

Rapport: Utvikling av ny eksamen i norsk for 10. trinn, vg2yf og vg3

Anne Holten Kvistad, Gustaf B. Skar, Kari Anne Nordgård, Linn Maria Magerøy-Grande, Marita Byberg Johansen, Trude Kringstad, Marthe Lønnum, Runa Eugenie Røttereng Greiner, Espen Tørset, Ingeborg Margrete Berge, Margrethe Sønneland, Tone Stavli, Ingrid Ertesvåg, Arne Johannes Aasen

14.02.24

Innhold

Rapport: Utvikling av ny eksamen i norsk for 10. trinn, vg2yf og vg3	1
1 Utvikling av ny eksamen i norsk for 10. trinn, vg2yf og vg3.....	4
1.1 Kort om oppdraget og denne rapporten	4
1.2 Arbeidsfordeling	4
1.3 Oppsummering av prosessen	5
1.3.1 Begrensninger og endrede rammer for Skrivesenterets arbeid	5
1.3.2 Endringer i norskeksamen som følge av Skrivesenterets arbeid	6
1.3.3 Skrivesenterets anbefalinger	6
2 Konstruktutvikling	7
3 Oppgaveutvikling	9
3.1 Forankring i rammeverk og konstrukt	9
3.2 Pilotering av oppgaver	10
3.2.1 Tilbakemelding fra lærere og elever om norskfaglig relevans og elevers forståelse av oppgavene.....	11
3.2.2 Tilbakemelding fra lærere og elever om antall oppgaver og tidsbruk	11
3.2.3 Lesing av elevbesvarelser for å vurdere forståelse og tidsbruk.....	12
3.2.4 Utprøving av sidemålsoppgaver	15
4 Kriterieutvikling	16
4.1. Utvikling og utprøving av analytiske kriterier	17
4.2. Vurderingskriterier for eksamen på 10. trinn.....	18
4.2.1 Utviklingsprosess	18
4.2.2. Utprøving av holistiske kriterier med og uten vektning 10. trinn	19
4.3. Vurderingskriterier for eksamen på vg2yf	20
4.3.1. Utviklingsprosess.....	20
4.4. Vurderingskriterier for eksamen på Vg3	21
4.4.1. Utviklingsprosess.....	21
4.4.2. Analytisk vs. holistisk vurdering vår 2022	22
4.4.3. Analytisk vs. holistisk vurdering høst 2023.....	23
5 Avslutning.....	25
6. Vedlegg	26
VEDLEGG 1: Testkonstrukt for norskeksamen på 10. trinn.....	26

1 Utvikling av ny eksamen i norsk for 10. trinn, vg2yf og vg3

1.1 Kort om oppdraget og denne rapporten

Skrivesenteret ved NTNU har hatt i oppdrag å bistå Utdanningsdirektoratet med å utvikle ny eksamen i norsk for 10. trinn, vg2yf og vg3 i perioden mars 2020 og ut 2023. I oppdraget var det spesifisert at den første nye eksamenen for 10. trinn (og ev. vg2yf) skulle avholdes våren 2022. Den første nye eksamenen for vg3 skulle avholdes våren 2023. Det ble avklart i august 2020 at eksamen for vg2yf skulle være sentralt gitt. Dermed ble denne eksamenen også en del av oppdraget. Oppdraget for 10. trinn og vg2yf hadde opprinnelig prosjektslutt i 2022, mens oppdraget for vg3 skulle avsluttes i 2023. Førstnevnte oppdrag ble imidlertid prolongert til årsskiftet 2023/2024 på grunn av at Utdanningsdirektoratet ønsket flere utprøvinger.

Oppdraget fra Utdanningsdirektoratet besto i å utvikle eksamensoppgaver, eksempeloppgaver, konstrukt for eksamen, beskrivelse av hvordan eksamen i norsk skal se ut, eksamensmateriell, vurderingskriterier og annet støttemateriell. En del av oppdraget var å heve kvaliteten på sensuren gjennom å utvikle støtteressurser for sensorer, og det var presisert at oppdragstaker skulle ta i bruk psykometriske metoder for å sikre reliabilitet og validitet i eksamen. Det var også presisert at eksamen i norsk og tilhørende dokumenter må samsvare med retningslinjene i dokumentet «Rammeverk for eksamen – LK20 og LK20S». Dette dokumentet var under revisjon ved prosjektoppstart.

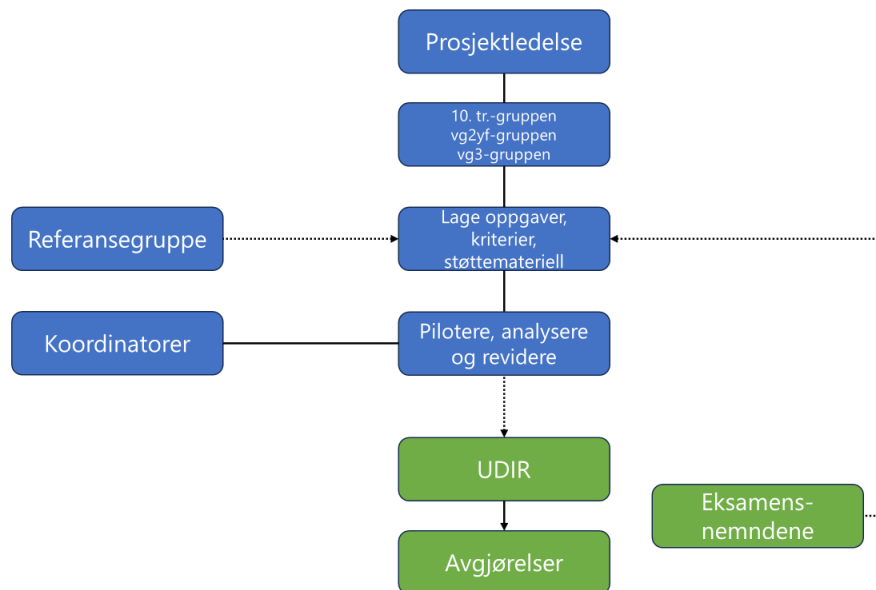
Denne rapporten beskriver Skrivesenterets arbeid med å levere produkter i henhold til avtalen mellom Utdanningsdirektoratet og NTNU/Skrivesenteret. Rapporten er, etter ønske fra Utdanningsdirektoratet, strukturert slik at vi i første del oppsummerer hvilke endringer Skrivesenterets arbeid har ført til, og hva som ble annerledes enn vi hadde tenkt. Vi avslutter denne første delen med våre videre anbefalinger. Delene som følger inneholder informasjon om konstruktutvikling, oppgaveutvikling og kriterieutvikling.

1.2 Arbeidsfordeling

Skrivesenteret dannet tre grupper, en for hvert trinn, for å levere og pilotere forslag til oppgaver, kriterier og støttemateriell (se figur 1). Disse gruppene rapporterte til prosjektleder og direkte til Utdanningsdirektoratet. Skrivesenteret hadde sin egen referansegruppe som gav innspill på oppgaver, kriterier og støttemateriell, oftest før disse produktene ble sendt til Utdanningsdirektoratet. Skrivesenteret fikk også innspill fra Utdanningsdirektoratets eksamensnemnder. Prosjektets koordinatorene hadde ansvar for å organisere utprøvinger (piloterings) og for å samle inn data fra disse. Etter endt utprøving leverte Skrivesenteret endelige forslag til Utdanningsdirektoratet, som tok avgjørelser om hvilke oppgaver, kriterier og støttemateriell som ville bli gjeldende. Følgende personell inngikk i prosjektet:

- Prosjektleder: Trude Kringstad (2020), Gustaf B. Skar (2020–2023)
- Koordinatorene: Anne Holten Kvistad (2020–2023), Marita Byberg Johansen (2021–2023)
- 10. trinns-gruppen: Trude Kringstad, Marthe Lønnum, Margrethe Sønneland, Tone Stavli
- vg2yf-gruppen: Runa Eugenie Røttereng Greiner, Kari Anne Nordgård, Ingrid Ertesvåg

- vg3-gruppen: Linn Maria Magerøy-Grande, Kari Anne Nordgård, Espen Tørset, Ingeborg Margrete Berge
- Skrivsesenterets referansegruppe: Henning Fjørtoft, Marte Blikstad-Balas, Therese Hopfenbeck, Leiv Inge Aa, Per Esben Svelstad, Toril Frafjord Hoem, Aslaug Gourvennec



Figur 1. Oversikt over arbeidsfordeling i prosjektet. Stiplede linjer betyr forslag.

1.3 Oppsummering av prosessen

1.3.1 Begrensninger og endrede rammer for Skrivsesenterets arbeid

I vårt tilbudsbrev til Utdanningsdirektoratet er det beskrevet hvordan Skrivsesenteret ønsket å prøve ut analytisk vurdering i arbeidet med å sikre reliabilitet og validitet i eksamen. Det var ønskelig fra Utdanningsdirektoratets side å prøve ut et system for vurdering som kan gi informasjon om eksamen på oppgavenivå (med såkalte «ankeroppgaver»). I rammeverket står det at oppgaver skal piloteres og analyseres med avanserte psykometriske metoder. I 2020 fikk vi imidlertid beskjed om å ikke bruke avanserte psykometriske metoder likevel. Årsaken til dette var at slike metoder ikke ville bli implementert i Utdanningsdirektoratets elektroniske prøvegjennomføringssystem (EPS) og at det derfor ville ha vært en uklar relasjon mellom resultatene på utprøvinger og reell eksamen. Skrivsesenteret arbeidet fram til desember 2021 med å pilotere og kvalitetssikre analytiske vurderingskriterier som kan gi informasjon på oppgave- og vurderingsområdenivå. På begynnelsen av 2022 ga Utdanningsdirektoratet beskjed om at det ikke skulle være analytisk vurdering på eksamen våren 2022. Noe senere ble det bestemt at våreksamen 2022 skulle avlyses på grunn av covid19-pandemien – med unntak av privatisteksamen. I arbeidet fram mot privatisteksamen 2022 var det Utdanningsdirektoratet som laget et forslag til vurderingskriterier for holistisk vurdering. Skrivsesenteret ga råd i prosessen, men understreket at Utdanningsdirektoratets forslag i stor

grad skiller seg fra senterets faglige råd og eget forslag. Ved prosjektets avslutning er det fremdeles holistisk vurdering på eksamen i norsk for både 10. trinn, vg2yf og vg3.

De nye eksamenene skulle gjennomføres i et nytt, elektronisk gjennomføringssystem (EPS). Dette systemet har lagt enkelte begrensninger for Skrivesenterets oppgavegrupper når det kommer til oppgaveformat. Lytting er en del av det norskfaglige konstruktet (se del 2 om konstruktutvikling), og i 2021 ble det pilotert oppgaver som krevde at elevene lyttet til lydklipp og filmer. I november 2021 ga Utdanningsdirektoratet beskjed om at oppgavene til våreksamen 2022 ikke kunne inneholde lydklipp eller film. Denne bestemmelsen fikk konsekvenser for teksttilfang og oppgaveformuleringer som allerede forelå som utkast til eksamen 2022. For eksempel var de fleste oppgavene som var piloterte på 10. trinn lite relevante å se til, da mange av dem inneholdt lyd eller film. Ved prosjektets avslutning er det fremdeles ikke mulighet for lydklipp eller film på eksamen.

I rammeverket for eksamen var det presisert ved prosjektoppstart at en eksamen med kun åpne oppgaver (slik som i norskfaget) skulle inneholde minst tre oppgaver. Tidlig i prosjektperioden ble det endret til at minst tre oppgaver er en sterk anbefaling (men altså ikke et absolutt krav). Skrivesenteret ønsket – i tråd med anbefalingen i rammeverket – å prøve ut oppgavesett med minst tre oppgaver (vi har prøvd ut både tre og fire). Ved prosjektslutt er status at 10. trinn og vg2yf har tre oppgaver på eksamen, mens vg3 har to. Eksamen for vg3 er altså ikke i tråd med anbefalingen i rammeverket. Enkelte stemmer har uttrykt sterk motstand mot tre oppgaver på vg3 og har blitt hørt av Utdanningsdirektoratet.

1.3.2 Endringer i norskeksamen som følge av Skrivesenterets arbeid

I perioden fra mars 2020 til utgangen av 2023 har Skrivesenteret utført oppdraget i tett samarbeid med Utdanningsdirektoratet og hensyntatt de bestemmelsene oppdragsgiver har gjort underveis i prosessen.

Skrivesenteret har utviklet konstrukt for norskeksamen og manualer med føringer for oppgavelaging. I tråd med rammeverket for eksamen har vi lagt stor vekt på å pilotere oppgaver og oppgavesett. Dette har vi gjort i til sammen elleve omganger. Skrivesenterets arbeid har resultert i at det nå er tre oppgaver på norskeksamen for 10. trinn og vg2yf, og at disse eksamenene derfor er i tråd med anbefalingen om antall oppgaver i rammeverket.

Skrivesenteret har bidratt til utviklingen av de foreliggende oppgavespesifikke kriteriene for 10. trinn, vg2yf og vg3. Dette innebærer at det er utviklet egne vurderingskriterier med relevante vurderingsområder for hver oppgavetype. Under hver eksamensoppgave er disse vurderingsområdene listet opp. Formålet med dette er at kandidatene skal få kjennskap til hvilke kriterier som skal ligge til grunn for vurdering av besvarelsen, jamfør føringer i rammeverket.

I tillegg har Skrivesenteret bidratt med mange utprøvinger og analyser som har gitt informasjon om hvordan analytisk vurdering kan fungere i norskfaget (mer om utvikling av vurderingskriterier i del 4).

1.3.3 Skrivesenterets anbefalinger

Når det gjelder vurderingsform, altså analytisk eller holistisk vurdering, har våre utprøvinger vist at det ikke er store forskjeller mellom vurderingsformene. Selv om analytisk vurdering har vist seg å føre til noe høyere vurderersamsvar, kan vi ikke konkludere med at den ene vurderingsformen er bedre enn den andre. Analytisk vurdering har imidlertid flere fordeler ut

over vurderersansvar, noe vi drøfter utfyllende i del 4. En av fordelene er at vurderingene i større grad vil være etterprøvbare, fordi en vil kunne ettergå hvordan besvarelser er skåret på oppgave- og vurderingsområdenivå.

Uansett hvilken vurderingsform en velger, vil det være sentralt å satse på å utvikle læreres vurderingskompetanse og tolkningsfellesskap. Dette kan gjøres gjennom obligatorisk sensorskolering. Videre vil Skrivesenteret på bakgrunn av analyser anbefale at en bruker gjennomsnittet av to uavhengige sensorers vurderinger for å sikre reliabiliteten i sensuren.

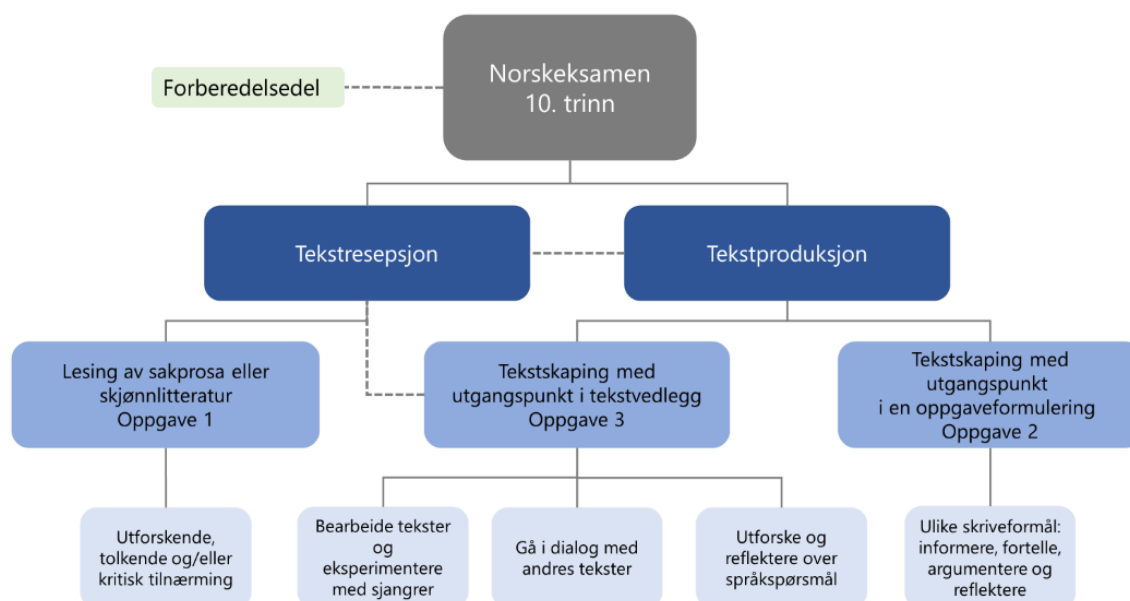
Når det gjelder antall oppgaver på eksamen for vg3, har tilbakemeldinger etter utprøvinger vært sprikende. Mens våre data fra 10. trinn og vg2yf viser at lærere og elever i stor grad rapporterer at det er nok tid til å besvare (minst) tre oppgaver, har lærere på vg3 i gjentatte utprøvinger med tre oppgaver vært delte i sine meninger om hvorvidt elevene får nok tid. Generelt ser det ut til at vg3-elever er mer positive enn lærere når det gjelder prøvesett med tre oppgaver. Skrivesenteret anbefaler at det arbeides mot at sidemålseksamen på vg3 får tre oppgaver og dermed blir i tråd med anbefalingene i rammeverket (se del 3.2.4 om utprøvinger med tre oppgaver på sidemål).

2 Konstruktutvikling

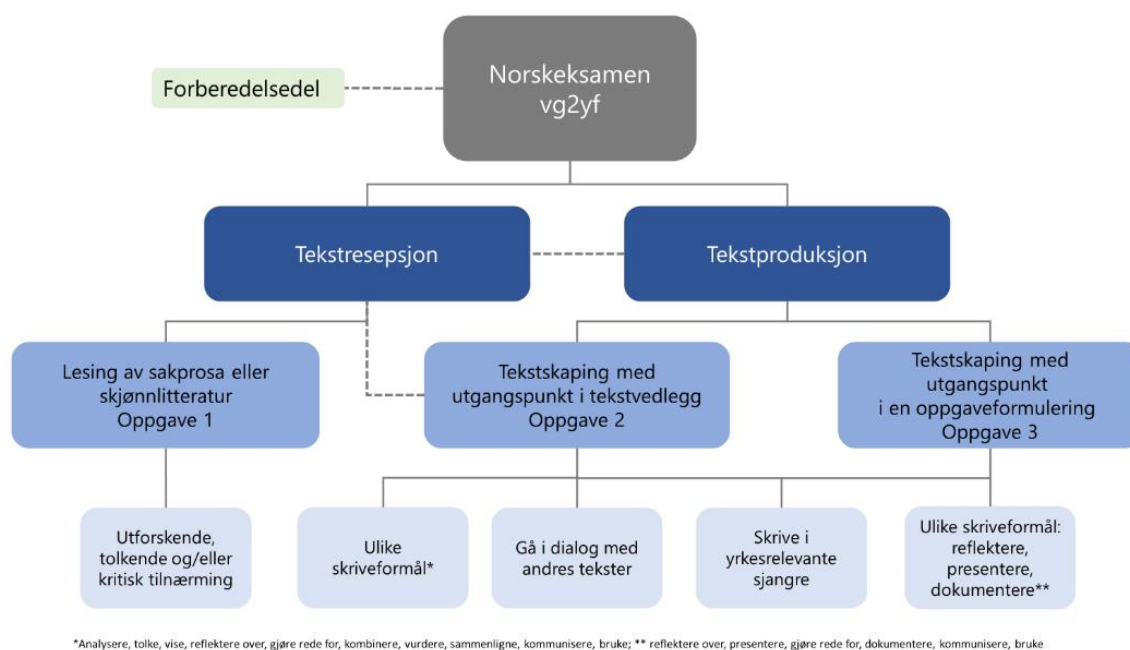
Arbeidet med å utvikle forslag til oppgaver og vurderingskriterier for samtlige fem eksamenskoder (NOR0218, NOR0219, NOR1262, NOR1267, NOR1268) startet med en definisjon av testkonstruktet for disse eksamenskodene. Vi utgikk fra følgende utdrag fra Utdanningsdirektoratets konkurransegrunnlag i forbindelse med konstruktdefinisjonene:

Eksamensoppgavene skal måle kandidatenes kompetanse i hovedmål og sidemål slik den kommer til uttrykk i kompetansemålene i det nye læreplanverket (LK20) for norsk etter 10. trinn og vg3 på studieforberedende utdanningsprogram, slik at eksamen blir valid. For vg2 yrkesfag skal eksamen måle kompetansen i norsk hovedmål. Eksamen må bygges opp på en slik måte at den ivaretar det nye kompetansebegrepet etter LK20. Oppdragstaker må ha et bevisst forhold til hvilke deler av læreplanen eksamen dekker, med begrunnelse for valgene som tas, samt kandidatenes tilgang på hjelpemidler.

Vi utgikk fra en forståelse av norskfaget som innebar at norskfaglighet kan sies å bestå av en *innholdsdimensjon*, som for eksempel litteratur- og språkvitenskap, og en *handlingsdimensjon*, som for eksempel det å kunne lese, skrive og uttrykke seg muntlig. Innholdsdimensjonen utgjøres av for eksempel bøker, kunnskap om dialekter, grammatisk kunnskap, som eleven forholder seg til når de handler i norskfaget, for eksempel ved å lese, skrive eller lytte. I vår konstruktdefinisjon utgikk vi fra skrivekompetansemodellen til Skar & Aasen (2021). Denne spesifiserer hva som kreves for å kunne skrive en tekst, noe som inkluderer sjangerkunnskap, kunnskap om hvordan tekster skal struktureres og hvordan ord staves. Vi utgikk også fra PISA sin modell av leseprosesser, som beskriver vesentlige faktorer i lesing, som for eksempel teksters vanskegrad, elevers kognitive kapasitet og rammer for lesning, som hvor lang tid en kan bruke på lesing. Til slutt lenket vi kompetansemål i faget til typer oppgaver. I vedlegg 1 gjengir vi konstruktet for 10. trinn, der strukturen er identisk med konstruktbeskrivelsene for vg2yf og vg3, men som bygger på andre kompetansemål. For hvert trinn ble det utviklet grafiske sammendrag av konstruktet. Disse gjengis i figurene nedenfor. I del 3 vil vi vise på hvilken måte vi brukte konstruktdefinisjonen i oppgavelagingen.

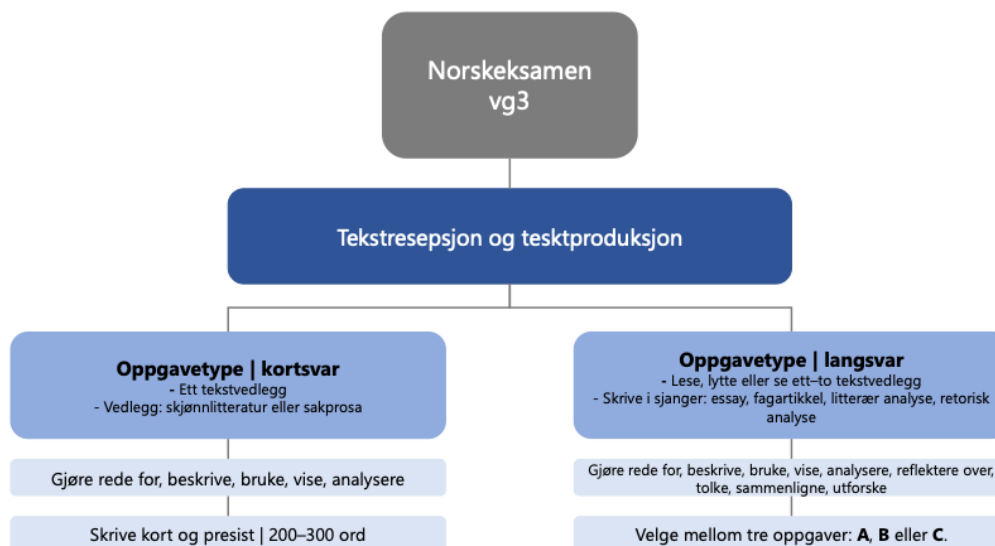


Figur 2. Konstruktmodell for norskeksamen på 10. trinn



*Analysere, tolke, vise, reflektere over, gjøre rede for, kombinere, vurdere, sammenligne, kommunisere, bruke; ** reflektere over, presentere, gjøre rede for, dokumentere, kommunisere, bruke

Figur 3. Konstruktmodell for norskeksamen på vg2yf



Figur 4. Konstruktmodell for norskeksamen på vg3

3 Oppgaveutvikling

3.1 Forankring i rammeverk og konstrukt

Skrivesenteret har prøvd ut oppgaver og oppgavesett med elever i elleve omganger. Oppgaver er prøvd ut i klasser både for 10. trinn, vg2yf og vg3 (se oversikt i tabell 1). Skrivesenteret har helt fra prosjektstart arbeidet bevisst mot at alle oppgaver skal være i tråd med føringene og anbefalingene i rammeverket for eksamen. For eksempel har punkt 5.2 i rammeverket («Lag oppgaver som er tydelige for kandidaten») stått svært sentralt i våre evalueringer etter piloteringer. I forbindelse med utprøvingene har vi spurt elever om deres forståelse av oppgavene, vi har lest elevbesvarelser for å avdekke eventuelle misforståelser og vi har spurt lærerne som var i klasserommet ved gjennomføring om i hvilken grad elevene så ut til å forstå hva de skulle gjøre. Piloteringene har bidratt med viktig innsikt som vi i stor grad har benyttet både i videreutvikling av enkeltoppgaver, og i utvikling av oppgavetyperne mer generelt. I denne delen vil vi oppsummere noen av funnene og erfaringene vi har gjort oss i forbindelse med utprøvinger.

I rammeverkets del 5.5. står det presisert at kandidaten må få kjennskap til hvilke kriterier som skal ligge til grunn for vurdering. Det skal altså framkomme tydelig hva som vektlegges i vurdering av elevens svar. Skrivesenteret har tatt dette til følge i norskeksamen for 10. trinn, vg2yf og vg3 ved å inkludere vurderingspunkter under hver oppgave. Disse er direkte koblet til vurderingsområdene som er aktuelle for den enkelte oppgavetyper. Elever har i ulike sammenhenger gitt uttrykk for at denne listen har vært nyttig og klargjørende. Skrivesenteret har også vektlagt tydelige instruksjonsverb i oppgavene, og på bakgrunn av utprøvinger har vi sett et behov for at klargjørende formuleringer som «bruk relevant fagspråk i svaret ditt» bør inkluderes i oppgaveformuleringene.

For å sikre at oppgavene er i tråd med rammeverket og det norskfaglige konstruktet, har Skrivesenteret fra prosjektstart arbeidet med å utvikle en såkalt *hvitbok* for hvert av trinnene, eller en manual for oppgavelaging. Denne manualen består av en oversikt over oppgavetyperne og hvilke kompetansemål og verb fra læreplanen som er aktuelle for hver av disse. I tillegg til mer generelle føringer for omfang, tekstvedlegg med mer, er det også laget egne føringer for hver oppgavetype når det gjelder aktuell tematikk for hver av oppgavetyperne. Denne listen har vært ett av verktøyene oppgaveutviklerne har tatt i bruk for å sikre at oppgavene er i tråd med føringene i rammeverket når det gjelder å unngå sensitive temaer, og å unngå emner som kan favorisere enkelte elevgrupper (jf. punkt 5.3: «Å lage en eksamen som er så rettferdig som mulig»). Hvitbøkene for oppgavelaging har vært under utvikling underveis i prosjektperioden etter hvert som Skrivesenteret har analysert data fra utprøvinger.

Nedenfor følger en oversikt over alle piloteringene som er gjennomført i regi av Skrivesenteret.

3.2 Pilotering av oppgaver

Tabell 1. Oversikt over gjennomførte piloteringer

Semester	10. trinn	vg2yf	vg3
Høst 2020	Prepilot Fire oppgaver (ca. 220 elever)	Prepilot Fire oppgaver (ca. 150 elever)	
Vår 2021	Hovedpilot hovedmål og sidemål Fire oppgaver (ca. 350 elever)	Hovedpilot Fire oppgaver (ca. 330 elever)	
Høst 2021	Pilotering av enkeltoppgaver (36 klasser)	Pilotering av enkeltoppgaver (18 klasser)	Pilotering hovedmål Tre oppgaver (ca. 740 elever)
Vår 2022			Pilotering hovedmål Tre oppgaver (ca. 1340 elever)
Høst 2022	Utprøving av eksamensoppgaver v22 Tre oppgaver (ca. 250 elever)		Pilotering sidemål Tre oppgaver (ca. 420 elever)
Høst 2023			Pilotering sidemål Metaspråksoppgaver (ca. 200 elever)

3.2.1 Tilbakemelding fra lærere og elever om norskfaglig relevans og elevers forståelse av oppgavene

Etter prepiloten på 10. trinn høsten 2020 kom det innspill på at oppgavene i det aktuelle oppgavesettet var for lite norskfaglige. Disse oppgavene var koblet til et av de tverrfaglige temaene i læreplanen. I hovedpiloten vår 2021 ble oppgavene derfor koblet direkte til norskfaglig kunnskap. Denne piloteringen prøvde ut oppgaver knyttet til språk, blant annet samisk og norske dialekter. Lærernes tilbakemeldinger viste at de anså oppgavene som norskfaglig relevante, men det kom innspill på at oppgavene burde revideres for å hjelpe elevene til å kunne gi gode svar på oppgavene (for eksempel at det bør presiseres at det forventes at elevene tar i bruk fagspråk i sine svar). I den siste utprøvingen i 2022 var lærere og elever stort sett enige om at oppgave 1 og 2 var forståelige, mens tallene var mer sprikende for oppgave 3. Konklusjonen etter denne piloteringen var at oppgavetype 3 nok er den oppgavetypen som kan oppleves som mest uvant og ny for elevene.

I spørreundersøkelsen til hovedpiloten på vg2yf våren 2021 svarte et flertall av lærerne at de opplevde oppgavene som norskfaglige relevante og at elevene forsto oppgavene. Flesteparten opplevde også at forberedelsesmateriellet var relevant for oppgavesettet, og at elevene forsto hvordan de kunne ta det i bruk. Svar fra lærerne på pilotering av enkeltoppgaver høst 2021 viser det samme; at et flertall av lærerne mente at elevene forstod oppgavene, og at elevene brukte forberedelsesdelen for å besvare oppgavene.

I tilbakemeldingene fra piloteringene på vg3 har lærerne i overveiende grad ment at oppgavene er norskfaglig relevante. Det har i liten grad kommet signaler som tyder på at oppgavene som har blitt gitt ikke har vært innenfor det man kan forstå som norskfaglig kompetanse. Når det gjelder elevenes forståelse av oppgavene, kom det fram en del sprik i tilbakemeldingene fra piloteringene på vg3. Kommentarene fra lærerne viser at det er nye oppgavetyper som kan forklare noe av det som skaper utfordringer med forståelse. Med LK20 kom essay inn som en klart uttrykt sjanger i norskfaget på studieforberedende og det tar naturlig nok litt tid før en ny sjanger er “under huden” på både lærere og elever. En annen utfordring har vært kortsvarsoppgaver som ber om å forklare grammatikk, siden dette også er oppgaver som hverken lærere eller elever nødvendigvis har trening i å skrive. Men dette er en del av norskfaget som har kommet sterkere til uttrykk gjennom LK20. Et siste aspekt om elevenes forståelse handler om tekstvedlegg og at de kan oppfattes som for vanskelige for elevene. Dette gjelder særlig tekster i eldre språkdrakt, der en del elever ikke har lesekompetansen som kreves for å forstå denne typen tekster.

3.2.2 Tilbakemelding fra lærere og elever om antall oppgaver og tidsbruk

Etter hovedpiloten på 10. trinn i 2021 viste svar fra spørreskjema til lærere at 11 av 12 opplevde at elevene deres fikk nok tid til å svare på alle oppgavene. Av disse var det åtte lærere som hadde elever som besvarte tre oppgaver, mens fire lærere hadde elever som besvarte fire oppgaver. I oktober 2021 kom det beskjed fra Utdanningsdirektoratet om at det var bestemt at det skulle være tre oppgaver til eksamen våren 2022. Skrivesenteret måtte derfor forholde seg til denne bestemmelsen, og vi har i de påfølgende piloteringene prøvd ut tre oppgaver i settene. I den siste piloteringen på 10. trinn i 2022 rapporterte samtlige lærere at elevene fikk nok tid til å gjøre tre oppgaver, og nesten samtlige elever (92,7 %) rapporterte

det samme. Elevene rapporterte at de brukte mest tid på å svare på oppgave 2 og minst tid på oppgave 1, noe som samsvarer med anbefalinger for tekstlengde.

Etter hovedpiloten på vg2yf i 2021 viste svar fra spørreskjema til lærere at 15 av 16 opplevde at elevene deres fikk nok tid til å svare på alle oppgavene. Av disse var det ti lærere som hadde elever som besvarte tre oppgaver, mens seks lærere hadde elever som besvarte fire oppgaver. Etter en gjennomlesing av tilfeldige besvarelser, kunne vi i eksamensgruppa ikke finne noen vesentlig forskjell på verken lengde, innhold eller tekstkvalitet mellom de som hadde svart på tre eller fire oppgaver. I oktober 2021 kom det beskjed fra Utdanningsdirektoratet om at det – i likhet med for 10. trinn – var bestemt at det skulle være tre oppgaver til eksamen våren 2022.

På vg3 var det stor enighet blant lærerne etter den første piloteringen høsten 2021 om at elevene ikke fikk tid til å gjøre seg ferdig med settet. Bare 35,7 % av lærerne mente at elevene hadde fått nok tid, og flere kommenterte at oppgavene ble for omfattende med tanke på antall tekstvedlegg og anbefalt tekstlengde for svarene. For de neste to piloteringene med tre oppgaver har vi i tillegg tilbakemeldinger fra elever, og et betydelig flertall av elevene rapporterte at de fikk nok tid til å gjøre oppgavene. Lærerne var mer delte i sine svar. I piloteringen fra høsten 2022 rapporterte for eksempel 75,7 % av elevene at de fikk nok tid, mens 59 % av lærerne mente det samme. I pilotering to og tre var anbefalt tekstlengde for svarene justert noe ned, og Skrivesenterets oppgavegruppe vektla gjenbruk av tekstvedlegg gjennom hele settene og mer presise og avgrensede oppgaveformuleringer. Forskjellene mellom piloteringene kan tyde på at det ikke nødvendigvis er antallet oppgaver i seg selv som er avgjørende for om man opplever tidspress, men kanskje heller det samlede omfanget når det kommer til for eksempel tekstvedlegg, samt oppgavenes vanskelighetsgrad og hvor vante elevene er til oppgavetypene.

3.2.3 Lesing av elevbesvarelser for å vurdere forståelse og tidsbruk

I flere omganger har Skrivesenterets oppgavegrupper lest et utvalg besvarelser fra pilotering for å vurdere vanskelighetsgrad og hvordan oppgavene har blitt forstått av elevene. Vi har brukt et dokument med påstander om oppgavene som en støtte i lesingen, oppsummert i følgende punkter:

- Oppgaven er norskfaglig relevant
- Oppgaven er forankret i læreplanen for gjeldende trinn
- Oppgaven er tydelig; kandidaten får en entydig instruks
- Oppgaven innebærer at kandidaten får mulighet til å vise sentrale deler av sin norskfaglige kompetanse
- Oppgaven passer for faglig sterke elever
- Oppgaven passer for faglig svake elever

Prosessene har foregått slik at et utvalg elevbesvarelser tilfeldig har blitt trukket ut, før gruppene har lest og diskutert opp mot påstandene over og eventuelt andre spørsmål som har vært relevante i enkelte av utprøvingene (f.eks. i hvilken grad de har inntrykk av at alle elever har rukket å besvare alle oppgavene i settet). De to siste påstandene i listen er med for å avdekke om oppgavene ser ut til å differensiere på en god måte, altså om svarene viser ulik grad av kompetanse.

På bakgrunn av arbeidet med lesing av elevbesvarelser har piloterte oppgaveformuleringer blitt endret og justert. Videre i denne delen vil vi vise noen eksempler på slike revisjoner fra 10. trinn og vg2yf.

I en tidlig utprøving for 10. trinn var det for eksempel en del elever som ikke hadde forstått at de skulle *sammenligne* to tekstvedlegg i oppgave 1a (se tabell under). De skrev heller hva forfatterne generelt var uenige i, altså hvilken sak de var engasjerte i (i dette tilfellet klimasaken). Oppgaven måtte derfor justeres, slik at det ble tydelig at de to tekstene skulle sammenlignes:

Oppgave 1a)

Pilotert formulering	«a) Forklar hva forfatterne ser ut til å være uenige om når det gjelder klimakrisen.»
Revidert formulering	«a) Thunberg og Ellefsen har ulike syn på klimakrisen. Forklar hva de ser ut til å være uenige med hverandre om.»

Et annet eksempel på revisjon er en oppgave fra 10. trinn om samisk språk og kultur i tabellen under:

Oppgave 1c, sett 1:

Pilotert formulering	«c) I forberedelsesdelen så du en film med Mari Boine som har skrevet dette diktet. Hvorfor tror du det har vært så viktig for henne å ta vare på den samiske kulturen? Skriv et sammenhengende avsnitt (ca. 100 ord).»
Revidert formulering	«c) I forberedelsesmateriellet så du en film med Mari Boine, og du har også lest diktet av henne. Hvorfor har det vært så viktig for Boine å ta vare på den samiske kulturen? Bruk kunnskap om holdninger til samisk språk i svaret ditt. Skriv et sammenhengende avsnitt (på ca. 100 ord).»

Grunnen til at denne oppgaven ble justert, var at flere elevbesvarelser handlet om språkendring og språkdød. Dette er relevante tema i norskfaget, men det kunne like gjerne ha handlet om andre minoritetsspråk. Intensjonen med oppgaven var at elevene skulle vise kunnskap om samisk språk og kulturhistorie. Dette ble derfor tydeliggjort i revidert versjon, ved at det ble lagt til en setning om hvilken fagkunnskap elevene er forventet å vise. Formuleringen «tror du» ble også tatt bort i andre setning. Slik oppgaven var formulert, la den i for stor grad opp til «synsing» i stedet for bruk av fagkunnskap.

Et annet eksempel på revisjon er fra en oppgave om dialekter, også fra 10. trinn:

Oppgave 4, sett 2:

Pilotert formulering	«Lytt til de to dialektprøvene. I denne oppgaven skal du vise kunnskap om dialekter. Skriv en tekst der du først beskriver tre kjennetegn ved hver av de to dialektene. Argumenter deretter for hvor i landet du mener man snakker de to dialektene. Svaret ditt skal skrives som en sammenhengende tekst på ca. 200-250 ord.»
----------------------	--

Revidert formulering	«Skriv en tekst der du forklarer hvordan vi kan høre at disse dialektprøvene er fra Stavanger og Molde. Vis til konkrete eksempler i lyttetekstene, og bruk norskfaglige begreper i forklaringen din. Skriv 200-250 ord, og oppgi eventuelle kilder du bruker.»
----------------------	---

Endringene i denne oppgaveformuleringen ble gjort basert både på tilbakemeldinger fra lærere og lesing av elevtekster. Lærere meldte om at denne oppgaven var for vanskelig, og at elever ble bekymret over å ikke få til å identifisere dialektene. Lesing av elevtekster bekreftet at det å lokalisere dialekter var vanskelig for svært mange. Oppgavegruppa løste dette ved å inkludere hvor dialektene kan plasseres geografisk i revidert oppgaveformulering. Elevene fikk i oppgave å heller forklare trekk som er typiske for disse dialektområdene (ut ifra lydprøvene). Det ble også tydeliggjort i oppgaveformuleringen at det forventes at det brukes relevante fagbegreper. Gjennom lesing av elevtekster ble vi bevisste på at dette er noe som vi ønsker å vurdere (og belønne), og det er derfor viktig at det framkommer tydelig i oppgaven at dette er forventet.

På vg2yf ble noen oppgaveformuleringer justert med tanke på å gjøre oppgaveformuleringene mer presise og mindre teksttunge. I eksempelet under gir den opprinnelige oppgaveteksten en ganske lang introduksjon til formålet med besvarelsen, men i den reviderte oppgaveformuleringen er introduksjonen kuttet ned og er mer presis.

Oppgavetype 1

Pilotert formulering	Du kjenner en vg1-elev på ditt utdanningsprogram som snart skal utføre en praktisk arbeidsoppgave som er vanlig i ditt yrkesfag. Det praktiske arbeidet skal vurderes med karakter av en programfagslærer. Eleven er redd for å gjøre feil og trenger hjelp av deg. Skriv en instruksjon til denne eleven der du gjør rede for hvordan han bør utføre arbeidsoppgaven på en best mulig måte. Skriv ca. 500 ord. <i>Kommentar: Det skal komme tydelig fram hvordan man skal utføre arbeidet og hvilket utstyr man eventuelt må bruke. Bruk relevant fagspråk fra ditt eget utdanningsprogram.</i>
Revidert formulering	Du kjenner en vg1-elev som snart skal utføre en praktisk arbeidsoppgave i sitt yrkesfag. Skriv en instruksjon til denne eleven der du gjør rede for hvordan han bør utføre arbeidsoppgaven på en best mulig måte. Skriv ca. 500 ord. <i>Kommentar: Det skal komme tydelig fram hvordan man skal utføre arbeidet og hvilket utstyr man eventuelt må bruke. Bruk relevant fagspråk fra ditt eget utdanningsprogram.</i>

På bakgrunn av innspill fra eksamensnemnden, referansegruppen og interne diskusjoner endret Skrivesenteret formuleringen av oppgaver der elevene ble oppfordret til å skrive en spesifikk sjanger, siden læreplanen i norsk for vg2yf ikke oppgir hvilke sjangre elevene skal skrive. Fagansvarlige i Utdanningsdirektoratet har kommunisert muntlig at vi ikke kan be elever skrive i bestemte sjangre i oppgavene for vg2yf eller 10. trinn, men at vi kan komme

med forslag til hva som kan være relevante sjangre. I revidert oppgaveformulering i eksempelet under er det foreslått novelle eller roman som sjanger, men dette er ikke et krav.

Pilotert formulering	Skriv begynnelsen av en skjønnlitterær tekst, en roman eller en fortelling, der du kombinerer virkemidler kreativt i teksten.
Revidert formulering	Skriv begynnelsen av en skjønnlitterær tekst, for eksempel en roman eller en fortelling, der du kombinerer virkemidler kreativt i teksten.

Noen oppgaver ble også forkastet etter pilotering. Et eksempel på dette er følgende oppgave fra en tidlig utprøving på vg2yf.

Oppgavetype 2

Vedlegg: Utdrag fra romanen *Anna. En fabel om klima og miljø* (2013) av Jostein Gaarder

Les vedlegget og svar på spørsmålene a, b og c. Svarene bør samlet sett være på 250-300 ord.

- Hva snakker Anna og psykiateren om? Forklar med egne ord.
- Gjør rede for hvordan psykiateren blir beskrevet i teksten. Bruk tre eksempler fra teksten som viser hvordan han blir beskrevet. Hvordan tolker du psykiateren ut fra personbeskrivelsene i teksten?
- Tror du samtalen med psykiateren kan være til hjelp for Anna? Begrunn svaret ditt.

Da oppgavegruppen leste elevsvarene til denne oppgaven, så de at det kom fram lite norskfaglig kompetanse i elevenes besvarelser. Dette var en oppgave der det ikke framkom tydelig hvilken norskfaglig kompetanse elevene skulle vise. Oppgaven ble derfor forkastet i det videre arbeidet mot ny norskeksamen.

3.2.4 Utprøving av sidemålsoppgaver

To piloteringer på vg3 har vært knyttet til sidemål. Dette gjaldt en pilotering i 2022 med tre oppgaver, og en i 2023 med utprøving av det vi har kalt metaspråksoppgaver.

Utdanningsdirektoratet har ønsket å prøve ut om sidemålseksamen kan prøve en litt annen norskfaglig kompetanse enn hovedmålseksamen. I dag er de to norskeksamenene på vg3 like, med en kortsvarsoppgave og en langsvarsoppgave. I samarbeid med Utdanningsdirektoratet har Skrivesenteret prøvd ut oppgavesett for sidemålseksamen med tre oppgaver, inkludert metaspråksoppgaver.

I utprøvingen med tre oppgaver høsten 2022 ble det laget oppgaver der kandidaten i større grad skulle skrive om språklige tema. Anbefalingen for tekstlengde i svarene var litt kortere enn på de tidligere piloteringene. Utprøvingen viser at omfanget var greit og at elevene fikk nok tid (se også 3.2.2). Elevene som prøvde ut settet meldte at oppgave 1, altså den språklig innrettede oppgaven, var en uvant oppgavetype. Det samme rapporterte lærerne som var med i piloteringen. Det var på denne oppgaven flest elever var usikre og spurte om hjelp. Det var også denne oppgavetypen flest lærere meldte at elevene ikke hadde fått relevant opplæring for å kunne svare på. Lærerne vurderte oppgavetype 1 som til dels norskfaglig relevant, men det var oppgavens vanskegrad som ble kommentert, ikke den faglige relevansen.

Høsten 2023 piloterte Skrivesenteret enkeltoppgaver. Det var igjen metaspråksoppgaver som ble prøvd ut, i denne omgangen oppgaver som gikk på å sammenligne norsk skriftspråk med dansk, norrønt eller sosiolekt. Igjen var lærerne enige om at oppgavene var norskfaglig relevante, men de var klare på at dette er oppgaver som elevene i liten grad har øvd på eller arbeidet med i klasserommet.

Vår vurdering er at tre oppgaver er gjennomførbart på sidemålsksamen. Det krever imidlertid at lengdeanbefalingene blir justert ned, særlig i langsvarsoppgaven, både fordi det er elevens sidemål og det er tre oppgaver. Man bør også i større grad tenke på å gjenbruke vedlegg, altså at ett eller flere vedlegg brukes i mer enn en oppgave. På denne måten vil den totale tekstmengden elevene skal forholde seg til bli justert ned. Skrivesenteret vil anbefale at en av oppgavene i settet bør være en språklig oppgave. I LK20 er kompetanse om språk og språkstruktur tydelig vektlagt, og å inkorporere dette i eksamensoppgaver vil føre til at eksamen måler sentrale deler av norskfaglig kompetanse.

4 Kriterieutvikling

En viktig del av arbeidet til Skrivesenteret har vært å foreslå nye vurderingskriterier. Skrivesenteret har også foreslått innføring av analytisk vurdering hvor elevbesvarelser vil få ett skår, som deretter «oversettes» til en karakter. UDIR har foreløpig valgt å ikke innføre analytisk vurdering, men Skrivesenteret har i flere omganger pilotert analytisk vurdering med formålet å undersøke reliabiliteten ved analytisk vurdering, og med formål å undersøke eventuelle forskjeller mellom analytisk og holistisk vurdering. Ved ett tilfelle har Skrivesenteret også undersøkt reliabilitetskonsekvensene av å angi hvor mye et område skal vektes i forhold til andre områder i forbindelse med holistisk vurdering. Tabell 2 nedenfor presenterer et sammendrag av de undersøkelser som er blitt gjort. Etter tabellen vil vi utdype resultatet av undersøkelsene noe.

Tabell 2. Oversikt over pilotering av vurderingskriterier

Trinn	Tidspunkt	Type utprøving	Elever (N)	Sensorer (N)	Resultat
10. tr.	Høst 2021	Reliabilitet ved analytisk vurdering	115	10	Analytisk vurdering ga høy reliabilitet ved bruk av Rasch-analyse
	Vår 2023	Konsekvenser av føringer for vektning av oppgaver i holistisk vurdering	20	8	Angitte vektingsprosenter hadde ikke betydning for vurderingene
vg3	Vår 2022	Sensorsamsvar ved henholdsvis analytisk og holistisk vurdering	541	111	Analytisk vurdering ga med liten margin høyere samsvar; korrelasjonen mellom analytisk og holistisk var $r = ,69$, noe som tilsier at vurderingsformene påvirker rangeringen av elevbesvarelser.

Trinn	Tidspunkt	Type utprøving	Elever (N)	Sensorer (N)	Resultat
	Vår 2023	Vurderingskriterier for sidemål (holistisk vurdering)	225	12	Reliabiliteten var noe lav; Vurderingen klarte å skille mellom to nivåer av elevkompetanse.
	Høst 2023	Sensorsamsvar ved henholdsvis analytisk og holistisk vurdering	221	12	Analytisk vurdering ga med liten margin høyere samsvar; korrelasjonen mellom analytisk og holistisk var $r = ,84$, noe som tilsier at vurderingsformene påvirker rangeringen av elevbesvarelser.

4.1. Utvikling og utprøving av analytiske kriterier

I innledningen til denne rapporten er det beskrevet hvordan Skrivesenteret ønsket å prøve ut analytisk vurdering i arbeidet med å sikre reliabilitet og validitet i eksamen. I rammeverket for eksamen står det at:

Vurderingskriteriene skal lages slik at analytisk vurdering er mulig. Med analytisk vurdering mener vi at vurderingen skal deles opp i ulike vurderingsområder for å vise tydelig hvilke deler som til sammen sier noe om kandidatens samlede kompetanse.

Skrivesenteret har gått empirisk til verks for å utvikle vurderingsområder og nivåbeskrivelser. Vi har utviklet kriterieutkast med utgangspunkt i reelle elevbesvarelser gjennom å samle elevtekster, gjøre initiale rangeringer av dem for å så lage beskrivelser av hva som kjennetegner tekster med ulik rangering. Disse beskrivelsene har deretter blitt gruppert tematisk i henhold til hvilke deler av en besvarelse de beskriver (f.eks. tekststruktur, formalia, etc.). De har også blitt organisert i henhold til kvalitetsnivå og slik blitt kvalitetskriterier. Deretter har Skrivesenteret internt pilotert kriteriene ved å bruke dem på et utvalg elevbesvarelser. Denne interne piloteringen har hatt som formål å gi svar på spørsmålet om kriteriene dekker sentrale trekk og kvaliteter i besvarelsene. Vi har også vurdert i hvilken utstrekning kriteriene har vært språklig kongruente og logiske. Som regel har vi justert kriteriene noe etter intern pilotering. Etter justeringer har kriteriene sammen med elevbesvarelser blitt sendt ut til sensorer.

Formålet med å sende kriterier til sensorer har vært å undersøke i hvilken utstrekning bruk av kriteriene fører til akseptabel reliabilitet, i de fleste fall målt som sensorsamsvar. Vi gjennomførte, for eksempel, en utprøving av kriterier for ungdomstrinnseksamen høsten 2021 med nettopp dette formålet. Vi sendte 115 elevbesvarelser til 10 sensorer og målte reliabiliteten i vurderingen. Dette arbeidet er nærmere beskrevet nedenfor, men hovedresultatet var at reliabiliteten var .93 og at det var større variasjon i resultatene relatert til oppgaver enn til sensorer (dvs. for eleven spilte det større rolle hvilken oppgave hen besvarte enn hvilken sensor som vurderte oppgaven). I andre tilfeller, som for eksempel ved utprøving av kriterier til vg3-eksamen høst 2023, har formålet vært å sammenligne vurderingsformer (dvs. holistisk og analytisk) ved å undersøke om formene gir lignende svar

på hvilke besvarelser som plasserer seg henholdsvis høyt og lavt. Nedenfor vil vi gjengi tekst fra rapporter vi har levert tidligere. Disse uttrekkene utdyper informasjonen om utprøvinger som gis i tabellen ovenfor.

4.2. Vurderingskriterier for eksamen på 10. trinn

4.2.1 Utviklingsprosess

Arbeidet med vurderingskriterier ble igangsatt i januar 2021. Førsteutkastet ble utviklet av eksamensgruppa på Skrivesenteret i samarbeid med den norskfaglige referansegruppa fra UH-sektoren. Måten dette ble gjort på var at det ble gjennomført en “comparative judgement”-prosedyre i programmet No More Marking. Prosedyren går ut på at hver vurderer blir presentert for to og to tekster – i dette tilfellet oppgavesvar fra prepiloteringen høsten 2020. Vurdererne skulle så trykke på den av de to tekstene de mente ga inntrykk av å være den beste besvarelsen. Når flere vurderere nominerer et stort antall tekster som den beste av to, blir det samlede resultatet en sortering fra “best” til “dårligst”. Dette ble brukt som grunnlag for å sortere tekstene i ulike nivåer - og som grunnlag for formulering av nivåbeskrivelser innenfor ulike vurderingsområder.

I etterkant av hovedpiloten i mai 2021 ble det i juni gjennomført en kvalitetssikringsrunde der oppgavegruppene og leder av oppgavenemnda for 10. trinn leste elevtekster og prøvde ut kriteriene. Nye runder med lesing av tekster fra pilotering og revisjon av vurderingskriteriene ble gjennomført i august. Vurderingsområdene med en kort forklaring til hvert område ble publiserte samtidig som eksempeloppgavene denne høsten.

Tidlig på høsten 2021 søkte Skrivesenteret etter medlemmer til et vurdererpanel for kvalitetssikring av vurderingskriterier. Utlysningen ble lagt ut på Skrivesenterets nettsider slik at alle lærere i landet hadde mulighet til å søke. Det ble lagt vekt på at paneldeltakerne skulle undervise i eller ha nylig undervisningserfaring fra 10. trinn (eller vg2yf/vg3). Søkerne skulle også skrive en kort tekst om hvorfor de var interesserte i å bidra til arbeidet. Arbeidet ble lønnet. 88 lærere meldte sin interesse totalt sett, og ti av disse ble valgt ut til å være vurderere for 10. trinn. Vurderingsarbeidet foregikk i oktober, og panelet hadde frist til å levere sine vurderinger og kommentarer tidlig i november. Vurdererdataene ga oss indikasjoner på hvordan skalaene så ut til å fungere. Den 11. november hadde vi digitalt møte med panelet der resultater ble presentert, og der vi fikk muntlige tilbakemeldinger fra deltakerne.

Også i andre sammenhenger i 2021 fikk Skrivesenteret tilbakemeldinger på utkastet til vurderingskriterier som da forelå. Til sammen vil vi anslå at vi i løpet av denne høsten fikk muntlige og/eller skriftlige tilbakemeldinger fra over 800 ungdomstrinns lærere, samt medlemmer fra ulike eksamensnemnder. Svært mange av tilbakemeldingene vi har fått underveis i kvalitetssikringen, har handlet om fagspråk, både når det gjelder navn på vurderingsområder og hvordan nivåbeskrivelsene har vært formulerte. Vi må ha et presist fagspråk for å kortfattet kunne uttrykke hva vi mener i nivåbeskrivelsene, samtidig som språket ikke skal være for akademisk. En av tilbakemeldingene vi hyppigst har fått er at vi må bruke et språk som faktisk brukes i skolen. Et eksempel på et begrep fra kriteriene som har blitt oppfattet som fremmed av mange, er “skilting”. Vurderingsområdet “Formalia” har også vært språklig fremmed for mange.

En av utfordringene i forbindelse med kvalitetssikring og revisjon av kriteriene har vært at tilbakemeldingene fra lærere og andre fagpersoner har vært til dels motstridende. Noen mener for eksempel at det er fint at vi kommer med eksempler på språklige virkemidler, mens andre

mener at dette er begrensende (at noen virkemidler fremstår som viktigere enn andre). Forholdet mellom å empirisk beskrive tekster på lavere nivå på en presis måte og det å ha “negative” formuleringer i nivåbeskrivelsene har også vært diskutert. Dette handler om i hvilken grad man kan peke på mangler/lav kvalitet.

På bakgrunn av tilbakemeldinger høsten 2021 arbeidet eksamensgruppa med revisjon av vurderingskriteriene fram til et ferdigstilt utkast ble oversendt Utdanningsdirektoratet i desember 2021. I januar 2022 fikk Skrivesenteret beskjed om at det ikke blir analytisk vurdering på eksamen våren 2022. I arbeidet fram mot (privatist)eksamen våren 2022 var det derfor Utdanningsdirektoratet som laget et forslag til vurderingskriterier for holistisk vurdering. Skrivesenteret ga råd i prosessen, men understreket at Utdanningsdirektoratets forslag i stor grad skilte seg fra senteres faglige råd og eget forslag. Vurderingskriteriene for 2022 og 2023 er å finne i eksamensveiledningene på Utdanningsdirektoratets nettsider.

Som en del av Utdanningsdirektoratets prolongering av oppdragsavtalen med Skrivesenteret, var det ønsket at vi undersøkte konsekvenser av vekting av oppgaver ved holistisk vurdering. Resultatene presenteres i det følgende.

4.2.2. Utpøving av holistiske kriterier med og uten vekting 10. trinn

På oppdrag av Utdanningsdirektoratet gjennomførte Skrivesenteret våren 2023 en undersøkelse hvor konsekvenser av vekting ble studert. Vi sendte ut to sett med kriterier til åtte sensorer. Ett av settene inneholdt beskrivelser av hvor mye vekt enkelte vurderingsområder skulle ha, mens det andre settet (som inneholdt samme kriterier) manglet denne opplysningen. Opplysningene var følgende:

Når du vurderer, skal du vekte oppgavene på følgende måte:

Oppgave 1: 20 %

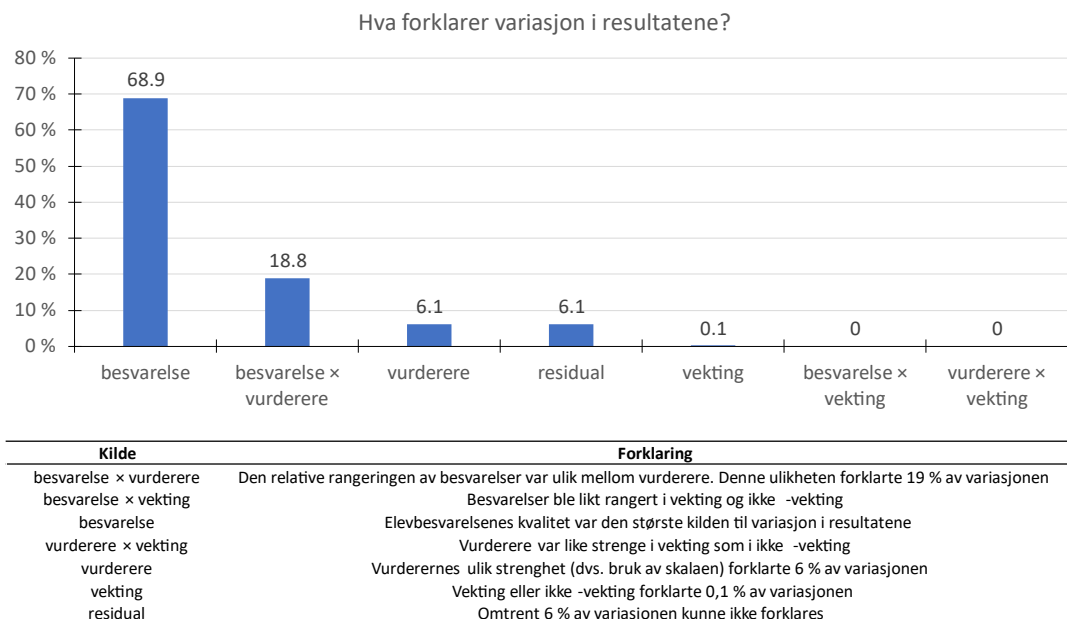
Oppgave 2: 45 %

Oppgave 3: 35 %

Disse vektingsprosentene hadde Skrivesenteret kommet fram til i samarbeid med fagansvarlig hos Utdanningsdirektoratet, blant annet ut ifra anbefalt tekstlengde og kunnskap fra piloteringen i 2022 om hvilke oppgaver elevene rapporterte å bruke mest tid på. De samsvarer også med vektingsprosentene som brukes for vg2yf.

Ved hjelp av *Generalizability theory* (Shavelson & Webb, 1991)¹ undersøkte vi hvor stor del av variasjonen i resultatene som kunne føres tilbake på vekting. Resultatene viste at den største kilden til variasjon var elevbesvarelsene; elevbesvarelsenes kvalitet var det som spilte størst rolle for variasjonen i vurderingene. Den nest største kilden til variasjon var vurderere; elevbesvarelsene ble ikke vurdert helt likt av vurderere. Vekting hadde ingen innvirkning på variasjonen. Se figur 5 nedenfor.

¹ Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Sage Publications, Inc.



Figur 5. Sammenstilling av undersøkelse av konsekvenser med vekting og ikke-vekting

Flere av vurdererne rapporterte til Skrivesenteret at de gjerne ville ha hatt mer runde tall i vektingsprosentene, og ikke 35/45 %. Skrivesenteret mottok også kommentarer og spørsmål når det gjelder forholdet mellom holistisk vurdering og matematisk utregning. Flere vurderere rapporterte at de lagde seg regneark der de satte delkarakterer og regnet ut den endelige karakteren matematisk. De kommenterte at de ved slik bruk av vektingsprosentene egentlig ikke vurderte holistisk.

Vurdererne rapporterte ellers i sine tilbakemeldinger til Skrivesenteret at oppgave 3 trakk ned karakteren for mange besvarelser. Dette kan sies å samsvare med funnene fra lærer- og elevundersøkelsene etter utprøving av oppgavene, der oppgave 3 framstår som noe mer utfordrende å forstå for elevene (se del 3.2.1).

4.3. Vurderingskriterier for eksamen på vg2yf

4.3.1. Utviklingsprosess

På grunn av at arbeidet med norskeksamen på vg2yf lå lenger bak i løypa enn 10. trinn (som en følge av at beskjeden om at norskeksamen på vg2yf skulle være sentralt gitt først kom i august 2020), var det tekster fra 10. trinn som la grunnlag for de aller første utkastene av vurderingskriterier for vg2yf. I oktober 2021 var det første utkastet til kriterier klart, basert på erfaringene fra det grundige arbeidet på 10. trinn (se del 4.2). Dette første utkastet var tilpasset eksamensoppgavene og læreplanen for vg2yf.

På samme måte som på 10. trinn ble det høsten 2021 etablert et vurdererpanel for kvalitetssikring av vurderingskriterier for vg2yf. Ti lærere med undervisningserfaring fra norsk på yrkesfag ble valgt ut til å være med i panelet. Hvert panelmedlem fikk tilsendt 45 elevbesvarelser fra hovedpilotingen på vg2 i mai 2021. Vurderingsarbeidet foregikk i oktober, og i november fikk Skrivesenteret tilbakemeldinger fra panelmedlemmene på hvordan vurderingskriteriene fungerte.

Også i andre sammenhenger i 2021 fikk Skrivesenteret tilbakemeldinger på utkastet til vurderingskriterier som da forelå, både fra lærere, referansegruppe og medlemmer av ulike eksamensnemnder. Til sammen vil vi anslå at vi fikk tilbakemeldinger fra over 100 norsklærere på videregående på utkast til vurderingskriterier for vg2yf denne høsten. På bakgrunn av alle disse tilbakemeldingene arbeidet eksamensgruppa med revisjon av vurderingskriteriene. Et ferdigstilt utkast til vurderingskriterier ble oversendt Utdanningsdirektoratet i desember 2021.

I januar 2022 fikk Skrivesenteret beskjed om at det ikke blir analytisk vurdering på eksamen våren 2022. I arbeidet fram mot (privatist)eksamen våren 2022 var det derfor Utdanningsdirektoratet som laget et forslag til vurderingskriterier for holistisk vurdering. Skrivesenteret ga råd i prosessen, men understreket at Utdanningsdirektoratets forslag i stor grad skilte seg fra senteres faglige råd og eget forslag. Vurderingskriteriene for 2022 og 2023 er å finne i eksamensveiledningene på Utdanningsdirektoratets nettsider.

4.4. Vurderingskriterier for eksamen på Vg3

4.4.1. Utviklingsprosess

Arbeidet med vurderingskriterier for vg3 bygger i stor grad på det grundige arbeidet med kriterieutvikling på 10. trinn og vg2yf. Det ble satt i gang senere på grunn av at eksamen på dette trinnet skulle ha første eksamensgjennomføring i 2023, altså et år etter de to andre trinnene. Noe av det som skiller læreplanen for vg3 fra de andre trinnene, er at det er oppgitt eksplisitte sjangre som elevene skal skrive. Dette medfører at kriterier ble utviklet for ulike oppgavetyper etter hovedsjangrene; essay, retorisk analyse, fagartikkel og analyse, tolking og sammenlikning.

Et vurdererpanel for kvalitetssikring av kriterier for vg3 ble rekruttert høsten 2021. Dette panelet besto av ti lærere med nylig undervisningserfaring fra norsk på vg3. Vurdererpanelet ble hyret til vurderingsarbeid ved to anledninger sammen med medlemmer fra eksamensnemnda for vg3. Høsten 2021 og våren 2022 gjennomførte prosjektgruppa flere kurs for norsklærere i regi av fylkeskommuner i flere deler av landet. I kursene ble nye vurderingskriterier utviklet for vg2 og vg3 testet ut, og mange lærere ga sin tilbakemelding på disse.

Første utkast av vurderingskriterier for vg3 etter analytisk modell forelå våren 2022. På dette tidspunktet var det allerede bestemt at det ikke skulle være analytisk vurdering på norskeksamen for de andre fagkodene denne våren. På oppdrag fra Utdanningsdirektoratet fortsatte Skrivesenteret å prøve ut kriterier for analytisk vurdering. Høsten 2023 ble det utviklet analytiske, oppgave- og sjangerspesifikke kriterier for kortsvarsoppgaven og de tre langvarsoppgavene (fagartikkel, essay og retorisk analyse). Her vurderes elevsvarene etter fem områder og svarene vurderes ut fra gitte poeng (2, 4, 8 eller 12). Våren 2023 piloterte Skrivesenteret på oppdrag fra Utdanningsdirektoratet holistiske kriterier for sidemåls eksamen. Denne piloteringen ble gjennomført for å gi et øyeblikkestimat på reliabiliteten og vil ikke bli kommentert videre i denne delen. I del 4.4.2 og 4.4.3 presenteres utprøvinger av analytisk vurdering sammenlignet med holistisk vurdering. Disse undersøkelsene kan bedre svare på konsekvensene av å bruke en gitt vurderingsform.

På våreksamen 2023 ble oppgave- og sjangerspesifikke kriterier tatt i bruk, men med holistisk vurdering. Vurderingskriteriene for 2023 er å finne i eksamensveiledningene på Utdanningsdirektoratets nettsider.

4.4.2. Analytisk vs. holistisk vurdering vår 2022

Våren 2022 gjennomførte vi et vurderingsarbeid med mål om å undersøke konsekvensene av å bruke henholdsvis holistisk og analytisk vurdering. I undersøkelsen ble holistisk vurdering definert som det å gi én helhetlig vurdering med utgangspunkt i kriterier, mens analytisk vurdering er definert som det å gi flere vurderinger, ett for hvert vurderingsområde.

111 norsklærere deltok som sensorer i undersøkelsen. Disse ble rekrutterte av Utdanningsdirektoratet. Sensorene ble delt i to grupper, en for holistisk vurdering og en for analytisk vurdering. Lærerne i gruppen som gjennomførte holistisk vurdering var noe eldre og hadde noe lengre undervisnings- og sensorerfaring, men gruppene var ellers like og representerte en bredde i lærerstanden.

Analysene i undersøkelsen baserer seg på vurderingene av 511 elevbesvarelser fra vg2sf og vg3sf. Besvarelsene var skrevet som svar på tre forskjellige oppgavesett laget som eksempeloppgaver til eksamen på vg3sf. Elevene hadde besvart tre oppgaver hver. Vurderingene er sett i sammenheng med vurderingskriterier for holistisk og analytisk vurdering. Skjemaet for holistisk vurdering ble utarbeidet av Utdanningsdirektoratet. Skjemaet for analytisk vurdering ble utarbeidet av Skrivesenteret i samarbeid med Utdanningsdirektoratet.

Hver elevbesvarelse ble sendt til fire sensorer, hvorav to sensorer hadde i oppdrag å vurdere besvarelsen analytisk og de andre to i oppdrag å vurdere besvarelsen holistisk. Hver sensor fikk 25 tekster å vurdere. Samtlige sensorer fikk seks identiske besvarelser i sine tekstbunker. Dette gjorde det mulig å «lenke» vurderinger på tvers av sensorpar og vurderingsform. All vurdering var uavhengig, det vil si at den skjedde uten at sensorene snakket sammen.

Det ble gjort flere ulike analyser i undersøkelsen. Det ble blant annet gjort undersøkelser for å undersøke sensorers samstemmighet og om besvarelsene ble rangert likt uavhengig av vurderingsform. Det ble også gjort analyser av hvor godt det var mulig å skille mellom kvaliteten til elevbesvarelsene når det er tatt hensyn til at sensorer er ulikt strenge og at skalaene er ulikt vanskelige. Til slutt ble det også gjennomført analyser for å undersøke om sensors strenghet henger sammen med erfaring fra sensurarbeid og tidligere gjennomført sensorskolering.

For begge vurderingsformene viste resultatene lavere samstemmighet enn man kunne ønske seg. Resultatene viste at samstemmighet for analytisk vurdering var noe høyere enn for holistisk vurdering. Forskjellen var imidlertid ikke stor, og den kan skyldes at den analytiske skalaen var lengre. En lengre skala har i tidligere forskning vist seg å ha positivt effekt på samstemmighet. For både holistisk og analytisk vurdering ble samstemmigheten høyere ved bruk av gjennomsnittet av to uavhengige sensorers vurderinger av samme besvarelse.

Resultatene viser også at rangeringen av elevbesvarelser ble ulik ved bruk av de to vurderingsformene. Det gjenstår et komplekst arbeid med å finne ut hva det kan skyldes at en besvarelse er blitt rangert høyt i den ene vurderingsformen, og ikke i den andre. Dette vil blant annet kreve at samme sensorer prøver begge vurderingsformene på samme elevbesvarelser, for å utelukke at det er ulikheter ved sensorene som fører til ulik rangering.

Et annet funn i undersøkelsen er at analytisk vurdering var bedre når det gjaldt å skille tydelig mellom kvalitetsgrupper. Dette er forventet, siden antallet målepunkter per elevbesvarelse er betraktelig høyere enn i holistisk vurdering. Ved analytisk vurdering var det mulig å skille mellom seks kvalitetsgrupper, mot fire ved holistisk vurdering.

Regresjonsanalysene viste at antall års erfaring med sensurarbeid var av betydning for strenghet i holistisk vurdering. Antall års undervisningserfaring og tidligere gjennomført sensorskolering hadde imidlertid ikke stor innvirkning på strengheten til den enkelte.

Oppsummerende kan man si at forskjellen mellom holistisk og analytisk vurdering i denne undersøkelsen viste seg å være liten. Et viktig funn i undersøkelsen er også at det var mulig å skille mellom seks kvalitetsgrupper med analytisk vurdering, noe som er i samsvar med karakterskalaen. Holistisk vurdering var kun egnet til å skille mellom fire kvalitetsgrupper. Selv om man med utgangspunkt i data fra denne undersøkelsen ikke kan si at den ene vurderingsformen er bedre enn den andre, finnes det noen argumenter som taler for analytisk vurdering:

- Vurderingen er i større grad etterprøvable fordi en vil kunne ettergå hvordan besvarelser har skåret på delområder.
- Vurderingen kan brukes formativt – skår på delområder er nyttig informasjon for elever og undervisende lærere.
- Vekting av enkelte områder vil bli gjort på forhånd og være kjent for elever.

Uansett hvilken vurderingsform en velger, anbefaler Skrivesenteret at en bruker gjennomsnittet av to uavhengige sensorers vurderinger.

4.4.3. Analytisk vs. holistisk vurdering høst 2023

Et panel bestående av 12 sensorer vurderte til sammen 221 elevbesvarelser. Disse besvarelsene kom fra eksamen vår 2023 og vi var kjent med vurderingene som ble gjort av sensorer da. Skrivesenteret fikk tilsendt 50 besvarelser per karakter fra våreksamen, sammen med informasjon om karakterforslag, endelig karakter og eventuell klagekarakter. Vi valgte i hovedsak besvarelser hvor forskjellen mellom karakterforslag, endelig karakter og eventuell klagekarakter var null. Dette ga 100 besvarelser (i gjennomsnitt 16,7 per karakternivå) og vi kompletterte således med 121 besvarelser hvor det var noe forskjell i vårsensuren. Hver besvarelse ble vurdert av to uavhengige sensorer. Praktisk ordnet vi dette ved å sende ut seks «vurderingspakker», ett til hvert par. I hver pakke inngikk noen besvarelser som var felles for samtlige sensorer. Dette ble gjort for å muliggjøre many-facet Rasch measurement-analyse. Hovedproblemstillingen dreide seg om analytisk vurdering ga lignende resultater som den holistiske (dvs. den som ble brukt på våreksamen). Vi svarte på dette spørsmålet ved å undersøke korrelasjonen mellom resultater på henholdsvis holistisk og analytisk vurdering og ved å analysere besvarelser som ble rangert ulikt avhengig av vurderingsform.

Vi starter med reliabilitetsestimat for hver vurderingsform. Tabell 3 nedenfor inneholder reliabilitetsestimat for analytisk vurdering for hver pakke og for to modus: 1) hvor besvarelsen blir vurdert av én sensor (ICC single) og 2) hvor resultatet for besvarelsen blir gjennomsnittet av to uavhengige sensorer (ICC average). I gjennomsnitt var reliabiliteten .79 for single og .88 for average. For holistisk vurdering var ICC Single = .73 og ICC Average =

.85. En many-facet Rasch measurement-analyse viste at den analytiske vurderingen i en Rasch-kontekst klarte å skille mellom syv nivåer av ferdighet (reliabilitet = .97).

Korrelasjonen mellom analytisk var $r = .84$ ($R^2 = .70$). Den analytiske vurdering klarte sånn sett å «reprodusere» den holistiske rangeringen av elevbesvarelser noenlunde vel. Samtidig var det flere besvarelser som ble rangert ulikt avhengig av vurderingsform.

For å videre undersøke forskjellen mellom holistisk og analytisk vurdering analyserte vi noen av tekstene som ble rangert ulikt. Dette presenterer vi nedenfor.

Tabell 3. Oversikt over pilotering av vurderingskriterier

Pakke	ICC Single			ICC Average		
	estimate	lower	upper	estimate	lower	upper
vupa1	.75	.58	.85	.85	.73	.92
vupa2	.70	.40	.85	.82	.57	.92
vupa3	.79	.50	.90	.88	.67	.95
vupa4	.81	.40	.92	.90	.57	.96
vupa5	.91	.84	.95	.95	.91	.97
vupa6	.81	.65	.89	.89	.79	.94
gj.snitt	.79			.88		

Vurderingsskjemaet som ble brukt på denne utprøvingen av analytisk vurdering, ble utviklet med utgangspunkt i vurderingsskjema fra IB-linjen, som har lang tradisjon for å i større grad vurdere analytisk.

I tilbakemeldingene fra vurdererpanelet kom det fram at det var en uvant måte å vurdere på, men at kriteriene stort sett var enkle å forstå. En del av tilbakemeldingene når det gjaldt vurderingsskjemaet, gikk på at de bør være enda mer oppgavespesifikke og at det bør være enda flere kolonner for å få et mer finmasket nett. Dette kan være utfordrende med tanke på å kunne skriftliggjøre de ulike nivåene av kompetanse på de ulike nivåene. Et annet dilemma er når kandidaten har levert en besvarelse som skårer høyt på språk og struktur, men som ikke svarer på oppgaven. Problemet ligger da i vektingen mellom de ulike vurderingsområdene. Skal da kandidaten automatisk strykes fordi oppgaven ikke er besvart? Eller skal kandidaten få uttelling for språk og struktur og kanskje få en ståkarakter selv om besvarelsen ikke er relevant for oppgaveinstruksen?

Vi så nærmere på de tekstene der det var størst sprik mellom vurderingen gitt på eksamen og den vurderingen som ble gitt på den analytiske utprøvingen. Da vi sammenlignet holistisk og analytisk vurdering, så vi at den analytiske vurderingen kan gi mer detaljert informasjon om hvilke vurderingsområder sensorene er uenige eller enige om. Dette vil være nyttig for å utvikle vurderingsfellesskapet videre for å sikre en mer helhetlig vurderingspraksis til eksamen. En annen side ved dette er at elevene ved analytisk vurdering kan få uttelling for sider ved eksamenssvaret som ikke vil bli fanget opp i en holistisk vurdering. Ved analytisk vurdering vil det i tillegg være enklere å gjøre en vekting av de ulike sidene ved en besvarelse. Dette er et evig tilbakevendende tema på sensorskolering og i fellessensur. Men

ved å bruke analytisk vurdering, vil denne vektingen være bakt inn i den vurderingen som sensorene gjør. På denne måten kan vi unngå muligheten for at personlig smak hos sensoren spiller inn i vurderingen, noe som er en fare ved holistisk vurdering.

Uavhengig av om man i framtida går for analytisk eller holistisk vurdering på eksamen, er det aller viktigste å satse på å utvikle lærernes vurderingskompetanse og tolkningsfellesskap. Dette kan løses gjennom blant annet obligatorisk sensorskolering.

5 Avslutning

Å utvikle forslag til ny norskeksamen på oppdrag fra Utdanningsdirektoratet har vært et spennende og utfordrende oppdrag. Vi har forsøkt å ivareta føringer fra oppdragsgiver og bestemmelser i læreplan og rammeverk for eksamen, og samtidig hensynta tilbakemeldinger fra ulike interesseorganisasjoner, referansegruppen, eksamensnemnder, oppmenn, lærere, sensorer og elever som har prøvd ut oppgaver. Som vi har redegjort for i denne rapporten, har tilbakemeldingene vi har fått underveis i prosessen ikke alltid vært like entydige. Mange har meninger om hvordan en eksamen best skal utformes, og engasjementet har tidvis vært stort rundt vårt arbeid med ny norskeksamen.

Alle valg vi har gjort i prosjektet har basert seg på et ønske om å lage en eksamen som gir elever en mest mulig valid, reliabel, håndterbar og rettferdig eksamen innenfor de rammene som er satt for en skriftlig eksamen på fem timer. I innledningen til denne rapporten har vi gjort rede for hva vi har lyktes med og hva som ble annerledes enn vi hadde tenkt.

Avslutningsvis vil vi berømme vår oppdragsgiver for å ha gitt oss rom for å prøve ut nye oppgaver og vurderingskriterier, og for at det i skrivende stund arbeides med å prøve ut alternativer til dagens eksamensformer i flere fag. Det blir spennende å følge den videre veien for norskeksamen, og vi håper at denne rapporten kan komme til nytte for andre eksamensutviklere.

6. Vedlegg

VEDLEGG 1: Testkonstrukt for norskeksamen på 10. trinn

Testkonstrukt for norskeksamen på 10. trinn

Dette dokumentet beskriver testkonstruktet for norskeksamen på 10. trinn. Testkonstrukt er en helhetlig beskrivelse av kompetanse og innhold som prøves på eksamen. Testkonstruktet består av en konstruktdefinisjon som definerer den norskfaglige kompetansen, som eksamen i norsk for 10. trinn skal og kan måle. Med konstrukt mener vi altså definisjonen av den norskfaglige kompetansen som kreves for å løse oppgaver på eksamen. Dette konstruktet vil, gitt begrensingene som ligger i eksamen, være noe smalere enn hele den norskfaglige kompetanse som læreplanen og opplæringen legger opp til. Testkonstruktet består også av en *domenebeskrivelse*, som redegjør for det faglige innholdet som det vil være aktuelt å prøve. I dokumentet presenteres også oppgavetyper, som er lenket opp mot konstruktdefinisjonen og domenebeskrivelsen.

Konstruktdefinisjonen og domenebeskrivelsen tar utgangspunkt i læreplanen for norsk ([NOR01-06](#)) og da særlig kompetansemålene etter 10. trinn. Konstruktdefinisjonen tar videre utgangspunkt i den forståelse av lesing og literacy som presenteres i OECD (2019) og den forståelse av skrivekompetanse som presenteres i Skar og Aasen (2020), Graham (2018) og Berge et al. (2016).

Konstruktdefinisjon: Definisjon av norskfaglig kompetanse

Norskfaglig kompetanse, slik den er mulig å måle i forbindelse med skriftlig eksamen, innebærer at eleven skal:

- interagere med tekster og
- ta i bruk formålstjenlige lese- og skriveprosesser for å løse eksamensoppgaven.

For å vise norskfaglig kompetanse må eleven ta i bruk leseprosesser som innebærer å lokalisere informasjon, å tolke tekster og å evaluere tekster (se eksempelvis OECD, 2019). Videre må eleven ha kunnskap om skrivning og være i stand til å aktivisere formålstjenlige skriveprosesser (Skar & Aasen, 2020). For å måle bredden i norskfaglig kompetanse må oppgavesett på eksamen innebære lesing og skrivning av ulike typer tekster og med en rekke ulike formål.

Lese- og skriveprosesser i sammenheng

Under presenteres en modell som viser relasjonen mellom lese- og skrivekompetanse og faktorer som påvirker hvordan en enkeltelev håndterer en oppgave. Denne modellen illustrerer hvilke faktorer som påvirker elevens bevisste og ubevisste valg av lese- og skriveprosesser. Modellen er viktig for å forstå hva som påvirker den kompetansen som eleven gir uttrykk for når han eller hun løser oppgaver.

I modellen framkommer fire faktorer: *Elevfaktorer*, *lese-/skrivebestilling*, *formålsfaktorer* og *rammefaktorer*.

- *Formålsfaktorer* handler om formålet med lesingen (f.eks. finne, tolke eller evaluere) og/eller skrivingen (f.eks. å informere, fortelle, argumentere og reflektere).
- *Lesebestilling* handler eksempelvis om tekstformat, teksters kompleksitet og antall tekster. *Skrivebestilling* handler for eksempel om krav til innhold, sjanger og modaliteter.
- *Elevfaktorer* inkluderer motivasjon, tidligere kunnskap (språkkunnskap, kunnskap om verden og tekstkunnskap) og kognitive egenskaper.
- *Rammefaktorer* dreier seg om eksempelvis omfanget av oppgaver, tid til å løse oppgavene og hjelpemiddel.

En prøveutvikler styrer ikke over elevfaktorer, men vil gjennom de rammene som gis for å løse en oppgave påvirke hvordan en elev engasjerer seg i oppgaven og hvilke tolkinger som er mulige å gjøre basert på et elevsvar.

Oppsummert kan det sies at modellen impliserer at oppgaveutvikler må ta hensyn til hvordan oppgaver konfigureres med tanke på faktorene i bunnen av modellen, slik at endelig oppgavesett gjenspeiler flest mulig relevante konfigurasjoner. Dette kan innebære at oppgavesettet ber eleven om å skrive i ulike sjangre for ulike kommunikative formål, om å lese med ulike formål og så videre.



Figur 1. Lese- og skriveprosesser i sammenheng. Modellen en adaptert versjon av en tilsvarende modell for prøveutvikling i OECD (2019).

For utvikling av oppgaver til eksamen, impliserer modellen at oppgaveutviklere må ta stilling til og pilotere:

- hvilke tekster som skal inngå i eksamen,
 - hvor mange tekster som skal inngå i eksamen,
 - hvor lange tekster som skal inngå i eksamen,
 - hvor vanskelige tekster som skal inngå i eksamen,
 - hvilke formål elever skal lese tekstene med,
 - hvilke typer av svarformat elever skal bruke for å demonstrere lesekompetanse
-
- hvilke type formål som skal inngå i skrivebestillinger til elever,
 - hvilket innhold elever skal generere i egne tekster,
 - hvilke sjangre (dvs. tekstetiketter, som f.eks. "kronikk") elever skal skrive
-
- hvor lang tid elever bruker på å svare på de enkelte oppgavene
 - hvor langt et svar må/kan være
 - hvilke hjelpemiddel som skal være tillatt og
 - hva som vil kjennetegne elevsvar som representerer ulike nivåer av vist fagkompetanse.

Domenebeskrivelse for norskfaget

I tråd med forskrift til opplæringsloven defineres det eksamensrelevante fagdomenet i norsk av innholdet i kompetansemålene. Det norskfaglige domenet kan sies å bestå av en *innholdsdimensjon*, som for eksempel litteratur- og språkvitenskap, og en *handlingsdimensjon*, som for eksempel det å kunne lese, skrive og uttrykke seg muntlig.

I kompetansemålene er innholds- og handlingsdimensjonene vevd tett sammen, for eksempel i kompetansemålet «sammenligne og tolke romaner, noveller, lyrikk og andre tekster ut fra historisk kontekst og egen samtid». Dette kompetansemålet kan sies å beskrive et visst innhold (romaner, noveller, lyrikk, litteraturhistorisk analyse) og handlinger (sammenligne, tolke).

Handlings- og innholdsdimensjonene som inngår i læreplanen for norsk danner grunnlag for oppgaver i sentralt gitt skriftlig eksamen. Eksamens praktiske forutsetninger legger føringer for hvilke handlings- og innholdsdimensjoner som er aktuelle å prøve. Disse forutsetningene er at eksamen skal utføres på maksimalt fem timer, at eksamen er skriftlig og at den skal være standardisert og mulig å vurdere av en ekstern sensor, som ikke har anledning til å stille oppfølgingsspørsmål til kandidaten eller observere kandidaten på andre måter enn gjennom kandidatens skriftlige svar.

I oppstillingen nedenfor framkommer de delene av fagdomenet som er rimelige å inkludere på norskeksamen, gitt de praktiske forutsetningene beskrevet over. *Kursiv* markerer handlingsdimensjonen, mens **fet** markerer innholdsdimensjonen. En klammeparentes med tre prikker ([...]) markerer at noe er utelatt. Handlinger som innebærer muntlig aktivitet er ikke tatt med. Muntlig aktivitet er per nå ikke en del av forutsetningene for eksamen.

1. *lese* **skjønnlitteratur og sakprosa på bokmål og nynorsk og i oversettelse fra samiske og andre språk**, og *reflektere over* tekstenes **formål, innhold, sjangertrekk og virkemidler**
2. *sammenligne og tolke* **romaner, noveller, lyrikk og andre tekster ut fra historisk kontekst og egen samtid**

3. [...] lese **tekster på svensk og dansk** og *gjøre rede for innhold og språklige trekk*
4. *utforske og reflektere over hvordan tekster framstiller unges livssituasjon*
5. *gjenkjenne og bruke språklige virkemidler og retoriske appellformer*
6. *bruke kilder på en kritisk måte, markere sitater og vise til kilder på en etterrettelig måte i egne tekster*
7. *utforske og vurdere hvordan digitale medier påvirker og endrer språk og kommunikasjon*
8. *bruke fagspråk og argumentere saklig [...] skriftlige framstillinger om norskfaglige og tverrfaglige temaer*
9. *informere, fortelle, argumentere og reflektere i ulike [...] skriftlige sjangre og for ulike formål tilpasset mottaker og medium*
10. *skrive tekster med funksjonell tekstbinding og riktig tegnsetting og mestre rettskriving og ordbøying på hovedmål og sidemål*
11. *bruke fagspråk og kunnskap om grammatikk, tekststruktur og sjanger i [...] bearbeiding av tekster*
12. *uttrykke seg i ulike sjangre og eksperimentere med sjangre på kreative måter*
13. *lage sammensatte tekster og begrunne valg av uttrykksformer**
14. *forklare den historiske bakgrunnen for bokmål og nynorsk og reflektere over statusen til de offisielle språkene i Norge i dag*
15. *Utforske språklig variasjon og mangfold i Norge og reflektere over holdninger til ulike språk og talespråkvarianter*

Oppgavetyper

I tabellen nedenfor presenteres oppgavetyper og hvordan de er koblet til aktuelle kompetansemål. Oppgavetyperne er ment å favne bredt, slik at flest mulig av kompetansemålene blir gjenstand for vurdering til eksamen. Nedenfor angir vi oppgavetyper og hvilke kompetansemål de kan kobles til.

Område og oppgavetype	Mest aktuelle kompetansemål
Tekstresepsjon; lesing av sakprosa eller skjønnlitteratur Denne oppgavetyper innebærer at kandidatene skal lese/lytte til sakprosa eller skjønnlitterære tekst (en eller to tekster) med en utforskende, tolkende og/eller kritisk tilnærming. I vurderingen av denne oppgavetyper vil hovedvekten ligge på kandidatens lesekompetanse, selv om språkføring også vil bli vurdert.	1, 2, 3, 4, 5, 6, 7, 14, 15
Tekstskaping med utgangspunkt i en oppgaveformulering: Denne oppgavetyper innebærer at kandidatene skal skrive en sammenhengende tekst med utgangspunkt i en oppgaveformulering. Kandidatene får anledning til å velge mellom to oppgaver. I vurderingen av denne oppgavetyper vil det vektlegges hvor godt kandidaten tilpasser teksten sin til formålet med skrivingen og hvor god språkføringen er.	6, 8, 9, 10, 12, 13*
Tekstskaping med utgangspunkt i tekstvedlegg: Denne oppgavetyper innebærer at kandidatene skal skrive med utgangspunkt i annen tekst. Dette kan for eksempel være å bearbeide andres tekster, eksperimentere med sjangre eller	1, 6, 9, 10, 12, 13*, 14, 15

utforske og reflektere over språkspørsmål. I vurderingen av denne oppgavetypen vil det vektlegges hvor godt kandidaten tilpasser teksten sin til formålet med skrivingen og hvor god språkføringen er.

* Dette kompetansemålet har vi ikke hatt mulighet til å prøve ut i prosjektperioden.

Litteratur

- Berge, K. L., Evensen, L. S., & Thygesen, R. (2016). The Wheel of Writing: a model of the writing domain for the teaching and assessing of writing as a key competency. *The Curriculum Journal*, 27(2), 172–189. <https://doi.org/10.1080/09585176.2015.1129980>
- Graham, S. (2018). A Revised Writer(s)-Within-Community Model of Writing. *Educational Psychologist*, 53(4), 258–279. <https://doi.org/10.1080/00461520.2018.1481406>
- OECD (2019). *PISA 2018 Assessment and Analytical Framework*. OECD Publishing. <https://doi.org/10.1787/b25efab8-en>.
- Skar, G. B., & Aasen, A. J. (2021). School writing in Norway: Fifteen years with writing as key competence. In J. V. Jeffery & J. M. Parr (Eds.), *International Perspectives on Writing Curricula and Development. A Cross-Case Comparison* (pp. 192–216). Routledge. <https://doi.org/10.4324/9781003051404-10>