

Rapport fra piloteringer og gjennomføring av den nasjonale prøven i lesing på 5. trinn 2023



Jostein Andresen Ryen, Solveig Aspelund, Anna Eriksen, Tove Stjern Frønes,
Tove Hårsaker, Eva K. Narvhus, Morten Skar og Cecilie Weyergang

Enhet for kvantitative utdanningsanalyser

Institutt for lærerutdanning og skoleforskning

Universitetet i Oslo

14.03.24

Innhold

1. Innledning	1
1.1 Måling av leseforståelse i de nasjonale prøvene i lesing	1
1.2 Tekster og oppgaveformater	2
1.3 Leseprosesser og eksempeltekst	4
1.4 Generaliserbarhet	6
1.5 Analyser i rapporten	6
2. Nasjonal prøve i lesing for 5. trinn 2023.....	9
2.1 Prøveutvikling i perioden 2021–2023.....	9
2.2 Innholdet i prøven.....	9
3. Første pilotering.....	12
3.1 Innholdet i prøve A, B og C.....	12
3.2 Klassisk item-analyse av prøve A, B og C.....	15
3.3 IRT-analyse av prøve A, B og C.....	19
3.4 Oppgavene fordelt på mestringsnivåer	21
3.5 Utvalg av tekster til andre pilotering	24
3.6 Endringer etter første pilotering.....	26
4. Andre pilotering.....	28
4.1 Innholdet i prøve D.....	28
4.2 Klassisk item-analyse av prøve D.....	30
4.3 IRT-analyse av prøve D	31
4.4 Endringer fra andre pilotering til endelig prøve.....	33
5. Forslag til endelig prøve 2023.....	35
5.1 Interessante leseutfordringer i forslaget til endelig prøve	36
5.2 IRT-analyse av prøvoforslaget.....	38
5.3 Oppgaver med lav diskriminering og DIF	39
5.4 Oppgavene fordelt på mestringsnivåer	39
5.5 Testinformasjon og måleusikkerhet	40
5.6 Prøvens TRF (Test respons function)	41
6. Gjennomføring.....	43
6.1 Resultater for prøven som helhet	44
6.2 Kjønnforskjeller	46
6.3 Klassisk item-analyse av prøven 2023.....	46
6.4 Oppgaver med lav diskriminering	48
6.5 Oppgavene fordelt på mestringsnivåer	49
6.6 Testinformasjon og måleusikkerhet	51
6.7 Prøvens TRF (Test respons function)	51
7. Referanser	53
8. Vedlegg 1: Beskrivelser av lesekompetanse på de tre ulike nivåene	54
9. Vedlegg 2: Oversikt over endringer i oppgaveformuleringer.....	55
10. Vedlegg 3: Spørreskjema til første pilot	58

1. Innledning

Denne rapporten beskriver og dokumenterer de lesefaglige og psykometriske vurderingene som ligger til grunn for utvikling av den nasjonale prøven i lesing på 5. trinn 2023, fra første pilotering til endelig prøve.

I innledningen gir vi først en oversikt over prøvespesifikasjonen og viser en eksempeltekst med oppgaver. Deretter gir vi en vurdering av generaliserbarhet fra årets leseprøve og presenterer kort hvilke analyser som er gjort – deriblant validitet, reliabilitet, utvalg til piloteringer, ankertekster og IRT-analyser. Videre dokumenterer vi utviklings- og kvalitetssikringsprosessen, før vi til slutt beskriver resultatene fra gjennomføringen av prøven i september 2023.

1.1 Måling av leseforståelse i de nasjonale prøvene i lesing

Å kunne lese er en grunnleggende ferdighet og en forutsetning for læring i alle fag. Lesing handler om å kunne forstå, bruke, reflektere over, vurdere og engasjere seg i tekster. I dagens informasjonssamfunn kreves det at vi leser for flere formål, at vi forholder oss til mange ulike typer tekster og at vi ofte må lese flere tekster som handler om det samme, for å tilegne oss ny kunnskap. Tekster har ulik utforming i forskjellige medier, og kan ved siden av verbalspråk også bestå av illustrasjoner, grafiske framstillinger, kart og andre visuelle elementer. Kunnskap om hva som kjennetegner ulike typer tekster og sjangre, er en viktig del av det å kunne lese, fordi sjangertrekk aktiviserer leserens forventninger og forkunnskaper om hvordan teksten skal leses og forstås.

Lesing er en sammensatt og komplisert aktivitet, som blant annet påvirkes av avkodingsferdigheter, lesehastighet, flyt, vokabular- og begrepsforståelse, samt kunnskap om tekststruktur og tekstens tema. God leseforståelse krever dessuten at lesere har evne til å identifisere forfatterens hensikt og vinkling, forstå hvilke bevis som brukes for å belegge påstander, og bestemme hvorvidt teksten er troverdig. I tillegg må lesere være bevisste på den påvirkningskraften tekster har i ulike kontekster. Lesere må ha kunnskap om hvilke strategier som er hensiktsmessige for å forstå ulike tekster, og tilpasse lesingen til det de leser og formålet med lesingen. Noen ganger leter lesere etter konkret informasjon i teksten, mens de andre ganger tolker, reflekterer eller foretar vurderinger av teksten (Utdanningsdirektoratet 2022).

De nasjonale prøvene i lesing måler i hvilken grad elevenes leseforståelse er i samsvar med beskrivelsen av lesing som grunnleggende ferdighet i læreplanene og i rammeverket for grunnleggende ferdigheter. Siden disse beskrivelsene er relativt generelle, uttrykker tekstutvalget og oppgavene i prøven prøveutviklernes fortolkning av dem. Kunnskap om den aktuelle aldersgruppas lesekompetanse ligger også til grunn for tekstutvalget og utformingen av prøven. Prøvene skal også gjenspeile de kravene som stilles til den leseforståelsen som elevene trenger for å klare seg, både på skolen og i samfunnet ellers.

Lesing som grunnleggende ferdighet i alle fag er beskrevet i Læreplanverket (Utdanningsdirektoratet 2020) og spesifiseres nærmere i «Rammeverk for grunnleggende ferdigheter» (Utdanningsdirektoratet 2017). Lesefaglige og psykometriske krav til de nasjonale leseprøvene defineres nærmere av rammeverket for nasjonale prøver (Utdanningsdirektoratet 2022), både generell del (del 1), om prøvenes innhold (del 2.4) og de tekniske kravene (del 3). I rammeverket defineres lesing slik:

Å lese handler om å kunne forstå, bruke, reflektere over og engasjere seg i innholdet i tekster. Tekster inkluderer alt som kan leses i ulike medier, ikke bare ord, men også illustrasjoner, symboler eller andre uttryksmåter. Kunnskap om hva som kjennetegner ulike typer tekster og deres funksjon, er en viktig del av lesing.

Prøvene skal måle lesing i alle fag i tråd med læreplanen, og dette skal gjenspeiles i et bredt utvalg av tekster. I beskrivelsen av lesing som grunnleggende ferdighet i ulike fag legges det vekt på elevenes evne til å finne fram til informasjon, forstå og tolke tekster, samt reflektere over form og innhold i et bredt spekter av tekster fra ulike fagområder. Disse utdragene fra læreplanen i naturfag, norsk, matematikk og samfunnsfag er eksempler på hvordan lesing som grunnleggende ferdigheter er nedfelt i fagplanene (Utdanningsdirektoratet, 2022):

- Å kunne lese i naturfag er å kunne forstå naturfaglige begreper, symboler, figurer og argumenter gjennom arbeid med naturfaglige tekster.
- Å kunne lese i matematikk vil si å sortere informasjon, analysere og vurdere form og innhold og sammenfatte informasjon i sammensatte tekster.
- Å kunne lese i norsk innebærer å kunne lese og reflektere over skjønnlitteratur og sakprosa, å beherske lesestrategier tilpasset formålet med lesingen og kunne vurdere tekster kritisk. Lesing i norsk innebærer også å lese sammensatte tekster som kan inneholde skrift, bilder, tegninger, tall og andre uttrykksformer.
- Å kunne lese i samfunnsfag inneber å utforske, tolke og reflektere over ulike historiske, geografiske og samfunnskunnskaplege kjelder.

I eksemplene ovenfor går det fram at den kritiske og analytiske lesekompetansen, samt evne til å integrere informasjon og trekke slutninger på bakgrunn av ulike tekster og tekstformater er sentral.

1.2 Tekster og oppgaveformater

Tekstene i nasjonale prøver i lesing skal gjenspeile det tekstmangfoldet som elevene møter i de ulike fagene og i samfunnet ellers. Ingen av tekstene er skrevet spesielt for prøvene, men har stått på trykk i ulike bøker, tidsskrifter, brosjyrer, nettsider etc. Det er ikke plass til alle fag i hver prøve, men over tid er det et mål at tekster fra alle skolefagene er representert, sammen med tekster fra hverdagslivet forøvrig. Tekstene kan ha tematisk tilknytning til faget eller være strukturert på en måte som stiller krav til den fagspesifikke leseforståelsen. Tekster som inneholder grafiske framstillinger, kan for eksempel være kategorisert under matematikk, selv om tekstens tema kan knyttes til andre fag. De aller fleste tekstene er aktuelle i mer enn ett fag og kan brukes tverrfaglig i etterarbeidet med prøveresultatene. Prøvene skal inneholde både skjønnlitterære tekster og sakprosaetekster. De skjønnlitterære tekstene består hovedsakelig av sammenhengende verbaltekst, mens sakprosaetekstene ofte vil være sammensatt av

ulike elementer, som for eksempel illustrasjoner, tabeller, figurer, diagram, kart osv. I tråd med læreplanen kan prøvene inneholde tekster både på bokmål og nynorsk.

Til hver tekst er det mellom tre og åtte oppgaver som tar utgangspunkt i relevante måter å lese teksten på. Et oppgavesett berører leseutfordringer knyttet både til innholdet i teksten og særtrekkene i måten den er skrevet og bygd opp på. Oppgavene i prøvene legger blant annet vekt på meningsinnhold som er implisitt uttrykt, bortgjemt informasjon eller opplysninger som står i motsetning til det forventede. Eksempler på andre leseutfordringer i prøvene er å avgjøre hva som er formålet med teksten og hvem som er målgruppen, hvilke stemmer som kommer til uttrykk, eller hvilke roller ulike aktører har i teksten. Det stilles også krav til at elevene skal sammenligne innholdet i flere tekster som handler om det samme, og identifisere og håndtere motstridende informasjon.

Oppgaver til tekster i leseprøvene skal både fungere som forbilder for hvordan tekster kan leses og forstås, og ikke minst skal de oppleves som relevante målinger av elevenes leseferdigheter. Et oppgavesett vil ha en logisk innholdsmessig oppbygning og en spredning i vanskegrad. Det vil være både lette og vanskelige oppgaver til alle tekster.

I Eksamen- og prøvegjennomføringssystemet (EPS) legges teksten inn som delt innhold (shared stimuli) i hver oppgave. For å unngå at elevenes lesing blir styrt av at første oppgave dukker opp med en gang, og for å gi elevene et overblikk over teksten, blir teksten vist uten en tilknyttet oppgave ved første visning. Elevene kan selv velge om de vil lese teksten her, eller om de vil klikke seg videre og starte med oppgavene med en gang.

I EPS er det flere ulike oppgaveformater, og disse har vi gitt navn som beskriver hva oppgavene går ut på. I leseprøven på 5. trinn er det flest flervalgsoppgaver med et spørsmål og fire svaralternativer der bare ett svar er riktig. I tillegg er det tabell- og rekkefølgeoppgaver, klikk ord- og klikk bilde-oppgaver, og oppgaver der elevene skal fylle inn ord eller fullføre en setning. Tabell-oppgavene består av en tabell med påstander eller opplysninger som elevene skal ta stilling til ved å krysse av for eksempel «riktig» eller «galt» eller «nevnt i tekst 1», «nevnt i tekst 2» eller «nevnt i begge tekstene». Rekkefølge-oppgaver inneholder tre eller fire utvalgte hendelser som elevene blir bedt om å plassere i riktig rekkefølge, fra det som skjedde først til det som skjedde sist. Dette formatet er særlig aktuelt i tekster der kronologien er brutt. I klikk ord-oppgaver får elevene mulighet til å trykke på ett eller flere ord i en tekst, mens de i klikk bilde-oppgavene skal velge mellom markerte områder i et bilde. Fyll inn ord-oppgavene går ut på å fullføre én eller flere setninger ved å velge riktige ord i en nedtrekkmeny.

De nevnte oppgavetypene blir skåret automatisk og er velegnet for leseprøver, både fordi elevens skriveferdigheter ikke påvirker resultatet, og fordi oppgavene vurderes uten lærerens skjønn. For 8. trinn brukes også åpne oppgaver der elevene kort kan formulere egne svar på en liten andel av oppgavene. Dette er et viktig format for å kunne måle elevenes evne til å reflektere over teksters form og innhold. Vi antar at det er en progresjon i elevens evne til å reflektere gjennom egen skriving, og har derfor i samråd med Utdanningsdirektoratet valgt å ikke bruke dette formatet i leseprøvene på 5. trinn inntil videre.

Når det gjelder oppgavene som skåres automatisk, er det et mål at elevene skal få svaralternativer og påstander presentert i rekkefølger som *enten* gjenspeiler den logiske orden i teksten (tallrekker eller andre logiske sammenhenger) *eller* i en tilfeldig rekkefølge som ikke fremmer gjetting (tilfeldig plassering av riktig svar A-D). I det endelige prøveforslaget er logiske rekkefølger kontrollert, og for øvrig er plassering av riktige svar tilfeldig. Vi gjør analyser av svarmønstre og dyktighet for alle oppgaver, og for komplekse oppgavetyper gir vi begrunnelser for poengsettingen støttet av analyseresultater.

1.3 Leseprosesser og eksempeltekst

Alle oppgavene i prøven er knyttet til en lesemåte som sier noe om hvilken leseoperasjon som kreves for å løse oppgaven. En oppgave skal enten måle elevenes evne til å finne informasjon, tolke og sammenholde informasjon eller å reflektere over og vurdere teksters form og innhold. Disse tre lesemåtene er omtalt som *aspekter* i rammeverket for nasjonale prøver, og er beskrevet slik:

Å finne informasjon i tekster innebærer å finne fram til informasjon som er eksplisitt eller implisitt uttrykt. Selv om informasjon i tekster er eksplisitt, kan det være vanskelig å lokalisere informasjonselementer i tekster som inneholder konkurrerende og til dels motstridende informasjon.

Å tolke og sammenholde informasjon innebærer å trekke slutninger på bakgrunn av én eller flere tekster. Dette leseaspektet spenner fra å trekke enkle slutninger mellom innholdet i to setninger som følger etter hverandre, til de mest kompliserte slutningene vi vanligvis forbinder med å «lese mellom linjene».

Å reflektere over og vurdere teksters form og innhold innebærer å forholde seg selvstendig til tekster. Dette leseaspektet spenner fra å kommentere teksters innhold med utgangspunkt i egne meninger til å forholde seg kritisk til tekster som helhet og begrunne egne synspunkter, analyser eller vurderinger av tekster.

De tre leseaspektene finnes ikke som atskilte ferdigheter hos leserne. Det er derfor ikke et mål å vurdere disse som avgrensede delkompetanser. Alle oppgavene i de nasjonale prøvene i lesing skal likevel være kategorisert etter hvilket leseaspekt som er den mest fremtredende i oppgavene. Alle de tre leseaspektene skal være representert i en prøve, men ikke nødvendigvis i samme omfang.

(Rammeverk for nasjonale prøver, pkt. 2.4)

Tre eksempeloppgaver til teksten «Fantastisk bambus!» som var en del av leseprøven for 5. trinn i 2020, illustrerer de tre ulike leseprosessene. De riktige svarene er markert med en sirkel.

Teksten nedenfor er hentet fra Miljøagentrapporten 4/2018. Miljøagentene er en miljøvernorganisasjon der barn og unge arbeider for et bedre miljø. Bruk teksten når du svarer på oppgavene.

FANTASTISK BAMBUS!



Bambusplanten er spesiell fordi den vokser kjempefort og kan bli hele 40 meter høy. Den kan vokse nesten en halv meter i døgnet.

Bambus er mye hardere enn mange tresorter og kan brukes til mye. Av bambus kan man lage både hus, møbler, gulv, kurver og matter, hatter, kopper, tallerkener, og også redskaper til jakt, fiske og jordbruk. I tillegg lager man skistaver, tannbørster, klær og papir av bambus. Frukten kan spises og er en viktig del av kostholdet i Kina. I hultrommene på eldre stammer lages et stoff som i Øst-Asia brukes i folkemedisin. Og visste du at man kan bygge hele broer av bambus?



Bro laget av bambus i Kambodsja

Mat til pandaen

Myke bambusskudd, stengler og blader er den viktigste maten til den store pandaen. Faktisk er 99 prosent av det de spiser, bambus. En stor panda spiser mellom 12 og 38 kilo bambus hver dag! Pandaen må spise så mye fordi bambus ikke inneholder så mye energi. Fordi den spiser så mye, må den bæsje opptil 40 ganger om dagen! Den store pandaen må ha minst to forskjellige bambusarter i sitt nærområde for å unngå å sulte, fordi bambus av samme art vokser opp, blomstrer og dør samtidig.



Pandaen elsker bambus!

Andre dyr spiser også bambus, for eksempel den røde pandaen i Nepal og bambuslemurene fra Madagaskar. Fjellgorillaer, sjimpanser og elefanter spiser også deler av bambusplanten.

Pandaen er truet

Den store pandaen, som lever villt bare i Kina, er et av verdens mest sjeldne dyr. Det finnes under to tusen ville, store pandaer igjen. Pandaen har blitt jaktet på og har mistet leveområder fordi bambusskogene har blitt kuttet ned eller ryddet til jordbruk. Kina har gjort mye for å ta vare på pandaer. Blant annet er verneområdene for pandaen utvidet de siste 20 årene, og skogsdrift i pandaens leveområder er redusert.

Miljøvennlig plante

Bambus er miljøvennlig fordi det er den raskest voksende skogplanten på jorda og fordi nye skudd kommer årlig. Bambus er en god erstatning for tømmer både fra tempererte skoger, som vi har i Norge, og fra tropiske skoger. Det er mye bedre å bruke bambus enn å sage ned gamle trær som vokser sakte og som det finnes få av.

Produktene av bambus som du får kjøpt, kommer ikke fra naturreservatene der pandaene bor, så vi bruker ikke opp maten til pandaene! Så neste gang noen i familien din skal kjøpe for eksempel en tannbørste eller møbler, kan du fortelle dem at de kan kjøpe disse tingene lagd av bambus, som er mye bedre enn plast og tropisk tømmer!

OM BAMBUS

Bambus er navnet på cirka tusen arter som hører til i gressfamilien. Bambus vokser i Asia, Mellom-Amerika og Sør-Amerika, og noen arter i deler av Afrika og Australia. De fleste artene av bambus ligner på trær, og de kan danne tette skoger.

ORDLISTE

Skudd: en ny vekst på en plante

Stengel: den delen av planten der bladene er festet

Tømmer: hele trestammer eller stokker

Temperert klima: klima der sommene er om lag like lange som vintrene

Tropisk klima: klima med mye varme og regn



1 Vi bruker ikke opp maten til pandaen selv om vi kjøper ting laget av bambus. Hva er grunnen til det, ifølge teksten?

- A Pandaen kan spise andre planter hvis det blir lite bambus.
- ☒ B Bambusen i tingene vi kjøper, kommer ikke fra leveområdene til pandaen.
- C Tingene vi kjøper, er laget av de delene av bambusen som pandaen ikke spiser.
- D I leveområdene til pandaen er det nok bambus til både pandaer og mennesker.

Leserne må finne informasjon i tekstens siste avsnitt om at bambusproduktene som vi får kjøpt, ikke kommer fra naturreservatene der pandaene bor.

2 Hvilken annen overskrift kunne passet til det siste avsnittet, som heter «Miljøvennlig plante»?

- A Plant mer bambus i Norge!
- B Ikke bruk opp maten til pandaene!
- ☒ C Kjøp bambus-produkter!
- D Tygg bambus og få renere tenner!

Leserne må tolke meningsinnholdet i det siste avsnittet og trekke ut det viktigste poenget ved å oppsummere fagstoffet. De må se at første avsnitt argumenterer for at bambus er en

bærekraftig erstatning for annet tropisk treverk og at leserne i andre avsnitt forsikres om at produksjon av ulike bambusprodukter ikke truer pandaens leveområder.

3 Den gule tekstboksen skal først og fremst ...

- A oppsummere innholdet i hovedteksten.
- B få oss til å bruke mer bambus.
- C vise hva bambus kan brukes til.
- ☒ D gi ekstra informasjon om bambus.

Leserne må reflektere over tekstboksens form og innhold og skille denne fra andre, frekvente tekstboksfunksjoner. Leserne må oppsummere fagstoffet i boksen og se at den inneholder tilleggsinformasjon som ikke er en del av brødteksten.

1.4 Generaliserbarhet

Vurderinger av leseprøvenes validitet er undersøkelser av om det som måles, korresponderer med fenomener i virkeligheten og om det kan trekkes gyldige slutninger om elevenes leseforståelse i andre kontekster (Silverman 2011: 367). Tekst- og oppgaveutvalget i nasjonal leseprøve for 5. trinn 2023 er i tråd med beskrivelser av lesing som grunnleggende ferdighet og kompetansemål i fagene etter 4. trinn, noe som blir bekreftet gjennom møter med lærere og andre eksterne kvalitetssikrere. Prøven inneholder både lette og vanskelige oppgaver, og måler dermed både elever med svake og sterke leseferdigheter. Prøvens innhold ser ikke ut til å favorisere verken gutter eller jenter i stor grad, men måler balansert. Denne rapporten dokumenterer alle detaljer i de undersøkelsene av reliabilitet og validitet som er foretatt i løpet av prøveutviklingen.

Reliabilitet er forbundet med målesikkerhet, altså hvorvidt vi måler det ønskede konstruktet på en konsistent måte. Vurderingsreliabilitet spiller ikke inn i en prøve med automatisk rettede flervalgsoppgaver. Det endelige prøveforslagets reliabilitetskoeffisient er 0,880 (Cronbachs alpha), som regnes som bra for en prøve av denne typen (Nunnally 1994).

Prøvens validitet handler om i hvor stor grad vi klarer å måle det vi ønsker å måle, nemlig elevens leseferdighet. Den digitale prøveplattformen som er valgt, samt kravene til universell utforming, begrenser til en viss grad både hvilke tekster det er mulig å bruke, og hvordan tekstene kan utformes. Prøven gir likevel et godt mål på elevens leseferdighet, men det er viktig å både utforske og utfordre den digitale prøveplattformen slik at prøveopplevelsen for elevene blir best mulig og at leseferdigheten blir målt i tråd med beskrivelsene av lesing som grunnleggende ferdighet.

1.5 Analyser i rapporten

I tråd med Rammeverk for nasjonale prøver (Utdanningsdirektoratet, 2022) skal prøvene utformes med beste mulige standarder for prøvekonstruksjon, ved hjelp av egnede psykometriske metoder. Oppgaveutviklingen skal skje med klassiske mål for reliabilitet og

oppgaveegenskaper. I tillegg skal enkeltoppgaver og hele prøven analyseres ved hjelp av item response theory (IRT). For 2023-prøven er klassisk analyse utført med programmet SPSS, mens IRT-analysene er utført med henholdsvis XCalibre 4.2 etter første pilotering og R 4.2.2 og R-pakken «mirt» fra og med andre pilotering.

Ved hjelp av IRT kan man estimere oppgavens vanskegrad og beregne elevenes ferdighet uavhengig av hvilke oppgaver de besvarte, forutsatt at elevene i utvalget representerer hele ferdigheten. IRT-analysen leverer resultatet i form av theta-verdier (θ), som går fra omtrent -3 til +3 med et gjennomsnitt på 0. Oppgavens vanskegrad blir også uttrykt på samme skala. Etter en enkel omregningsnøkkel ($\theta * 10 + 50$) regnes theta-verdiene om til positive verdier, såkalte skalapoeng.

IRT et godt verktøy i prøveutviklingen fordi analysene av pilotresultatene estimerer hvordan oppgavene vil fungere på et større antall elever. Når det kommer til gjennomføringen av prøven, handler det ikke om å estimere resultater, men å analysere data for hele populasjonen, og da er klassisk analyse mer relevant enn IRT.

IRT-analysen leverer en sum av informasjonsverdier og målefeil for hele ferdighetsfordelingen i en prøve (TIF) og en Test Respons Funksjon (TRF). Sistnevnte er en beskrivelse av sammenhengen mellom elevenes estimerte ferdighet (theta) og andel riktige besvarte oppgaver i prøven. Vi bruker TRF til å sammenligne årets prøve med fjorårets.

I IRT-analysene av oppgavene i prøven brukes en to-parameter-modell (2PL). Modellen viser oppgavens vanskegrad (b) og oppgavens diskriminering (a og point biserial correlation). En oppgaves diskrimineringsgrad sier noe om i hvilken grad oppgaven skiller mellom elever med svake og sterke leseferdigheter. Diskrimineringsmålene viser korrelasjon mellom skår på hver enkelt oppgave og total skår på prøven. Oppgaver med høy diskriminering gir høy reliabilitet for prøven. En oppgave med lav diskriminering ($< 0,250$ point biserial eller $< 0,8$ a-parameter) kan være «bortkastet» ved ikke å skille godt mellom lesere med ulik ferdighet, men oppgaven kan likevel forsvares pedagogisk eller lesefaglig. For eksempel kan en svært lett oppgave som nesten alle lesere mestrer, gi en god og oppklarende introduksjon til teksten og oppgavesettet. Polytome oppgaver, som skåres med 0, 1, 2 eller flere poeng, gir som regel høyere diskriminering, fordi de gir mer informasjon. Dersom en prøve inneholder mange oppgaver med for høy diskriminering, kan det være en indikasjon på lav validitet ved at oppgavene er for like og dermed dekker konstruktet for dårlig. Tilsvarende kan en prøve med mange lavt diskriminerende oppgaver også være en indikasjon på lav validitet ved at oppgavene måler noe annet enn leseforståelse.

IRT-analysen gir også svar på om oppgavene, og dermed prøven som helhet, favoriserer én bestemt gruppe framfor en annen gjennom en analyse av Differential Item Functioning (DIF). For eksempel om en oppgave favoriserer jenter framfor gutter, gitt at representanter for begge grupper har samme underliggende ferdighet (Utdanningsdirektoratet 2022).

Resultatene fra klassisk analyse blir rapportert i form av prosentvis fordeling av elevsvar (p-verdier), gjennomsnittlig dyktighet til gruppene av elever som valgte de ulike

svaralternativene, mål på diskriminering (point biserial correlation), og mål for reliabilitet/indre konsistens (Cronbachs alpha).

P-verdi refererer til prosentandel elever som har riktig svar på en oppgave. For polytome oppgaver, som skåres med 0, 1 eller 2 poeng, beregnes p-verdien ved å legge sammen andelen elever som fikk full skår (2 poeng) og 50 prosent av elevene som fikk delvis riktig svar (1 poeng).

Chronbach's alpha er et reliabilitetsmål som gir en indikasjon på hvor konsistent prøven måler den latente variabelen, leseforståelse, som et uttrykk for sann varians. En korrelasjonskoeffisient som Chronbachs alpha beregnes fra gjennomsnittet av alle mulige korrelasjonskoeffisienter mellom to «halvdeler» av prøven. En tilfredsstillende alpha for leseprøven bør ligge mellom 0,7 og 0,9.

2. Nasjonal prøve i lesing for 5. trinn 2023

2.1 Prøveutvikling i perioden 2021–2023

For å få et så representativt utvalg som mulig ved pilotering av nye oppgaver, inviterer vi skoler med ulikt antall elever fra hele landet. Skolene er tilfeldig utvalgt fra en liste gitt av Utdanningsdirektoratet. Prøven for 2023 er pilotert i to omganger: første gang i februar 2022 og andre gang i august 2022. I alle tilfeller deltok elever på 5. trinn, henholdsvis et halvt år eldre og like gamle som målgruppen. Vi vet at elevene blir mer modne og utvikler bedre leseferdigheter i løpet av skoleåret, og det er derfor viktig å være klar over denne aldersforskjellen når man tolker resultatene fra de to piloteringene.

Prøvene er kvalitetssikret av eksterne lesefaglige eksperter ved Uppsala universitet, av psykometrisk ekspert ved NTNU og av en referansegruppe bestående av lærere på mellomtrinnet.

2.2 Innholdet i prøven

Den nasjonale prøven i lesing for 2023 består av 5 tekster med til sammen 30 oppgaver som måler elevenes leseforståelse, i tråd med læreplanen og kompetansemålene etter 2.–4. trinn. Tekstene er presentert i tabellen nedenfor.

Tabell 1: Oversikt over tekstene i forslaget til nasjonal prøve i lesing på 5. trinn 2023.

Tekstnavn i rapport	Original tittel	Fagtilhørighet	Teksttype	Antall ord: 4197	Antall oppg.: 30	
					Flervalg: 25	Andre: 5
Fritidsaktiviteter i vikingtiden	«Fritidsaktiviteter i vikingtiden»	Samfunnsfag	Fagartikkel	743	5	
Disney-filmer får kjeft	«Disney-filmer får kjeft»	Samfunnsfag, krle	Nyhetsartikkel	793	5	1
Ikke lov å le	«Ikke lov å le»	Norsk	Skjønnlitterært utdrag	995	5	1
Sjøku	«Et skipsforlis og en sjøku», «Stellers sjøku»	Naturfag, samfunnsfag	Fagtekst, informasjonstekst	878	4	3
Paranøtteffekten	«Hvorfor havner de største nøttene øverst?»	Naturfag	Fagartikkel	788	6	

Det har vært et mål at prøven skal gjenspeile det tekstmangfoldet som elevene møter i ulike fag, og de utvalgte tekstene kan knyttes til fagene norsk, samfunnsfag, naturfag og krle. Tekstene er tildelt ulike plasser i prøven, i tråd med plan for teksttyper i ankeret. Den første teksten i prøven bør være engasjerende for elevene, og «Fritidsaktiviteter i vikingtiden» er valgt for at flest mulig elever skal få en god start på prøven. Dette er et av tiltakene for å forhindre at elevene mister motivasjonen og gir opp under prøvegjennomføring. Det er som vanlig én skjønnlitterær tekst med i utvalget. Flertallet av sakprosa tekstene er satt sammen av verbaltekst og andre tekstressurser, som figurer, faktabokser og grafiske framstillinger.

Prøven har en tradisjonell fordeling av oppgavene knyttet til de tre aspektene med 11 finne-, 15 tolke- og 4 refleksjonsoppgaver.

Ved siden av at alle oppgaver skal tilfredsstillende lesefaglige og psykometriske krav, legges det stor vekt på at prøven som helhet representerer en bredde i fagtilhørighet, sjangre, tema og leseutfordringer. Den følgende beskrivelsen av leseutfordringer i prøven viser at årets prøve ivaretar alle disse målsettingene, og at prøven måler lesing som grunnleggende ferdighet i flere fag:

«**Fritidsaktiviteter i vikingtiden**» er hentet fra bladet Nysgjerrigper og er skrevet av Kjersti Busterud. Teksten handler i hovedsak om brettspillet hnefatafl som var populært i vikingtiden. Spillet kan minne om sjakk. Begge spillene har firkantete brett delt inn i ruter og i begge spill er det om å gjøre å fange kongen, men reglene for hnefatafl som man har rekonstruert i ettertid, ligner ikke på reglene i sjakk. Fra spørreskjemaet som ble gjennomført etter første pilotering, var den blant tekstene som elevene rapporterte at de likte best i utprøvingen. For å løse oppgavene til teksten må elevene blant annet forstå hvordan vi kan vite at hnefatafl ble spilt i vikingtiden, og hvorfor det var viktig at barn lærte å spille dette spillet.

«**Disney-filmer får kjeft**» er en tekst fra Aftenposten Junior og handler om at flere av de gamle, klassiske Disney-filmene har fått kritikk for å være rasistiske og ha dårlig kvinnesyn. Blant annet blir det trukket fram at Kong Louie i «Jungelboken» er en rasistisk karikatur av afroamerikanere og at Snehvit ikke gir prinsen tillatelse til å kysse henne. Tema for teksten er tidsaktuelt, og filmene bør være kjent for de fleste barn. Flere av oppgavene til teksten handler om å få en helhetlig forståelse av temaet i teksten. Elevene må også kunne reflektere over hvem som får uttale seg i teksten og hva disse mener.

«**Ikke lov å le**» er prøvens skjønnlitterære tekst og er et utdrag fra en roman med samme navn av Kine Jeanette Solberg. I utdraget møter vi Kaja og moren hennes. De har nettopp flyttet til et nytt sted, og nå vil moren at Kaja skal invitere de nye klassekameratene sine til bursdagsfest. Men Kaja gruer seg, hun er utrygg og liker ikke mye oppmerksomhet. Utdraget var blant tekstene som elevene rapporterte at de likte best i utprøvingen. For å løse oppgavene til denne teksten må elevene tolke sammenhenger i fortellingen, forstå hvorfor aktørene sier eller gjør som de gjør, og holde orden på kronologien i teksten.

«**Sjøku**» består av to tekster som handler om Stellers sjøku. Den første teksten er hentet fra boka *Dyrene som forsvant* som er skrevet av Line Renslebråten. Denne boka er skrevet for aldersgruppa og handler om dyr som har blitt utryddet. Den andre teksten er hentet fra wikipedia. Stellers sjøku ble oppdaget av George Steller da ekspedisjonen til Vitus Bering strandet på en øde øy utenfor Kamtsjatka i Russland. Sjøkua var lett å fange, og kjøttet smakte godt. Bare 27 år etter at den ble oppdaget, var arten utryddet. For å svare riktig på oppgavene må elevene holde orden på i hvilken tekst de finner informasjonen. De må også forstå sammenhenger i teksten som ikke er tydelig uttrykt eller som er komplisert forklart.

«**Paranøtteffekten**» hentet fra bladet Nysgjerrigper og er skrevet av Veronica Danielsen. Teksten handler om et naturvitenskapelig fenomen som kalles paranøtteffekten. Fenomenet er vanskelig å forklare, men handler om at de største nøttene i en nøttepose som regel havner

øverst i posen når den ristes. Paranøtteffekten kan skape problemer når medisiner skal blandes, men den har også fordeler, for eksempel i skredredningsutstyr. Under pilotering rapporterte elevene at teksten var blant de vanskeligste. Selv om teksten er skrevet for aldersgruppa er innholdet komplekst og krevende å forstå. En av de vanskeligste oppgavene til teksten handler om å forstå hva som har skjedd i de tilfellene der de største nøttene ikke har havnet øverst, noe som er det motsatte av det forventete.

3. Første pilotering

I den første piloteringen ble det prøvd ut 13 tekster og 89 oppgaver fordelt på prøve A, B og C. Pilotprøven ble gjennomført i den nye, digitale plattformen EPS, og prøvetiden var 90 minutter, som ved ordinær gjennomføring. 1473 elever fra 39 skoler i 9 fylker (etter regionreformen 2020) deltok i den første piloteringen. Prøvene ble pilotert på 5. trinn i februar 2022, og det betyr at elevene som deltok i piloteringen, var ca. et halvt år eldre enn de elevene som gjennomførte prøven like etter skolestart høsten 2023.

Tabell 2: Oversikt over utvalget av elever (N=1473) i første pilotering.

Antall inviterte elever	4006 (fra 94 skoler)
Antall oppmeldte elever	1940 (fra 39 skoler)
Antall elever som gjennomførte (N)	Totalt: 1473 Prøve A: 482 (227 gutter, 255 jenter) Prøve B: 474 (237 gutter, 237 jenter) Prøve C: 517 (286 gutter, 249 jenter)
Antall skoler	39

3.1 Innholdet i prøve A, B og C

Tekstene fordeler seg på fag og teksttyper som vist i tabell 3, 4 og 5. Tabellene oppgir både navn for tekstene som brukes i rapporten («Tekstnavn i rapport») og original tittel for hver av tekstene. For hver tekst er det her oppgitt en p-verdi og kjønnsforskjell, basert på den gjennomsnittlige p-verdien til hver enkelt oppgave som er knyttet til teksten. Oppgavenes vanskegrad avhenger både av kvaliteter ved teksten og oppgaveformuleringen, og det vil som regel være både lette og vanskelige oppgaver knyttet til én og samme tekst. Den gjennomsnittlige p-verdien gir derfor ikke et presist bilde av hvor lett eller vanskelig en tekst er for elevene, men i vurderingen av hvilke tekster som passer sammen i en prøve, kan det være nyttig å se på gjennomsnittlig p-verdi for hver tekst og slik beregne vanskegraden totalt sett.

Antall ord som står oppgitt i tabellen, viser hvor mange ord hver enkelt tekst består av, inkludert oppgavene. Dette sier ikke noe om vanskegraden på tekstene, men gir en indikasjon på lengden på prøven og hvor mye elevene må lese. Prøve A, B og C har henholdsvis 4201, 4162 og 4222 ord, altså litt over gjennomsnittlige antall ord for nasjonal leseprøve på 5. trinn de siste 5 årene, som er 3880 ord.

Tabell 3: Oversikt over tekstene i prøve A.

Tekstnavn i rapport	Original tittel	Fagtilhørighet	Teksttype	Antall ord: 4201	Antall oppg.: 33		Gj. snittlig p-verdi	Kjønnsforskj. (j-g)*
					Flervalg: 32	Andre: 1		
Fritidsaktiviteter i vikingtiden	«Fritidsaktiviteter i vikingtiden»	Samfunnsfag	Fagartikkel	882	6		0,68	0,03
Tett på melbille	«Melbille-larver er perfekt middag»	Naturfag	Fagtekst	598	6		0,57	0,01
Blå flaggermus	«Blå flaggermus»	Norsk	Skjønnlitteratur	1304	8		0,58	0,08
Tasmansk pungulv	«Benjamin», «Den utryddede arten pungulv skal ha blitt observert åtte ganger»	Naturfag, samfunnsfag	Multippel	1092	7	1	0,40	0,01
Når gammel kjærlighet ruster	«Når gammel kjærlighet ruster»	Norsk, samfunnsfag, mat og helse	Reklame	325	5		0,35	-0,03

* Positive verdier er i jentenes favør, negative verdier i guttenes favør.

Tabell 4: Oversikt over tekstene i prøve B.

Tekstnavn i rapport	Original tittel	Fagtilhørighet	Teksttype	Antall ord: 4162	Antall oppg.: 34		Gj. snittlig p-verdi	Kjønnsforskj. (j-g)*
					Flervalg: 33	Andre: 1		
Påvirkere	«– Jeg synes det er dumt at folk blir påvirket»	Samfunnsfag	Junior	818	7		0,49	0,04
Tett på melbille	«Melbille-larver er perfekt middag»	Naturfag	Fagtekst	605	6		0,58	0,02
Ikke lov å le	«Ikke lov å le»	Norsk	Skjønnlitteratur	1318	6	1	0,60	0,12
Stripemysteriet	«Stripemysteriet», «Hvorfor har sebraen striper?»	Naturfag	Multippel	804	7		0,46	0,01
Kråkebolle	«Den korte historien om kråkeballen»	Mat og helse	Oppskrift	617	7		0,52	0,02

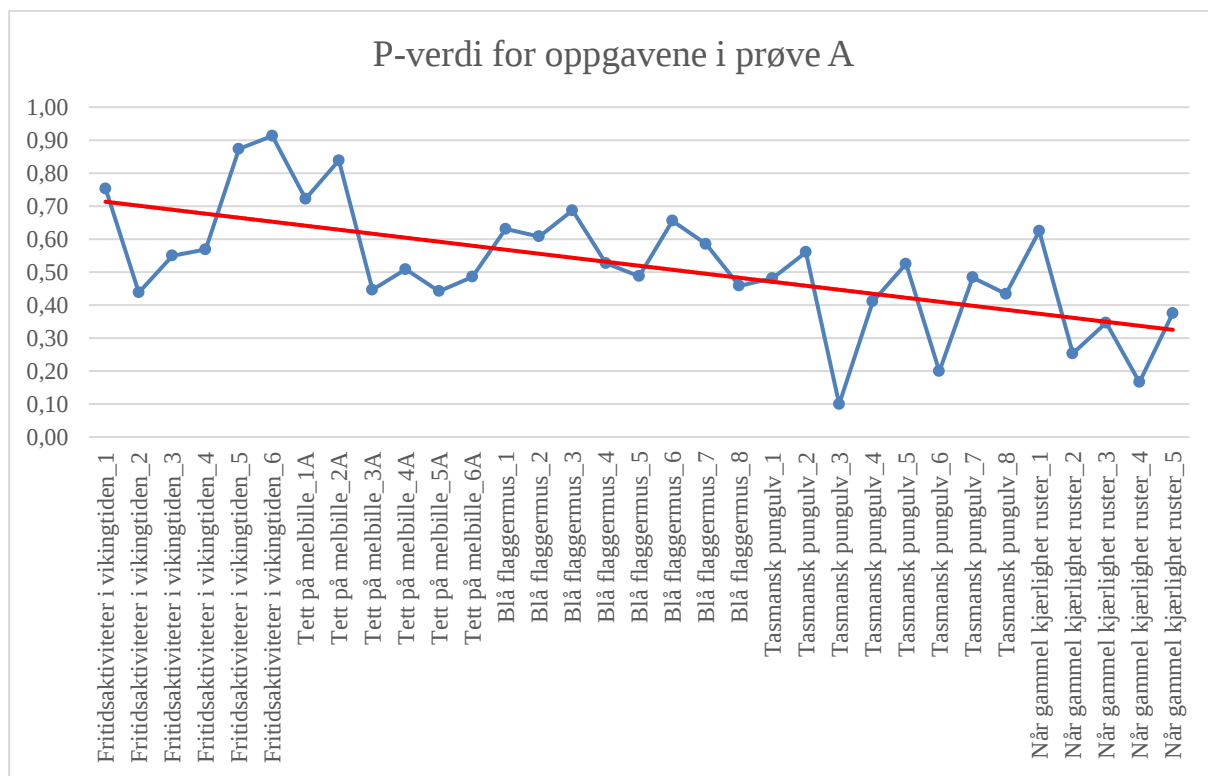
* Positive verdier er i jentenes favør, negative verdier i guttenes favør.

Tabell 5: Oversikt over tekstene i prøve C.

Tekstnavn i rapport	Original tittel	Fagtilhørighet	Teksttype	Antall ord: 4222	Antall oppg.: 34		Gj. snittlig p-verdi	Kjønnsforskj. (j-g)*
					Flervalg: 33	Andre: 1		
Kraken	«Kraken – sjømennenes skrekk»	Naturfag, samfunnsfag	Fagtekst	949	7		0,54	0,03
Tett på melbille	«Melbille-larver er perfekt middag»	Naturfag	Fagtekst	605	6		0,59	0,05
Verdens rareste jobb	«Min bror er konge!»	Norsk	Skjønnlitteratur	1006	7		0,50	0,09
Sjøku	«Et skipsforlis og en sjøku», «Stellers sjøku»	Naturfag	Multippel	812	6	1	0,50	0,05
Paranøtteffekten	«Hvorfor havner de største nøttene øverst?»	Naturfag	Fagartikkel	850	7		0,47	0,03

* Positive verdier er i jentenes favør, negative verdier i guttenes favør.

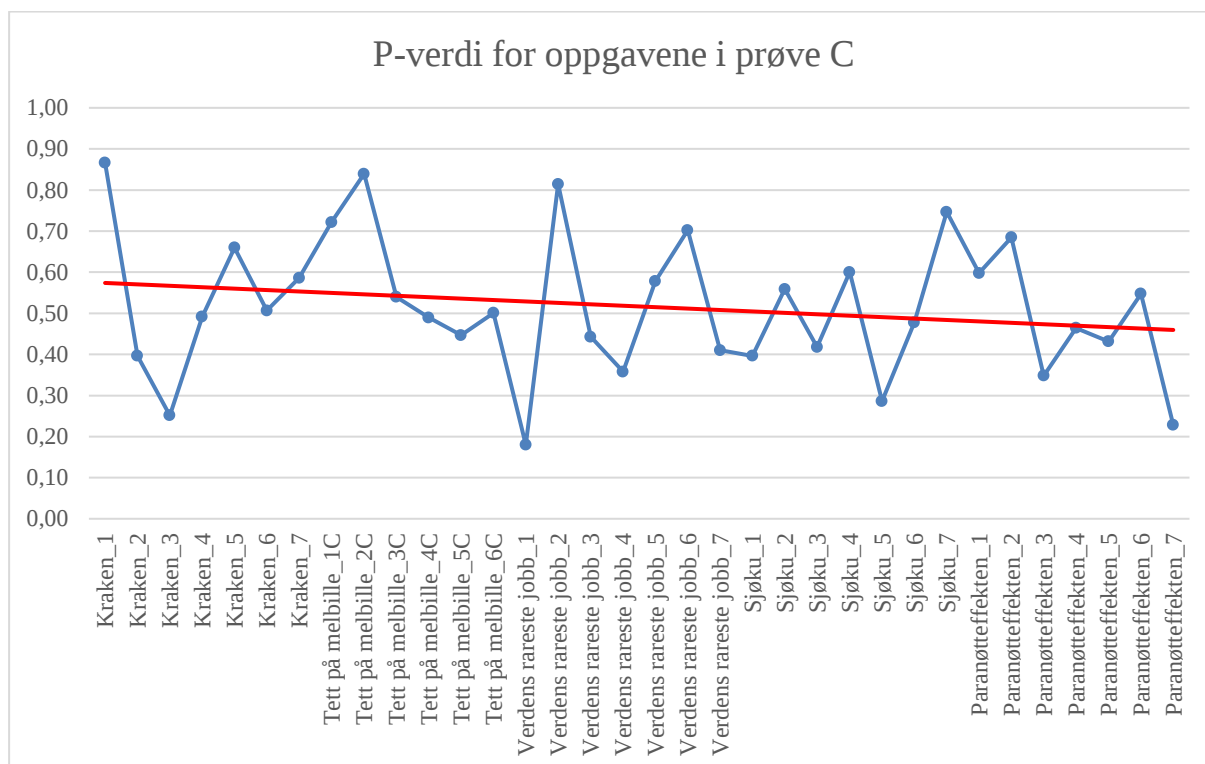
Figur 1, 2 og 3 viser p-verdier for hver oppgave i henholdsvis prøve A, B og C. Oppgavene presentert i samme rekkefølge som i prøvene.



Figur 1: Andelen elever med riktig svar (p-verdi) på oppgavene i prøve A, organisert etter rekkefølgen i prøven.



Figur 2: Andelen elever med riktig svar (p-verdi) på oppgavene i prøve B, organisert etter rekkefølgen i prøven.



Figur 3: Andelen elever med riktig svar (p-verdi) på oppgavene i prøve C, organisert etter rekkefølgen i prøven.

Figur 1, 2 og 3 viser at det er både lette og vanskeligere oppgaver knyttet til alle tekstene i prøven. Trendlinjen viser en tendens til at p-verdien blir noe lavere mot slutten av alle prøvene. I alle prøvene er det en tydelig økning i andelen ubesvarte fra og med tekst 4. Andelen ubesvarte er på 8-9 prosent mot slutten av prøvene, noe som er mer enn det som er vanlig etter en ordinær gjennomføring. Det kan altså virke som om heftene har vært noe for lange og tungleste enn det som er ønskelig, og at flere elever ikke har klart å besvare oppgaver til de to siste tekstene i prøvene.

3.2 Klassisk item-analyse av prøve A, B og C

Data fra både klassisk analyse og IRT-analyse vil være grunnlag for utvalg av tekster og oppgaver til andre pilot. Tabell 6, 7 og 8 viser resultater fra klassisk item-analyse av henholdsvis prøve A, B og C. Oppgavene er her gitt et løpenummer, og hver oppgave er tildelt en item-ID og et navn. Navnet viser hvilken tekst oppgaven tilhører og den nummerplasseringen oppgaven har innenfor tekstenheten under utprøving. Item-ID er et unikt nummer som følger oppgaven uavhengig av endringer og omplasseringer som eventuelt blir gjort mellom de ulike piloteringene.

Resultatene i tabellene viser prosentvis fordeling på de ulike svaralternativene for hver oppgave og gjennomsnittlig antall poeng på prøven for elevene som har valgt de ulike svaralternativene. Kolonnen «diskriminering» viser en biserial korrelasjon, som ifølge fastsatte mål bør være over 0,250 for at oppgaven skal differensiere godt mellom lesere med svake og sterke leseferdigheter. De siste kolonnene viser de ulike oppgavenes p-verdi og kjønnsforskjell, der positive verdier er i jentenes favør og negative verdier i guttenes favør. P-

verdien til de sammensatte flervalgsoppgavene er regnet ut ved å legge sammen p-verdien for 2 poeng og 50 % av p-verdien for 1 poeng.

Prøve A

Tabell 6: Resultater fra klassisk item-analyse, prøve A.

N: 482 (227 gutter, 255 jenter). Reliabilitet (Cronbachs Alpha): 0,833.

Gjennomsnittlig p-verdi: 0,52 (gutter: 0,51/ jenter: 0,53). Sum svarfordeling kan avvike fra 100 pga. avrunding.

Nr	Item-ID	Prøve A	Svarfordeling i % (frekvens)					Gj.snittlig poengsum på prøven per svaralt.				Diskrim. (biserial)	P-verdi	Kjønns-forskj. (j-g)	Merk
			A	B	C	D	Blank	A	B	C	D				
			0p	1p	2p										
1	L148S01	Fritidsaktiviteter i vikingtiden_1	9	5	75	10	1	12	12	19	13	0,397	0,75	0,02	
2	L148S02	Fritidsaktiviteter i vikingtiden_2	44	16	27	13	1	19	15	15	17	0,251	0,44	0,07	
3	L148S03	Fritidsaktiviteter i vikingtiden_3	31	6	55	6	1	15	13	19	14	0,351	0,55	0,02	
4	L148S04	Fritidsaktiviteter i vikingtiden_4	27	11	57	4	1	16	13	19	12	0,296	0,57	-0,04	
5	L148S05	Fritidsaktiviteter i vikingtiden_5	5	4	3	87	1	12	12	11	18	0,327	0,87	0,04	
6	L148S06	Fritidsaktiviteter i vikingtiden_6	91	3	3	2	1	18	10	12	9	0,324	0,91	0,04	
7	L149S01	Tett på melbille_1A	72	6	12	9	1	19	13	11	15	0,404	0,72	0,02	
8	L149S02	Tett på melbille_2A	84	6	4	6	1	18	9	11	12	0,450	0,84	0,04	
9	L149S03	Tett på melbille_3A	20	45	7	27	1	15	20	11	16	0,393	0,45	0,06	
10	L149S04	Tett på melbille_4A	8	11	29	51	1	13	14	15	20	0,440	0,51	0,03	
11	L149S05	Tett på melbille_5A	47	44	4	4	2	16	19	14	13	0,199	0,44	-0,06	*
12	L149S06	Tett på melbille_6A	23	13	49	14	1	15	15	20	15	0,316	0,49	0,00	
13	L150S01	Blå flaggermus_1	22	5	9	63	2	13	11	12	20	0,546	0,63	0,08	
14	L150S02	Blå flaggermus_2	11	61	13	14	2	16	18	15	16	0,175	0,61	0,14	*
15	L150S03	Blå flaggermus_3	69	8	7	15	2	19	12	12	13	0,467	0,69	0,12	
16	L150S04	Blå flaggermus_4	23	52	9	13	2	16	19	11	15	0,333	0,53	0,01	
17	L150S05	Blå flaggermus_5	25	5	20	49	2	14	11	15	20	0,461	0,49	0,07	
18	L150S06	Blå flaggermus_6	11	66	10	11	3	13	20	12	12	0,572	0,66	0,08	
19	L150S07	Blå flaggermus_7	8	27	6	59	2	11	15	11	20	0,503	0,59	0,10	
20	L150S08	Blå flaggermus_8	22	46	10	19	3	17	20	13	14	0,407	0,46	0,01	
21	L144S01	Tasmansk pungulv_1	48	37	7	4	4	19	17	13	13	0,151	0,48	0,01	*
22	L144S02	Tasmansk pungulv_2	15	16	9	56	5	14	16	13	20	0,383	0,56	0,00	
23	L144S03	Tasmansk pungulv_3	82	10			8	17	23			0,287	0,10	0,01	Rekkefølge
24	L144S04	Tasmansk pungulv_4	20	41	23	10	6	15	20	17	14	0,329	0,41	0,03	
25	L144S05	Tasmansk pungulv_5	53	25	7	10	6	20	16	12	13	0,435	0,52	-0,03	*
26	L144S06	Tasmansk pungulv_6	38	20	12	23	7	19	19	14	17	0,066	0,20	-0,08	
27	L144S08	Tasmansk pungulv_7	25	14	48	6	7	16	14	20	10	0,417	0,48	0,06	
28	L144S07	Tasmansk pungulv_8	24	43	15	9	9	17	20	14	16	0,307	0,43	0,06	
29	L151S01	Når gammel kjærlighet ruster_1	4	5	20	62	8	14	11	16	19	0,316	0,62	0,01	
30	L151S02	Når gammel kjærlighet ruster_2	15	25	21	30	8	15	19	15	19	0,155	0,25	-0,06	*
31	L151S04	Når gammel kjærlighet ruster_3	51	2	35	4	9	18	11	19	12	0,115	0,35	-0,08	*
32	L151S05	Når gammel kjærlighet ruster_4	17	17	26	31	10	17	18	17	19	-0,039	0,17	0,00	*
33	L151S06	Når gammel kjærlighet ruster_5	27	13	38	13	9	17	14	21	14	0,381	0,38	-0,04	

* Lav diskriminering (under 0,250).

Prøve B

Tabell 7: Resultater fra klassisk item-analyse, prøve B.

N: 474 (237 gutter, 237 jenter). Reliabilitet (Cronbachs Alpha): 0,860.

Gjennomsnittlig p-verdi: 0,53 (gutter: 0,51 / jenter: 0,55). Sum svarfordeling kan avvike fra 100 pga. avrunding.

Nr	Item-ID	Prøve B	Svarfordeling i % (frekvens)					Gj.snittlig poengsum på prøven per svaralt.				Diskrim. (biserial)	P-verdi	Kjønnsforskj. (j-g)	Merk
			A	B	C	D	Blank	A	B	C	D				
			0p	1p	2p										
1	L152S01	Påvirkere_1	89	7	2	3	0	19	12	9	15	0,256	0,89	0,08	
2	L152S02	Påvirkere_2	18	34	39	8	0	14	16	22	16	0,439	0,39	0,03	
3	L152S03	Påvirkere_3	24	15	51	10	0	16	14	21	14	0,346	0,51	0,00	
4	L152S04	Påvirkere_4	49	28	8	14	0	20	17	14	15	0,259	0,49	0,08	
5	L152S05	Påvirkere_5	36	39	14	11	0	17	20	18	14	0,204	0,39	0,04	*
6	L152S06	Påvirkere_6	37	41	18	4	0	20	17	16	15	0,210	0,36	0,00	*
7	L152S07	Påvirkere_7	25	31	5	38	0	17	15	10	22	0,422	0,38	0,04	
8	L149S01	Tett på melbille_1B	78	8	7	6	0	19	13	12	13	0,355	0,78	-0,05	
9	L149S02	Tett på melbille_2B	83	6	5	6	0	19	11	11	14	0,362	0,82	0,02	
10	L149S03	Tett på melbille_3B	25	49	7	19	1	14	21	12	16	0,446	0,49	0,05	
11	L149S04	Tett på melbille_4B	8	18	25	49	0	13	15	16	21	0,364	0,49	0,03	
12	L149S05	Tett på melbille_5B	51	41	4	4	1	17	20	15	14	0,226	0,41	0,03	*
13	L149S06	Tett på melbille_6B	23	13	50	14	1	15	15	21	14	0,448	0,50	0,03	
14	L153S01	Ikke lov å le_1	79	6	8	6	2	19	11	17	14	0,293	0,79	0,13	
15	L153S02	Ikke lov å le_2	12	6	77	3	3	14	13	19	11	0,363	0,77	0,17	
16	L153S03	Ikke lov å le_3	48	2	1	48	3	16	13	12	21	0,290	0,47	0,15	
17	L153S04	Ikke lov å le_4	8	59	7	23	3	12	20	12	17	0,367	0,59	0,12	
18	L153S05	Ikke lov å le_5	3	77	5	12	3	12	20	15	13	0,403	0,77	0,11	
19	L153S06	Ikke lov å le_6	87	9			3	18	25			0,292	0,09	0,03	Rekkefølge
20	L153S07	Ikke lov å le_7	4	70	14	9	3	12	20	14	16	0,398	0,70	0,14	
21	L154S01	Stripemysteriet_1	12	6	32	46	4	13	14	18	21	0,341	0,46	-0,10	
22	L154S02	Stripemysteriet_2	7	13	43	32	5	14	13	20	19	0,253	0,43	0,00	
23	L154S03	Stripemysteriet_3	10	12	12	61	5	12	14	15	21	0,538	0,61	0,03	
24	L154S04	Stripemysteriet_4	56	11	13	14	6	21	14	16	16	0,376	0,56	-0,03	
25	L154S05	Stripemysteriet_5	30	14	14	36	7	16	17	17	22	0,407	0,36	0,07	
26	L154S06	Stripemysteriet_6	16	45	15	18	6	14	21	18	16	0,382	0,45	0,08	
27	L154S07	Stripemysteriet_7	21	12	35	24	7	18	14	21	18	0,305	0,35	0,04	
28	L155S01	Kråkebolle_1	12	32	39	10	7	13	17	23	16	0,531	0,39	0,04	
29	L155S02	Kråkebolle_2	5	11	66	11	8	13	14	20	13	0,515	0,66	0,09	
30	L155S03	Kråkebolle_3	50	7	25	11	8	20	13	18	18	0,259	0,50	0,02	
31	L155S04	Kråkebolle_4	26	52	11	3	8	16	21	14	14	0,452	0,52	-0,04	
32	L155S05	Kråkebolle_5	9	14	25	44	8	14	14	17	22	0,495	0,44	0,02	
33	L155S06	Kråkebolle_6	23	49	19	1	8	15	21	17	14	0,388	0,49	0,00	
34	L155S07	Kråkebolle_7	8	15	4	65	8	16	16	11	20	0,373	0,65	0,00	

* Lav diskriminering (under 0,250).

Prøve C

Tabell 8: Resultater fra klassisk item-analyse, prøve C.

N: 517 (286 gutter, 249 jenter). Reliabilitet (Cronbachs Alpha): 0,881.

Gjennomsnittlig p-verdi: 0,52 (gutter: 0,49 / jenter: 0,54). Sum svarfordeling kan avvike fra 100 pga. avrunding.

Nr	Item-ID	Prøve C	Svarfordeling i % (frekvens)					Gj.snittlig poengsum på prøven per svaralt.				Diskrim. (biserial)	P-verdi	Kjønns-forskj. (j-g)	Merk
			A	B	C	D	Blank	A	B	C	D				
			0p	1p	2p										
1	L102S01	Kraken_1	6	4	2	87	1	12	13	13	19	0,282	0,87	-0,01	*
2	L102S03	Kraken_2	22	29	40	9	1	17	15	22	14	0,390	0,40	0,06	
3	L102S04	Kraken_3	15	32	27	25	1	16	18	17	22	0,222	0,25	-0,02	*
4	L102S05	Kraken_4	16	16	49	18	2	13	18	21	17	0,250	0,49	0,13	
5	L102S06	Kraken_5	66	12	12	9	1	19	14	17	19	0,128	0,66	-0,05	*
6	L102S07	Kraken_6	13	51	15	20	1	15	21	15	16	0,353	0,51	0,11	
7	L102S09	Kraken_7	14	13	59	13	2	14	16	21	15	0,387	0,59	0,00	
8	L149S01	Tett på melbille_1C	72	6	11	9	2	20	13	12	14	0,448	0,72	0,03	
9	L149S02	Tett på melbille_2C	84	4	4	5	2	20	10	13	13	0,390	0,84	0,04	
10	L149S03	Tett på melbille_3C	19	54	6	19	3	14	22	11	16	0,515	0,54	0,12	
11	L149S04	Tett på melbille_4C	8	15	26	49	2	12	15	16	22	0,461	0,49	0,06	
12	L149S05	Tett på melbille_5C	45	45	3	5	2	17	20	17	16	0,195	0,45	-0,02	*
13	L149S06	Tett på melbille_6C	20	12	50	15	3	14	14	22	16	0,480	0,50	0,07	
14	L156S01	Verdens rareste jobb_1	3	18	69	6	4	13	23	18	19	0,260	0,18	0,01	
15	L156S02	Verdens rareste jobb_2	11	2	81	2	4	12	11	20	11	0,474	0,81	0,12	
16	L156S03	Verdens rareste jobb_3	13	15	44	24	4	15	16	22	17	0,381	0,44	0,08	
17	L156S04	Verdens rareste jobb_4	12	16	33	36	4	14	14	19	22	0,336	0,36	0,10	
18	L156S05	Verdens rareste jobb_5	14	16	58	9	4	15	16	21	13	0,423	0,58	0,18	
19	L156S06	Verdens rareste jobb_6	70	11	7	8	4	20	14	12	15	0,416	0,70	0,02	
20	L156S07	Verdens rareste jobb_7	17	23	41	15	4	16	18	22	15	0,353	0,41	0,15	
21	L157S01	Sjøku_1	38	40	8	9	5	18	22	15	15	0,323	0,40	0,06	
22	L157S02	Sjøku_2	56	17	14	7	6	22	14	16	15	0,509	0,56	0,06	
23	L157S03	Sjøku_3	18	16	18	42	6	15	14	19	22	0,396	0,42	0,00	
24	L157S04	Sjøku_4	17	46	36			11	16	24		0,623	0,60	0,06	Tabell
		Sjøku_4a	68	5	12	8	8	21	13	15	15				
		Sjøku_4b	8	18	55	6	13	15	16	21	14				
		Sjøku_4c	55	7	6	17	15	22	13	14	16				
25	L157S05	Sjøku_5	8	34	29	22	7	16	18	22	17	0,309	0,29	0,09	
26	L157S06	Sjøku_6	17	17	11	48	7	15	15	14	21	0,546	0,48	0,05	
27	L157S07	Sjøku_7	75	6	5	7	8	21	14	13	12	0,504	0,75	0,05	
28	L158S01	Paranøtteffekten_1	7	12	11	60	10	12	14	14	22	0,605	0,60	0,09	
29	L158S02	Paranøtteffekten_2	8	8	6	69	10	14	11	13	21	0,552	0,68	0,11	
30	L158S03	Paranøtteffekten_3	19	35	20	17	10	15	22	19	17	0,371	0,35	0,04	
31	L158S04	Paranøtteffekten_4	12	46	16	15	11	15	22	15	17	0,446	0,46	-0,05	*
32	L158S05	Paranøtteffekten_5	43	10	14	22	11	22	14	18	17	0,396	0,43	0,02	
33	L158S06	Paranøtteffekten_6	12	55	17	6	11	14	22	16	14	0,497	0,55	0,00	
34	L158S07	Paranøtteffekten_7	24	19	22	23	12	17	17	20	22	0,219	0,23	-0,01	*

* Lav diskriminering (under 0,250).

3.3 IRT-analyse av prøve A, B og C

IRT-analysen av oppgavene i prøve A, B og C er utført ved hjelp av programvaren XCalibre 4.2. Tabell 9 viser parameterne for alle oppgavene i prøvene.

De tre prøvene har en felles tekst, «Tett på melbille», og med hjelp av denne teksten har vi samkalibrert de tre prøvene slik at alle verdier er på samme skala. Bortsett fra teksten «Tett på melbille», som er plassert til sist, ligger tekster og oppgaver i den rekkefølgen som de lå i under gjennomføringen, og fargene viser hvilke oppgaver som hører til henholdsvis prøve A, B og C.

Prøve ABC

Tabell 9: Parametere fra IRT-analyse av oppgavene i prøve A, B og C.

A: N=482 (227 gutter, 255 jenter). B: N=474 (237 gutter, 237 jenter). C: N=517 (268 gutter, 249 jenter).

Lav diskriminering (point biserial < 0,250 samt a < 0,5) er merket (*).

Oppg.	Item-ID	Oppgavenavn	Diskriminering (point biserial)	a	b	DIF (disfavør) Flagg
1	L148S01	Fritidsaktiviteter i vikingtiden_1	0,397	0,765	-1,177	
2	L148S02	Fritidsaktiviteter i vikingtiden_2	0,251	0,407*	0,345	
3	L148S03	Fritidsaktiviteter i vikingtiden_3	0,351	0,564	-0,301	
4	L148S04	Fritidsaktiviteter i vikingtiden_4	0,296	0,473*	-0,439	
5	L148S05	Fritidsaktiviteter i vikingtiden_5	0,327	0,781	-1,913	
6	L148S06	Fritidsaktiviteter i vikingtiden_6	0,324	0,922	-2,072	
13	L150S01	Blå flaggermus_1	0,546	1,121	-0,508	F
14	L150S02	Blå flaggermus_2	0,175*	0,370*	-0,797	gutter
15	L150S03	Blå flaggermus_3	0,467	0,899	-0,781	
16	L150S04	Blå flaggermus_4	0,333	0,574	-0,184	
17	L150S05	Blå flaggermus_5	0,461	0,816	-0,007	
18	L150S06	Blå flaggermus_6	0,572	1,203	-0,583	F
19	L150S07	Blå flaggermus_7	0,503	0,933	-0,372	
20	L150S08	Blå flaggermus_8	0,407	0,685	0,128	
21	L144S01	Tasmansk pungulv_1	0,151*	0,333*	0,094	
22	L144S02	Tasmansk pungulv_2	0,383	0,613	-0,335	
23	L144S03	Tasmansk pungulv_3	0,287	0,755	2,125	
24	L144S04	Tasmansk pungulv_4	0,329	0,555	0,402	
25	L144S05	Tasmansk pungulv_5	0,435	0,715	-0,160	
26	L144S06	Tasmansk pungulv_6	0,066*	0,346*	2,438	
27	L144S08	Tasmansk pungulv_7	0,417	0,685	0,019	
28	L144S07	Tasmansk pungulv_8	0,307	0,511	0,307	
29	L151S01	Når gammel kjærlighet ruster_1	0,316	0,527	-0,710	
30	L151S02	Når gammel kjærlighet ruster_2	0,155*	0,382*	1,747	
31	L151S04	Når gammel kjærlighet ruster_3	0,115*	0,331*	1,144	
32	L151S05	Når gammel kjærlighet ruster_4	-0,039*	0,311*	3,083	K, Hb
33	L151S06	Når gammel kjærlighet ruster_5	0,381	0,635	0,531	
34	L152S01	Påvirkere_1	0,256	0,640	-2,238	
35	L152S02	Påvirkere_2	0,439	0,741	0,446	
36	L152S03	Påvirkere_3	0,346	0,554	-0,054	
37	L152S04	Påvirkere_4	0,259	0,436*	0,031	
38	L152S05	Påvirkere_5	0,204*	0,385*	0,766	
39	L152S06	Påvirkere_6	0,210*	0,406*	0,876	
40	L152S07	Påvirkere_7	0,422	0,697	0,513	
41	L153S01	Ikke lov å le_1	0,293	0,597	-1,563	
42	L153S02	Ikke lov å le_2	0,363	0,719	-1,274	

43	L153S03	Ikke lov å le_3	0,290	0,486*	0,131	
44	L153S04	Ikke lov å le_4	0,367	0,618	-0,420	
45	L153S05	Ikke lov å le_5	0,403	0,827	-1,164	
46	L153S06	Ikke lov å le_6	0,292	0,723	2,320	
47	L153S07	Ikke lov å le_7	0,398	0,743	-0,874	
48	L154S01	Stripemysteriet_1	0,341	0,543	0,172	
49	L154S02	Stripemysteriet_2	0,253	0,439*	0,413	
50	L154S03	Stripemysteriet_3	0,538	1,055	-0,422	
51	L154S04	Stripemysteriet_4	0,376	0,615	-0,301	
52	L154S05	Stripemysteriet_5	0,407	0,669	0,619	
53	L154S06	Stripemysteriet_6	0,382	0,606	0,217	
54	L154S07	Stripemysteriet_7	0,305	0,530	0,793	
55	L155S01	Kråkebolle_1	0,531	0,990	0,378	
56	L155S02	Kråkebolle_2	0,515	1,025	-0,596	
57	L155S03	Kråkebolle_3	0,259	0,458*	0,004	
58	L155S04	Kråkebolle_4	0,452	0,775	-0,096	
59	L155S05	Kråkebolle_5	0,495	0,873	0,194	
60	L155S06	Kråkebolle_6	0,388	0,634	0,053	
61	L155S07	Kråkebolle_7	0,373	0,648	-0,717	
62	L102S01	Kraken_1	0,282	0,645	-2,065	
63	L102S03	Kraken_2	0,390	0,556	0,559	
64	L102S04	Kraken_3	0,222*	0,418*	1,737	
65	L102S05	Kraken_4	0,250	0,391*	0,082	
66	L102S06	Kraken_5	0,128*	0,323*	-1,252	
67	L102S07	Kraken_6	0,353	0,521	-0,019	
68	L102S09	Kraken_7	0,387	0,595	-0,413	
69	L156S01	Verdens rareste jobb_1	0,260	0,493*	2,117	
70	L156S02	Verdens rareste jobb_2	0,474	1,143	-1,228	
71	L156S03	Verdens rareste jobb_3	0,381	0,541	0,319	
72	L156S04	Verdens rareste jobb_4	0,336	0,505	0,825	
73	L156S05	Verdens rareste jobb_5	0,423	0,633	-0,361	
74	L156S06	Verdens rareste jobb_6	0,416	0,706	-0,925	
75	L156S07	Verdens rareste jobb_7	0,353	0,512	0,515	
76	L157S01	Sjøku_1	0,323	0,487*	0,614	
77	L157S02	Sjøku_2	0,509	0,835	-0,238	
78	L157S03	Sjøku_3	0,396	0,564	0,440	
79	L157S04	Sjøku_4	0,712	1,244	-1,12, 0,41	
80	L157S05	Sjøku_5	0,309	0,486*	1,311	
81	L157S06	Sjøku_6	0,546	0,932	0,083	
82	L157S07	Sjøku_7	0,504	1,086	-0,940	
83	L158S01	Paranøtteffekten_1	0,605	1,203	-0,355	
84	L158S02	Paranøtteffekten_2	0,552	1,141	-0,677	
85	L158S03	Paranøtteffekten_3	0,371	0,574	0,809	
86	L158S04	Paranøtteffekten_4	0,446	0,665	0,176	
87	L158S05	Paranøtteffekten_5	0,396	0,588	0,359	
88	L158S06	Paranøtteffekten_6	0,497	0,791	-0,194	
89	L158S07	Paranøtteffekten_7	0,219*	0,437*	1,871	
7	L149S01	Tett på melbille_1	0,404	0,735	-1,104	
8	L149S02	Tett på melbille_2	0,396	0,908	-1,481	
9	L149S03	Tett på melbille_3	0,458	0,712	0,010	
10	L149S04	Tett på melbille_4	0,420	0,641	-0,001	
11	L149S05	Tett på melbille_5	0,204*	0,315*	0,541	
12	L149S06	Tett på melbille_6	0,420	0,641	0,005	

I tabell 9 er oppgavenes diskrimineringssevne oppgitt både som et klassisk mål (point biserial) og en a-verdi. Oppgaver med diskriminering under 0,250 (point biserial) samt a-parameter under 0,5 er merket med stjerne. Oppgaver som er merket for begge diskriminerings-

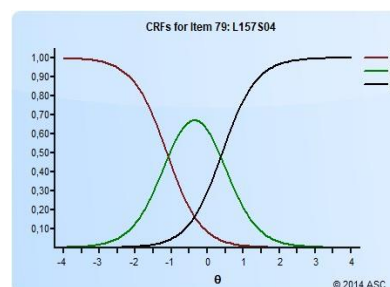
parameterne, er uthevet i tabellen, og de oppgavene som blir med videre i prøven blir kommentert i kapitlet 3.5, «Utvalg av tekster og oppgaver til andre pilotering», og kapittel 3.6, «Endringer etter første pilotering». Det samme gjelder oppgavene med DIF (Differential Item Functioning) for kjønn.

En av oppgavene i piloten er polytom, altså en oppgave hvor det er det mulig å få mer enn ett poeng. I analysene etter piloten gir vi gjerne 1 og 2 poeng på slike oppgaver, for å se i hvilken grad det er forskjell på elever som får 0, 1 og 2 poeng. Utfra disse analysene velger vi om vi behandler oppgaven som polytom eller dikotom i neste pilotering og endelig gjennomføring.

Oppgave Sjøku_4 (L157S04) er en paringstabell (match) med tre påstander og fire svarkolonner. Vi har rekodet oppgaven til 2 poeng, 1 poeng og 0 poeng etter følgende omregningsnøkkel: 3 poeng gir 2 poeng, 2 eller 1 poeng gir 1 poeng, og 0 poeng gir 0 poeng. IRT-analysen som vises i tabell 10 viser at oppgaven skiller godt mellom elever som har fått 0, 1 eller 2 poeng.

Tabell 10: Parametere fra IRT-analyse av oppgave 79 Sjøku_4 (L157S04)

Category	N	Prop.	Mean	SD	Boundary Location (b)	b SE
0	89	0,172	-1,012	0,680	-1,123	0,037
1	240	0,464	-0,265	0,616	0,413	0,032
2	188	0,364	0,880	0,754		
Omit	0					
Not Admin	956		-0,027	0,904		



(NB: på grunn av en feil i dataprogrammet blir Category 0 vist som 1, rød linje, i grafen, Category 1 vist som 2, grønn linje, osv.)

3.4 Oppgavene fordelt på mestringsnivåer

Tabell 11 viser oppgavene i det første piloteringsforslaget, fordelt på mestringsnivåer. Nivågrensene som er brukt, ble bestemt etter gjennomføringen av nasjonale leseprøver i 2016. I theta-verdiene som kommer ut av denne IRT-analysen, er sentrert på theta-fordelingen i piloteringen og er ikke de samme som ved selve gjennomføringen. Piloteringen er altså ikke ankret på startpunktet av trenden i 2016. Det er derfor en viss usikkerhet knyttet til nivågrensene. Etter gjennomføring av prøven i 2022 ble det satt nye nivågrenser, og disse vil brukes når resultatene etter andre pilot analyseres.

Tabell 11: Oppgavene i piloteringsforslaget, fordelt på mestringsnivåer.

Nivå	Løpenummer	Item-ID	Oppgave	Skalapoeng
Nivå 1	34	L152S01	Påvirkere_1	27,618
	6	L148S06	Fritidsaktiviteter i vikingtiden_6	29,282
	62	L102S01	Kraken_1	29,347
	5	L148S05	Fritidsaktiviteter i vikingtiden_5	30,872

	41	L153S01	Ikke lov å le_1	34,367
	8	L149S02	Tett på melbille_2	35,188
	42	L153S02	Ikke lov å le_2	37,258
	66	L102S06	Kraken_5	37,485
	70	L156S02	Verdens rareste jobb_2	37,724
	1	L148S01	Fritidsaktiviteter i vikingtiden_1	38,234
	45	L153S05	Ikke lov å le_5	38,365
	79_1p	L157S04	Sjøku_4_1p	38,77
	7	L149S01	Tett på melbille_1	38,958
	82	L157S07	Sjøku_7	40,605
	74	L156S06	Verdens rareste jobb_6	40,751
	47	L153S07	Ikke lov å le_7	41,258
	14	L150S02	Blå flaggermus_2	42,033
	15	L150S03	Blå flaggermus_3	42,189
	61	L155S07	Kråkebolle_7	42,828
	29	L151S01	Når gammel kjærlighet rustet_1	42,896
Nivå 2	84	L158S02	Paranøtteffekten_2	43,228
	56	L155S02	Kråkebolle_2	44,036
	18	L150S06	Blå flaggermus_6	44,171
	13	L150S01	Blå flaggermus_1	44,92
	4	L148S04	Fritidsaktiviteter i vikingtiden_4	45,606
	50	L154S03	Stripemysteriet_3	45,78
	44	L153S04	Ikke lov å le_4	45,799
	68	L102S09	Kraken_7	45,869
	19	L150S07	Blå flaggermus_7	46,283
	73	L156S05	Verdens rareste jobb_5	46,39
	83	L158S01	Paranøtteffekten_1	46,447
	22	L144S02	Tasmansk pungulv_2	46,646
	51	L154S04	Stripemysteriet_4	46,99
	3	L148S03	Fritidsaktiviteter i vikingtiden_3	46,992
	77	L157S02	Sjøku_2	47,618
	88	L158S06	Paranøtteffekten_6	48,057
	16	L150S04	Blå flaggermus_4	48,162
	25	L144S05	Tasmansk pungulv_5	48,4
	58	L155S04	Kråkebolle_4	49,037
	36	L152S03	Påvirkere_3	49,459
	67	L102S07	Kraken_6	49,811
	17	L150S05	Blå flaggermus_5	49,931
	10	L149S04	Tett på melbille_4	49,989
	57	L155S03	Kråkebolle_3	50,044
	12	L149S06	Tett på melbille_6	50,052
	9	L149S03	Tett på melbille_3	50,098
	27	L144S08	Tasmansk pungulv_7	50,19
	37	L152S04	Påvirkere_4	50,311
	60	L155S06	Kråkebolle_6	50,532
	65	L102S05	Kraken_4	50,818
	81	L157S06	Sjøku_6	50,827
	21	L144S01	Tasmansk pungulv_1	50,939
	20	L150S08	Blå flaggermus_8	51,283
	43	L153S03	Ikke lov å le_3	51,314
	48	L154S01	Stripemysteriet_1	51,723
	86	L158S04	Paranøtteffekten_4	51,758
	59	L155S05	Kråkebolle_5	51,939
	53	L154S06	Stripemysteriet_6	52,17
	28	L144S07	Tasmansk pungulv_8	53,068
	71	L156S03	Verdens rareste jobb_3	53,187
	2	L148S02	Fritidsaktiviteter i vikingtiden_2	53,446
	87	L158S05	Paranøtteffekten_5	53,589
	55	L155S01	Kråkebolle_1	53,783
	24	L144S04	Tasmansk pungulv_4	54,024

	49	L154S02	Stripemysteriet_2	54,127
	79_2p	L157S04	Sjøku_4_2p	54,134
	78	L157S03	Sjøku_3	54,405
	35	L152S02	Påvirkere_2	54,46
	40	L152S07	Påvirkere_7	55,132
	75	L156S07	Verdens rareste jobb_7	55,154
	33	L151S06	Når gammel kjærlighet rustet_5	55,306
	11	L149S05	Tett på melbille_5	55,41
	63	L102S03	Kraken_2	55,591
	76	L157S01	Sjøku_1	56,144
	52	L154S05	Stripemysteriet_5	56,188
	38	L152S05	Påvirkere_5	57,657
	54	L154S07	Stripemysteriet_7	57,926
Nivå 3	85	L158S03	Paranøtteffekten_3	58,085
	72	L156S04	Verdens rareste jobb_4	58,248
	39	L152S06	Påvirkere_6	58,758
	31	L151S04	Når gammel kjærlighet rustet_3	61,445
	80	L157S05	Sjøku_5	63,112
	64	L102S04	Kraken_3	67,371
	30	L151S02	Når gammel kjærlighet rustet_2	67,467
	89	L158S07	Paranøtteffekten_7	68,707
	69	L156S01	Verdens rareste jobb_1	71,174
	23	L144S03	Tasmansk pungulv_3	71,254
	46	L153S06	Ikke lov å le_6	73,203
	26	L144S06	Tasmansk pungulv_6	74,384
	32	L151S05	Når gammel kjærlighet rustet_4	80,827

Ideelt sett skal prøven måle tilfredsstillende og med et tilstrekkelig antall oppgaver i alle deler av skalaen, jevnt fordelt mellom mestringsnivåene. Dette har imidlertid vist seg å være vanskelig å få til, og i prøveforslaget for 2023 ser vi at det er færre oppgaver på det øverste nivået, enn på nivå 1 og 2. I tabell 12 under vises diskrimineringstallene til oppgavene på nivå 3. Det er viktig å sikre at prøven inneholder oppgaver på dette nivået. Tabellen viser at mange av oppgavene på dette nivået har en lav diskriminering (merket med stjerne og grå utheving), og vi prøver derfor å reparere oppgavene uten å gjøre oppgavene lettere. Vi følger derfor disse oppgavene nøye i det videre arbeidet med prøven.

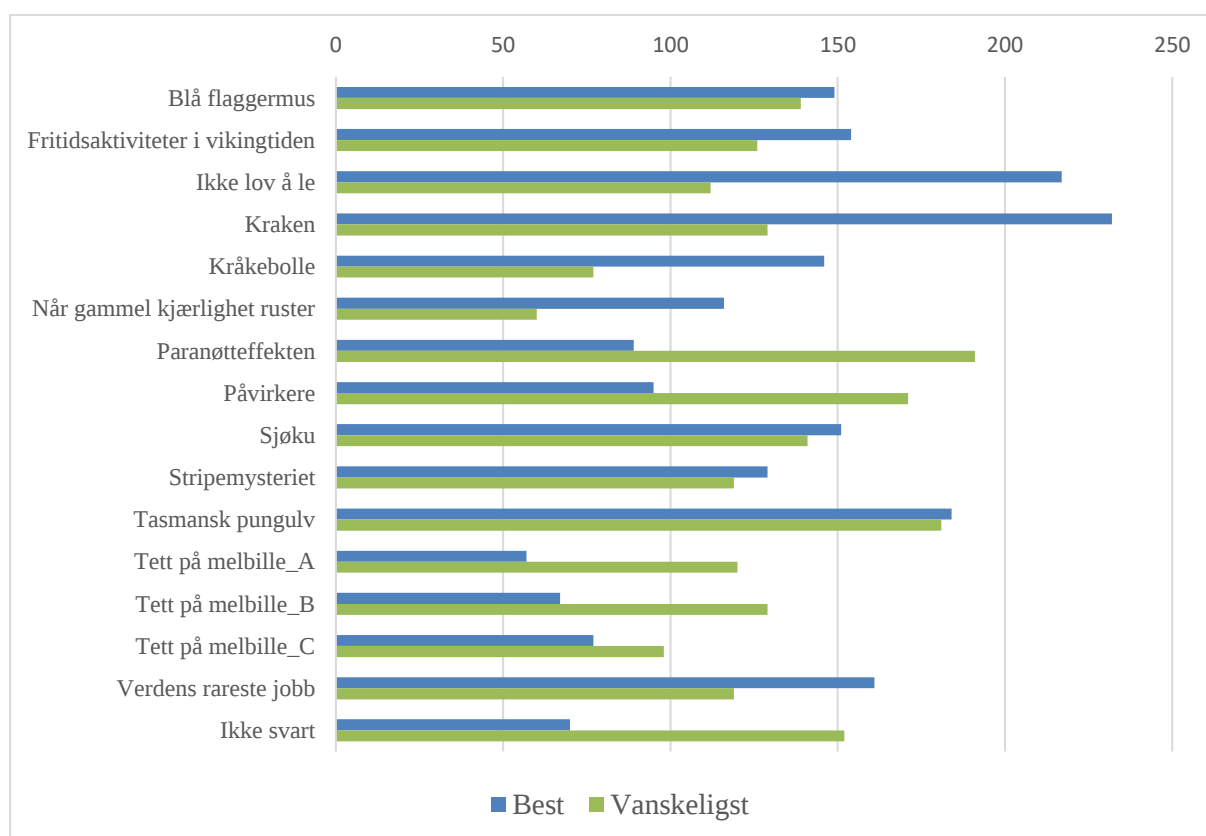
Tabell 12: Oppgavene på nivå 3 med diskrimineringstall

	Item-ID	Oppgave	Diskrim. (point biserial)	Diskrim. a	Skalapoeng
Nivå 3	L158S03	Paranøtteffekten_3	0,371	0,574	58,085
	L156S04	Verdens rareste jobb_4	0,336	0,505	58,248
	L152S06	Påvirkere_6	0,210*	0,406*	58,758
	L151S04	Når gammel kjærlighet rustet_3	0,115*	0,331*	61,445
	L157S05	Sjøku_5	0,309	0,486*	63,112
	L102S04	Kraken_3	0,222*	0,418*	67,371
	L151S02	Når gammel kjærlighet rustet_2	0,155*	0,382*	67,467
	L158S07	Paranøtteffekten_7	0,219*	0,437*	68,707
	L156S01	Verdens rareste jobb_1	0,260	0,493*	71,174
	L144S03	Tasmansk pungulv_3	0,287	0,755	71,254
	L153S06	Ikke lov å le_6	0,292	0,723	73,203
	L144S06	Tasmansk pungulv_6	0,066*	0,346*	74,384
	L151S05	Når gammel kjærlighet rustet_4	-0,039*	0,311*	80,827

3.5 Utvalg av tekster til andre pilotering

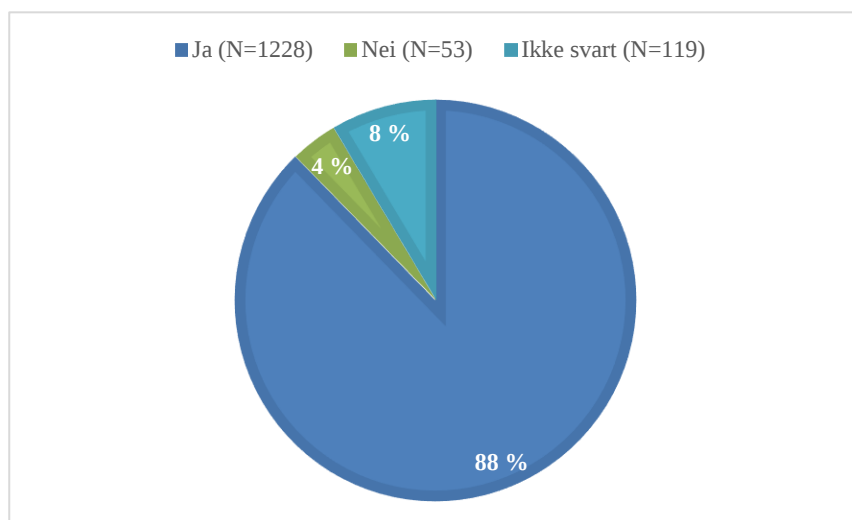
I prosessen med å velge ut tekster til andre pilotering ligger ikke bare resultatene og analysene fra første pilotering til grunn, men også tilbakemeldinger fra de eksterne, lesefaglige kvalitetssikrerne og lærere i referansegruppen. Kvalitetssikrerne fra Uppsala vurderte først og fremst kvaliteten på oppgaveformuleringene og svaralternativene, mens lærerne vurderte tekster og oppgaver med tanke på vanskegrad og egnethet for aldersgruppa, samt potensialet for anvendelse og etterarbeid i klasserommet. Referanselærerne syntes at de fleste piloterte tekstene egnet seg som prøvetekster for femtetrinn.

Elevene som deltok på den første piloteringen, ga også tilbakemeldinger på hvilke tekster de likte best og hvilke de syntes var vanskeligst. Tilbakemeldingene ble innhentet ved hjelp av et enkelt spørreskjema (vedlegg 3) som lå bakerst i prøven. Figur 4 viser hvor mange ganger en tekst ble nevnt som svar på spørsmålene «Hvilken tekst likte du best?» og «Hvilken tekst var vanskeligst?». Ettersom elevene kunne oppgi mer enn én tekst på begge spørsmålene, er det til sammen flere «stemmer» enn antall elever. Det er nyttig for oss å få informasjon om hva elevene selv tenker om tekstene de har lest i prøven. For eksempel er «Tasmansk pungulv» den nest vanskeligste teksten, men også den tredje best likte.



Figur 4: Elevsvar fra spørreundersøkelsen, «hvilken tekst likte du best» (blå søyler) og «hvilken tekst var vanskeligst» (røde søyler).

Elevene ble også spurt om de fikk god nok tid til å svare på oppgavene. Figur 5 viser at et stort flertall av elevene opplevde å ha god nok tid på prøven.



Figur 5: Elevsvar fra spørreundersøkelsen, «fikk du god nok tid til å svare på alle oppgavene».

De fem tekstene som gikk videre til andre pilotering, er presentert nærmere i «Nasjonal leseprøve for 5. trinn 2023» i delkapittel 2.2. Her gir vi noen korte begrunnelser for hvorfor de følgende tekstene **ikke** ble brukt i den videre piloteringen.

- **Tett på melbille:** Kvalitetssikrerne syntes teksten hadde et spennende tema og et interessevekkende bilde, men at selve teksten ikke var så godt skrevet. Teksten er lagt på vent, og kan eventuelt brukes seinere hvis teksten bearbeides.
- **Blå flaggermus:** Kvalitetssikrerne mente dette var en morsom og gjenkjennelig tekst for elevene. Teksten fungerer godt som utdrag, men blir lagt på vent for å unngå å ha samme forfatteren i prøven for ofte.
- **Tasmansk pungulv:** Denne teksten ble vurdert til å være for vanskelig på 5. trinn. Den ble også pilotert på 8. trinn og går videre til andre pilotering i den prøven. Kvalitetssikrerne syntes både tema og tekst var spennende, men konkluderte med at det er for vanskelig for elever på 5. trinn å holde oversikt i teksten.
- **Når gammel kjærlighet rustet:** Denne teksten mente kvalitetssikrerne var lite aldersadekvat, selv om det er fint med reklametekster i prøvene.
- **Påvirkere:** Kvalitetssikrerne likte denne teksten godt. Den har et viktig og aktuelt tema, men den har også noen vanskelige ord, og det er mulig at ikke alle elever er kjent med eller kjenner seg igjen i problematikken. Teksten er lagt på vent.
- **Stripemysteriet:** Kvalitetssikrerne var enige om at det er et spennende tema, men at det ikke utnyttes godt nok at det er to tekster. Tekstene er litt for like, og det er ikke nok motstridene eller konkurrerende informasjon.

- **Kråkebolle:** Tilbakemeldingene til denne teksten gikk ut på at det er gøy med oppskrift og at det er en spennende matrett, men at teksten inneholder en del vanskelige begreper og verb. Teksten er lagt på vent.
- **Verdens rareste jobb:** Kvalitetssikrerne var noe delt i tilbakemeldingene på denne teksten. Noen mente at utdraget fungerte selv om ikke poenget i boka kommer fram, mens andre mente at utdraget virket ufullendt. Denne ble valgt bort til fordel for andre tekster og er lagt på vent.

3.6 Endringer etter første pilotering

Nedenfor beskriver vi hvilke endringer som ble gjort i tekstene og oppgavene som endte opp i den endelige prøven, før de ble pilotert på nytt. Før andre pilotering ble det også gjort andre, mindre endringer i de utvalgte tekstene. I tråd med innspill fra kvalitetssikrerne og referanselærerne ble en del av oppgaveformuleringene og svaralternativene til tekstene justert og endret. Endringene er dokumentert og finnes tilgjengelige som ulike tekstversjoner av prøvetekstene, lagret i prøveutviklernes arkiver. Endringer i oppgavene kan også ses i vedlegg 2 «Oversikt over endringer i oppgaveformuleringer».

- **Fritidsaktiviteter i vikingtida:** Det er gjort noen mindre endringer i teksten for å tydeliggjøre noen vanskelige ord og setninger. I tillegg er det gjort noen kutt i teksten fordi vi ønsket å ha en kortere tekst. Oppgave L148S02 ble endret for å gjøre spørsmålsformuleringen mer presis. To av svaralternativene ble også endret av samme grunn. I oppgave L148S04 har vi lagt til «ifølge teksten». Oppgave L148S06 er tatt ut fordi den var veldig lett og bidro med lite informasjon.
- **Disney-filmer får kjeft:** Denne teksten ble tatt inn i piloten fordi vi ønsket en nyhetsartikkel med et enklere innhold enn i «Påvirkere». Denne teksten er tidligere pilotert for osloprøven på 4. trinn, og hadde gode parametere fra denne utprøvingen.

Ikke lov å le: Teksten ble vurdert til å være veldig lang, og det er derfor tatt ut en større passasje fra teksten. Teksten fungerer godt uten denne passasjen, og kuttet påvirker ikke noen av oppgavene til teksten. I oppgave L153S03 ble en distraktor byttet ut fordi det var få elever som valgte det. I oppgave L153S04 ble to distraktorer byttet ut fordi de var lite plausible og få elever valgte dem. Oppgave L153S06 var en svært vanskelig oppgave i første pilotering. Her har vi fjernet ett svaralternativ, slik at elevene nå skal sortere tre hendelser etter når de inntraff, istedenfor fire.

- **Stellers sjøku:** Det er ikke gjort noen endringer i teksten etter første pilotering. I oppgave L157S01 er en distraktor omformulert for å gjøre den mer tydelig feil. I oppgave L157S02, L175S03 og L157S05 er det gjort noen mindre endringer i spørsmålsformuleringene for å tydeliggjøre hvilke sjøkuer det spørres om, i tillegg til noen mindre endringer i et par svaralternativer. I matriseoppgaven L157S04 er det fjernet en svarkolonne fordi vi ønsket å gjøre oppgaven litt mindre omfattende, og

denne kolonnen var mindre interessant enn de andre. Oppgave L157S07 har blitt omgjort fra flervalgsoppgave til matriseoppgave hvor elevene skal krysse av for om informasjonen i oppgavene er hentet fra tekst 1, tekst 2 eller begge. Svaralternativene er derfor byttet ut med annen informasjon enn i flervalgsoppgaven. Det er også laget en ny oppgave, L157S08, hvor elevene skal sortere fire hendelser etter når de inntraff.

- **Paranøtteffekten:** Vi har fjernet bildeteksten fordi denne gjentok deler av teksten og ikke inneholdt ny informasjon. Oppgave L158S01 har fått ny og mer presis spørsmålsformulering. I oppgave L158S02 er ordet «partiklene» byttet ut med «nøttene», og en distraktor er byttet ut fordi den innholdsmessig var veldig lik en annen distraktor. En distraktor til oppgave L158S03 ble byttet ut fordi den kunne tenkes å være litt riktig. I oppgave L158S04 har vi lagt til «ifølge teksten» i spørsmålet og byttet ut en distraktor som kunne tenkes å være litt riktig. I oppgave L158S05 er det riktige svaret endret slik at det er mer konkret og presist. En distraktor er byttet ut. Spørsmålsformuleringen i oppgave L158S06 har blitt endret for å tydeliggjøre hva det spørres etter. Denne endringen førte til at det måtte gjøres mindre endringer i alle svaralternativ. I oppgave L158S07 ble det fjernet ett ord i en distraktor.

4. Andre pilotering

De fem utvalgte tekstene ble satt sammen til en ny prøve: prøve D. Denne prøven inneholdt 33 oppgaver, og piloteringen ble gjennomført digitalt. Tekstene og oppgavene ble prøvd ut i den rekkefølgen som var planlagt for den ordinære gjennomføringen. 1 311 elever fra 28 skoler i 9 fylker deltok. Prøvetiden var 90 minutter. Utprøvingen ble gjennomført i august 2022.

Tabell 13: Oversikt over utvalget av elever (N=1 311) pilot prøve D.

	Alle
Antall inviterte elever	3 604
Antall oppmeldte elever	1 558
Antall elever som gjennomførte (N)	1 311 (654 gutter, 657 jenter)
Antall skoler	28

4.1 Innholdet i prøve D

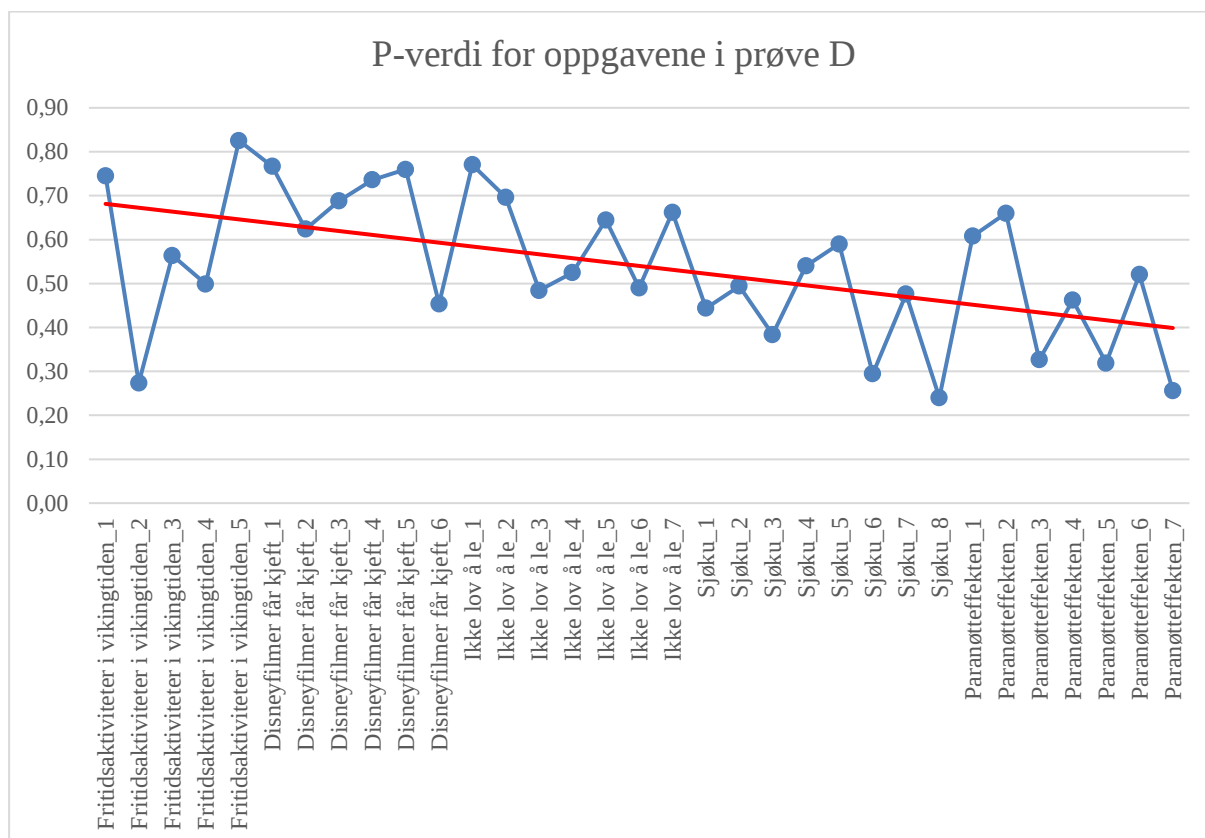
Tekstene fordeler seg på fag og teksttyper som vist i tabell 14. Kolonnen lengst til venstre viser hvilken prøve og plassering den enkelte tekst hadde i første pilot.

Tabell 14: Oversikt over tekstene i prøve D

Plass i første pilot	Tekstnavn i rapport	Original tittel	Fagtilhørighet	Teksttype	Antall ord: 4 326	Antall oppg.: 33		Gj. snittlig p-verdi	Kjønnsforskj. (j-g)*
						Flervalg: 28	Andre: 5		
A1	Fritidsaktiviteter i vikingtiden	«Fritidsaktiviteter i vikingtiden»	Samfunnsfag	Fagartikkel	748	5		0,58	0,00
—	Disneyfilmer får kjeft	«Disney-filmer får kjeft»	Samfunnsfag	Nyhetsartikkel	792	5	1	0,67	0,04
B3	Ikke lov å le	«Ikke lov å le»	Norsk	Skjønnlitterært utdrag	1041	6	1	0,61	0,08
C4	Sjøku	«Et skipsforlis og en sjøku», «Stellers sjøku»	Naturfag	Fagtekst, informasjonstekst	918	5	3	0,43	0,05
C5	Paranøtteffekten	«Hvorfor havner de største nøttene øverst?»	Naturfag	Fagartikkel	827	7		0,45	0,02

*Positive verdier i jentenes favør, negative i guttenes favør.

Figur 6 viser p-verdien for hver av oppgavene i prøve D. Oppgavene er presentert i samme rekkefølge som i prøven.



Figur 6: Andelen elever med riktig svar (p-verdi) på oppgavene i prøve D, organisert etter rekkefølgen i prøven.

Figur 6 viser at det også her er både lette og vanskeligere oppgaver knyttet til alle tekstene i prøven. Trendlinjen viser at p-verdiene blir noe lavere mot slutten av prøven. Som vi så etter 1. pilot er det en tydelig økning i andelen ubesvarte fra og med den fjerde teksten «Sjøku». Andelen ubesvarte er på 10 prosent mot slutten av prøven, noe som er mer enn det som er vanlig etter en ordinær gjennomføring. Av erfaring vet vi at andelen ubesvarte mot slutten av prøven blir mindre under ordinær gjennomføring. For 2022-prøven gikk andelen ubesvart ned fra rundt 12 prosent på de siste oppgavene i andre pilot til rundt 4 prosent på endelig prøve. Pilotprøven er fortsatt litt lang, men det vil bedre seg noe når vi reduserer antall oppgaver i det endelige prøveforslaget.

4.2 Klassisk item-analyse av prøve D

Tabell 15 viser resultater fra klassisk item-analyse av prøve D.

Tabell 15: Resultater fra klassisk item-analyse, prøve D.

Antall elever (N): 1 311 (654 gutter, 657 jenter). Reliabilitet (Cronbachs Alpha): 0,881.

Gjennomsnittlig p-verdi: 0,54 (gutter: 0,52 / jenter: 0,56). Sum svarfordeling kan avvike fra 100 pga. avrunding.

Nr	Item-ID	Prøve D	Svarfordeling i % (frekvens)					Gj.snittlig poengsum på prøven per svaralt.				Diskrim. (biseriell)	P-verdi	Kjønns-forskj. (j-g)	Merk
			A	B	C	D	Blank	A	B	C	D				
			0p	1p	2p			0p	1p	2p					
1	L148S01	Fritidsaktiviteter i vikingtiden_1	9	6	75	10	1	14	13	22	15	0,411	0,75	0,33	
2	L148S02	Fritidsaktiviteter i vikingtiden_2	27	17	38	17	1	23	18	19	19	0,247	0,27	0,03	*
3	L148S03	Fritidsaktiviteter i vikingtiden_3	31	5	56	7	1	18	13	22	14	0,346	0,56	0,22	
4	L148S04	Fritidsaktiviteter i vikingtiden_4	30	13	50	6	1	18	16	23	14	0,324	0,50	0,18	
5	L148S05	Fritidsaktiviteter i vikingtiden_5	5	8	4	83	1	12	12	13	21	0,422	0,83	0,40	
6	L165S01	Disneyfilmer får kjeft_1	15	77	3	4	1	15	22	12	11	0,429	0,77	0,34	
7	L165S02	Disneyfilmer får kjeft_2	6	25	5	62	1	13	18	11	22	0,364	0,62	0,26	
8	L165S03	Disneyfilmer får kjeft_3	14	6	9	69	2	15	12	16	22	0,428	0,69	0,26	
9	L165S04	Disneyfilmer får kjeft_4	74	7	6	12	2	22	13	13	14	0,481	0,74	0,26	
10	L165S05	Disneyfilmer får kjeft_5	7	76	11	5	2	14	22	14	14	0,421	0,76	0,34	
11	L165S06	Disneyfilmer får kjeft_6	45	29	15	9	2	22	20	14	18	0,256	0,45	0,20	Klikk bilde
12	L153S01	Ikke lov å le_1	77	8	7	7	2	22	14	16	14	0,394	0,77	0,38	
13	L153S02	Ikke lov å le_2	14	8	70	6	3	16	15	22	12	0,475	0,70	0,22	
14	L153S03	Ikke lov å le_3	43	1	6	48	2	18	12	15	22	0,268	0,48	0,22	
15	L153S04	Ikke lov å le_4	6	53	16	22	3	12	23	14	20	0,379	0,52	0,15	
16	L153S05	Ikke lov å le_5	7	65	11	15	3	13	22	17	17	0,377	0,64	0,27	
17	L153S06	Ikke lov å le_6	26	43	28		3	17	19	24		0,347	0,49	0,14	Rekkefølge
18	L153S07	Ikke lov å le_7	3	66	17	11	3	11	22	15	16	0,457	0,66	0,20	
19	L157S01	Sjøku_1	29	44	14	9	4	18	24	17	17	0,402	0,44	0,04	
20	L157S02	Sjøku_2	49	20	13	12	5	23	16	18	17	0,434	0,49	0,06	
21	L157S03	Sjøku_3	18	13	25	38	6	17	17	20	23	0,316	0,38	0,07	
22	L157S08	Sjøku_4	21	39	35		6	15	18	27		0,635	0,54	-0,10	Rekkefølge
23	L157S04	Sjøku_5	10	49	34		7	14	18	26		0,633	0,59	-0,04	Fyll inn ord
24	L157S05	Sjøku_6	7	36	29	21	7	17	20	23	19	0,212	0,29	0,08	*
25	L157S06	Sjøku_7	18	17	11	48	7	16	18	16	24	0,518	0,48	-0,04	
26	L157S07	Sjøku_8	50	36	6		8	18	23	29		0,461	0,24	-0,22	Tabell
27	L158S01	Paranøtteffekten_1	9	10	12	61	8	13	15	17	23	0,566	0,61	0,04	
28	L158S02	Paranøtteffekten_2	9	10	7	66	9	15	14	13	23	0,564	0,66	0,10	
29	L158S03	Paranøtteffekten_3	17	33	21	19	10	17	24	19	20	0,321	0,33	0,01	
30	L158S04	Paranøtteffekten_4	16	46	18	10	10	17	24	18	17	0,445	0,46	0,02	
31	L158S05	Paranøtteffekten_5	32	12	16	30	10	22	18	19	21	0,161	0,32	0,16	*
32	L158S06	Paranøtteffekten_6	11	52	16	10	11	16	24	16	17	0,520	0,52	0,00	
33	L158S07	Paranøtteffekten_7	22	21	22	26	10	18	19	21	24	0,269	0,26	-0,01	

*Lav diskriminering (under 0,250).

4.3 IRT-analyse av prøve D

Oppgavene i prøve D er analysert ved hjelp av IRT, og tabellen nedenfor viser parameterne for alle oppgavene i prøven. Vi gjør oppmerksom på at IRT-analysen er utført ved hjelp av programvaren R. Dette er en forskjell fra første pilot, der analysene ble utført ved hjelp av Xcalibre. I tabell 16 er nedre grense for a-verdiene 0,8, og de er derfor ikke direkte sammenlignbare med a-verdiene etter første pilot (kapittel 3.3).

Tabell 16: Parametere fra IRT-analyse av oppgavene i prøve C. Reliabilitet (Cronbachs Alpha): 0,879. N = 1 311 (654 gutter, 657 jenter). Lav diskriminering (point biserial < 0,250 og/eller a < 0,8) er merket (*).

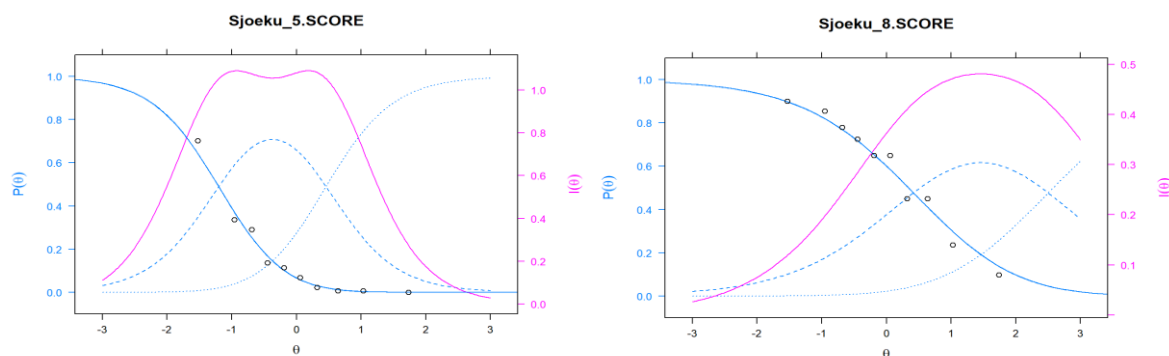
Nr.	Item-ID	Oppgave	Diskriminering (point biserial)	a	b	DIF (disfavør)
1	L148S01	Fritidsaktiviteter i vikingtiden_1	0.448	1.336	-1.068	0
2	L148S02	Fritidsaktiviteter i vikingtiden_2	0.287	0.638*	1.659	0
3	L148S03	Fritidsaktiviteter i vikingtiden_3	0.384	0.869	-0.354	-0,326
4	L148S04	Fritidsaktiviteter i vikingtiden_4	0.358	0.773*	-0.004	-0,397
5	L148S05	Fritidsaktiviteter i vikingtiden_5	0.458	1.725	-1.311	0
6	L165S01	Disney-filmer får kjeft_1	0.471	1.533	-1.089	0
7	L165S02	Disney-filmer får kjeft_2	0.396	0.944	-0.647	0
8	L165S03	Disney-filmer får kjeft_3	0.467	1.305	-0.810	0
9	L165S04	Disney-filmer får kjeft_4	0.519	1.740	-0.892	0
10	L165S05	Disney-filmer får kjeft_5	0.458	1.434	-1.095	0,285
11	L165S06	Disney-filmer får kjeft_6	0.294	0.606*	0.324	-0,616
12	L153S01	Ikke lov å le_1	0.432	1.323	-1.203	0,382
13	L153S02	Ikke lov å le_2	0.518	1.606	-0.761	0,403
14	L153S03	Ikke lov å le_3	0.294	0.605*	0.105	0
15	L153S04	Ikke lov å le_4	0.424	0.979	-0.136	0
16	L153S05	Ikke lov å le_5	0.412	1.015	-0.720	0
17	L153S06	Ikke lov å le_6_1p	0.383	0.558*	-0.969	0,335
		Ikke lov å le_6_2p			1.044	0,406
18	L153S07	Ikke lov å le_7	0.499	1.416	-0.665	0
19	L157S01	Sjøku_1	0.455	1.068	0.243	0
20	L157S02	Sjøku_2	0.484	1.187	0.005	0
21	L157S03	Sjøku_3	0.366	0.800	0.664	0
22	L157S08	Sjøku_4_1p	0.702	1.649	-0.732	0
		Sjøku_4_2p			0.385	0
23	L157S04	Sjøku_5_1p	0.713	1.917	-1.195	0
		Sjøku_5_2p			0.457	0
24	L157S05	Sjøku_6	0.240*	0.513*	1.801	0
25	L157S06	Sjøku_7	0.590	1.690	0.054	0
26	L157S07	Sjøku_8_1p	0.541	1.119	0.419	0
		Sjøku_8_2p			2.503	-0,233
27	L158S01	Paranøtteffekten_1	0.618	2.096	-0.384	0
28	L158S02	Paranøtteffekten_2	0.611	2.216	-0.547	0
29	L158S03	Paranøtteffekten_3	0.385	0.874	0.951	0
30	L158S04	Paranøtteffekten_4	0.500	1.243	0.137	0
31	L158S05	Paranøtteffekten_5	0.181*	0.374*	2.093	0
32	L158S06	Paranøtteffekten_6	0.575	1.634	-0.102	-0,326
33	L158S07	Paranøtteffekten_7	0.328	0.759*	1.566	-0,397

I tabell 16 er oppgavenes diskrimineringssevne oppgitt både som et klassisk mål (point biserial) og en a-verdi. Oppgaver med diskriminering under 0,250 (point biserial) og a-parameter under 0,8 er merket med stjerne, og de oppgavene som er merket for begge parametre, er uthevet med grått og vil bli kommentert i kapittel 4.4 «Endringer fra andre pilotering til endelig prøve» og 5.3 «Oppgaver med lav diskriminering». Vi har valgt å merke oppgaver som har en DIF (Differential Item Functioning) for kjønn på mer enn 0,5 ESSD. Det er én oppgave som er markert, og den blir kommentert nedenfor.

Kolonnen til høyre i tabell 16 viser at det er noen oppgaver med DIF (Differential Item Functioning) for kjønn. Dette betyr at det er noe ved oppgaven som favoriserer det ene kjønn, gitt at begge har samme ferdighet. DIF skiller seg fra måling av kjønnsforskjell, som viser hvilket kjønn som presterer best totalt sett. DIF mellom 0,5 og 0,8 regnes som en moderat DIF, mens oppgaver som ligger over 0,8 har en stor DIF. Oppgaver som slår ut på en DIF-analyse med en effektstørrelse som er lavere enn 0,5, tar vi ikke hensyn til ettersom det er knyttet noe usikkerhet til DIF-analysene, og dette regnes som svært små utslag.

Det er 11 oppgaver i prøveforslaget som slår ut på DIF-analysen, men av disse er det bare én som er over 0,5 ESSD. Oppgave Disney-filmer får kjeft (L165S06) har utslag på DIF-analysen i guttenes favør. DIF-en er på -0,628, noe som regnes som et middels sterkt utslag. Vi har valgt å beholde oppgaven i prøveforslaget til tross for utslaget på DIF, fordi vi mener at denne oppgaven representerer noe unikt i prøven ved at det er en klikk bilde-oppgave. I tillegg ser vi at de DIF-utslagene som ligger mellom 0,2 og 0,5 har en ganske jevn fordeling i favør mellom kjønnene. Teksten som oppgaven er knyttet til, ser heller ikke ut til å favorisere gutter spesielt.

Det er fire oppgaver i andre pilotering som er polytome. Vi har siden andre pilotering lagt oss på en linje hvor rekkefølge-oppgaver alltid skal være dikotome. Begrunnelsen for dette er at det å sortere hendelsene i en tekst er én og samme oppgave, og at det gir liten mening å få poeng for å plassere én eller to av tre eller fire hendelser. Det er heller ikke mulig å få til tre av fire plasseringer fordi man nødvendigvis har feil rekkefølge på to, hvis man plasserer én feil. De to rekkefølge-oppgavene i prøven vil derfor rekodes til dikotome i prøveforslaget. Sjøku_5 (L157S04) og Sjøku_8 (L157S07) beholdes med polytom koding. Figur 7 nedenfor viser informasjonskurve, responsfunksjon og empiriske plot fra IRT-analysen for Sjøku_5 (L157S04) og Sjøku_8 (L157S07).



Figur 7: Parametere fra IRT-analyse av oppgave Sjøku_5 (L157S04) og Sjøku_8 (L157S07). Den rosa linjen er informasjonskurven, mens de blå linjene viser 0, 1 og 2 poeng.

I tillegg til å skille godt mellom elever som får 0, 1 og 2 poeng er det kvalitative grunner for å beholde en polytom koding. I oppgave Sjøku_5 (L157S04) skal elevene holde orden på hva tre ulike dyr i teksten spiser. Informasjonen er kompleks, og det gir mening å gi poeng også til elever som får delvis rett. Vi har likevel valgt å stramme inn grensene for å få 1 poeng i det endelige prøveforslaget slik at to riktige gir 1 poeng, mens én riktig og ingen riktige gir 0 poeng. For oppgave Sjøku_8 (L157S07) der elevene skal kunne sortere i hvilken tekst ulike opplysninger er nevnt, har vi beholdt samme polytome koding som i andre pilot, altså at 4 riktige gir 2 poeng, 3 og 2 riktige gir 1 poeng og resten får 0 poeng.

4.4 Endringer fra andre pilotering til endelig prøve

På bakgrunn av resultater fra den andre piloteringen har vi foretatt noen endringer i det endelige prøveforslaget. Endringene er gjort for å sørge for god spredning i oppgavenes vanskegrad, for at alle spørsmål og svaralternativ i prøven er så godt formulerte som mulig, samt at ingen av disse inneholder ord eller formuleringer som kan virke kompliserende for elevene.

Avsnittene nedenfor beskriver kort endringene for hver enkelt av oppgavene som er endret etter piloteringen. Endringer i oppgavene kan også ses i vedlegg 2 «Oversikt over endringer i oppgaveformuleringer».

- **«Fritidsaktiviteter i vikingtiden»:** Det er gjort noen mindre språklige endringer i teksten, og i oppgave L148S01. I oppgave L148S02 ble spørsmålsendringen endret slik at det ble tydeligere at det var reglene for spillet elevene skulle fokusere på. I oppgave L148S03 ble «at» flyttet fra spørsmålet til hvert av svaralternativene.
- **«Disney-filmer får kjeft»:** Det ble gjort noen mindre språklige endringer i teksten, blant annet ble «urbefolkning» endret til «urfolk». Denne endringen ble også gjort i et svaralternativ i oppgave L165S05. I oppgave L165S03 er to svaralternativer byttet om.
- **«Ikke lov å le»:** Det ble gjort en mindre språklig endring i et svaralternativ til oppgave L153S01. I oppgave L153S03 endret vi rekkefølgen på svaralternativ A og B. Oppgave L153S04 ble tatt ut for å redusere antallet oppgaver til teksten. I oppgave L153S05 ble det en endring i spørsmålet for å gjøre dette mer presist, og dermed også noen endringer i alle svaralternativene. Rekkefølge-oppgave L153S06 gjøres dikotom.
- **«Sjøku»:** Det er gjort noen mindre språklige endringer i teksten, blant annet omtales sjøkua nå som «stellersjøku» istedenfor «Stellers sjøku». Dette har konsekvenser for oppgave L157S02, L157S03, L157S04 og L157S07. Oppgave L157S04 ble endret til et nedtrekkformat, og spørsmålsformuleringen ble tilpasset det nye formatet. Oppgave L157S05 ble tatt ut fordi den diskriminerer dårlig. Rekkefølge-oppgave L157S08 gjøres dikotom. Det er gjort en mindre språklig endring i L157S06, slik at spørsmålet refererer til tekst 2.

- **«Paranøtteffekten»:** Det er gjort noen mindre språklige endringer i teksten. I oppgave L158S02 endret vi spørsmålet slik at det ikke lenger er en setning som elevene skal fullføre, men et spørsmål og fire helsetninger i svaralternativene. Det er gjort en mindre språklig endring i svaralternativ C i L158S03 og svaralternativ A og D til L158S04. Oppgave L158S05 ble tatt ut fordi den diskriminerte for dårlig. Det ble gjort en mindre, språklig endring i et svaralternativ til oppgave L158S06. I oppgave L158S02 er to svaralternativer byttet om.

5. Forslag til endelig prøve 2023

Forslaget til endelig prøve inneholder de samme tekstene som i prøve D. Det er i alt 30 oppgaver til tekstene. Alle oppgavene er kategorisert etter den leseprosessen som er mest framtreddende, og i antall er det 11 finne-oppgaver, 15 tolke- og 4 refleksjonsoppgaver. Tabell 17 gir en oversikt over oppgavene med tilhørende p-verdier. Den gjennomsnittlige p-verdien for prøveforslaget er beregnet til 0,56.

Tabell 17: Oppgavene i forslaget til endelig prøve, riktig svar og leseaspekt.

Nr.	Item-ID	Tekst	P-verdi	Riktig svar	Leseprosess
1	L148S01	Fritidsaktiviteter i vikingtiden_1	0,75	C	Finne
2	L148S02	Fritidsaktiviteter i vikingtiden_2	0,27	A	Finne
3	L148S03	Fritidsaktiviteter i vikingtiden_3	0,56	C	Tolke
4	L148S04	Fritidsaktiviteter i vikingtiden_4	0,50	C	Tolke
5	L148S05	Fritidsaktiviteter i vikingtiden_5	0,83	D	Tolke
6	L165S01	Disney filmer får kjeft_1	0,77	B	Tolke
7	L165S02	Disney filmer får kjeft_2	0,62	D	Reflektere
8	L165S03	Disney filmer får kjeft_3	0,69	C	Tolke
9	L165S04	Disney filmer får kjeft_4	0,74	A	Finne
10	L165S05	Disney filmer får kjeft_5	0,76	B	Finne
11	L165S06	Disney filmer får kjeft_6	0,45	A	Finne
12	L153S01	Ikke lov å le_1	0,77	A	Tolke
13	L153S02	Ikke lov å le_2	0,70	C	Tolke
14	L153S03	Ikke lov å le_3	0,48	D	Finne
15	L153S05	Ikke lov å le_4	0,64	B	Tolke
16	L153S06	Ikke lov å le_5	0,49	B – C – A	Reflektere
17	L153S07	Ikke lov å le_6	0,66	B	Tolke
18	L157S01	Sjøku_1	0,44	B	Reflektere
19	L157S02	Sjøku_2	0,49	A	Tolke
20	L157S03	Sjøku_3	0,38	D	Tolke
21	L157S08	Sjøku_4	0,54	D – B – A – C	Reflektere
22	L157S04	Sjøku_5	0,59	A – C – A	Tolke
23	L157S06	Sjøku_6	0,48	D	Tolke
24	L157S07	Sjøku_7	0,24	B – C – A – C	Finne
25	L158S01	Paranøtteffekten_1	0,61	D	Finne
26	L158S02	Paranøtteffekten_2	0,66	C	Finne
27	L158S03	Paranøtteffekten_3	0,33	B	Tolke
28	L158S04	Paranøtteffekten_4	0,46	B	Finne
29	L158S06	Paranøtteffekten_5	0,52	B	Finne
30	L158S07	Paranøtteffekten_6	0,26	D	Tolke
Gjennomsnittlig p-verdi:			0,56		

5.1 Interessante leseutfordringer i forslaget til endelig prøve

I det følgende viser vi et utdrag av støtteressursene som ligger ute i PAS – Prøver. Disse ressursene gir et lite innblikk i hvilken informasjon som man kan hente ut av prøvene, hvis man arbeider med resultatene og reflekterer over hva de kan fortelle om interessante og relevante leseutfordringer for elevene.

Oppgaver fordelt på lesetekster og mestringsnivå Klikk på oppgavene for å se en besvarelse

Lesetekst	Mestringsnivå 1	Mestringsnivå 2	Mestringsnivå 3
☒ Fritidsaktiviteter i vikingtiden	1 5	3 4	2
☒ Disney-filmer får kjøft	5 8 9 10	7 11	
Ankertekst	A	A A A A	A
☒ Sjøku		18 19 21 22 23	20 24
☒ Paranøtteffekten		25 26 28 29	27 30

☒ Riktig svar ☒ Delvis riktig ☒ Feil svar ☒ Ikke besvart ☒ Ankeroppgave

Elevfanen: Nora sine resultater

Nora har fått til mange oppgaver til tekstene «Disney-filmer får kjøft» og «Paranøtteffekten». Dette tyder på at Nora mestrer disse teksttypene. Til «Fritidsaktiviteter i vikingtiden» derimot har hun bare fått til én oppgave. Hun har også svart feil på flere av oppgavene til «Sjøku».

Lærerne trykker på «Fritidsaktiviteter i vikingtiden» for å se teksten. Når de leser den, forstår de hvorfor teksten har vært vanskelig for mange av elevene i klassen, ikke bare for Nora. De ønsker å se nærmere på oppgavene og går videre til oppgavevisningen.

Elevvisningen gir en oversikt over hvilke oppgaver hver enkelt elev har fått til, og hvilke mestringsnivå de ulike oppgavene er på. Ved å trykke på titlene til venstre, vises tekstene i prøven. Oppgavene vises ved å trykke på oppgavenumrene. Alle elevene har fått en ankeretekst. Ankereteksten og oppgavene til denne er ikke synlige.

Ferdig? Klikk her for å avslutte Forrige Neste

Oppgaveoversikt for klasse 5A

Oppgave	Tekst	Gruppens løsningsprosent	Nasjonal løsningsprosent	Avvik	Mestingsnivå	Formål	Fag	Teksttype
23	Sjøku	24	48	-24	3	Finne	Nat, Saf	Fagtekst, informasjonstekst
2	Fritidsaktiviteter i vikingtiden	8	27	-19	3	Finne	Saf	Fagartikkel
4	Fritidsaktiviteter i vikingtiden	33	50	-17	2	Tolke	Saf	Fagartikkel
11	Disney-filmer får kjøft	30	45	-15	2	Finne	Saf, Krlø	Nyhetsartikkel
3	Fritidsaktiviteter i vikingtiden	42	56	-14	2	Tolke	Saf	Fagartikkel
20	Sjøku	26	38	-12	3	Tolke	Nat, Saf	Fagtekst, informasjonstekst
21	Sjøku	46	54	-8	2	Tolke	Nat, Saf	Fagtekst, informasjonstekst

Oppgavefanen: 5A sin skår på ulike oppgavetyper

Lærerne ser videre på oppgavefanen. De ser hvordan elevene i klassen har prestert sammenliknet med den nasjonale løsningsprosenten. De sorterer etter både positivt og negativt avvik. Oppgaver med stort negativt avvik viser leseutfordringer som mange elever trenger å arbeide videre med. Lærerne bestemmer seg for å undersøke noen av disse oppgavene nærmere. Slik kan de få informasjon om hva som er utfordrende for elevene, og bruke dette til å tilpasse undervisningen i tiden framover.

- Oppgavevisningen gir oversikt over alle oppgaver i prøven. Her kan man se hvor vanskelige oppgavene har vært, og hvor stor andel av elever i klassen som har svart riktig. Avvik mellom elevgruppens resultater og det nasjonale gjennomsnittet vises i en egen kolonne.



Ferdig? Klikk her for å avslutte

Forrige

Neste



Lærerne ser at flere oppgaver handler om hvordan vi i dag kan ha troverdig kunnskap om fortiden. Det kan virke som mange elever ikke har forstått hva som er kildene til kunnskapen. I teksten om fritidsaktiviteter i vikingtiden må elevene forstå hvordan vi kan vite at hnefatafl ble spilt i vikingtiden og hvordan vi har klart å rekonstruere reglene til spillet. I teksten om sjøkua må elevene forstå hvordan vi kan vite at stellersjøkua levde ved Bering-øya.

Lærerne bestemmer seg for å arbeide videre med hvordan man utvikler kunnskap om fortiden basert på vitenskapelige framgangsmåter og hvordan elever kan vurdere ulike kilder til kunnskap.



Ferdig? Klikk her for å avslutte

Forrige

Neste

5.2 IRT-analyse av prøveforslaget

Oppgavene i prøveforslaget er analysert ved hjelp av programvaren R. Forslaget til endelig prøve er analysert ved hjelp av IRT, og tabell 18 viser parameterne for alle oppgavene i prøveforslaget.

Tabell 18: Parametere fra IRT-analyse av oppgavene i prøveforslaget 2023.

N = 1311 (654 gutter, 657 jenter). Gjennomsnittlig p-verdi = 0,57. Alpha for hele prøven = 0,880.

Lav diskriminering (point biserial < 0,250, og/eller a-parameter < 0,8) er merket (*).

Oppg.	Item-ID	Oppgave	Diskriminering (point biserial)	a	b	DIF (disfavør)
1	L148S01	Fritidsaktiviteter i vikingtiden_1	0.453	1.347	-1.064	0
2	L148S02	Fritidsaktiviteter i vikingtiden_2	0.282	0.624*	1.691	0
3	L148S03	Fritidsaktiviteter i vikingtiden_3	0.390	0.882	-0.351	-0,325
4	L148S04	Fritidsaktiviteter i vikingtiden_4	0.360	0.773*	-0.004	-0,396
5	L148S05	Fritidsaktiviteter i vikingtiden_5	0.456	1.688	-1.328	0
6	L165S01	Disney filmer får kjeft_1	0.475	1.543	-1.087	0
7	L165S02	Disney filmer får kjeft_2	0.398	0.943	-0.648	0
8	L165S03	Disney filmer får kjeft_3	0.469	1.301	-0.813	0
9	L165S04	Disney filmer får kjeft_4	0.524	1.754	-0.891	-0,01
10	L165S05	Disney filmer får kjeft_5	0.465	1.459	-1.087	0,28
11	L165S06	Disney filmer får kjeft_6	0.290	0.594*	0.330	-0,628
12	L153S01	Ikke lov å le_1	0.434	1.323	-1.204	0,382
13	L153S02	Ikke lov å le_2	0.516	1.568	-0.771	0,412
14	L153S03	Ikke lov å le_3	0.295	0.606*	0.105	0
15	L153S05	Ikke lov å le_4	0.412	1.009	-0.723	0
16	L153S06	Ikke lov å le_5	0.353	0.809	1.342	0,439
17	L153S07	Ikke lov å le_6	0.502	1.419	-0.665	0,407
18	L157S01	Sjøku_1	0.455	1.060	0.245	0
19	L157S02	Sjøku_2	0.487	1.188	0.005	0
20	L157S03	Sjøku_3	0.369	0.804	0.662	0
21	L157S08	Sjøku_4	0.675	2.324	0.461	0
22	L157S04	Sjøku_5_1 p	0.715	1.904	-1.202	0
		Sjøku_5_2 p			0.461	0
23	L157S06	Sjøku_6	0.596	1.707	0.055	-0,014
24	L157S07	Sjøku_7_1 p	0.540	1.107	0.425	0
		Sjøku_7_2 p			2.516	0
25	L158S01	Paranøtteffekten_1	0.619	2.066	-0.386	0
26	L158S02	Paranøtteffekten_2	0.613	2.189	-0.550	-0,231
27	L158S03	Paranøtteffekten_3	0.386	0.874	0.952	0
28	L158S04	Paranøtteffekten_4	0.500	1.230	0.139	0
29	L158S06	Paranøtteffekten_5	0.573	1.604	-0.101	0
30	L158S07	Paranøtteffekten_6	0.329	0.759	1.568	0

I tabell 18 er oppgavenes diskriminering oppgitt både som et klassisk mål (point biserial) og en a-verdi. Oppgaver med diskriminering under 0,250 (point biserial) og a-parameter under 0,8 er merket med stjerne. De oppgavene som er merket for begge parametere, blir uthevet med grått. Oppgaver med lave diskrimineringsverdier og oppgaver som har en DIF (Differential Item Functioning) for kjønn på mer enn 0,5 ESSD, vil bli kommentert nedenfor.

5.3 Oppgaver med lav diskriminering og DIF

Etter at vi har fjernet noen oppgaver etter andre pilotering og kjørt nye analyser på prøveforslaget, er det ingen oppgaver som har for lav diskriminering på både klassisk mål og på a-verdien. Det er én oppgave som slår ut med DIF over 0,5 ESSD. Denne oppgaven er kommentert i kapittel 4.3, «IRT-analyse av prøve D».

5.4 Oppgavene fordelt på mestringsnivåer

Tabell 19 viser oppgavene i forslag til endelig prøve, fordelt på mestringsnivåer. Nivå-grensene som er brukt, ble bestemt etter gjennomføringen av nasjonale leseprøver i 2022. Theta-verdiene som kommer ut av denne IRT-analysen, er sentrert på theta-fordelingen i piloteringen og er ikke de samme som ved selve gjennomføringen. Det er derfor en viss usikkerhet knyttet til nivågrensene.

Tabell 19: Oppgavene i forslaget til endelig prøve, fordelt på mestringsnivåer

Nivå	Nr.	Item-ID	Oppgave	Skalapoeng
Nivå 1	5	L148S05	Fritidsaktiviteter i vikingtiden_5	36,72
	12	L153S01	Ikke lov å le_1	37,96
	22	L157S04	Sjøku_5_1p	37,98
	6	L165S01	Disney filmer får kjeft_1	39,13
	10	L165S05	Disney filmer får kjeft_5	39,13
	1	L148S01	Fritidsaktiviteter i vikingtiden_1	39,36
	9	L165S04	Disney filmer får kjeft_4	41,09
	8	L165S03	Disney filmer får kjeft_3	41,87
	13	L153S02	Ikke lov å le_2	42,29
Nivå 2	15	L153S05	Ikke lov å le_4	42,77
	17	L153S07	Ikke lov å le_6	43,35
	7	L165S02	Disney filmer får kjeft_2	43,52
	26	L158S02	Paranøtteffekten_2	44,5
	25	L158S01	Paranøtteffekten_1	46,14
	3	L148S03	Fritidsaktiviteter i vikingtiden_3	46,49
	29	L158S06	Paranøtteffekten_5	48,99
	4	L148S04	Fritidsaktiviteter i vikingtiden_4	49,96
	19	L157S02	Sjøku_2	50,05
	23	L157S06	Sjøku_6	50,55
	14	L153S03	Ikke lov å le_3	51,05
	28	L158S04	Paranøtteffekten_4	51,39
	18	L157S01	Sjøku_1	52,45
	11	L165S06	Disney filmer får kjeft_6	53,3
	24	L157S07	Sjøku_7_1p	54,25
	21	L157S08	Sjøku_4	54,61
	22	L157S04	Sjøku_5_2p	54,61
Nivå 3	20	L157S03	Sjøku_3	56,62
	27	L158S03	Paranøtteffekten_3	59,52
	16	L153S06	Ikke lov å le_5	63,42
	30	L158S07	Paranøtteffekten_6	65,68
	2	L148S02	Fritidsaktiviteter i vikingtiden_2	66,91
	24	L157S07	Sjøku_7_2p	75,16

Ideelt sett skal prøven måle likt i alle deler av skalaen, og antallet oppgaver bør med andre ord være det samme innenfor alle mestringsnivåene. Som i tidligere prøver er det flest oppgaver på mestringsnivå 2, men antall oppgaver på det øverste og nederste nivået er forholdsvis likt i dette prøveforslaget. Det er seks oppgaver på det øverste mestringsnivået, noe som bedre enn det vi har klart i tidligere prøveforslag.

I kapittel 3.4 så vi at mange av oppgavene på nivå 3 etter første pilotering hadde lav diskriminering. Flere av disse oppgavene forsvant ut fordi tekstene de var knyttet til, ikke ble med videre i prosessen. Andre oppgaver er reparert slik at de nå diskriminerer bedre uten at vi har mistet oppgavens vanskegrad. Av de oppgavene vi står igjen med er det bare oppgave Fritidsaktiviteter i vikingtiden_2 (L148S02) som har a-parameter under målet på 0,8. I denne oppgaveformuleringen har vi også endret spørsmålsformuleringen for å tydeliggjøre hva elevene skulle fokusere på i teksten. Det er altså grunn til å tro at vi har klart å utvikle oppgaver som måler godt i den øvre delen av ferdighetsskalaen

Diskrimineringstallene for oppgavene på det høyeste nivået i prøveforslaget vises i tabell 20.

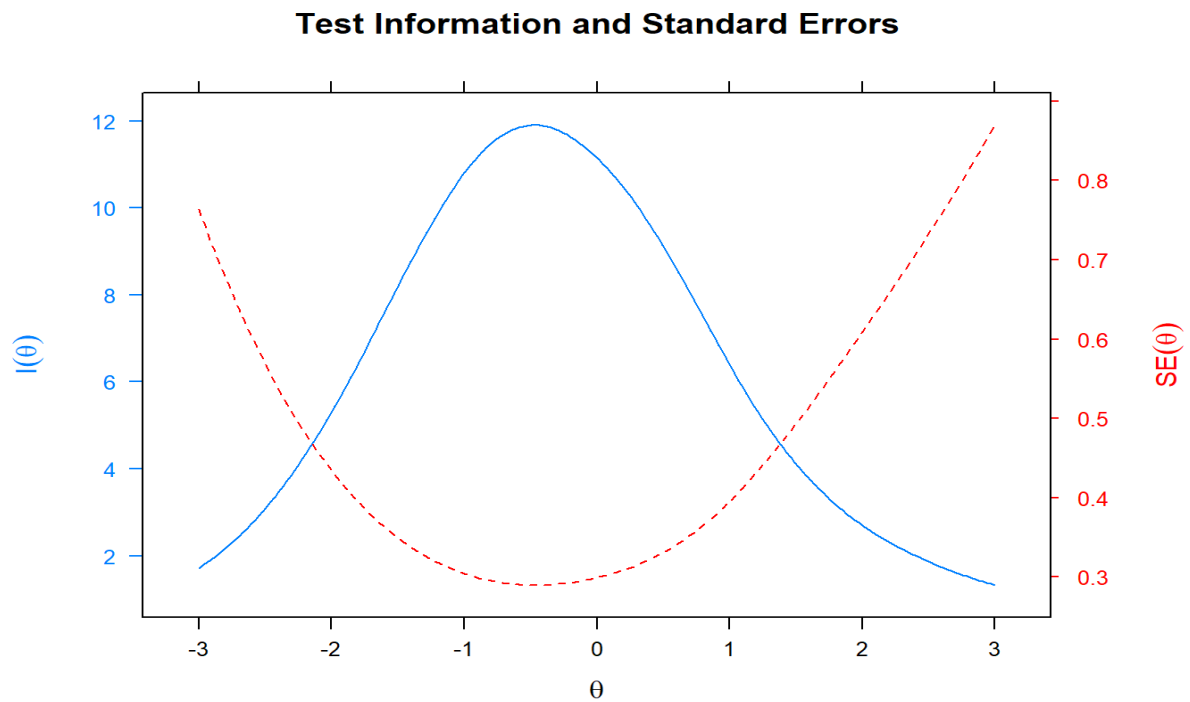
Tabell 20: Oppgavene på nivå 3 med diskrimineringstall

Nivå 3	Item-ID	Oppgave	Diskriminering (point biserial)	a	b	Skalapoeng
	L157S03	Sjøku_3	0.369	0.804	0.662	56,62
	L158S03	Paranøtteffekten_3	0.386	0.874	0.952	59,52
	L153S06	Ikke lov å le_5	0.353	0.809	1.342	63,42
	L158S07	Paranøtteffekten_6	0.329	0.759	1.568	65,68
	L148S02	Fritidsaktiviteter i vikingtiden_2	0.282	0.624*	1.691	66,91
	L157S07	Sjøku_7_2p	0.540	1.107	2.516	75,16

5.5 Testinformasjon og måleusikkerhet

IRT-analysen av prøveforslaget er basert på resultatene fra andre pilotering, men tre oppgaver er tatt ut, og flere av oppgavene som ligger til grunn for analysen, er endret. Med andre ord er dette kun et estimat av prøveforslaget, og vi regner med at IRT-analysen av prøven vil endre seg noe etter gjennomføringen.

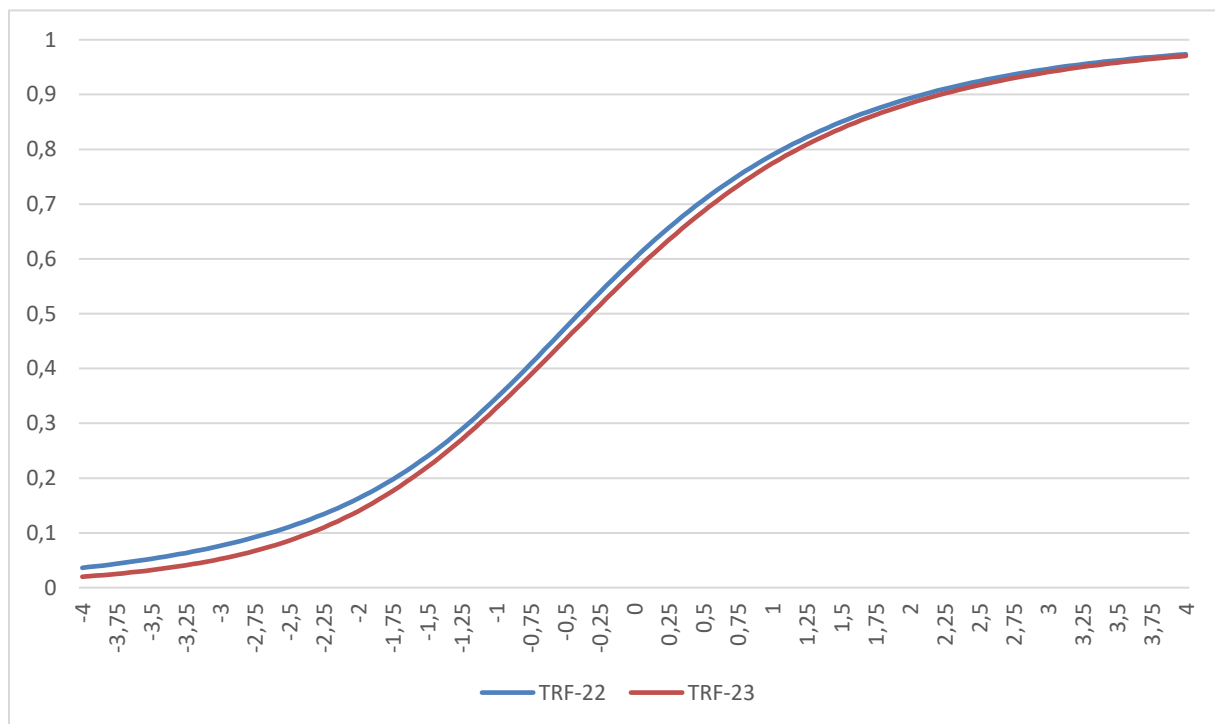
Figur 8 viser at testinformasjonen for prøveforslaget (TIF) er høyest på theta lik -0,5, og at måleusikkerheten (CSEM) er minst i det samme ferdighetsområdet i prøven.



Figur 8: Informasjonsverdi og måleusikkerhet for prøveforslaget 2023

5.6 Prøvens TRF (Test respons function)

Figur 9 viser testrespons-funksjonen (TRF) for prøveforslaget (rød linje), sammenlignet med 2022-prøven (blå linje). Ettersom 2022-prøven og prøveforslaget er analysert med ulike analyseverktøy, er ikke variasjonen i distribusjonen direkte sammenlignbar.



Figur 9: TRF for prøveforslaget 2023 (rød linje), sammenlignet med prøven 2022 (blå linje)

TRF er en beskrivelse av sammenhengen mellom estimert ferdighet (θ) og andel riktig besvarte oppgaver i prøven. Figur 9 viser med andre ord hvor stor andel av oppgavene en elev med en gitt ferdighet (θ) må klare for å få samme resultat på de to prøvene.

Det går fram av figuren at det er liten forskjell mellom prøvene. Prøveforslaget for 2023 forkorter målingen på ferdighetsskalaen i noen grad sammenliknet med 2022-prøven.

Der er viktig å understreke at sammenligningen som er gjort her, er forbundet med noe usikkerhet, fordi 2022-prøven inneholder et sett av ankeroppgaver som ikke er inkludert i prøveforslaget 2023. Når oppgavene blir ankret og prøven gjennomføres og ekvivaleres med 2022-prøven, er det sannsynlig at grafene vil falle mer sammen.

6. Gjennomføring

Den nasjonale prøven i lesing for 5. trinn 2023 ble gjennomført i tidsrommet 4. september – 29. september 2023, i prøvegjennomføringssystemet PAS – Prøver. Totalt 56 492 elever gjennomførte leseprøven. Hver elev fikk en prøve med fem tekster.

På 5. trinn blir endringer i elevenes leseferdigheter målt over tid ved hjelp av ankertekster, det vil si tekster som blir brukt om igjen fra år til år. Det er i alt seks ankertekster, og i 2022 ble disse for første gang distribuert slik at alle elevene fikk en prøve bestående av én ankertekst og fire nye tekster. Med dette designet ble hver ankertekst distribuert til ca. 10 000 elever, mens hver av de nye tekstene ble distribuert til ca. 47 000 elever. Dermed har ingen elever gjennomført årets prøve i sin helhet.

Ankertekstene gjør det mulig å måle endring i elevers leseferdigheter over tid. Overgangen til nytt prøvegjennomføringssystem har ført til at målingen av utvikling over tid, trendmålingen, startet på nytt i 2022. Grensene for mestringsnivåene ble satt på nytt, med et nasjonalt gjennomsnitt på 50 skalapoeng. Resultatene fra gjennomføringen i 2023 er derfor ikke direkte sammenlignbare med resultater fra tidligere gjennomføringer.

Tabell 1 i kapittel 2.2 gir en oversikt over tekstene i den endelige prøven. Tabell 21 under presenterer vi resultatene fra de fem tekstene og 30 oppgavene som utgjorde den nye delen av prøven. Siden ankertekstene skal brukes om igjen, blir ikke resultatene fra disse omtalt.

Tabell 21: Oversikt over fordelingen av tekster og prøveversjoner i årets gjennomføring

	Tekst 1	Tekst 2	Tekst 3	Tekst 4	Tekst 5
	N = 47 058	N = 47 046	N = 37 605	N = 47 104	N = 47 167
Prøve 1	Fritidsaktiviteter i vikingtiden	Disney-filmer får kjeft	Ikke lov å le	Sjøku	Ankertekst 1
Prøve 2	Fritidsaktiviteter i vikingtiden	Disney-filmer får kjeft	Ikke lov å le	Ankertekst 2	Paranøtteffekten
Prøve 3	Fritidsaktiviteter i vikingtiden	Disney-filmer får kjeft	Ankertekst 3	Sjøku	Paranøtteffekten
Prøve 4	Fritidsaktiviteter i vikingtiden	Ankertekst 4	Ikke lov å le	Sjøku	Paranøtteffekten
Prøve 5	Ankertekst 5	Disney-filmer får kjeft	Ikke lov å le	Sjøku	Paranøtteffekten
Prøve 6	Fritidsaktiviteter i vikingtiden	Disney-filmer får kjeft	Ankertekst 6 (pilot)	Sjøku	Paranøtteffekten

I dette kapittelet beskriver vi resultatene etter gjennomføring av årets leseprøve. Resultatene blir presentert i form av p-verdier, det vil si prosentandel elever som fikk poeng på hver oppgave. Det blir rapportert gjennomsnittlige p-verdier for alle elevene og for jenter og gutter, både samlet for hele prøven og for hver enkelt oppgave. Der det er relevant, blir samsvar og ulikheter mellom årets prøve og tidligere prøver kommentert.

6.1 Resultater for prøven som helhet

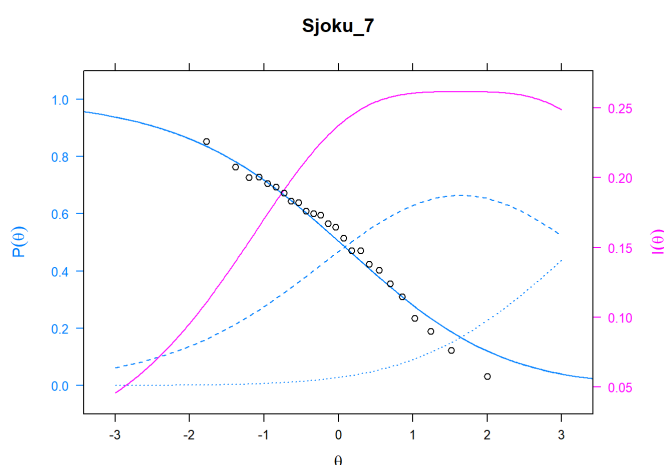
Tabell 22 nedenfor inneholder en oversikt over alle oppgavene i prøven med beskrivelser av oppgaveformat og leseprosess, gjennomsnittlige resultater i form av p-verdier for jenter og gutter, samt differansen mellom disse. Tabellen viser også samlet p-verdi for hver av tekstene, og den gjennomsnittlige p-verdien for hele prøven.

Tabell 22: Leseprosess, oppgaveformat og gjennomsnittlige p-verdier for hver av oppgavene i prøven. Summene er avrundet til to desimaler i Excel.

Oppg.	Tekst	Leseprosess	Oppgave- format	Alle	Jenter	Gutter	Kjønns- forskj. (j-g)
1	Fritidsaktiviteter i vikingtiden_1	Finne	Flervalg	0,73	0,75	0,72	0,03
2	Fritidsaktiviteter i vikingtiden_2	Finne	Flervalg	0,28	0,28	0,27	0,01
3	Fritidsaktiviteter i vikingtiden_3	Tolke	Flervalg	0,59	0,59	0,60	0,00
4	Fritidsaktiviteter i vikingtiden_4	Tolke	Flervalg	0,49	0,48	0,51	-0,03
5	Fritidsaktiviteter i vikingtiden_5	Tolke	Flervalg	0,80	0,82	0,79	0,03
Gjennomsnitt for Fritidsaktiviteter i vikingtiden				0,58	0,58	0,58	0,01
6	Disney filmer får kjeft_1	Tolke	Flervalg	0,77	0,79	0,75	0,04
7	Disney filmer får kjeft_2	Reflektere	Flervalg	0,60	0,62	0,58	0,04
8	Disney filmer får kjeft_3	Tolke	Flervalg	0,64	0,69	0,60	0,09
9	Disney filmer får kjeft_4	Finne	Flervalg	0,70	0,72	0,68	0,04
10	Disney filmer får kjeft_5	Finne	Flervalg	0,77	0,83	0,72	0,10
11	Disney filmer får kjeft_6	Finne	Klikk bilde	0,42	0,40	0,44	-0,04
Gjennomsnitt for Disney-filmer får kjeft				0,65	0,67	0,63	0,04
12	Ikke lov å le_1	Tolke	Flervalg	0,81	0,85	0,77	0,08
13	Ikke lov å le_2	Tolke	Flervalg	0,72	0,78	0,65	0,13
14	Ikke lov å le_3	Finne	Flervalg	0,54	0,56	0,51	0,05
15	Ikke lov å le_4	Tolke	Flervalg	0,70	0,73	0,67	0,07
16	Ikke lov å le_5	Reflektere	Rekkefølge	0,28	0,32	0,24	0,08
17	Ikke lov å le_6	Tolke	Flervalg	0,65	0,71	0,59	0,12
Gjennomsnitt for Ikke lov å le				0,61	0,66	0,57	0,09
18	Sjøku_1	Reflektere	Flervalg	0,44	0,45	0,42	0,03
19	Sjøku_2	Tolke	Flervalg	0,51	0,53	0,49	0,04
20	Sjøku_3	Tolke	Flervalg	0,41	0,39	0,43	-0,04
21	Sjøku_4	Reflektere	Rekkefølge	0,34	0,35	0,33	0,02
22	Sjøku_5	Tolke	Fyll inn ord	0,52	0,52	0,51	0,01
23	Sjøku_6	Tolke	Flervalg	0,52	0,54	0,49	0,05
24	Sjøku_7	Finne	Tabell	0,27	0,27	0,26	0,01
Gjennomsnitt for Sjøku				0,43	0,44	0,42	0,02
25	Paranøtteffekten_1	Finne	Flervalg	0,64	0,66	0,62	0,04
26	Paranøtteffekten_2	Finne	Flervalg	0,68	0,70	0,67	0,03
27	Paranøtteffekten_3	Tolke	Flervalg	0,35	0,36	0,34	0,02
28	Paranøtteffekten_4	Finne	Flervalg	0,47	0,47	0,48	-0,01
29	Paranøtteffekten_5	Finne	Flervalg	0,53	0,53	0,52	0,01
30	Paranøtteffekten_6	Tolke	Flervalg	0,29	0,28	0,29	-0,01
Gjennomsnitt for Paranøtteffekten				0,49	0,50	0,49	0,01
Gjennomsnitt for hele prøven				0,55	0,57	0,53	0,03

Den gjennomsnittlige p-verdien for hele leseprøven er 0,55. Det vil si at en gjennomsnittselev på 5. trinn klarer 55 prosent av oppgavene.

Prøven har en spredning i vanskegrad, både totalt og innen hver tekstenhet. Den vanskeligste oppgavene i 2023-prøven er Sjøku_7 (L157S07), som har en samlet, gjennomsnittlig p-verdi på 0,27. Dette er en polytom oppgave, bare 5 prosent av elevene har fått 2 poeng på denne oppgaven. 44 prosent fikk ett poeng. For å svare riktig på denne oppgaven må elevene finne ut om fire opplysninger om sjøku finnes i tekst 1, tekst 2 eller begge tekstene. For å få 2 poeng må elevene ha alt riktig. For å få 1 poeng må elevene ha klart tre eller to påstander. Elever som har fått til én eller null påstander, har fått 0 poeng. Grafen nedenfor viser at oppgaven er svært vanskelig, og at den diskriminerer godt mellom elever som har fått 0, 1 og 2 poeng.



Figur 10: Utdrag fra IRT-rapport med analysen etter gjennomføringen.

Den letteste oppgaven i prøven er Ikke lov å le_1 (L153S01). Denne oppgaven hadde en gjennomsnittlig p-verdi på 0,81. For å løse denne oppgaven må elevene forstå hva som har skjedd like før handlingen i teksten starter. Det riktige svaret er at Kaja og moren har flyttet til et nytt sted. Denne opplysningen er sentral for å forstå teksten som helhet, og den blir nevnt flere ganger. Det er altså ikke så rart at mange elever svarer riktig på dette spørsmålet.

Hvis vi sammenligner resultatet fra gjennomføringen med resultatet etter andre pilotering, ser vi at den gjennomsnittlige p-verdien på prøven er omtrent helt lik og ligger på 0,55. I 2022 var den gjennomsnittlige p-verdien på den ordinære nasjonale prøven 0,60. Årets prøve er altså noe vanskeligere enn fjorårets. Samtidig klarte vi for første gang å utvikle seks oppgaver på det høyeste mestringsnivået. Det er ikke nødvendigvis en sammenheng mellom en prøves vanskegrad og antall oppgaver på det høyeste nivået, men vi tror at sjansen for å lykkes med å utvikle vanskelige oppgaver som utfordrer elevene med den høyeste lesekompetansen, øker hvis prøven totalt sett blir vanskeligere.

Prøven i år er litt lengre enn den var i fjor, men andelen ubesvarte oppgaver på slutten av prøven er omtrent som i fjor og tidligere år, noe som tilsier at de fleste elevene har kommet igjennom prøven.

6.2 Kjønnforskjeller

Tabell 23 viser de gjennomsnittlige p-verdiene for jenter og gutter, samt kjønnforskjellene. Effektstørrelsen er beregnet som prosentandeler av standardavvikene.

Tabell 23: Kjønnforskjeller i lesekompetanse, uttrykt som p-verdier og effektstørrelse

Kjønnforskjeller i lesekompetanse	
Jenter	0,57
Gutter	0,55
Kjønnforskjell (jenter-gutter)	0,03
Effektstørrelser i form av standardavvik	0,16

Det er generelt små kjønnforskjeller i leseprøvene på 5. trinn, og prøven i lesing for 2023 ligger på samme nivå som de fleste tidligere år. Prøven for 2022 skiller seg noe fra de andre prøvene ved at den har en noe høyere kjønnforskjell, mens 2019 skiller seg ut ved at det ikke var noen kjønnforskjell. I 2023-prøven er det noen tilfeller hvor det ene kjønnnet presterer bedre enn det andre på oppgavenivå, men den gjennomsnittlige kjønnforskjellen for hele prøven er 0,03 i jentenes favør. Tabell 24 under viser en oversikt over gjennomsnittlig kjønnforskjell for prøvene de siste seks årene.

Tabell 24: Oversikt over kjønnforskjell i lesekompetanse de seneste årene

År	Kjønnforskjell (jenter-gutter)
2017	0,04
2018	0,04
2019	0,00
2020	0,03
2021	0,03
2022	0,06
2023	0,03

6.3 Klassisk item-analyse av prøven 2023

Tabell 25 viser resultater fra klassisk item-analyse av den nasjonale prøven i lesing 2023. På grunn av det nye prøveoppsettet finnes det ikke lenger en N for årets prøve som helhet, men vi vet at 56 492 elever gjennomførte leseprøven. Det er heller ikke mulig å gjøre reliabilitetsanalyse i SPSS. Se derfor diskrimineringsverdiene fra IRT-analysene i kapittel 5.3 «Oppgaver med lav diskriminering».

Tabell 25: Resultater fra klassisk item-analyse, endelig prøve, 5. trinn.

Gjennomsnittlig p-verdi: 0,55 (gutter: 0,53/jenter: 0,57). Sum svarfordeling kan avvike fra 100 pga. avrunding.

Nr	Item-ID	Oppgaver	Svarfordeling i % (frekvens)					Gj.snittlig poengsum på prøven per svaralt.				P-verdi	Kjønns-forskj. (j-g)	Merk
			A	B	C	D	Blank	A	B	C	D			
			0p	1p	2p									
1	L148S01	Fritidsaktiviteter i vikingtiden_1	10	6	73	11	0	12	12	19	13	0,73	0,03	
2	L148S02	Fritidsaktiviteter i vikingtiden_2	28	17	39	15	0	20	17	16	16	0,28	0,01	
3	L148S03	Fritidsaktiviteter i vikingtiden_3	27	6	59	7	1	15	12	20	12	0,59	0,00	
4	L148S04	Fritidsaktiviteter i vikingtiden_4	30	13	49	7	1	16	13	20	12	0,49	-0,03	
5	L148S05	Fritidsaktiviteter i vikingtiden_5	6	8	6	80	1	11	11	11	19	0,80	0,03	
6	L165S01	Disney filmer får kjeft_1	16	77	3	4	0	13	19	9	10	0,77	0,04	
7	L165S02	Disney filmer får kjeft_2	7	28	5	60	1	11	16	10	20	0,60	0,04	
8	L165S03	Disney filmer får kjeft_3	18	6	64	11	1	14	11	20	15	0,64	0,09	
9	L165S04	Disney filmer får kjeft_4	70	9	8	13	1	20	11	13	13	0,70	0,04	
10	L165S05	Disney filmer får kjeft_5	7	77	10	5	1	13	19	12	12	0,77	0,10	
11	L165S06	Disney filmer får kjeft_6	42	32	16	10	1	20	19	12	16	0,42	-0,04	Klikk bilde
12	L153S01	Ikke lov å le_1	80	6	6	7	1	19	11	13	12	0,81	0,08	
13	L153S02	Ikke lov å le_2	13	8	71	5	2	13	12	19	11	0,72	0,13	
14	L153S03	Ikke lov å le_3	1	39	5	53	1	9	16	12	19	0,54	0,05	
15	L153S05	Ikke lov å le_4	70	6	7	16	2	19	11	13	14	0,70	0,07	
16	L153S06	Ikke lov å le_5	71	28			1	16	22			0,28	0,08	Rekkefølge
17	L153S07	Ikke lov å le_6	3	64	20	11	2	10	20	13	14	0,65	0,12	
18	L157S01	Sjøku_1	29	44	13	11	2	16	21	15	14	0,44	0,03	
19	L157S02	Sjøku_2	51	21	16	10	3	21	14	16	14	0,51	0,04	
20	L157S03	Sjøku_3	18	15	24	41	3	15	14	17	20	0,41	-0,04	
21	L157S08	Sjøku_4	64	34			3	15	23			0,34	0,02	Rekkefølge
22	L157S04	Sjøku_5	33	30	37			12	16	23		0,52	0,01	Fyll inn ord
23	L157S06	Sjøku_6	17	17	11	52	4	13	14	13	21	0,52	0,05	
24	L157S07	Sjøku_7	48	44	5		4	15	20	26		0,53	0,02	Tabell
25	L158S01	Paranøtteffekten_1	10	10	12	64	4	11	13	14	20	0,64	0,04	
26	L158S02	Paranøtteffekten_2	9	11	68	7	4	13	12	20	11	0,68	0,03	
27	L158S03	Paranøtteffekten_3	18	35	22	20	5	15	21	16	17	0,35	0,02	
28	L158S04	Paranøtteffekten_4	16	47	18	14	5	15	21	15	15	0,47	-0,01	
29	L158S06	Paranøtteffekten_5	14	52	19	10	5	13	21	15	14	0,53	0,01	
30	L158S07	Paranøtteffekten_6	24	22	21	29	5	16	16	18	21	0,29	-0,01	

6.4 Oppgaver med lav diskriminering

Tabell 26 under viser diskrimineringsverdiene for alle oppgavene i prøven. Oppgavenes diskrimineringsevne er oppgitt både som et klassisk mål (point biserial) og en a-verdi. Oppgaver med diskriminering under 0,250 (point biserial) og a-parameter under 0,8 er merket med stjerne, og oppgaver som er merket for begge parametre, blir uthevet med grått.

Tabell 26: Oversikt over diskrimineringsverdier i prøven.

Oppg.	Item-ID	Oppgave	Diskriminering (point biserial)	a
1	L148S01	Fritidsaktiviteter i vikingtiden_1	0,642	1,420
2	L148S02	Fritidsaktiviteter i vikingtiden_2	0,350	0,520*
3	L148S03	Fritidsaktiviteter i vikingtiden_3	0,549	0,980
4	L148S04	Fritidsaktiviteter i vikingtiden_4	0,481	0,770*
5	L148S05	Fritidsaktiviteter i vikingtiden_5	0,655	1,590
6	L165S01	Disney filmer får kjeft_1	0,642	1,440
7	L165S02	Disney filmer får kjeft_2	0,515	0,880
8	L165S03	Disney filmer får kjeft_3	0,546	0,990
9	L165S04	Disney filmer får kjeft_4	0,673	1,510
10	L165S05	Disney filmer får kjeft_5	0,642	1,470
11	L165S06	Disney filmer får kjeft_6	0,316	0,420*
12	L153S01	Ikke lov å le_1	0,586	1,280
13	L153S02	Ikke lov å le_2	0,667	1,550
14	L153S03	Ikke lov å le_3	0,433	0,660*
15	L153S05	Ikke lov å le_4	0,556	1,080
16	L153S06	Ikke lov å le_5	0,565	1,000
17	L153S07	Ikke lov å le_6	0,611	1,240
18	L157S01	Sjøku_1	0,515	0,880
19	L157S02	Sjøku_2	0,597	1,140
20	L157S03	Sjøku_3	0,460	0,730*
21	L157S08	Sjøku_4	0,791	2,010
22	L157S04	Sjøku_5	0,762	1,380
23	L157S06	Sjøku_6	0,756	1,890
24	L157S07	Sjøku_7	0,542	0,880
25	L158S01	Paranøtteffekten_1	0,731	1,810
26	L158S02	Paranøtteffekten_2	0,746	1,970
27	L158S03	Paranøtteffekten_3	0,515	0,900
28	L158S04	Paranøtteffekten_4	0,562	1,040
29	L158S06	Paranøtteffekten_5	0,646	1,320
30	L158S07	Paranøtteffekten_6	0,388	0,600

Det er ingen grå felter i tabellen over, fordi ingen oppgaver i prøven har både diskrimineringsverdi under 0,250 (point biserial) og a-parameter under 0,8. Alle oppgavene har point-biserial-verdier over 0,250, men fem oppgaver i prøven har a-parameter under 0,8. Av disse hadde fire oppgaver a-parameter under grenseverdien også i prøveforslaget. For å se endringene som er gjort i disse oppgavene gjennom utviklingsprosessen, se vedlegg 2 «Oversikt over endringer i oppgaveformuleringer».

Fritidsaktiviteter i vikingtiden_2 (L148S02) ble endret både før og etter andre pilotering. Selv om oppgaven har hatt noe lave a-verdier gjennom hele prosessen, berører oppgaven et viktig poeng i teksten, og den er på mestringsnivå 3. Vi har derfor sørget for å bevare oppgaven og vanskegraden. Fritidsaktiviteter i vikingtiden_4 (L148S04) hadde a-verdi rett under grensa i andre pilotering, og det ble ikke gjort noen endringer etter andre pilotering.

Disney filmer får kjeft_6 (L165S06) er heller ikke blitt endret etter andre pilotering. Denne oppgaven representerer et nytt oppgaveformat (klikk-bilde), og det var en motivasjon for å bevare oppgaven. En tredjedel av elevene har valgt filmen «Peter Pan», som har fått kritikk for å framstille amerikansk urbefolkning på en uheldig måte. Det virker som mange elever blander sammen begrepene «amerikansk urbefolkning» med «afroamerikanere» – som har blitt uheldig framstilt i filmen «Jungelboken», og som er det riktige svaret.

I Ikke lov å le_3 (L153S03) spør vi hva moren serverte da hun fortalte Kaja at de skulle flytte. Det nevnes flere matretter i teksten, men det nevnes bare to måltider. Derfor er det andre måltidet, alternativ B «pannekaker», et attraktivt feilsvar, og 39% av elevene valgte dette. I andre pilotering var «pannekaker» svaralternativ A, og 43% av elevene valgte svaret. Vi flyttet dette svaret for å gjøre det mindre attraktivt, men det er fortsatt mange som svarer dette.

Sjøku_3 (L157S03) har hatt akseptable diskrimineringsverdier i alle utprøvingene og i prøveforslaget lå a-verdien rett over grensa. Det ble gjort en ubetydelig endring i oppgaven («Stellers sjøku» til «stellersjøkua») etter andre pilotering, men dette burde ikke ha noen innvirkning på oppgavens diskriminering. Oppgaven er nokså vanskelig og ligger helt øverst på mestringsnivå 2. Vi vet fra tidligere prøver at oppgaver med høy vanskegrad har en tendens til å diskriminere dårligere enn andre oppgaver, og grunnen til dette kan være at IRT-modellen ikke klarer å skille godt mellom elever med høyest dyktighet.

6.5 Oppgavene fordelt på mestringsnivåer

Tabell 27 viser oppgavene i den nasjonale prøven i lesing fordelt på mestringsnivå. Det er satt nye grenser for mestringsnivåer i 2022, slik at resultatene fra gjennomføringen i 2023 ikke er direkte sammenlignbare med resultater fra tidligere gjennomføringer.

Antall oppgaver på hvert mestringsnivå er det samme som i estimatet. På nivå 3 er det seks oppgaver. Sjøku_3 (L157S03) lå på nivå 3 i forslag til endelig prøve, men ligger nå på nivå 2, mens det motsatte gjelder for Disney-filmer får kjeft_6 (L165S06).

På nivå 1 er det 9 oppgaver, like mange som i estimatet for endelig prøve. Selv om noen av oppgavene har havnet på et annet nivå, eller har endret plassering innad på nivåene, stemmer

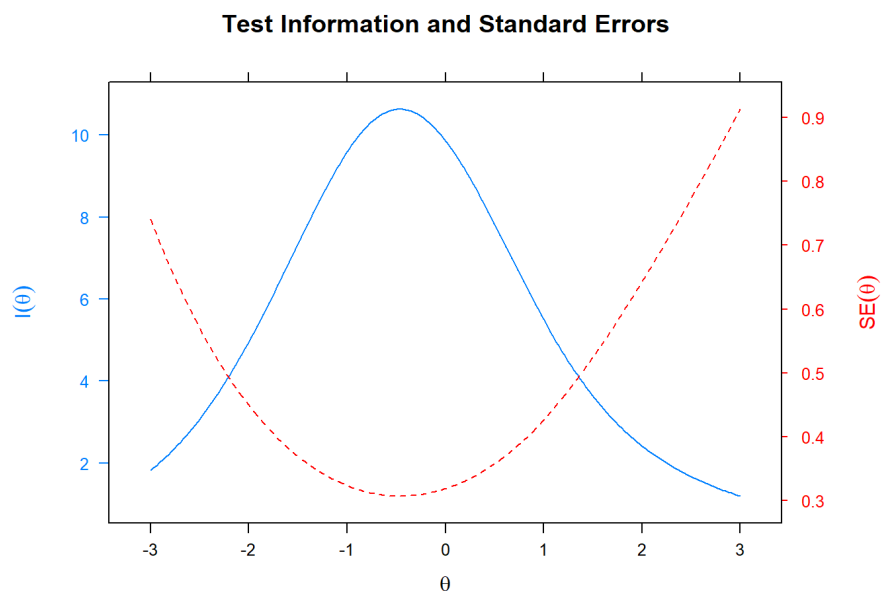
estimatet i hovedsak ganske godt overens med hvilke oppgaver som har vært lette, og hvilke som har vært vanskelige i den endelige gjennomføringen.

Tabell 27: Oppgavene i den nasjonale prøven i lesing 2023 fordelt på mestringsnivå

Nivå	Item-ID	Oppgave	Skalapoeng
Nivå 1	L153S01	Ikke lov å le_1	35,2
	L148S05	Fritidsaktiviteter i vikingtiden_5	37
	L165S05	Disney filmer får kjeft_5	37,8
	L165S01	Disney filmer får kjeft_1	37,9
	L148S01	Fritidsaktiviteter i vikingtiden_1	39,7
	L153S05	Ikke lov å le_4	39,8
	L153S02	Ikke lov å le_2	40,9
	L165S04	Disney filmer får kjeft_4	41,4
	L165S03	Disney filmer får kjeft_3	42,1
Nivå 2	L158S02	Paranøtteffekten_2	42,9
	L153S07	Ikke lov å le_6	43
	L165S02	Disney filmer får kjeft_2	43,8
	L158S01	Paranøtteffekten_1	44,1
	L148S03	Fritidsaktiviteter i vikingtiden_3	44,7
	L157S04	Sjøku_5_1p	45,8
	L153S03	Ikke lov å le_3	47,1
	L158S06	Paranøtteffekten_5	48,2
	L157S06	Sjøku_6	48,4
	L157S02	Sjøku_2	48,8
	L148S04	Fritidsaktiviteter i vikingtiden_4	49,8
	L158S04	Paranøtteffekten_4	50,7
	L157S04	Sjøku_5_2p	50,9
	L157S07	Sjøku_7_1p	50,9
	L157S01	Sjøku_1	52,5
	L157S03	Sjøku_3	54,6
	L157S08	Sjøku_4	54,7
Nivå 3	L165S06	Disney filmer får kjeft_6	57,5
	L158S03	Paranøtteffekten_3	57,6
	L153S06	Ikke lov å le_5	60,7
	L158S07	Paranøtteffekten_6	65,6
	L148S02	Fritidsaktiviteter i vikingtiden_2	69
	L157S07	Sjøku_7_2p	82

6.6 Testinformasjon og måleusikkerhet

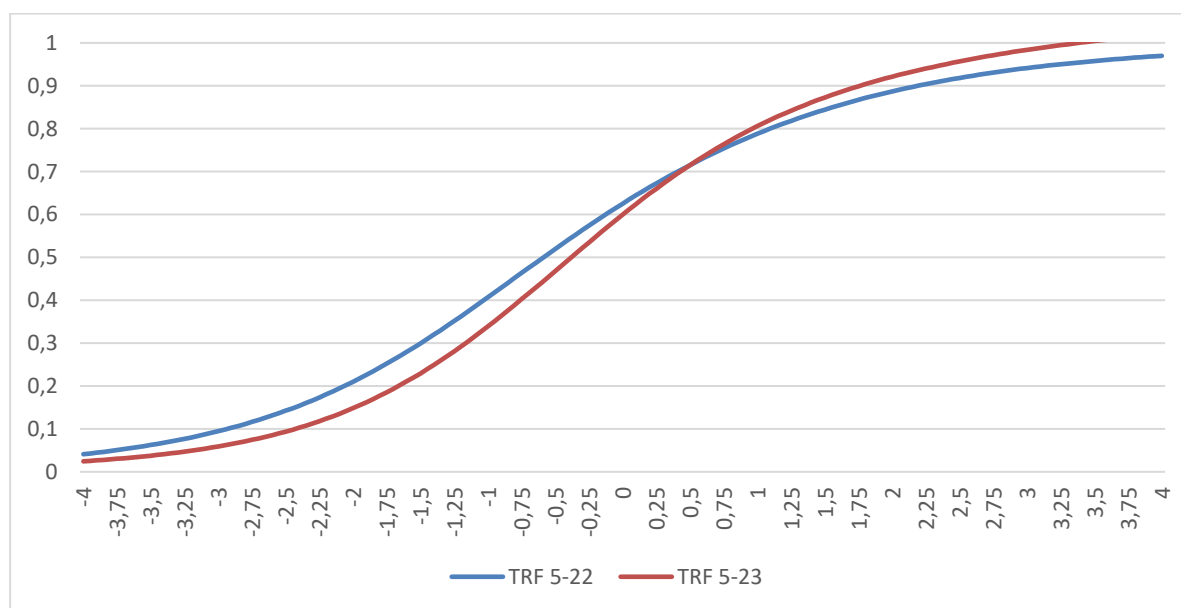
En IRT-analyse av den endelige prøven viser at testinformasjonen (TIF) som prøven leverer i sin helhet, er høyest på theta lik -0,5 og at måleusikkerheten (CSEM) er minst i det samme ferdighetsområdet i prøven. Figur 11 viser testinformasjon og måleusikkerhet for prøven.



Figur 11: Informasjonsverdi og måleusikkerhet for 2023-prøven

6.7 Prøvens TRF (Test respons function)

Figur 12 viser testrespons-funksjonen (TRF) for 2023-prøven (rød linje), sammenlignet med 2022-prøven (blå linje). Ettersom 2022-prøven og 2023-prøven er analysert med ulike analyseverktøy, er ikke variasjonen i distribusjonen direkte sammenlignbar.



Figur 12: TRF for endelig prøve 2023, sammenlignet med endelig prøve 2022

TRF er en beskrivelse av sammenhengen mellom estimert ferdighet (θ) og andel riktig besvarte oppgaver i prøven. Figur 12 viser med andre ord hvor stor andel av oppgavene en elev med en gitt ferdighet (θ) må klare for å få samme resultat på de to prøvene.

Det går fram av figuren at det er noe forskjell mellom prøvene. For elever med ferdighet mellom -4 og 0,5 ligger kurven for årets prøve under kurven for 2022-prøven. Det betyr at elevene i dette ferdighetsområdet må få til en mindre andel oppgaver i årets prøve enn i prøven fra 2022 for å oppnå samme resultat. For elever med ferdighet mellom 0,5 og 4 er det motsatt.

7. Referanser

Chalmers RP (2012). “mirt: A Multidimensional Item Response Theory Package for the R Environment.” *Journal of Statistical Software*, 48(6), 1–29. doi:10.18637/jss.v048.i06.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Silverman, D. (2011). *Interpreting qualitative data. Methods for analysing talk, text and interaction* (3. utg.). Thousand Oaks: SAGE Publications.

Utdanningsdirektoratet (2020). *Læreplanverket*. <https://www.udir.no/laring-og-trivsel/lareplanverket/>

Utdanningsdirektoratet (2017). Rammeverk for grunnleggende ferdigheter. Lastet ned fra: <https://www.udir.no/laring-og-trivsel/rammeverk/rammeverk-for-grunnleggende-ferdigheter/> [09.03.2023]

Utdanningsdirektoratet (2022). Rammeverk for nasjonale prøver. Lastet ned fra: <https://www.udir.no/eksamen-og-prover/prover/rammeverk-for-nasjonale-prover2/> [03.02.2023]

8. Vedlegg 1: Beskrivelser av lesekompetanse på de tre ulike nivåene

	Finne	Tolke	Reflektere og vurdere
Nivå 1	lokalisere tydelig uttrykte elementer i en tekst med lite konkurrerende informasjon	trekke enkle slutninger eller oppfatte hovedtemaet i en tekst når innholdet er tydelig uttrykt i teksten	bruke personlige meninger til å kommentere form eller innhold i en tekst
Nivå 2	lokalisere tydelig uttrykte elementer i en tekst med klart konkurrerende informasjon	oppfatte hovedtemaet og forstå sammenhenger som ikke er tydelig uttrykt i teksten	identifisere formelle trekk ved tekster og ta stilling til eller vurdere tekstens innhold
Nivå 3	skille sterkt konkurrerende informasjon fra informasjon som er relevant for oppgaven	forstå motsetningsfylt innhold og komplekse sammenhenger i teksten	bruke kunnskap om språk og tekst til å identifisere og forholde seg til mer komplekse trekk ved tekstens form og innhold

9. Vedlegg 2: Oversikt over endringer i oppgaveformuleringer

Item-ID	Første pilot	Andre pilot	Endelig prøve
Fritidsaktiviteter i vikingtiden			
L148S01	Hvordan kan man vite at hnefatafl ble spilt i vikingtiden?		Hvordan kan vi vite at hnefatafl ble spilt i vikingtiden?
L148S02	I teksten står det at man ikke vet nøyaktig hvordan hnefatafl ble spilt i vikingtiden. Hvordan kan det likevel være mulig å forklare spillereglene til dette spillet?	Man vet ikke nøyaktig hvordan reglene for å spille hnefatafl var i vikingtiden. Hvordan er det likevel mulig å spille dette brettspillet i dag? A) Man kjenner reglene til et gammelt spill som ligner på hnefatafl. D) Man har tenkt seg fram til reglene ut fra utformingen av brettet og brikkene.	Vi vet ikke nøyaktig hvordan spillereglene for hnefatafl var i vikingtiden. Hvilke regler brukes når man spiller hnefatafl i dag? A) reglene fra et gammelt spill som ligner på hnefatafl B) regler som er beskrevet i sagaene fra vikingtiden C) de samme reglene som i sjakk D) regler man har funnet på ut fra hvordan brettet og brikkene ser ut
L148S03	Når kongen til en spiller kommer ut til kanten av brettet i hnefatafl, betyr det at ...		<i>«at» er flyttet fra spørsmålet og ned til svaralternativene</i>
L148S04	Hva var grunnen til at kongen Knut den mektige fikk motstanderen sin drept etter en omgang brettspill?	Hva var grunnen til at kongen Knut den mektige fikk motstanderen sin drept etter en omgang brettspill, ifølge teksten?	
L148S05	Hvorfor var det viktig at barn lærte å spille hnefatafl i vikingtiden?		
L148S06	Hvem spilte mest brettspill i vikingtiden?	<i>Tatt ut.</i>	
Disney-filmer får kjeft			
L165S01		Hvorfor får noen gamle Disney-filmer kjeft?	
L165S02		Hvorfor er Gry Rustad intervjuet i denne teksten?	
L165S03		Hva mener Gry Rustad om de gamle Disney-filmene?	
L165S04		Hvordan har Disney reagert på kritikken?	
L165S05		Hvorfor har filmen «Snehvit» fått kritikk?	A) fordi amerikanske urfolk blir framstilt som dumme
L165S06		Hvilken av de omtalte filmene har fått kritikk for å framstille afroamerikanere på en negativ måte?	
Ikke lov å le			

L153S01	Hva har skjedd like før handlingen i denne teksten starter?		A) Kaja har oppført seg dårlig på den nye skolen.
L153S02	Hvorfor synes Kaja pannekakene som moren serverer, er mistenkelige?		
L153S03	Hva serverte moren da hun fortalte Kaja at de skulle flytte?	C) paco	Endret rekkefølgen på A og B
L153S04	Hva synes Kaja om å skulle feire bursdagen sin?	A) Hun gruer seg til å få gaver hun ikke vil ha. C) Hun gruer seg til å spise p-mat.	Tatt ut
L153S05	Hva synes Kaja om å flytte til et nytt sted?		Endret fra presens til preteritum i spørsmål og svaralternativer
L153S06	Plasser hendelsene nedenfor i den rekkefølgen de faktisk skjedde, fra først (øverst) til sist (nederst).	Alternativ A er fjernet	
L153S07	Hvilken setning beskriver best forholdet mellom Kaja og moren?		
Stellers sjøku			
L157S01	Hva er hovedpoenget med tekst 1?	A) å informere om forskjellige dyr som lever på Bering-øya	
L157S02	Hvordan kan vi vite at det levde sjøkuer ved Bering-øya, ifølge tekst 1?	Hvordan kan vi vite at Stellers sjøku levde ved Bering-øya, ifølge tekst 1? <i>Liten endring i B</i>	Hvordan kan vi vite at stellersjøkua levde ved Bering-øya, ifølge tekst 1? <i>Liten endring i C</i>
L157S03	Hva var den viktigste grunnen til at sjøkuene ble utryddet, ifølge tekst 1?	Hva var den viktigste grunnen til at Stellers sjøku ble utryddet, ifølge tekst 1?	Hva var den viktigste grunnen til at stellersjøkua ble utryddet, ifølge tekst 1?
L157S04	Hvem spiser hvem, ifølge tekst 2? Sett kryss i tabellen for å vise hvem som spiser hvem.	Dra boksene til riktig plass slik at setningene viser hva sjøkuer, havotere og kråkeboller spiser.	Oppgaven endret til nedtrekk og spørsmålsformuleringen ble tilpasset det nye formatet.
L157S05	Da George Steller oppdaget sjøkuene i 1741, var det ...	Da George Steller oppdaget de spesielle sjøkuene i 1741, var det ... C) ganske få sjøkuer igjen i havet.	Tatt ut
L157S06	Jakt på havotere førte til at det ble flere kråkeboller i havet. Dette førte til ...		
L157S07	Hvilken påstand om sjøkuer stemmer med informasjonen i begge tekstene?	I tabellen nedenfor står det fire opplysninger om Stellers sjøku. Sett kryss for om opplysningene finnes i tekst 1, tekst 2 eller begge. A) Jakt på havoter bidro til at Stellers sjøku ble utryddet. B) Stellers sjøku kunne veie opptil 10 tonn. C) Stellers sjøku hadde ingen tenner. D) Stellers sjøku ble utryddet 27 år etter at den ble oppdaget.	I tabellen nedenfor står det fire opplysninger om stellersjøkua. Kryss av for om opplysningene finnes i tekst 1, tekst 2 eller begge. <i>«Stellers sjøku» er endret til «stellersjøku» i alle svaralternativer</i>

L157S08		Plasser hendelsene fra tekst 1 i den rekkefølgen de faktisk skjedde, fra først (øverst) til sist (nederst).	<i>Liten endring i C</i>
Paranøtteffekten			
L158S01	Hva er grunnen til at fenomenet som beskrives i denne teksten, er oppkalt etter paranøtten?	Paranøtteffekten har fått navnet sitt fra paranøtten. Hvorfor det?	
L158S02	Når man rister en blanding med nøtter av ulik størrelse, vil ...	B) nøttene blande seg helt tilfeldig. C) de største nøttene knuse de minste nøttene.	Hva skjer når man rister en blanding med nøtter av ulik størrelse? A) De største nøttene vil havne nederst. B) Nøttene vil blande seg helt tilfeldig. C) De største nøttene vil knuse de minste nøttene. D) De største nøttene vil havne øverst.
L158S03	På hvilken måte kan paranøtteffekten gjøre at medisiner ikke virker som de skal?	C) Noen av ingrediensene kan bli ødelagt når kapslene ristes kraftig.	<i>En mindre, språklig endring i svaralternativ C.</i>
L158S04	Skiløpere bruker utstyr som blåser seg opp som en ballong hvis de blir tatt av et snøskred. På hvilken måte virker ballongen?	Skiløpere bruker utstyr som blåser seg opp som en ballong hvis de blir tatt av et snøskred. På hvilken måte virker ballongen, ifølge teksten? D) Ballongen blåser seg opp slik at skiløperen flyr i sikkerhet.	A) Skiløpere bruker ballongen til å vise hvor de er i snøskredet. D) Ballongen blåser seg opp slik at skiløperen flyr ut av skredet
L158S05	Hva er det forskere fortsatt ikke forstår med paranøtteffekten?	A) hvorfor de store nøttene beveger seg oppover D) hvorfor det bare er paranøttene som havner øverst	<i>Tatt ut</i>
L158S06	Hvordan kan forskere vite at store nøtter beveger seg oppover i en nøttepose når man rister på den?	Forskere har funnet ut hvordan store nøtter beveger seg oppover i en nøttepose når man rister på den. Hvordan fant de ut dette? A) De snakket med mennesker som arbeider på nøttefabrikker. B) De filmet hvordan nøttene beveget seg mens de ristet på posen. C) De merket nøttene for å se hvor de havnet etter at posen ble ristet. D) De gjorde matematiske beregninger av hvordan nøttene kom til å bevege seg.	D) De beregnet matematisk hvordan nøttene kom til å bevege seg.
L158S07	Hvis du finner noen store nøtter nederst i en nøttepose, betyr det sannsynligvis at disse nøttene har ...	A) beveget seg fra toppen av posen og ned.	

10. Vedlegg 3: Spørreskjema til første pilot

Takk for at du gjennomførte prøven!

Nå vil vi som lager prøven, vite hva du syns om tekstene.

De tre oppgavene under får du ikke poeng for, og du kan krysse av for så få eller mange tekster du vil.

Hvilken tekst likte du best? Du kan krysse av for flere tekster.

- ☐ Er humlemorderen funnet?
- ☐ Verdens lengste dyr
- ☐ Eksperimentet
- ☐ Da skateboard var forbudt
- ☐ Farlig katt

Hvilken tekst syns du var vanskelig? Du kan krysse av for flere tekster.

- ☐ Er humlemorderen funnet?
- ☐ Verdens lengste dyr
- ☐ Eksperimentet
- ☐ Da skateboard var forbudt
- ☐ Farlig katt

Fikk du god nok tid til å svare på alle oppgavene?

- ☐ Ja
- ☐ Nei