshell*. <https://www.oph.fi/en/statistics-and-publications/publications/finnish-education-nutshell>
- Utbildningsstyrelsen. (2019). *Grunderna för gymnasiets läroplan 2019*. https://www.oph.fi/sites/default/files/documents/grunderna_for_gymnasiet_s_laroplan_2019.pdf
- Utdanningsdirektoratet [Udir]. (2017). *Overordnet del – verdier og prinsipper for grunnopplæringen*. Utdanningsdirektoratet. <https://www.udir.no/lk20/overordnet-del/?lang=nob>
- Utdanningsdirektoratet [Udir]. (2020). *Vurderingsordning, Norsk (NOR01-06)*. Utdanningsdirektoratet. <https://www.udir.no/lk20/nor01-06/vurderingsordning?lang=nob>

Utdanningsdirektoratet [Udir]. (2021, 10. mai). *Individuell vurdering*.

Utdanningsdirektoratet.

<https://www.udir.no/regelverkstolkninger/opplaring/Vurdering/udir-2-2020-individuell-vurdering/>

Utdanningsdirektoratet [Udir]. (2022, 28. januar). *Trekkordning ved eksamen for grunnskole og videregående opplæring. Udir-2-2018*. Utdanningsdirektoratet.

<https://www.udir.no/regelverkstolkninger/opplaring/eksamen/trekkordning-ved-eksamen-for-grunnskole-og-videregaende-opplaring-udir-2-2018/>

*Waldow, F. (2014). Conceptions of Justice in the Examination Systems of England, Germany and Sweden: A Look at Safeguards of Fair Procedure and Possibilities of Appeal. *Comparative Education Review*, 58(2), 322-343.

<https://doi.org/10.1086/674781>

Wollscheid, S., Bergene, A.C. og Sutherland D. O. (2020). *Fleksibel opplæring for voksne. En kunnskapsoppsummering*. NIFU-rapport 2020:29. NIFU.

<https://hdl.handle.net/11250/2728210>

Vedlegg 1

Skjema: Koder/ dataekstraksjon

- Reviewer
- Forfatternavn
- År
- Tittel
- Publiseringstype
- Navn publiseringskanal
- Land
- Språk
- Metode/forskningsdesign
- Temaer
- Subtemaer
- Populasjon
- Kunnskapshull
- Kommentar

Vedlegg 2

Tabell V1: Studier med lav relevans etter temaklynger

| Individuelle påvirkende faktorer på eksamen | | |
|--|--|---------|
| Nr. | Referanse | Land |
| 1 | Bensnes, 2016, You sneeze, you lose: The impact of pollen exposure on cognitive performance during high-stakes high school exams | Norge |
| 2 | Berg et al., 2020, Gender grading bias in junior high school mathematics | Sverige |
| 3 | Boman, 2021, Regional Differences in Educational Achievement among Swedish Grade 9 Students | Sverige |
| 4 | Brown & Eklöf, 2018, Swedish student perceptions of achievement practices: The role of intelligence. | Sverige |
| 5 | Elstad & Bakken, 2015, The effects of parental income on Norwegian adolescents' school grades: A sibling analysis | Norge |
| 6 | Ericsson & Cederberg, 2015, Physical activity and school performance: a survey among students not qualified for upper secondary school | Sverige |
| 7 | Evensen, 2019, Adolescent mental health problems, behaviour penalties and distributional variation in educational achievement | Norge |
| 8 | Heiskala et al., 2020, Compensatory and Multiplicative Advantages: Social Origin, School Performance, and Stratified Higher Education Enrolment in Finland | Finland |
| 9 | Holopainen et al., 2017, Implications of Overlapping Difficulties in Mathematics and Reading on Self-Concept and Academic Achievement | Finland |

| | | |
|--|--|---------|
| 10 | Hotulainen et al., 2020, Development of learning to learn competence across secondary education and its association with attainment in Finnish/Swedish high-stake exit exam | Finland |
| 11 | Ollfors & Andersson, 2021, Influence of Personality Traits, Goals, Academic Efficacy, and Stressload on Final Grades in Swedish High School Students | Sverige |
| 12 | Sæle et al., 2016, Demographic and psychological predictors of grade point average (GPA) in North-Norway: a particular analysis of cognitive/ school-related and literacy problems | Norge |
| 13 | Sagtun et al., 2014, Mental health problems in the 10th grade and non-completion of upper secondary school: the mediating role of grades in a population-based longitudinal study | Norge |
| 14 | Sivertsen et al., 2016, Cognitive fatigue influences students' performance on standardized tests | Danmark |
| 15 | Vestheim et al., 2019, A Population Study of Relative Age Effects on National Tests in Reading Literacy | Norge |
| Eksterne påvirkende faktorer på eksamen (innvandrerkonsentrasjon på skolen, etc.) | | |
| 1 | Bogetoft et al., 2015, The efficiency of educational production: A comparison of the Nordic countries with other OECD countries | Norden |
| 2 | Hardoy et al., 2018, Immigrant concentration and student outcomes in upper secondary schools: Norwegian evidence | Norge |
| 3 | Karlson, 2015, Expectations on Track? High School Tracking and Adolescent Educational Expectations | Danmark |
| 4 | Kortelainen & Manninen, 2019, Effectiveness of Private and Public High Schools: Evidence from Finland | Finland |
| Omtale av form på eksamen, tester, vurderinger | | |
| 1 | Christenson & Rundgren, 2015, A framework for teachers' assesment of socio-scientific argumentation: An example using the GMO issue | Sverige |
| 2 | Dahlin, 2012, Comparison of communication skills between medical students admitted after interviews or on academic merits | Sverige |

| | | |
|---|---|---------|
| 3 | Granbom, 2016, Formative Assessment and Increased Student Involvement Increase Grades in an Upper Secondary School Biology Course | Sverige |
| 4 | Jalava et al., 2014, Grades and rank: Impacts of non-financial incentives on test performance | Sverige |
| 5 | Johannesson & Bjarnadottir, 2016, Meaningful education for returning-to-school students in a comprehensive upper secondary school in Iceland | Island |
| 6 | Mäkipää, 2021, What Kind of Feedback is Perceived as Encouraging by Finnish General Upper Secondary School Students? | Finland |
| 7 | Pitkänen, 2022, The politics of pupil self-evaluation: A case of Finnish assessment policy discourse | Finland |
| 8 | Scherer & Siddiq, 2015, The Big-FisheLittle-Pond-Effect revisited: Do different types of assessments matter? | Norge |
| 9 | Sikveland et al., 2021, How teachers use prosody to guide students towards an adequate answer | Norge |
| Kroppsøving | | |
| 1 | Hagen et al., 2022, Association Between Motor Competence, Physical Fitness, and Academic Achievement in Physical Education in 13- to 16-Year-Old School Children | Norge |
| 2 | Leirhaug & Annerstedt, 2015, Assessing with new eyes? Assessment for learning in Norwegian physical education | Norge |
| 3 | Lyngstad et al., 2020, Norwegian upper secondary students' experiences of their teachers' assessment of and for learning in physical education: examining how assessment is interpreted by students of different physical abilities | Norge |
| 4 | Svennberg, 2017, Swedish PE teachers' understandings of legitimate movement in a criterion-referenced grading system | Sverige |
| Omhandler nasjonale prøver og/eller andre former for standardiserte prøver (PISA, TIMSS) | | |
| 1 | Gustafsson & Erickson, 2013, To trust or not to trust?-teacher marking versus external marking of national tests | Sverige |

| | | |
|--------------|--|---------|
| 2 | Larsson & Olin-Scheller, 2020, Adaptation and resistance: washback effects of the national test on upper secondary Swedish teaching | Sverige |
| 3 | Lofgren et al., 2018, Pupils' enactments of a policy for equivalence: Stories about different conditions when preparing for national tests | Sverige |
| 4 | Novak, 2017, Juridification of examination systems: extending state level authority over teacher assessments through regrading of national tests | Sverige |
| 5 | Tanner & Prieto, 2017, " ... When It Is Us the Tests Are Made for". Students' Argumentations in a Performative Education System | Sverige |
| 6 | Wiberg, 2019, The relationship between TIMSS mathematics achievements, grades, and national test scores | Sverige |
| 7 | Wiberg & Rolfsman, 2019, The association between science achievement measures in schools and TIMSS science achievements in Sweden | Sverige |
| Annet | | |
| 1 | Thomsen, 2018, Test-Based Admission to Selective Universities: A Lever for First-Generation Students or a Safety Net for the Professional Classes? | Danmark |

Nordisk institutt for studier av
innovasjon, forskning og utdanning

Nordic institute for Studies in
Innovation, Research and Education

www.nifu.no