



Matematikksenteret
Nasjonalt senter for matematikk i opplæringen

TEKNISK RAPPORT

NASJONAL PRØVE I REGNING
5. TRINN 2021
Kapittel 1–5

Morten Svorkmo
Olav Dalsegg Tokle
Monica Rehaug
Camilla N. Justnes
Stian Tømmerdal

MATEMATIKKSENTERET, NTNU 2022

Nasjonal prøve i regning for 5. trinn 2021

Morten Svorkmo, Olav Dalsegg Tokle, Monica Rehaug, Camilla N. Justnes og Stian Tømmerdal

Nasjonalt senter for matematikk i opplæringen (Matematikksenteret)

NTNU mars 2022

Innhold

Innledning	4
Kapittel 1: Prepiloter NP5 - 2021	5
Kapittel 2: 1. piloter NP5 – 2021	6
Kapittel 3: 2. piloter NP5 - 2021 (siste utprøving)	6
Hvordan 2. pilotene ble satt sammen.....	6
Gjennomføring av 2. pilotene	7
Kapittel 4: Forslag til Endelig prøve NP5 - 2021.....	8
Innledning	8
Psykometri	8
Forklaring til noen oppgaver i kohortprøven.....	11
Oppgaver med lav diskriminering	11
Oppgaver med kjønnsforskjeller	13
Validitet.....	16
Generelle vurderinger av prøven som helhet.....	19
Kapittel 5: Analyse av nasjonal prøve i regning 2021	20
Innledning	20
Psykometri	20
Oversikt over oppgaver 2021.....	21
Testinformasjon og måleusikkerhet	22
Tolkninger av resultatene	22
Resultater til enkeltoppgaver	23
Kjønnsforskjeller i årets prøve	23
Ubesvarte oppgaver.....	25
Sammenligning av prøvene over tid	25
Fordeling på mestringsnivåer 2014–2020	26
Vurdering av prøven	27
Oppgaveanalyse NP5 2021	28
Kilder:.....	29

Innledning

Denne rapporten beskriver prosessen for utvikling, utprøvinger og analyse av resultatene til nasjonal prøve i regning for 5. trinn 2021. Vanligvis gjennomfører vi tre utprøvingsrunder, men på grunn av koronasituasjonen har vi dette året måtte nøye oss med to utprøvinger. Den siste utprøvingen ble derfor annerledes og omfattet flere oppgaver enn det vi har hatt tidligere år. Det kommer vi tilbake til i kapittel 3.

De tre første kapitlene beskriver prosessen rundt utprøvingene og formålet med og utfordringer i de ulike fasene av utprøvingene.

Kapittel 4 beskriver prosessen og begrunner sammensetningen av forslag til endelig prøve. Der er også noen utvalgte oppgaver mer detaljert beskrevet. Kapittel 5 er en analyse og betraktninger rundt resultatene etter at nasjonal prøve for 2021 er gjennomført.

Gjennom hele utviklingsprosessen brukes klassisk test teori (CTT) og Item Response Theory (IRT) til å analysere resultatene. IRT analysene er foretatt med Xcalibre, med spenn i b-verdier fra -3 til 3 og en D-konstant på 1,7 i beregning av a-verdier. De utvalgte skolene som deltar i utprøvingene, er trukket av Utdanningsdirektoratet.

Kapittel 1: Prepiloter NP5 - 2021

Prepiloter for 2021-prøven ble gjennomført i februar 2020 med sju prøvesett og 32 oppgaver i hvert sett. Av disse var det 174 kortvarsoppgaver der elevene fyller inn svar uten at de har svaralternativer å velge blant. De fleste av disse oppgavene er nye oppgaver til 2021, men det er også med oppgaver fra tidligere år som er endret eller omarbeidet såpass mye at de igjen må prøves ut som åpne. Prepilotene ble gjennomført med totalt 1701 elever på 5. trinn.

Oppgavene vi utvikler skal være regning i fag på fagenes premisser og måle grunnleggende ferdigheter slik dette er beskrevet i Rammeverk for nasjonale prøver (UDIR, 2017) og i fagplanene til LK06 (UDIR, 2006). Fagtilknytningene til en oppgave er derfor viktig allerede når oppgavene blir produsert. Oppgavene skal være i kontekster som elevene kan kjenne seg igjen i, og favne alle vanskegrader innen de matematiske områdene som Rammeverket beskriver. En utfordring med å lage oppgaver for 2021 er at den nasjonale prøven i 2021 vil lages etter kompetansemål fra LK06, samtidig som elevene har hatt undervisning etter ny læreplan i ett år når prøven gjennomføres. Den grunnleggende ferdigheten å regne har ikke endret seg vesentlig, men det er noen endringer i beskrivelsen av ferdigheten i de ulike fagene. Beskrivelsen av den grunnleggende ferdigheten å regne har ikke endret seg vesentlig i de ulike fagene i den nye læreplanen sammenlignet med tidligere, men det er noen endringer i kompetansemål som kan føre til endringer for nasjonale prøver.

Vi har forsøkt å fordele oppgaver både etter vanskegrad og innhold i prepilotene, men det er i første rekke resultater fra enkeltoppgaver vi er ute etter i denne utprøvingen. Hovedhensikten med prepilotene er å se hvordan oppgavene fungerer og samtidig finne høyfrekvente feilsvar, og spesielt høyfrekvente feilsvar som tyder på misoppfatninger.

For å kunne sammenligne resultater mellom prepilotene har vi også i denne utprøvingen med åtte oppgaver som vi betegner prepilotanker. En av oppgavene er felles i alle sju piloter, mens de sju andre oppgavene fordeles med én oppgave i hver pilot. Alle prepilotankeroppgaver er prøvd i tidligere prepiloter og gir en pekepinn på forskjellene i resultater mellom de ulike pilotene. Disse oppgavene setter vi på samme plass i oppgavesettet, noe som også gjør det mulig å sammenligne resultater fra den første utprøvingen fra år til år.

Prepiloter NP5 2021	Prepilot A	Prepilot B	Prepilot C	Prepilot D	Prepilot E	Prepilot F	Prepilot G
Antall gutter	113	124	130	127	122	121	126
Antall jenter	125	103	119	134	114	123	120
Totalt antall elever	238	227	249	261	236	244	246
Gj.snitt Theta gutter	0,068	0,173	0,195	0,194	0,152	0,167	0,028
Gj.snitt Theta jenter	-0,070	-0,182	-0,161	-0,190	-0,118	-0,149	-0,064
Gj.snitt Theta totalt	0,000	0,012	0,025	-0,003	0,022	0,008	-0,017
Standardavvik Theta	0,589	0,938	0,963	0,966	0,945	0,890	0,891
SE til gjennomsnitt Theta	0,038	0,062	0,061	0,060	0,062	0,057	0,057
Gj.snitt løsningsprosent gutter	52,0 %	58,5 %	53,9 %	58,5 %	56,5 %	55,7 %	48,8 %
Gj.snitt løsningsprosent jenter	44,1 %	50,8 %	46,7 %	50,8 %	51,0 %	49,3 %	46,9 %
Gj.snitt løsningsprosent totalt	47,9 %	55,0 %	50,4 %	54,6 %	53,8 %	52,5 %	47,9 %
Svarprosent	93,5 %	95,9 %	95,9 %	93,7 %	92,1 %	96,6 %	95,1 %
Cronbachs alpha	0,865	0,877	0,879	0,881	0,881	0,853	0,849

Antall items	32	32	32	32	32	32	32
Antall flervalgsoppgaver (MC)	6	7	6	4	5	8	10
Antall interaktive oppgaver (IA)	2	3	5	3	2	2	4

Tabell 1 Oppsummering prepiloter

Kapittel 2: 1. piloter NP5 – 2021

1. piloter for NP5 – 2021 ble ikke gjennomført i mai 2020 slik planen var på grunn av koronasituasjonen. I og med at alle skoler var stengt og elevene hadde hjemmeundervisning til slutten av mai, ble denne utprøvingen avlyst. Vi hadde et forslag klart til fire piloter med 35 oppgaver i hver slik vi har hatt tidligere, men disse ble aldri gjennomført. Det resulterte i en endring i den siste utprøvingen, noe som er beskrevet nærmere i kapittel 3.

Kapittel 3: 2. piloter NP5 - 2021 (siste utprøving)

Hvordan 2. pilotene ble satt sammen

Tidligere år har denne siste utprøvingen bestått av to oppgavesett som hver for seg skal tilfredsstillere kravene til en endelig prøve, både med hensyn til vanskegrad, format, innhold og reliabilitet. I og med at vi ikke har noen resultater fra 1. pilot å støtte oss til, ble det denne gangen en større utfordring å velge oppgaver til siste utprøving. Vi hadde blant annet få oppgaver som var prøvd ut i flervalgsformatet og som vi dermed var usikre på hvordan ville fungere når de ble endret fra åpne oppgaver der eleven selv fyller inn svaret, til flervalgsoppgaver med svaralternativer.

For å være sikre på at vi hadde nok oppgaver, spesielt i flervalgsformat, valgte vi derfor å utvide utprøvingen med ett oppgavesett, slik at vi hadde tre sett som hver for seg skulle representere validitet og vanskegrad tilsvarende en endelig prøve. Det ble også viktig at vi fikk et tilstrekkelig antall elever til utprøvingen i og med at vi hadde et ekstra oppgavesett.

Erfaringen vår fra tidligere er at løsningsprosenten til oppgavene øker noe fra 2. pilot til endelig prøve. Tilsvarende endring skjer også med oppgavens b-verdi, som er et annet mål på vanskegrad. Dette må vi ta i betraktning når vi velger oppgaver til denne siste utprøvingen. Når vi da ikke helt vet hvordan nye flervalgsoppgaver oppfører seg, måtte vi støtte oss til de verdier oppgavene hadde etter første utprøving. For å kunne ha en viss oversikt over vanskegraden i hvert av de tre settene, hadde vi 10 fellesoppgaver som ble plassert på samme sted i de tre prøvene. Det ga en indikasjon på forskjeller mellom de tre settene, samtidig som vi kunne bruke parametere fra disse oppgavene til å skalere verdiene til de øvrige oppgavene på en felles skala. Vi vet av erfaring at vi trenger å ha noen vanskelige oppgaver å velge blant til den endelige prøven, derfor må vi passe på at det er nok oppgaver fra mestringsnivå 3 i de to pilotene til sammen uten at oppgavesettene oppleves for krevende.

Vi forsøkte å få fordelingen av oppgaver innen hovedområdene tall, måling og geometri og statistikk nokså lik den fordelingen vi så for oss i en endelig prøve. Noen oppgaver kan være vanskelig å kategorisere innenfor ett hovedområde i og med at de fleste oppgaver på en eller annen måte krever noe tallbehandling. Vi har likevel prøvd å få en fordeling som vi mener harmonerer med omfanget av kompetansemål som omhandler grunnleggende ferdighet i regning slik det er beskrevet i LK06 (UDIR, 2006). Vi forsøker å få med oppgaver med kontekster fra ulike fag som kan være gode eksempler på regning som grunnleggende ferdighet i faget. Det kan være oppgaver som egner seg i

en veiledning og som hjelper lærerne med å se sammenhengen mellom den grunnleggende ferdigheten å regne og kompetansemål i faget.

For 2021-prøven ble det vanskelig å ta mye hensyn til ankeroppgavene da vi satte sammen de tre 2. pilotene. Så lenge ankeroppgavene stort sett befinner seg mellom -1 og 1 i b-verdi, regner vi likevel med at vi i utprøvingen vil å ha nok oppgaver å velge blant som kan fungere som skyggeoppgaver til ankeret, spesielt siden vi har en utprøving med tre sett i stedet for to, i alt 115 oppgaver i siste utprøving.

Oppgaver med bias mot kjønnsbias fra prepilot vil i stor grad også få bias etter 2. pilot. Slike oppgaver prøver vi å unngå i siste utprøving så lenge det ikke er oppgaver som vi av helt spesielle hensyn ønsker skal være med i en endelig prøve. Vanskelige oppgaver med lav løsningsprosent og tilsvarende høy b-verdi har en tendens til å komme ut med bias etter utprøving, og det kan være en avveining om slike oppgaver skal tas ut fordi de har bias. Dette hensynet må i flere oppgaver vurderes opp mot behovet for å ha nok oppgaver å velge blant på høyeste mestringsnivå.

Vi satte sammen tre 2. piloter, A, B og C, som hver bestod av 45 oppgaver. I alt 35 oppgaver i hver pilot i tillegg til 10 fellesoppgaver som gikk i alle tre pilotene. Tabell 2 og 3 viser sammensetning før gjennomføring.

Hovedområde	2. pilot A	2. pilot B	2. pilot C	Fellesoppgaver	Sum
Tall (T)	20	19	18	5	62
Måling og geometri (M&G)	11	11	11	3	36
Statistikk (S)	4	5	6	2	17

Tabell 2 Fordeling av oppgaver innen hovedområder

Mestringsnivå	2. pilot A	2. pilot B	2. pilot C	Fellesoppgaver	Sum
Mestringsnivå 1	8	7	7	3	25
Mestringsnivå 2	16	14	17	3	50
Mestringsnivå 3	11	14	11	4	40

Tabell 3 Fordeling av oppgaver innen mestringsnivå

Gjennomføring av 2. pilotene

2. piloter ble gjennomført 17.–28. august 2020 med tre fullskala oppgavesett, hver med 45 oppgaver. Utprøvingen gjennomføres på 5. trinn, med elever som senere på høsten skal gjennomføre en ordinær nasjonal prøve i regning. Gjennomføringstidspunktet er så nær opptil tidspunktet for gjennomføring av endelig prøve som mulig.

Det var totalt 3164 elever som gjennomførte 2. pilotene, i overkant av 1000 elever på hver prøve. Det er omtrent det samme antallet som vi har hatt på hver prøve tidligere. Tabell 4 viser utvalg og resultater fra gjennomføringen. Her ser vi at løsningsprosenten i alle tre settene er under 50 prosent, mens de 10 fellesoppgavene har en gjennomsnittlig løsningsprosent på litt over 50 prosent.

Ellers viser analysen av resultatene at vi har flere oppgaver som diskriminerer lavere enn ønskelig sammenlignet med siste utprøving tidligere år. Siden vi mistet en utprøving i mai på grunn av stengte skoler og koronasituasjonen, måtte vi prøve oppgaver i siste utprøving som mest sannsynlig hadde blitt forkastet etter den utprøvingen vi mistet. Spesielt gjelder dette flervalgsoppgaver og interaktive oppgaver. Heldigvis valgte vi å prøve ett oppgavesett mer enn vi tidligere har gjort, slik at

vi har flere oppgaver å velge blant til en endelig prøve. Vi har likevel en del oppgaver med parametere som gjør at vi må vurdere dem nøyer for å se om de kan være aktuelle i en endelig prøve. Her blir det en avveining mellom psykometri, validitet og pedagogiske hensyn. Oppgaver med litt lav diskriminering eller kjønnsbias etter siste utprøving som vi likevel har valgt å ta med, er nærmere beskrevet i kapittel 4.

2. piloter NP5 2021	2. pilot A	2. pilot B	2. pilot C	10 Fellesoppgaver
Antall gutter	543	561	550	1654
Antall jenter	516	489	505	1510
Totalt antall elever	1059	1050	1055	3164
Gj.snitt Theta gutter	0,069	0,132	0,151	0,075
Gj.snitt Theta jenter	-0,179	-0,127	-0,09	-0,092
Gj.snitt Theta totalt	-0,052	0,011	0,036	-0,005
Standardavvik Theta	0,943	0,958	0,939	0,820
SE til gjennomsnitt Theta	0,029	0,030	0,029	0,015
Gj.snitt løsningsprosent gutter	48,3 %	45,4 %	48,4 %	52,9 %
Gj.snitt løsningsprosent jenter	43,8 %	40,5 %	43,7 %	48,5 %
Gj.snitt løsningsprosent totalt	46,1 %	42,9 %	46,2 %	50,8 %
Svarprosent	92,0 %	89,4 %	93,1 %	93,3 %
Cronbachs alpha, inkl fellesoppgaver	0,896	0,900	0,899	x
Antall items	35	35	35	10
Antall flervalgsoppgaver (MC)	19	14	18	8
Antall interaktive oppgaver (IA)	3	3	3	1

Tabell 4 Oppsummering av 2. piloter 2021

Kapittel 4: Forslag til Endelig prøve NP5 - 2021

Innledning

Høsten 2020 deltok til sammen 3164 elever fordelt på tre piloter til 2021-prøven. 2. pilot er den siste utprøvingen for oppgavene før den endelige prøven, og det er elevsvarene fra denne utprøvingen som danner grunnlag for tallene som blir løftet fram i dette kapitlet. I tillegg er det kjørt en IRT-analyse av oppgavene som utgjør den endelige prøven, kohortprøven. Tabeller eller diagram som henviser til disse, er kommentert med «tall fra preanalyse». Det betyr at tallene er basert på elevbesvarelser fra 2. pilot, men at det er foretatt en samlet analyse av de 45 oppgavene som utgjør forslag til endelig prøve og ankerprøve (tabell 5).

Nasjonal prøve for 5. trinn 2021 består av 45 oppgaver. Prøven foreligger i fire versjoner. Oppgavene i tre av versjonene er de samme, men rekkefølgen til de ti første oppgavene er ulik i de tre versjonene. Det er gjort for å redusere sannsynligheten for at to elever som sitter nær hverandre, arbeider med samme oppgave samtidig. Versjonen som er ulik de andre tre, er en ankerprøve som brukes til å måle utvikling over tid. Ankerprøven består av 20 ankeroppgaver som har verdier fra tidligere år, mens de øvrige oppgavene i ankerprøven er de samme som i de tre andre versjonene.

Psykometri

Tabell 5 viser sentrale verdier til prøven for 2021 ut fra en estimert preanalyse. Oppgavene er plukket ut fra tre 2. piloter, og samtlige oppgaver har vært gjennom to utprøvinger. Siden settet er satt sammen av oppgaver fra tre 2. piloter, som er besvart av tre uavhengige elevgrupper, er det

vanskelig å beregne en estimert reliabilitet for den endelige prøven. Basert på tidligere erfaringer med å sette sammen prøver etter denne modellen, kan vi anta at reliabiliteten til den endelige prøven kommer til å bli rundt 0,90. Dette tilfredsstillende kravet i Rammeverket (UDIR, 2017).

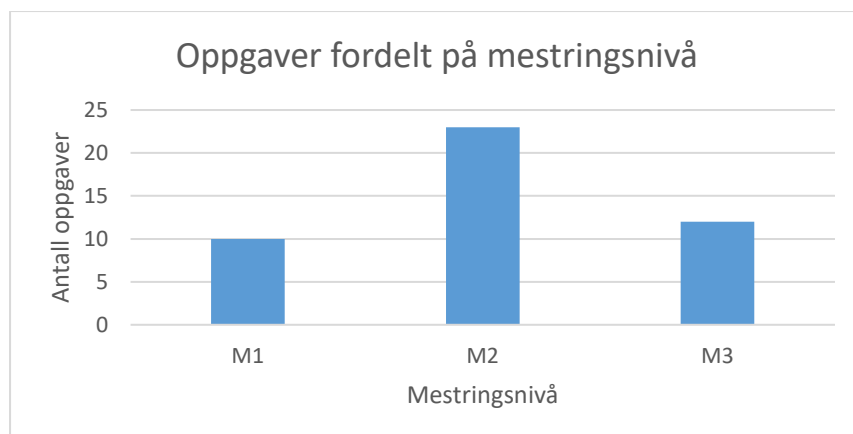
Forslag til prøve NP5 2021	Kohortprøve	Ankeret
Gj.snitt Theta gutter	0,095	0,133
Gj.snitt Theta jenter	-0,107	-0,147
Gj.snitt Theta totalt	-0,001	0,001
Standardavvik Theta	0,944	0,965
SE til gjennomsnitt Theta	0,004	0,016
Gj.snitt løsningsprosent gutter	50,6 %	50,0 %
Gj.snitt løsningsprosent jenter	46,0 %	42,4%
Gj.snitt løsningsprosent totalt	48,4 %	46,4 %
Svarprosent	91,2 %	91,7 %
Cronbachs alpha	0,900	0,872
Gj. Snitt T-Rpbis	0,490	0,516
Gj. Snitt a-verdi	0,711	0,759
Gj. Snitt b-verdi	0,084	0,154
Antall items	45	20
Antall flervalgsoppgaver (MC)	24	14
Antall interaktive oppgaver (IA)	4	3

Tabell 5 Sentrale verdier forslag til endelig prøve etter preanalyse. Ankeret er basert på verdier fra NP5 2020.

Ut fra tidligere erfaring kan vi vente at gjennomsnittlig løsningsprosent i den ordinære prøven vil stige noe, slik at vi antar den vil havne på noe over 50 prosent. Endringen i løsningsprosent fra siste utprøving til endelig prøve har de siste årene vært på 4–6 prosentpoeng, og tilsvarende har gjennomsnittlig b-verdi sunket 0,20–0,30.

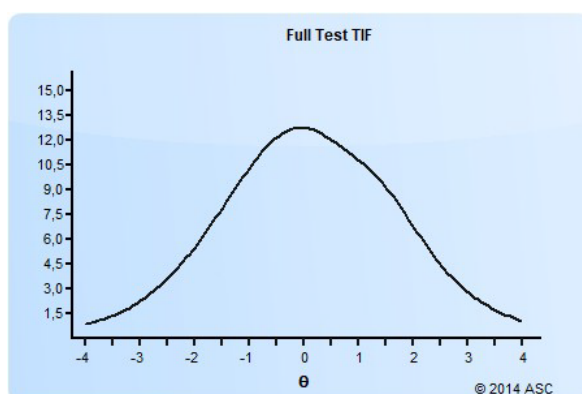
Andelen flervalgsoppgaver er 53 prosent. Det er litt færre flervalgsoppgaver enn tidligere år, noe som blant annet skyldes at vi mistet en utprøving der vi i første rekke tester flersvarsalternativer. Det medførte at vi i siste utprøving hadde med en del flervalgsoppgaver som diskriminerte lavere enn ønskelig, og som derfor ikke ble med i forslaget til prøve. Av samme årsak er det færre interaktive oppgaver i årets forslag til prøve enn det har vært tidligere.

Fordelingen av oppgaver på mestringsnivå kan forskyves noe. Oppgaver som ligger nær grensen mellom to mestringsnivå, har en tendens til å havne på det laveste mestringsnivået. Dette har vi forsøkt å ta hensyn til i forslaget som foreligger. Denne gangen har vi derfor 12 oppgaver som etter siste utprøving er på mestringsnivå 3. I tillegg har vi 10 oppgaver på mestringsnivå 1 og 23 oppgaver på mestringsnivå 2.

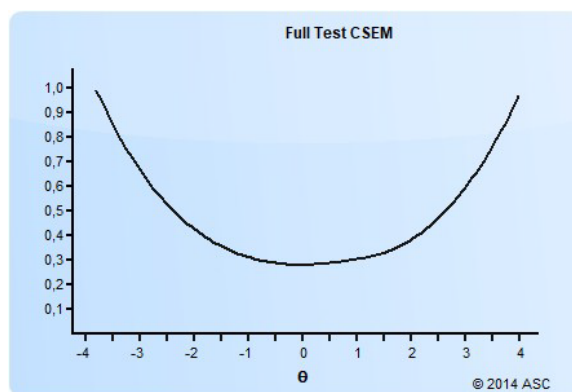


Figur 1 Oppgaver fordelt på mestringsnivå basert på tall fra preanalyse

En preanalyse av oppgavene til kohortprøven, viser at testinformasjonen (figur 2) er størst for elever med $\theta = 0,050$. Det vil si at prøven gir mest informasjon om elever som presterer omtrent midt på ferdighetskalaen. Det er også der vi har flest elever. I tillegg ser vi at prøven gir tilfredsstillende informasjon om elever som presterer lavt og høyt på prøven. Figur 5 viser at måleusikkerheten naturlig nok er lavest der det er flest elever.

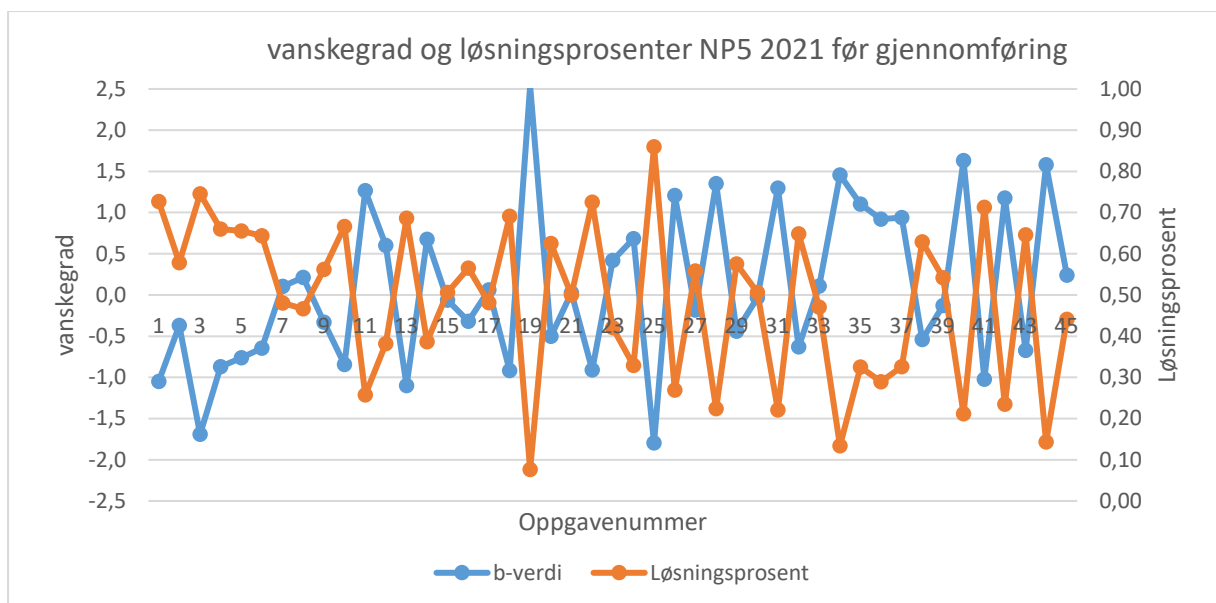


Figur 2 Testinformasjon til kohortprøve basert på tall fra preanalyse



Figur 3 Måleusikkerhet til forslag kohortprøve basert på tall fra preanalyse

Diagrammet i figur 4 viser oppgavenes vanskegrad og løsningsprosent i samme diagram. Hvis vi sammenligner verdiene, ser vi at de speiler hverandre, noe som tyder på god sammenheng mellom oppgavenes vanskegrad og gjennomsnittlig løsningsprosent. Vi forsøker med hensikt å ha en variasjon i vanskegrad gjennom hele prøven. Vi vet imidlertid ikke om oppgaver med vanskegrad og som vi definerer som vanskelige, alltid oppleves som mer krevende for elevene enn oppgaver med lavere vanskegrad. Vi prøver uansett å unngå at elevene får flere oppgaver med høy vanskegrad etter hverandre når vi setter sammen oppgavene i en prøve.



Figur 4 vanskegrad (b-verdi) og løsningsprosent etter oppgavenummer i kohortprøven NP5 2020. Tall fra preanalyse.

Forklaring til noen oppgaver i kohortprøven

Av de 45 oppgavene er det 40 oppgaver som tilfredsstiller anbefalingene om at oppgavene skal ha point-biserial-verdi $> 0,3$ og a-verdi $> 0,5$ i Rammeverk for nasjonale prøver (KILDE). De fem som ikke tilfredsstiller begge disse diskrimineringsverdiene, er likevel foreslått til prøven ut fra didaktiske og helhetlige hensyn.

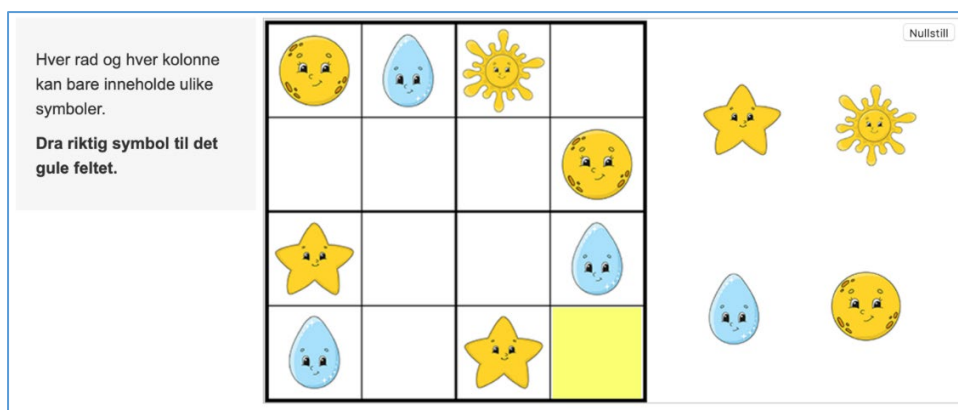
I IRT-preanalysen av kohortprøven får fem oppgaver kjønnsbias, to mot gutter (oppgave 13 og 32) og tre mot jenter (oppgavene 16, 31 og 34). Noen av disse oppgavene får ikke bias i analysen etter siste utprøving, men får bias når de blir prøvd i en preanalyse i et stipulert oppgavesett. I tabell 6–13 er det satt inn tall for vanskegrad og diskriminering fra både CTT- og IRT-analysene. I tillegg er gjennomsnittlig dyktighet (Theta) for elevene som har valgt de ulike svaralternativene i flervalgsoppgavene, tatt med.

Når det gjelder tall knyttet til forskjeller mellom gutter og jenter har vi tatt med Mantel-Haenszel-koeffisienten (M-H). I IRT-analysen er gutter satt som gruppe 1 og jenter som gruppe 2, noe som betyr at dersom $M-H > 1$, har guttene i gjennomsnitt over hele ferdighetsskalaen større odds for å løse oppgaven enn jentene. Dersom $M-H < 1$, har jentene større odds for å løse oppgaven enn guttene. MH-koeffisienten i oppgavene fordeler seg slik at 22 oppgaver i utgangspunktet går i jentefavør mens 23 oppgaver går i guttefavør. Kjønnsforskjellen i løsningsprosent mellom gutter og jenter er i den stipulerte kohortprøven 4,6 prosentpoeng i favør gutter.

Oppgaver med lav diskriminering

Av de 45 oppgavene i forslag til endelig prøve er det fire oppgaver som etter preanalyse har a-verdi $< 0,5$. Det er oppgave 3, 14, 21 og 35. Oppgave 14 og 21 har a-verdi så vidt under 0,5 og point biserial-verdi som er godt over 0,3. Vi velger derfor å ikke gå nærmere inn på en begrunnelse for hvorfor disse to oppgavene er med. Når det gjelder oppgave 3 og 35 har vi gitt en nærmere beskrivelse av hvorfor vi mener oppgavene bør være med i prøven. I tillegg har vi oppgave 19 som har $S-Rpbis < 0,3$ men samtidig a-verdi 0,76. Vi begrunner hvorfor vi ønsker den skal være med i prøven.

Oppgave 3 Sudoku for kids



Figur 5 Item 5196847 Sudoku for kids

Dette er en interaktiv oppgave som har høy svarprosent, men som samtidig kanskje er litt uvant for elever i en regneprøve. Det å gjenkjenne og forstå mønster er en kompetanse vi finner både i kunst og håndverk og i matematikk og som vi mener bør være med i en nasjonal prøve i regning. I og med at oppgaven skiller seg litt ut fra mange av de andre oppgavene, kan dette bidra til å gi noe lavere diskrimineringsverdier. Samtidig er det en relativt kjønnsnøytral oppgave med en kontekst vi tror mange elever vil kjenne seg igjen i.

Itemnr	S-Rpbis	T-Rpbis	a-verdi	b-verdi	Løsningsprosent	M-H koeffisient
5196847	0,26	0,30	0,41	-1,69	0,75	0,94

Tabell 6 Item 5196847 Sudoku for kids

Oppgave 19 Strikke skjerf

Guro strikker et skjerf som skal bli 150 cm langt.
Etter at hun har strikket 20 omganger, er skjerfet 15 cm.

Hvor mange flere omganger må Guro strikke for å gjøre skjerfet ferdig?

Svar:

Figur 6 Item 5125970 Strikke skjerf

Oppgave 19 er en oppgave i en praktisk kontekst og som er et godt eksempel på bruk av regning i kunst og håndverk. Det er en vanskelig oppgave, den vanskeligste i hele prøven basert på b-verdien. Til tross for en relativt høy MH-koeffisient kommer den ikke ut med bias i preanalysen. Det er ofte krevende å finne vanskelige oppgaver uten bias, så vi mener denne krevende oppgaven bør være med i prøven. Vi har diskutert om begrepet omganger er ukjent for elevene, men det er et fagbegrep som er vanskelig å erstatte. Vi tenker også at det er mulig å løse oppgaven uten å kjenne dette begrepet.

Itemnr	S-Rpbis	T-Rpbis	a-verdi	b-verdi	Løsning s-	M-H koeffisient
5125970	0,27	0,34	0,76	2,54	0,08	1,99

Tabell 7 Item 5125970 Strikke skjerf

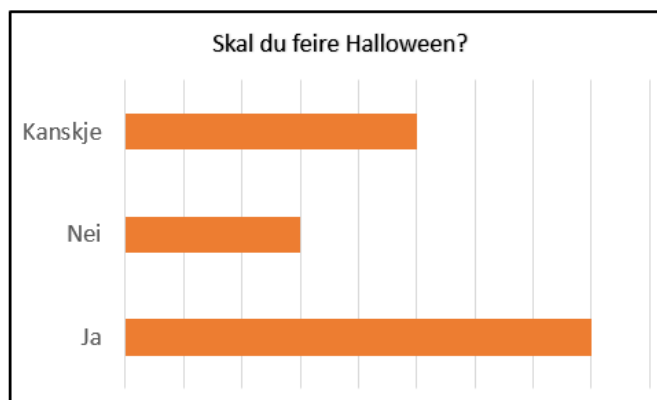
Oppgave 35 Undersøkelse Halloween

Pelle har spurt 80 elever på skolen om de skal feire Halloween.

Halvparten av dem svarte ja.

Hvor mange elever svarte nei?

- 3
- 15
- 30
- 40



Figur 7 Item 5196801 Undersøkelse Halloween

Oppgave 35 måler forståelse av et liggende søylediagram. I diagrammet er ikke enhetene angitt med tall, og derfor må elevene bruke opplysninger i teksten for å tolke søylene i diagrammet. Dette er en oppgavetype som kan være litt uvant for elevene, men vi synes den gir en god tilbakemelding på hvem som forstår hvilke tall de liggende søylene i diagrammet representerer.

Itemnr	S-Rpbis	T-Rpbis	a-verdi	b-verdi	Løsningsprosent	M-H koeffisient	Gjennomsnittlig theta			
							1	2	3	4
5196801	0,29	0,37	0,45	1,10	0,32	1,04	-0,34	0,63	0,04	-0,02

Tabell 8 Item 5196801 Undersøkelse Halloween

Oppgaver med kjønnsbias

Etter siste utprøving og preanalyser har vi fem oppgaver med kjønnsbias, og som vi derfor har vurdert ekstra nøye om de er aktuelle til 2021-prøven. Det er én oppgave med bias mot gutter på mestringsnivå 1, og én oppgave på mestringsnivå 2. Vi har tre oppgaver med bias mot jenter, én på mestringsnivå 2 og to på mestringsnivå 3. Det er vanskelig å følge anbefalingen om at oppgaver med bias mot kjønn bør være på samme vanskegrad, men vi mener likevel at 40 av 45 oppgaver uten bias etter preanalyse, bør gi en godt balansert prøve. De fem oppgavene med bias dekker områder vi mener bør være representert i en nasjonal prøve, og vi har derfor valgt å ta dem med.

Oppgave 13 Fødselsdag Mari

Mari ser på telefonen sin at det er lørdag 10. mars.
Hun har fødselsdag 18. mars.

Hvilken ukedag er 18. mars?

Svar:



Figur 8 Item 5133433 Fødselsdag Mari

Oppgaver med datoer, dager eller måneder har ofte en tendens til å gå i jentefavør. Dette opplever vi uavhengig av om konteksten i oppgaven dreier seg om jenter eller gutter. I denne oppgaven må eleven kunne ukedagene og rekkefølgen på dem for å finne løsningen. Dersom du kan dette, kan oppgaven løses med enkle strategier, men vi ser likevel at nesten en tredel av elevene ikke greier dette. For å balansere antallet oppgaver med bias, ønsker vi å beholde en oppgave som går i jentefavør. I tillegg mener vi at å kunne regne seg fram til en ukedag er en grunnleggende ferdighet.

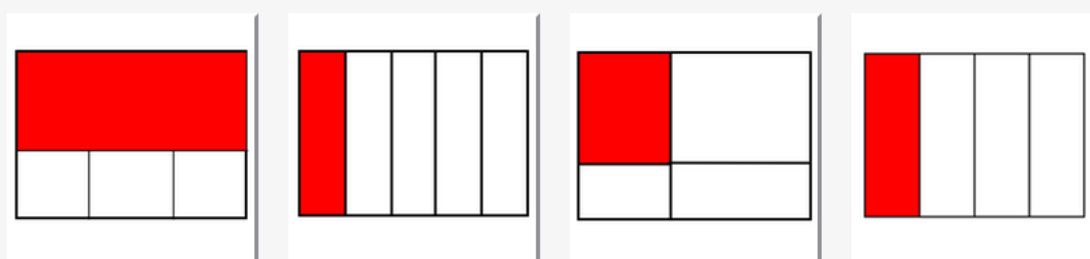
Itemnr	S-Rpbis	T-Rpbis	a-verdi	b-verdi	Løsningsprosent	M-H koeffisient	Gjennomsnittlig theta			
							1	2	3	4
5133433	0,32	0,39	0,51	-1,10	0,69	0,75	-0,38	-0,64	-0,54	0,32

Tabell 9 Item 5133433 Fødselsdag Mari

Oppgave 32 Farg en firedel

David har fargelagt en firedel av et rektangel.

Hvilket rektangel har David fargelagt?



Figur 9 Item 5196410 Farg en firedel

Dette er en oppgave som går i jentefavør. Vi ønsker å ha med oppgaver der elevene kan vise om de har en grunnleggende forståelse for representasjoner av helt enkle brøker. I denne oppgaven er brøken uttrykt som en firedel og ikke som brøkuttrykket $\frac{1}{4}$. Theta-verdien viser også at det er et relativt markant skille i dyktighet mellom de elevene som velger riktig alternativ i denne oppgaven og de som ikke gjør det. Å forstå brøkrepresentasjon mener vi er viktig for å kunne utvikle en grunnleggende brøkførståelse senere, også i andre fag enn i matematikk. Mat og helse er eksempel på et slikt fag.

Itemnr	S-Rpbis	T-Rpbis	a-verdi	b-verdi	Løsningsprosent	M-H koeffisient	Gjennomsnittlig theta			
							1	2	3	4
5196410	0,44	0,57	0,82	-0,63	0,65	0,77	-0,54	-0,43	-0,59	0,43

Tabell 10 Item 5196410

Oppgave 16 Vekt katt

Sara er hos dyrlegen med katten. Katten veier 4,5 kg.
Forrige gang katten var hos dyrlegen veide den 3,8 kg.


Hvor mange kilogram mer veier katten nå enn tidligere?

0,7 kg

1,3 kg

1,7 kg

7 kg



Figur 10 Item 5133799 Vekt katt

Konteksten i denne oppgaven mener vi ikke skal favorisere gutter, men vi ser at oppgaver som inneholder desimaltall har en tendens til å gå i guttefavør. Det gjelder uansett kontekst og valg av regneoperasjon. Dette er i utgangspunktet en subtraksjonsoppgave som kan løses med ulike strategier. Bruk av hensiktsmessige regnestrategier i subtraksjon er nyttig i hoderegning og kan være høyst aktuelt å behandle nærmere i en veiledning.

Itemnr	S-Rpbis	T-Rpbis	a-verdi	b-verdi	Løsningsprosent	M-H koeffisient	Gjennomsnittlig theta			
							1	2	3	4
5133799	0,40	0,50	0,62	-0,32	0,56	2,41	0,56	-0,45	-0,40	-0,19

Tabell 11 Item 5133799 Vekt katt

Oppgave 31 Dele planke i fem

Malin skal dele en planke i fem like lange biter.
Planken er 1,5 m.


Hvor lang blir hver bit?

3 cm

5 cm

25 cm

30 cm



Figur 11 Item 5162049 Dele planke i fem

Dette er en oppgavetype vi år etter år erfarer gir de samme kjønnsforskjeller. Oppgaver som krever omgjøring mellom meter og centimeter har en klar tendens til å gå i guttefavør. Vi tenker også at de to første svaralternativene bør kunne være mulig å utelukke gjennom en refleksjon. Det å lære elevene til å gjøre en vurdering av hva et mulig svar kan være, er en aktuell strategi i mange oppgaver, og denne oppgaven bør egne seg godt til dette. Det kan jo også hende at utfordringen for noen i oppgaven er å lese desimaltallet 1,5. Selv om vi vet at oppgaver som krever omgjøring mellom enheter går i guttefavør, kan vi ikke utelukke den typen oppgaver i en prøve i den grunnleggende ferdigheten å kunne regne.

Itemnr	S-Rpbis	T-Rpbis	a-verdi	b-verdi	Løsningsprosent	M-H koeffisient	Gjennomsnittlig theta			
							1	2	3	4
5162049	0,43	0,52	0,82	1,30	0,22	1,87	0,16	-0,34	0,12	1,08

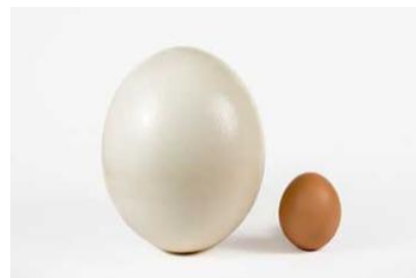
Tabell 12 Item 5162049 Dele planke i fem

Oppgave 34 Strutsegg

Hanna leser at et strutsegg kan veie 1,5 kg.
Hun veier et vanlig høneegg og finner at det veier 50 g.

Hvor mange vanlige høneegg veier til sammen like mye som et strutsegg?

Svar:



Figur 12 Item 5196403 Strutsegg

Denne oppgaven omhandler omgjøring mellom enhetene gram og kilogram. I tillegg kan det være relativt krevende regning hvis du ikke har strategier for å regne med store tall. Oppgaven diskriminerer godt, men den viser også en relativt klar forskjell mellom jenter og gutter. Det er en vanskelig oppgave på øverste mestringsnivå, og en oppgave som skiller godt. Vi har kun tre oppgaver i årets prøve som inneholder en form for omgjøring mellom enheter. Det er denne oppgaven som bruker enhetene gram og kilogram, og oppgavene 27 og 31 som prøver sammenhengen mellom centimeter og meter. Vi har ingen oppgaver i årets prøve med omgjøring mellom volumenheter eller tidsenheter. Mange regneoperasjoner i dagliglivet handler om regning der omgjøring mellom enheter er nødvendig. Vi mener derfor at slike oppgaver bør være med.

Itemnr	S-Rpbis	T-Rpbis	a-verdi	b-verdi	Løsningsprosent	M-H koeffisient
5196403	0,44	0,57	1,32	1,46	0,13	2,02

Tabell 13 Item 5196403 Strutsegg

Validitet

I Kunnskapsløftet (UDIR, 2006) presiseres det at regning er en grunnleggende ferdighet i alle fag. Nasjonal prøve i regning skal avdekke i hvilken grad elevenes regneferdigheter er i samsvar med beskrivelsene av regning som grunnleggende ferdighet, jf. Rammeverket (UDIR, 2017).

Nasjonal prøve i regning for 5. trinn måler elevenes kompetanse etter 4. trinn i grunnleggende ferdigheter i å kunne regne. Derfor må valideringen skje ut fra det en kan forvente av elever som nettopp har begynt i 5. trinn. Det betyr en vurdering av enkeltoppgaver og prøvesettet som helhet opp mot kravene i Rammeverk for nasjonale prøver, som handler om validitet, og opp mot beskrivelsene av den grunnleggende ferdigheten å kunne regne.

Oppgavene i prøven for 2021 knyttes til kompetansemål i fag som vist i tabell 14. En oppgave kan knyttes til kompetansemål i flere fag.

Fag	Antall oppgaver
Engelsk	0
KRLE	4
Kroppsøving	7
Kunst og håndverk	9

Matematikk	45
Mat og helse	10
Musikk	1
Naturfag	9
Norsk	2
Samfunnsfag	14

Tabell 14 Antall oppgaver som kan knyttes til kompetansemål i ulike fag etter 4.trinn

Som tabell 14 viser, varierer antall oppgaver som kan knyttes til kompetansemål i det enkelte faget. Utfordringen er å lage oppgaver som blir regning på fagenes premisser, samtidig som oppgavene skal ha gode nok psykometriske verdier til å tilfredsstille anbefalingene i rammeverket. Vi mener fordelingen slik den er i vårt forslag, er representativ for kompetansemål som kan knyttes til den grunnleggende ferdigheten å kunne regne i fagene.

Nasjonal prøve i regning for 5. trinn inneholder oppgaver innen områdene tall (T), måling og geometri (M&G) og statistikk (S). Oppgavene er knyttet til områdene i tråd med beskrivelsene i rammeverket (UDIR 2017). Prøven består av 20 oppgaver i området tall, 16 innenfor måling og geometri og 9 statistikkoppgaver (tabell 15). Denne fordelingen med litt flere oppgaver innen tall enn innenfor måling og geometri, samt omtrent halvparten av dette antallet i statistikk, gjenspeiler etter vår vurdering hvordan regning er synliggjort i kompetansemålene i læreplanen. Det er imidlertid viktig å påpeke at mange av oppgavene krever kompetanse fra flere fagområder. Alle oppgaver krever en form for tallbehandling, men vi har forsøkt å kategorisere oppgavene etter i hvilket område vi vurderer at hovedutfordringen ligger.

Opgaver fra fagområder	Mestringsnivå 1	Mestringsnivå 2	Mestringsnivå 3	Totalt
Tall	7	10	3	20
Måling og geometri	3	8	5	16
Statistikk	0	5	4	9
Totalt	10	23	12	45

Tabell 15 Fordeling av oppgaver på mestringsnivå basert på tall fra preanalyse

Det er ingen statistikkoppgaver på mestringsnivå 1 etter siste utprøving, men vi har et par oppgaver på mestringsnivå 2 som er såpass nær det laveste mestringsnivået at vi anser det som sannsynlig at de kan havne på laveste mestringsnivå etter endelig prøve. Inndeling i mestringsnivåer er basert på de samme kuttene til løsningsprosent som brukes i den nasjonale prøven i regning for 5. trinn.

I tillegg til å være knyttet til kategoriene i læreplanen i matematikk, er oppgavene knyttet til de ulike delferdighetene i den helhetlige problemløsningsprosessen som er beskrevet i kravene i Rammeverk for nasjonale prøver (UDIR, 2017). I noen oppgaver består hovedutfordringen i å identifisere situasjoner som involverer tall og formulere et regneuttrykk (gjenkjenne og beskrive), andre i å velge strategier for problemløsning og foreta beregninger (bruke og bearbeide) eller å vurdere gyldighet og reflektere over hva resultatene betyr for problemstillingen (reflektere og vurdere).

Tabell 16 viser en oversikt over fordeling av oppgaver knyttet til disse delferdighetene. Oppgavene er plassert der hvor hovedutfordringen i oppgaven er ment å ligge. Det er imidlertid viktig å være klar over at elever har ulike ferdighetsprofiler og at én elev kan ha utfordringer i å gjenkjenne og beskrive et problem, mens en annen elev kan ha utfordringer knyttet til å bruke og bearbeide i samme oppgave.

Oppgaver fra ferdighetsområder	Mestringsnivå 1	Mestringsnivå 2	Mestringsnivå 3	Totalt
Gjenkjenne og beskrive	7	5	3	15
Bruke og bearbeide	3	15	9	27
Reflektere og vurdere	0	3	0	3
Totalt	10	23	12	45

Tabell 16 Fordeling av oppgaver innen ferdighetsområder

Som tabell 16 viser, er det færrest oppgaver der hovedutfordringen er reflektere og vurdere. Det betyr ikke at dette er en delferdighet som i liten grad måles i prøven. Elevene bør i alle oppgaver vurdere og reflektere om det matematiske svaret de har funnet er et svar som er realistisk til konteksten i oppgaven.

Etter at forslaget til endelig prøve er ferdig, har en lærergruppe med fire personer gjennomført ekstern validering av prøvesettet. De gjør en vurdering av innhold, område, hovedutfordringer i løsningsprosessen, relevans til fag og mestringsbeskrivelser til hver enkelt oppgave. Tabell 17 viser hvilke kategorier vi ønsker å få tilbakemelding på, og et utdrag av valideringsskjemaet som lærerne bruker.

Nr.	Innhold ¹	Område ²	Kommentarer	Hovedutfordring i løsningsprosess ³	Relevans til fag ⁴ (LK06/20)	Mestringsnivå	Underpunkt mestringsnivå ⁵	Fasit	P-verdi 2. pilot
1	Plassverdisystemet	TA		GB	Ma, Samf	1	1.1	14 stk	0,73
2	Lese tabell	ST		BB	Ma, Nat, Samf, No	2	2.4 (2.9)	Alt.2: 2 fossefall	0,58
3	Koordinatsystem. Lese mønster.	MG		GB	Ma, Samf	1	1.7	Sol	0,75
4	Gjentatt addisjon	TA		BB	Ma	1	1.2	27 000 kr	0,66
5	Gjentatt addisjon	TA		BB	Ma, Krø	1	1.2	Alt.4: 17 poeng	0,66
6	Tid. Digital klokke og intervall.	MG		BB	Ma, Samf	2	2.5	20 min	0,64
7	Symmetri og speiling	MG		RV	Ma, Kuh	2	2.7	Alt.4: speiling	0,48
8	Kjøp og salg. S sammensatt	TA		RV	Ma, No	2	2.4	Alt.3: 150 kr	0,47
9	Mønster	MG		RV	Ma, Kuh	2	2.7	Alt.3: puslespillbrikke	0,56

Tabell 17 Utdrag fra skjema til ekstern validering NP5 2021

Mestringsnivå, fasit og løsningsprosent fra siste utprøving er verdier lærergruppen får før de starter sin egen validering, slik at de har noe å gå ut fra. De øvrige kolonnene fyller de ut. Tabell 17 viser et eksempel på hvordan den eksterne valideringen kan se ut etter at lærerne har gått gjennom prøven og gitt sine vurderinger. Vi får også anledning til å ta en prat med lærerne etter at valideringen er gjennomført, der de gir sine tilbakemeldinger. Dette arbeidet tar vi med oss når vi gjør våre vurderinger i de samme kategoriene.

I årets oppgavesett ble blant annet teksten i noen oppgaver kommentert. Spesielt er lærerne opptatt av at teksten ikke skal være for lang, og vi fikk noen konstruktive forslag til endringer som vi mener kan gjøres uten at oppgaven endres for mye.

Vi ser at lærerne har lettere for å ta med fagtilknytning til flere fag enn det vi har. Det kan være lett å se formuleringer under den grunnleggende ferdigheten å regne i ulike fag som passer til oppgaven, men som vi ikke finner igjen under kompetansemålene. Bruk av oppgaveformater, bilder og mestringsnivå er også områder lærerne har meninger om og som vi får anledning til å diskutere med dem. Denne prosessen er nyttig for oss i den videre prøveutviklingen.

Tekstlengde og at lesing påvirker resultatet for elevene er de hyppigste tilbakemeldingene vi får i møter med lærere som har gjennomført nasjonale prøver. Vi gleder oss derfor til overgangen til nytt prøvegjennomføringssystem i 2022, der lesestøtte for elever som ønsker det må være på plass for å styrke validiteten og reliabiliteten til nasjonale prøver i regning.

Generelle vurderinger av prøven som helhet

Gjennom utprøvinger har hver oppgave i forslag til endelig prøve vært prøvd ut to ganger. Vi måtte dessverre avlyse den andre utprøvingen på grunn av koronasituasjonen og stengte skoler, noe som gjorde utviklingen av 2021-prøven litt annerledes sammenlignet med tidligere år. Det gir oss en ekstra utfordring når vi skal sette sammen prøver og forsøke å estimere en vanskegrad til en endelig prøve. Det gjør også at vi får flere oppgaver i den siste utprøvingen som ikke fungerer så godt psykometrisk. Vi har prøvd å møte denne utfordringen med å ha tre prøvesett i stedet for to i siste utprøving.

Opgavene i prøven har stor bredde både når det gjelder vanskegrad, matematiske tema og kontekster. Prøven inneholder oppgaver som er relevante for mange fag. Dessuten er det en god blanding av oppgaver hvor elevene skal fylle inn svar eller velge svaralternativer.

Vi har 4 interaktive oppgaver i kohortprøven. Vi skulle gjerne hatt flere interaktive oppgaver, men det er også oppgaver som i tillegg til å fungere faglig må fungere psykometrisk. I og med at en utprøving falt ut, ble utvalget av interaktive oppgaver som fungerte godt noe mindre enn tidligere.

En prøve med 45 oppgaver er nokså omfattende, og vi må nok erkjenne at utholdenheten også testes i en prøvesituasjon. Mange elever på 5. trinn strever med å holde konsentrasjonen oppe gjennom en prøve med 45 oppgaver. Dette har vi også observert når vi har vært til stede i ulike klasserom ved utprøvinger. Det bør absolutt vurderes å gå ned til 40 oppgaver. Vi ser også at andelen ubesvart øker nokså kraftig mot slutten av prøven. Her er også kjønnsforskjellene tydelige. Det er langt flere jenter enn gutter i andelen ubesvarte oppgaver, spesielt i de 5-10 siste oppgavene i prøven.

I og med at vi høsten 2020 gikk i gang med en ny læreplan, kan det være et gunstig tidspunkt å redusere antallet oppgaver i prøven for 5. trinn fra og med 2022. Foruten ønsket om lydstøtte til elevene i gjennomføringen, er det opplevelsen av for mange oppgaver vi får de fleste tilbakemeldinger på fra lærere som har gjennomført prøven med sine elever. Dette kommer også tydelig fram i samtalen vi har med lærerne som gjennomfører en ekstern validering av prøvene. Prøven 2021 er den siste prøven basert på LK06. I og med det kom ny læreplan høsten 2020, ny prøveplattform og nytt ankerdesign fra 2022, mener vi det kan være et gunstig tidspunkt for å gjøre endringer i prøven generelt og antall oppgaver spesielt.

Kapittel 5: Analyse av nasjonal prøve i regning 2021

Innledning

Dette kapitlet er en analyse av resultatene til elevene som gjennomførte nasjonal prøve i regning på bokmål eller nynorsk for 5. trinn høsten 2021. Det var 62 167 elever på 5. trinn som gjennomførte prøven, fordelt på 58 588 på kohortprøven og 3 579 på ankerprøven. Tallene det henvises til er basert på en IRT-analyse av resultatene fra kohortprøven og ankerprøven, der parameterne til ankeroppgavene er fastsatt på forhånd slik at det sikrer at kohortoppgavene kommer på samme skala som oppgavene fra tidligere år. Oppgavenes vanskegrad gir et estimat av ferdighet som så er omregnet til skalapoeng. I praksis betyr dette at elever som løser oppgaver med høyere vanskegrad oppnår flere skalapoeng enn elever som løser oppgaver med lavere vanskegrad.

Det er tre mestringsnivå for prøven på 5. trinn, der mestringsnivå 3 er det høyeste nivået. Grensene til mestringsnivåene er faste fra år til år, og følgende grenser gjelder for nasjonal prøve i regning for 5. trinn:

Mestringsnivå 1:	42 eller færre skalapoeng
Mestringsnivå 2:	43–56 skalapoeng
Mestringsnivå 3:	57 eller flere skalapoeng

Psykometri

Tabell 18 viser de tekniske verdiene til kohortprøven og ankerprøven.

NP5 2021	Kohortprøve ¹	Ankerprøve
Antall gutter	29720	1839
Antall jenter	28868	1740
Totalt antall elever	58588	3579
Gj.snitt Theta gutter	0,162	0,060
Gj.snitt Theta jenter	-0,191	-0,201
Gj.snitt Theta totalt	-0,012	-0,067
Standardavvik Theta	0,958	0,940
SE til gjennomsnitt Theta	0,004	0,016
Gj.snitt løsningsprosent gutter	57,2 %	54,8 %
Gj.snitt løsningsprosent jenter	49,9 %	47,9 %
Gj.snitt løsningsprosent totalt	53,6 %	51,4 %
Svarprosent	95,3 %	96,1 %
Cronbach alpha	0,919	0,864
Antall item	45	45
Antall flervalgsoppgaver (MC)	24	26
Antall interaktive oppgaver (IA)	4	4

Tabell 18: Psykometriske verdier til kohortprøven og ankerprøven 2021

¹ Analysen er basert på alle elever som gjennomførte NPREG05, også de som tok ankerprøven, men ikke alle oppgavene i kohortprøven

Oversikt over oppgaver 2021

Oppgaver	Innhold	Område	Format	Tilknytning til fag	Løsningsprosent alle	Løsningsprosent jenter	Løsningsprosent gutter	Prosentpoeng forskjell J - G	b-verdi	Mestringsnivå (IRT)
1	Posisjonssystemet. Plassverdier	T	Åpen	Mat, Saf	0,76	0,69	0,83	-0,14	-1,052	1
2	Tolke og lese tabell	S	Flervalg	Mat, Saf, Nat	0,59	0,56	0,62	-0,06	-0,408	2
3	Egenskaper til mønstre	M&G	Flervalg	Mat, Kro, Saf	0,80	0,82	0,78	0,04	-1,867	1
4	Multiplikasjon. Hele tall	T	Åpen	Mat	0,70	0,66	0,73	-0,07	-1,134	1
5	Velge regnearter. Sammensatt	T	Flervalg	Mat, Kro	0,66	0,61	0,70	-0,09	-0,673	1
6	Tid. Tidsdifferanse	M&G	Åpen	Mat, Mhe	0,68	0,59	0,77	-0,18	-0,759	2
7	Speiling om en linje	M&G	Flervalg	Mat, Khv, Krl	0,42	0,43	0,42	0,01	0,514	2
8	Tolke tabell. Bruke tall i tabell	S	Flervalg	Mat, Saf	0,54	0,53	0,56	-0,03	-0,201	2
9	Egenskaper til figurer	M&G	Flervalg	Mat, Khv	0,71	0,71	0,71	0,00	-1,348	2
10	Vurdere rimeligheten av svar	T	Flervalg	Mat	0,70	0,69	0,71	-0,02	-0,928	1
11	Tid. Beregne tid	M&G	Flervalg	Mat, Nat, Kro	0,30	0,24	0,35	-0,12	0,943	3
12	Tolke tabell. Bruke tall i tabell	S	Flervalg	Mat, Saf	0,49	0,47	0,51	-0,04	0,043	2
13	Tid. Tidsintervall	M&G	Flervalg	Mat, Nat	0,73	0,75	0,70	0,04	-1,346	1
14	Divisjon. Halvering	T	Flervalg	Mat, Saf	0,43	0,39	0,46	-0,06	0,361	2
15	Addisjon. Desimaltall	T	Åpen	Mat	0,56	0,49	0,63	-0,13	-0,266	2
16	Vekt. Subtraksjon. Desimaltall	M&G	Flervalg	Mat, Mhe	0,55	0,45	0,65	-0,21	-0,218	2
17	Subtraksjon. Hele tall.	T	Åpen	Mat, Mhe, Saf	0,51	0,51	0,51	-0,01	-0,061	2
18	Divisjon. Hele tall	T	Åpen	Mat, Nat, Mhe	0,67	0,63	0,72	-0,09	-0,800	1
19	Velge regnearter. Sammensatt	T	Åpen	Mat, Khv	0,10	0,07	0,13	-0,06	2,010	3
20	Tallinje. Desimaltall	T	Åpen	Mat, Khv	0,65	0,56	0,73	-0,18	-0,574	2
21	Tolke og lese diagram	S	Åpen	Mat, Saf, Nat	0,59	0,53	0,66	-0,12	-0,439	2
22	Divisjon. Halvering	T	Åpen	Mat	0,75	0,72	0,77	-0,04	-0,920	1
23	Areal. Regne areal ved hjelp av	M&G	Flervalg	Mat, Khv	0,48	0,46	0,51	-0,05	0,070	2
24	Multiplikasjon. Hele tall	T	Åpen	Mat, Mhe	0,48	0,42	0,54	-0,11	0,060	2
25	Kart og koordinatsystem	M&G	Flervalg	Mat, Kro, Saf	0,94	0,94	0,94	-0,01	-2,659	1
26	Tolke og lese tabell. Utføre	S	Flervalg	Mat, Mhe	0,30	0,25	0,35	-0,09	0,835	3
27	Lengde. Forståelse av enheter	M&G	Åpen	Mat, Kro	0,67	0,59	0,75	-0,16	-0,623	2
28	Speiling om en linje	M&G	Åpen	Mat, Khv, Krl	0,27	0,29	0,26	0,03	1,138	3
29	Geometriske figurer. Egenskaper	M&G	Flervalg	Mat, Khv	0,67	0,62	0,72	-0,10	-0,923	2
30	Velge regnearter. Sammensatt	T	Flervalg	Mat, Saf	0,51	0,47	0,55	-0,08	-0,077	2
31	Lengde. Omgjøring mellom	M&G	Flervalg	Mat, Khv	0,24	0,16	0,31	-0,16	1,138	3
32	Brøk. Forståelse av brøk	T	Flervalg	Mat, Khv, Mhe	0,73	0,73	0,73	0,01	-1,051	2
33	Multiplikasjon. Hele tall	T	Åpen	Mat, Mus	0,55	0,53	0,58	-0,05	-0,218	2
34	Vekt. Omgjøring mellom	M&G	Åpen	Mat, Mhe	0,16	0,08	0,23	-0,15	1,460	3
35	Tolke og lese diagram. Utføre	S	Flervalg	Mat, Saf, Nat	0,42	0,38	0,45	-0,07	0,414	3
36	Tid. Beregne tid	M&G	Flervalg	Mat, Mhe	0,37	0,33	0,40	-0,07	0,470	3
37	Velge regnearter. Sammensatt	T	Åpen	Mat	0,36	0,33	0,39	-0,06	0,612	3
38	Tid. Beregne tid	M&G	Flervalg	Mat, Nat	0,68	0,66	0,70	-0,04	-0,918	2
39	Multiplikasjon. Hele tall	T	Åpen	Mat	0,58	0,56	0,60	-0,04	-0,322	2
40	Tolke og lese diagram. Utføre	S	Flervalg	Mat, Krl, Saf, Nat	0,27	0,23	0,31	-0,09	1,049	3
41	Addisjon. Hele tall	T	Åpen	Mat, Kro	0,76	0,73	0,80	-0,06	-1,297	1
42	Kart. Beregne avstand	M&G	Åpen	Mat, Kro, Saf	0,30	0,25	0,35	-0,10	0,822	3
43	Lese av tabell. Tolke diagram	S	Flervalg	Mat, Krl, Saf, Nat	0,73	0,70	0,76	-0,07	-0,999	2
44	Subtraksjon. Hele tall	T	Åpen	Mat	0,20	0,17	0,24	-0,07	1,202	3
45	Multiplikasjon. Hele tall	T	Åpen	Mat, Mhe	0,55	0,49	0,61	-0,12	-0,200	2
Gjennomsnitt					0,54	0,50	0,57	-0,07	-0,20	

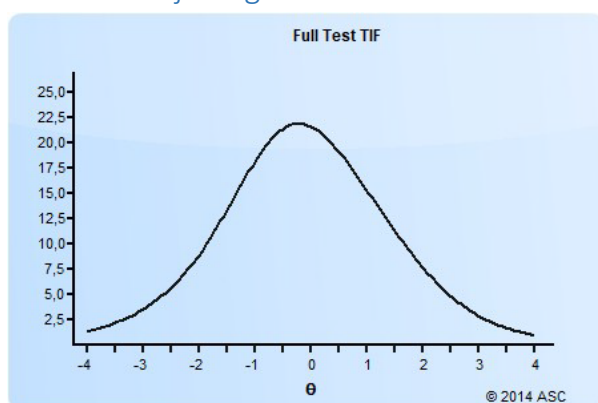
Tabell 19 Oppgaverekkefølge i versjon 1 av Nasjonal prøve i regning, NP5 2021, N = 58 588

Tabell 22 viser rekkefølgen til oppgavene i versjon 1 av kohortprøven, hva oppgavene måler, fagområde, oppgaveformat, oppgavens tilknytning til fag, løsningsprosent for alle og for jenter og gutter hver for seg, kjønnsforskjeller i løsningsprosent, oppgavens vanskegrad (b-verdi ut fra IRT-analyse) og mestringsnivå.

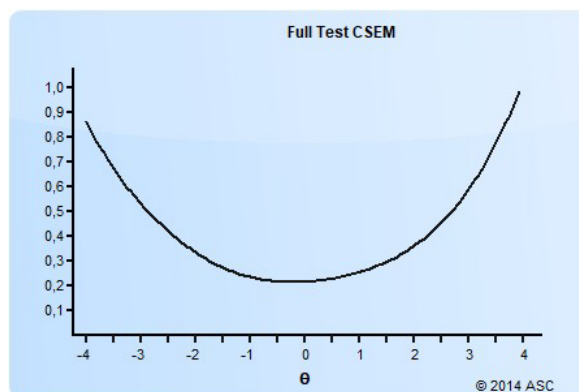
Nasjonal prøve i regning for 5. trinn måler elevenes kompetanse etter 4. trinn i grunnleggende ferdigheter i å kunne regne. Derfor må valideringen skje ut fra det en kan forvente av elever som nettopp har begynt i 5. trinn.

En lærergruppe med fire personer har gjennomført ekstern validering av prøvesettet. Det vil si at de har vurdert hver enkelt oppgave og prøvesettet som helhet opp mot kravene til validitet som står i rammeverket. Denne prosessen er nærmere beskrevet i kapittel 4. Endelig godkjenning av prøven ble gjort av Utdanningsdirektoratet og deres egne interne og eksterne kvalitetssikrere.

Testinformasjon og måleusikkerhet



Figur 14 Testinformasjon NP5 2021



Figur 13 Måleusikkerhet NP5 2021

Prøven har stor informasjonsverdi med lav måleusikkerhet for elever som ligger midt på ferdighetsskalaen og innen ett standardavvik resultatmessig (figur 14). Maksimum testinformasjon var 21,9 ved theta verdi $-0,200$. Måleusikkerheten er imidlertid større for elever i begge ender av ferdighetsskalaen (figur 13). Dette er imidlertid en naturlig konsekvens av at det er flest oppgaver i prøven som ligger midt på ferdighetsskalaen, som igjen er en konsekvens av hvordan elevene fordeler seg utover ferdighetsskalaen. Minimum CSEM var 0,21 ved theta verdi $-0,200$.

Tolkninger av resultatene

En enkeltoppgave kan inneholde elementer fra både tall, måling og geometri og statistikk, og det kan derfor være vanskelig å plassere oppgaver i ett bestemt fagområde. I alle oppgavene behandler vi tall i en eller annen sammenheng. Det typiske for måling og geometri er imidlertid behandling av enheter for masse, lengde og volum, beregning av areal, kjøp og salg, egenskaper til geometriske figurer og regning med temperatur og tid. Statistikk handler om å lage og lese tabeller og diagram og gjøre beregninger. Av den grunn vil både tallbehandling og riktig bruk av enheter også ofte inngå i statistikkoppgaver.

Alle områdene har oppgaver med ulik vanskegrad og på alle tre mestringsnivåer. I preanalysen før prøven hadde vi flere oppgaver som ut fra vanskegrad burde havne på høyeste mestringsnivå, men som i gjennomføringen ble noe enklere enn det verdiene fra utprøvingene viste. Oppgaver som etter utprøvinger så vidt er på mestringsnivå 3, har en tendens til å havne på mestringsnivå 2. Tilsvarende

kan oppgaver lavt på mestringsnivå 2 havne på mestringsnivå 1 etter gjennomføring. Dette prøver vi å ta med i vurderingen når den endelige prøven settes sammen.

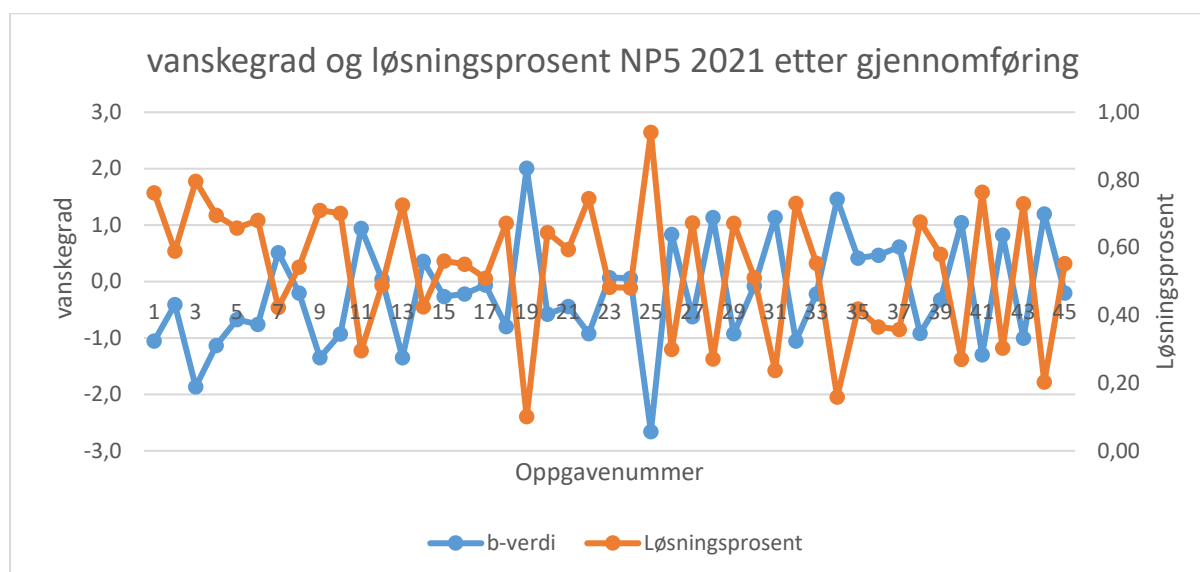
Oppgaver fra fagområder	Mestringsnivå 1	Mestringsnivå 2	Mestringsnivå 3	Totalt
Tall	8	10	3	21
Måling og geometri	4	6	5	15
Statistikk	3	5	1	9
Totalt	15	21	9	45

Tabell 20 Fordeling av oppgaver på mestringsnivå etter gjennomført prøve

Tidligere år har løsningsprosenten for prøven for 5. trinn gjerne endret seg i størrelsesorden 4–6 prosentpoeng fra siste utprøving til endelig prøve. I 2020 var endringen 4,0 prosentpoeng, mens den i årets prøve var 5,2 prosentpoeng. Som vi tidligere har vært inne på, er inntrykket at en endelig prøve blir behandlet mer seriøst enn en utprøving. Det viser blant annet svarprosenten som er 4,1 prosentpoeng høyere i den endelige gjennomføringen enn den var i den siste utprøvingen.

Resultater til enkeltoppgaver

Den gjennomsnittlige løsningsprosenten for hele prøven er 53,6 prosent. Det betyr at elevene i gjennomsnitt løste 53,6 prosent av alle oppgavene riktig. Oppgaven med høyest løsningsprosent er oppgave 25, som ble løst riktig av 94 prosent av elevene. Oppgave 19 har en løsningsprosent på 10, og er oppgaven med lavest løsningsprosent i prøven.



Figur 15 vanskegrad (b-verdi) og løsningsprosent for hver oppgave

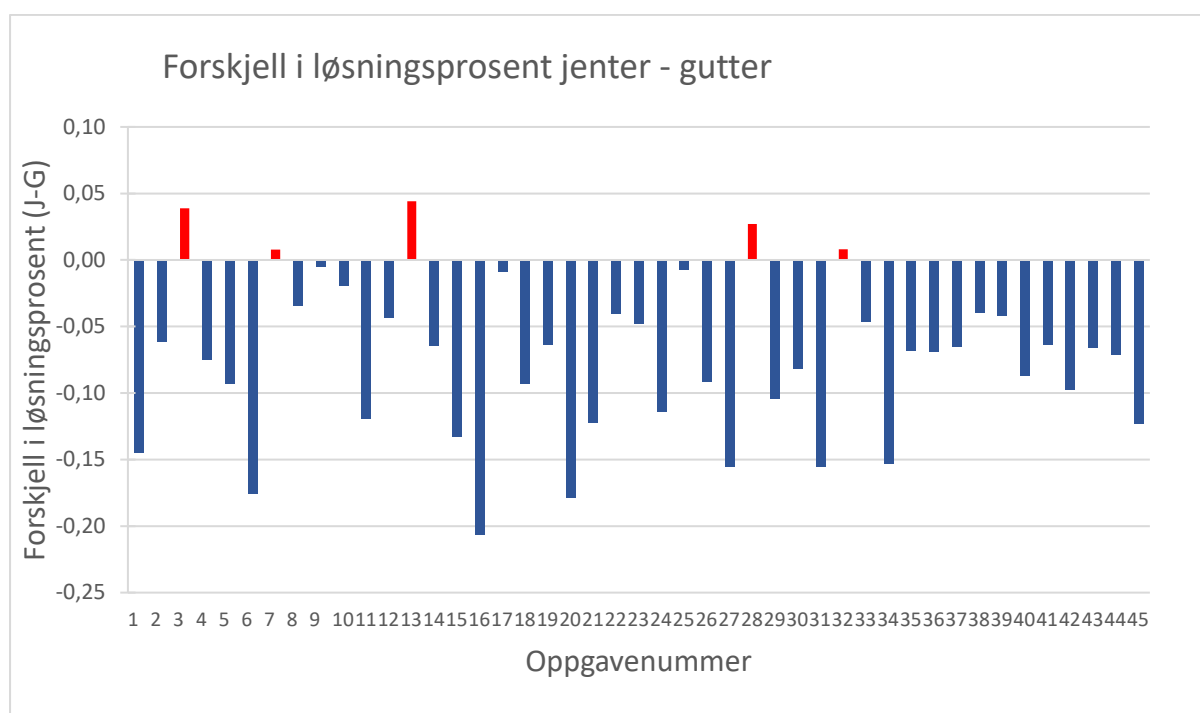
Figur 15 viser hvordan oppgavene i prøven fordeler seg ut fra vanskegrad og løsningsprosent. Vi ser at vanskegrad og løsningsprosent speiler hverandre og viser, som forventet, stort samsvar mellom oppgavens vanskegrad og løsningsprosent.

Kjønnsforskjeller i årets prøve

Guttene gjør det bedre enn jentene på prøven som helhet. Den gjennomsnittlige kjønnsforskjellen i løsningsprosent er 7,3 prosentpoeng i favør gutter i prøven for 2021. Forskjellen mellom kjønnene er større enn den var i 2020 da forskjellen var 4,1 prosentpoeng. Resultatene til de nasjonale prøvene i regning viser hvert år signifikante forskjeller i guttefavør. Vi ser også at det er de samme

oppgavetyperne som hvert år viser forskjeller, og forskjellene er nokså stabile. I utprøvningsprosessen ser vi at oppgavekontekst har lite å si for om oppgaven går i jente- eller guttefavør.

I likhet med tidligere år, ser vi en klar tendens til at guttene i gjennomsnitt presterer bedre enn jentene i oppgaver som krever omgjøring mellom de mest kjente enheter i masse, lengde og volum. Dette gjelder både rene omgjøringsoppgaver og oppgaver der konteksten kan være en hjelp til å vurdere størrelse eller mengde. Resultatene viser også at oppgaver som inneholder desimaltall uten at det nødvendigvis er omgjøring mellom enheter, er oppgaver som har en tendens til å gå i guttefavør. Oppgaven med størst kjønnsforskjell i løsningsprosent i årets prøve, 21 prosentpoeng i guttefavør, er oppgave 16 som er en subtraksjonsoppgave med desimaltall og som handler om massen til en katt. Her er det ingen omgjøring mellom enheter, og oppgavetekster som handler om dyr mener vi ikke skal favorisere ett av kjønnene. Oppgaven er omtalt tidligere i rapporten.



Figur 16 Forskjell i løsningsprosent jenter-gutter. Favør jenter røde søyler, favør gutter blå søyler

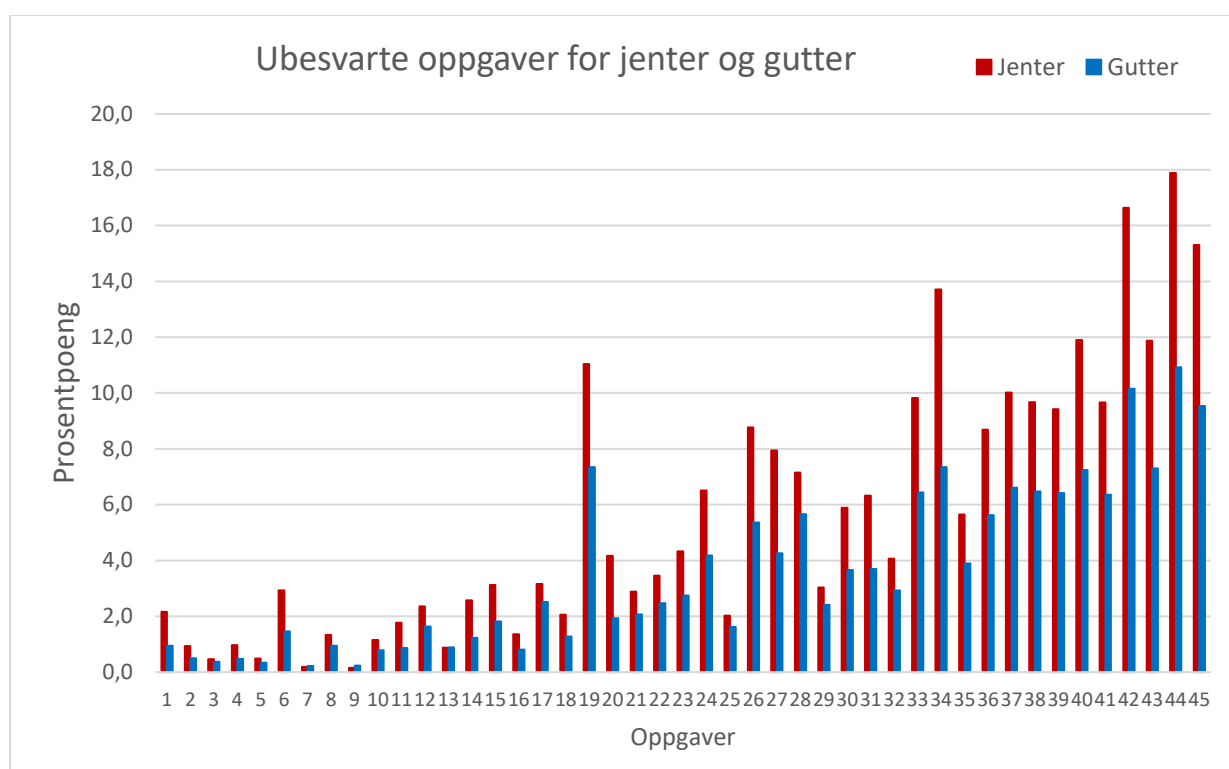
Når det gjelder oppgaver i jentefavør, ser vi av diagrammet i figur 16 at det kun er fem oppgaver der løsningsprosenten hos jentene er høyere enn hos guttene. Diagrammet viser også at forskjellene er større i oppgaver som går i guttefavør enn de som går i jentefavør. Oppgave 13 har en kjønnsforskjell på 4 prosentpoeng og er oppgaven med størst løsningsprosent i favør jenter. Det er en flervalgsoppgave der eleven skal bestemme ukedag ut fra gitte datoer.

Kjønnsforskjellen i løsningsprosent har variert noe siden 2014. Tendensen har hele tida vært at gutter har en høyere løsningsprosent enn jenter. Valget av oppgaver vil ha betydning for kjønnsforskjeller, men valg av oppgaver ut fra kjønnsforskjeller kan ikke overstyre validiteten til prøven. Det er mulig å lage en prøve med både mindre og større kjønnsforskjeller. Men så lenge vi skal lage en prøve med bredde både i innhold, ferdighetsområder og vanskegrad, med variasjon i formater og kontekster og der den enkelte oppgave og prøven som helhet skal svare på psykometriske krav, er dette utfordringer som ikke alltid lar seg forene.

Ubesvarte oppgaver

I gjennomsnitt er 4,7 prosent av oppgavene i prøven ubesvart. Det betyr at gjennomsnittlig svarprosent på årets prøve er høyere enn i fjor da andelen ubesvart var 6,1 prosent. Årets prøve var også noe lettere enn fjorårets hvis vi sammenligner gjennomsnittlig b-verdi.

Ikke overraskende er det større andel ubesvart blant kortsvarsoppgavene enn blant flervalgsoppgavene. Diagrammet viser at andelen ubesvart øker utover i prøven, og at andelen ubesvart er høyere blant jentene enn blant guttene i nesten alle oppgaver (figur 17). Det er kun tre oppgaver der andelen ubesvart er marginalt høyere blant guttene enn jentene. Det er litt bekymringsfullt at andelen ubesvart har utviklet seg og stadig blitt større, og at forskjellen mellom kjønnene er såpass tydelig som den er. Dette vil også påvirke løsningsprosenten i oppgavene, spesielt utover i prøven.

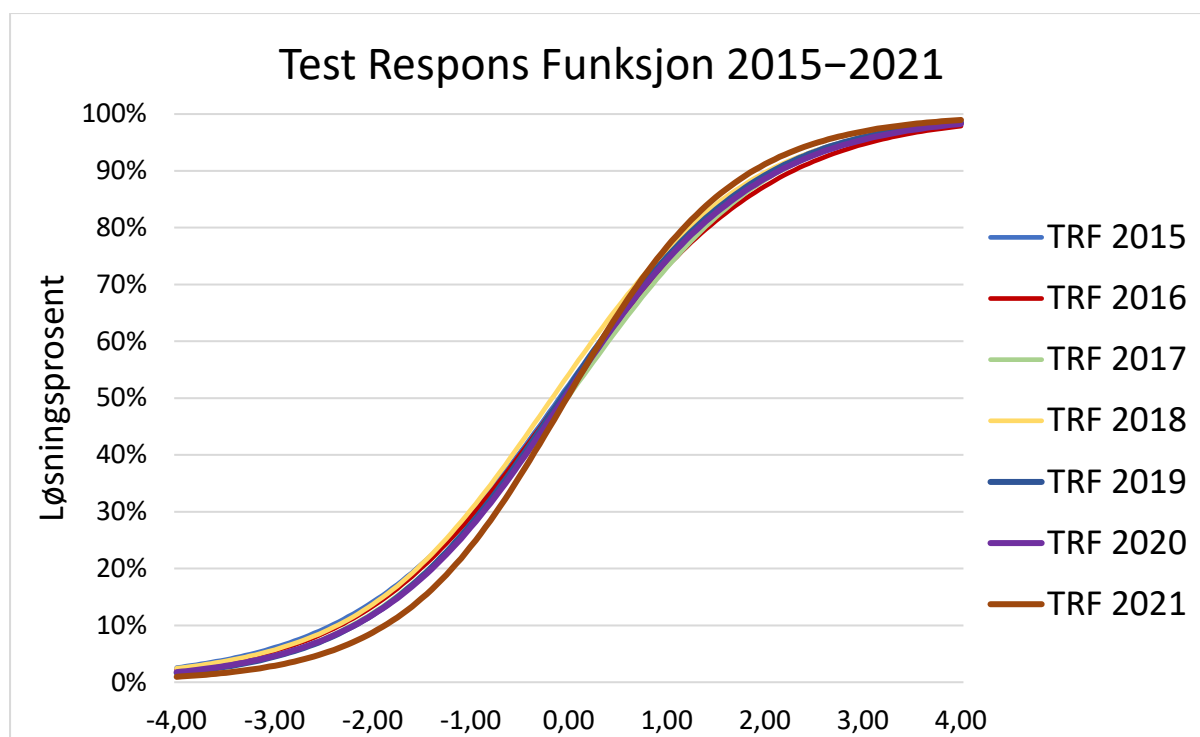


Figur 17 Prosent jenter og gutter som ikke har svart på oppgavene. Rødt for jenter og blått for gutter. Oppgavenummer langs horisontal akse, prosent langs vertikal akse

Sammenligning av prøvene over tid

Fra 2014 har prøvene vært laget på en måte som gjør det mulig å måle utvikling over tid. Det er seks prosent av elevene som gjennomfører en ankerprøve hvert år. Omtrent halvparten av oppgavene i ankerprøven har vært med i tidligere ankerprøver, mens resten av oppgavene er identiske med oppgaver fra den ordinære prøven. Ankeroppgavene kan ikke offentliggjøres, og elevene som tar ankerprøven, får ikke vite hvordan oppgavene er vurdert.

2014 er år null for vurdering av utvikling over tid. Ved hjelp av ankeroppgavene er det mulig å sammenligne resultatene for 2021 med de sju foregående år og vurdere om elevenes ferdighet har endret seg. Sju år er fortsatt kort tid til å kunne trekke noen klare konklusjoner, og endringene er nokså små. Prøven for 5. trinn i 2021 har samme nasjonale gjennomsnitt som prøvene fra og med 2014, 50 skalapoeng.



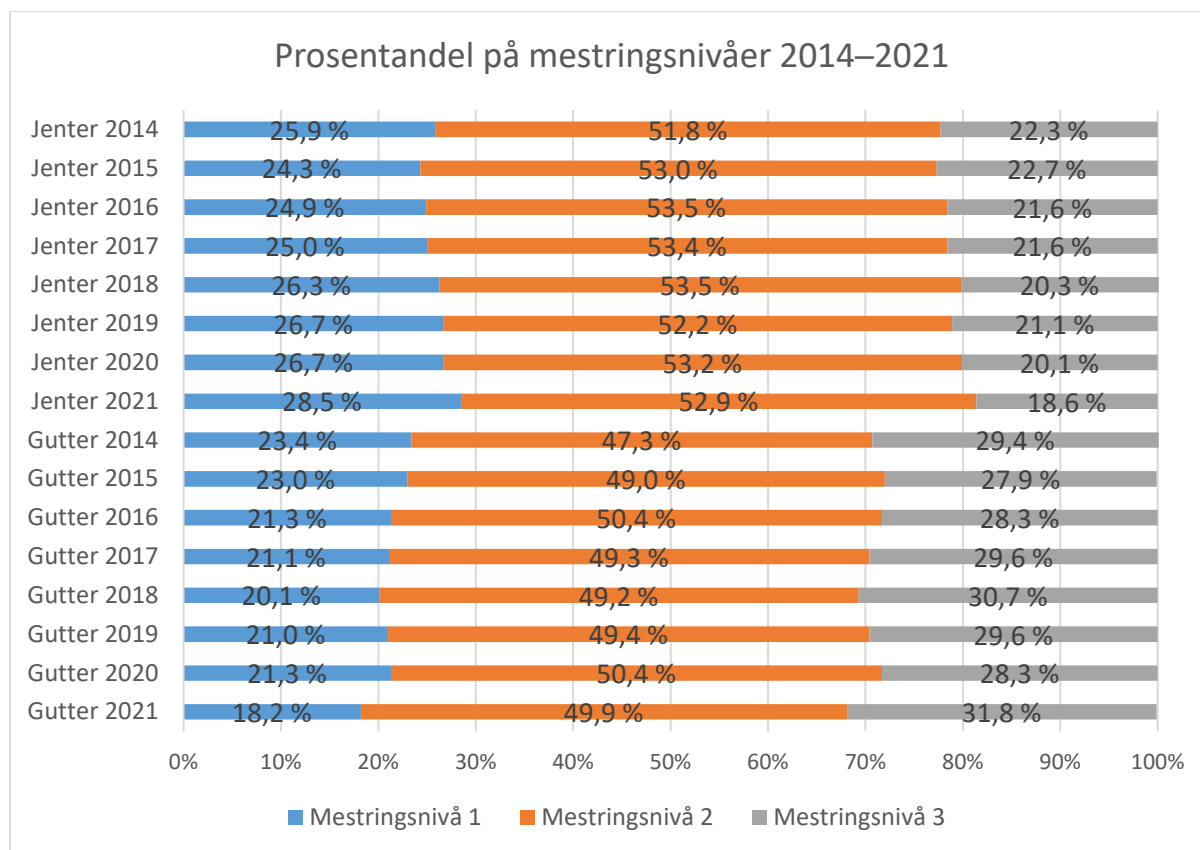
Figur 18 Test-respons funksjonen for NP5 2015-2020

Test respons funksjonen (TRF) viser sammenhengen mellom ferdigheten til elevene med en theta verdi mellom -4 og 4, og prosentandel riktige besvarte oppgaver på prøven. Theta verdien ligger til grunn for skalapoengene som hver elev får. Diagrammet i figur 18 viser en sammenligning mellom de sju siste årene. Vi ser at prøvene er nokså like i vanskegrad. Det er kun små forskjeller fra år til år. Dette må vi også se som et resultat av de grundige utprøvinger hver enkelt oppgave har vært gjennom.

Fordeling på mestringsnivåer 2014–2020

Elevene er plassert på tre mestringsnivåer etter oppnådde skalapoeng. Diagrammet i figur 19 viser at det hvert år er prosentvis flere jenter enn gutter på mestringsnivå 1 og 2, hhv. 10,2 og 3,0 prosentpoeng i 2021, mens det er omvendt på mestringsnivå 3 med 13,2 prosentpoeng flere gutter enn jenter. 31,8 prosent av guttene havner på mestringsnivå 3, mens dette gjelder 18,6 prosent av jentene. Tallene fra 2014 til 2021 viser at forskjellen mellom prosentandelen jenter og gutter på mestringsnivå 3 helt siden 2014 har vært i guttefavør, men at forskjellen i 2021 er større enn tidligere år.

Tilsvarende er andelen jenter på mestringsnivå 1 større enn andelen gutter på dette mestringsnivået. Figur 19 viser utviklingen for begge kjønn gjennom de sju siste årene.



Figur 19 Prosentandel jenter og gutter på mestringsnivåer 2014 – 2021

Vurdering av prøven

Prøven for 2021 totalt sett viser ingen store endringer i resultatene sammenlignet med tidligere år, men det mest påfallende er at vi ser at kjønnsforskjellene ser ut til å øke. Det er oppgaver innen de samme regneområder som år etter år gir kjønnsforskjeller, ganske uavhengig av selve konteksten. Valg av oppgaver har selvsagt en betydning, men vi ser en klar tendens til at antallet oppgaver som går i guttefavør er flere og gir større forskjeller mellom kjønnene enn oppgaver som går i jentefavør. Vi finner ingen umiddelbar årsak til dette, men det er jo et tema det hadde vært spennende og studert litt nærmere. Vi ser jo også en tydelig tendens til at guttene svarer på flere oppgaver enn jentene. Spesielt øker andelen ubesvart mot slutten av prøven, og også der er forskjellen mellom jenter og gutter ganske tydelig.

En prøve med 45 oppgaver er nokså omfattende, og vi må erkjenne at utholdenheten også testes i en prøvesituasjon. Elever på 5. trinn strever med å holde konsentrasjonen oppe gjennom en prøve med såpass mange oppgaver. Fra og med 2022 blir antallet oppgaver i prøven redusert slik at det vil bli 35 oppgaver for 5. trinn. Vi håper at dette kan redusere andelen ubesvart, og at flere opplever at de får tid nok til å svare på oppgavene innen 90 minutter, som fortsatt er prøvetiden.

Prøven for 2022 vil være første prøve etter ny læreplan og samtidig en prøve som skal gjennomføres i nytt prøvesystem. Selv om den grunnleggende ferdigheten å kunne regne ikke er veldig endret i ny læreplan, vil nok prøven oppleves litt annerledes. Det er også lovet lesestøtte til elevene i gjennomføringen, noe vi foreløpig ikke har hatt erfaring med i utprøvinger, men som mange lærere har etterspurt i flere år. Både færre oppgaver og lesestøtte vil vi tro blir tatt godt imot av lærerne. Det blir ekstra spennende å se om endringene i prøven for 2022 vil gi noen endringer i resultatene etter gjennomføring, også med tanke på kjønnsforskjeller.

Oppgaveanalyse NP5 2021

NP5 2021		Svarfordeling i prosent						Ferdighet (Theta means), uankrede verdier					Point biserial (Trpbis)	IRT-analyse		Løsningsprosent	Forskjell i løsningsprosent	Signifikans 5%	
Nr.	FV Å	A 0	B 1	C	D	E	Ubesv art	A 0	B 1	C	D	E		Ubesv art	a- verdi				b- verdi
1	Å	0,22	0,76				0,02	-0,84	0,38				-0,77	0,516	0,96	-1,052	0,76	-0,14	J
2	FV	0,08	0,59	0,14	0,17		0,01	-0,13	0,51	-0,49	-0,71		-0,46	0,485	0,71	-0,408	0,59	-0,06	G
3	FV	0,20	0,80				0,00	-0,55	0,25				-0,29	0,315	0,49	-1,867	0,80	0,04	G
4	Å	0,30	0,70				0,01	-0,45	0,33				-0,72	0,359	0,50	-1,134	0,70	-0,07	J
5	FV	0,17	0,07	0,10	0,66		0,00	-0,70	-0,64	-0,46	0,46		-0,35	0,502	0,79	-0,673	0,66	-0,09	
6	Å	0,30	0,68				0,02	-0,63	0,43				-0,56	0,505	0,82	-0,759	0,68	-0,18	J
7	FV	0,18	0,33	0,07	0,42		0,00	0,08	-0,29	-0,28	0,45		-0,24	0,300	0,37	0,514	0,42	0,01	G
8	FV	0,10	0,12	0,54	0,22		0,01	-0,10	-0,44	0,55	-0,62		-0,32	0,481	0,69	-0,201	0,54	-0,03	G
9	FV	0,06	0,11	0,71	0,11		0,00	-0,34	-0,54	0,30	-0,37		-0,59	0,322	0,44	-1,348	0,71	0,00	G
10	FV	0,16	0,70	0,08	0,05		0,01	-0,41	0,39	-0,96	-0,75		-0,33	0,457	0,71	-0,928	0,70	-0,02	G
11	FV	0,08	0,30	0,10	0,51		0,01	-0,12	0,78	0,03	-0,26		0,01	0,445	0,67	0,943	0,30	-0,12	J
12	FV	0,49	0,49				0,02	-0,31	0,52				-0,54	0,423	0,57	0,043	0,49	-0,04	G
13	FV	0,13	0,06	0,08	0,73		0,01	-0,29	-0,67	-0,61	0,30		-0,50	0,345	0,49	-1,346	0,73	0,04	G
14	FV	0,16	0,43	0,09	0,30		0,02	-0,44	0,57	-0,54	-0,11		-0,08	0,419	0,57	0,361	0,43	-0,06	
15	Å	0,41	0,56				0,03	-0,47	0,54				-0,58	0,510	0,76	-0,266	0,56	-0,13	J
16	FV	0,55	0,26	0,13	0,05		0,01	0,56	-0,44	-0,54	-0,59		-0,37	0,531	0,81	-0,218	0,55	-0,21	J
17	Å	0,46	0,51				0,03	-0,35	0,53				-0,48	0,444	0,61	-0,061	0,51	-0,01	G
18	Å	0,31	0,67				0,02	-0,56	0,41				-0,62	0,453	0,68	-0,800	0,67	-0,09	J
19	Å	0,81	0,10				0,09	-0,02	1,21				-0,19	0,375	0,84	2,010	0,10	-0,06	J
20	Å	0,32	0,65				0,03	-0,63	0,48				-0,40	0,543	0,89	-0,574	0,65	-0,18	J
21	Å	0,38	0,59				0,03	-0,45	0,47				-0,58	0,465	0,67	-0,439	0,59	-0,12	J
22	Å	0,22	0,75				0,03	-0,85	0,40				-0,63	0,557	1,08	-0,920	0,75	-0,04	G
23	FV	0,23	0,48	0,16	0,09		0,04	-0,61	0,53	-0,07	-0,04		-0,31	0,436	0,59	0,070	0,48	-0,05	G
24	Å	0,46	0,48				0,06	-0,42	0,64				-0,47	0,548	0,85	0,060	0,48	-0,11	J
25	FV	0,04	0,94				0,02	-0,79	0,14				-0,45	0,277	0,76	-2,659	0,94	-0,01	
26	FV	0,10	0,13	0,28	0,12	0,30	0,07	-0,54	-0,43	-0,13	0,02	0,83	-0,18	0,492	0,78	0,835	0,30	-0,09	
27	Å	0,27	0,67				0,06	-0,76	0,48				-0,49	0,600	1,14	-0,623	0,67	-0,16	J
28	Å	0,66	0,27				0,06	-0,15	0,75				-0,23	0,410	0,61	1,138	0,27	0,03	G
29	FV	0,67	0,10	0,09	0,12		0,03	0,34	-0,66	-0,60	-0,10		-0,36	0,390	0,55	-0,923	0,67	-0,10	J
30	FV	0,14	0,17	0,51	0,13		0,05	-0,12	-0,37	0,42	-0,24		-0,30	0,373	0,48	-0,077	0,51	-0,08	
31	FV	0,24	0,22	0,26	0,24		0,05	0,06	-0,54	-0,04	0,91		-0,24	0,470	0,79	1,138	0,24	-0,16	J
32	FV	0,11	0,07	0,06	0,73		0,04	-0,63	-0,61	-0,62	0,34		-0,33	0,455	0,73	-1,051	0,73	0,01	G
33	Å	0,36	0,55				0,08	-0,52	0,57				-0,42	0,570	0,93	-0,218	0,55	-0,05	G
34	Å	0,74	0,16				0,10	-0,08	1,13				-0,30	0,467	0,94	1,460	0,16	-0,15	J
35	FV	0,16	0,42	0,14	0,23		0,05	-0,42	0,53	-0,07	-0,16		-0,28	0,400	0,53	0,414	0,42	-0,07	
36	FV	0,56	0,37				0,08	-0,34	0,83				-0,27	0,587	1,02	0,470	0,37	-0,07	G
37	Å	0,55	0,36				0,09	-0,24	0,69				-0,26	0,476	0,70	0,612	0,36	-0,06	
38	FV	0,02	0,02	0,68	0,08	0,04	0,08	-0,81	-0,96	0,31	-0,34	-0,37	-0,26	0,403	0,58	-0,918	0,68	-0,04	
39	Å	0,34	0,58				0,08	-0,53	0,51				-0,32	0,549	0,88	-0,322	0,58	-0,04	G
40	FV	0,27	0,13	0,32	0,18		0,10	0,77	-0,23	-0,04	-0,31		-0,22	0,447	0,69	1,049	0,27	-0,09	J
41	Å	0,16	0,76				0,08	-0,59	0,27				-0,25	0,415	0,67	-1,297	0,76	-0,06	J
42	Å	0,56	0,30				0,13	-0,21	0,78				-0,19	0,490	0,77	0,822	0,30	-0,10	J
43	FV	0,73	0,08	0,06	0,04		0,10	0,32	-0,82	-0,64	-0,56		-0,22	0,480	0,80	-0,999	0,73	-0,07	
44	Å	0,65	0,20				0,14	-0,13	1,03				-0,22	0,502	0,95	1,202	0,20	-0,07	
45	Å	0,32	0,55				0,12	-0,65	0,59				-0,22	0,626	1,14	-0,200	0,55	-0,12	J
Gjennomsnitt ubesvart:							0,05	Gjennomsnitt:					-0,36	0,46	0,73	-0,20	0,54	-0,07	

Tabell 21 Oppgaveanalyse for alle oppgavene i kohortprøven. Svarfordeling angitt i prosent. Gjennomsnittlig ferdighet for hvert svaralternativ og for rett eller galt i de åpne oppgavene. Point-biserial betyr oppgavens korrelasjon med gjennomsnittlig ferdighet. a-verdi betyr oppgavens diskriminering i IRT-analyse, og b-verdien er oppgavens vanskegrad på en skala fra -4 til 4 ut fra IRT-analyse. Guttene har høyest løsningsprosent i de oppgavene hvor differansen (j-g) er negativ.

G) Bias mot gutter, signifikant

J) Bias mot jenter, signifikant

Kilder:

Olsen, R. V. & Björnsson, J. K. (2018). Fødselsmåned og skoleprestasjoner. I R. V. Olsen & J. K. Björnsson (Red.), *Tjue år med TIMSS og PISA i Norge. Trender og nye analyser* (s. 76-93). Oslo: Universitetsforlaget.

Utdanningsdirektoratet. (2006). Læreplanverket. Hentet 24.10 2019 fra <https://www.udir.no/laring-og-trivsel/lareplanverket/>

Utdanningsdirektoratet. (2017). Rammeverk for nasjonale prøver. Hentet 13.12 2019 fra <https://www.udir.no/eksamen-og-prover/prover/rammeverk-for-nasjonale-prover/>

Vinje, B. & Tokle, O. D. (2018). *How can we develop multiple choice items that support assessment for learning?* Innlegg presentert ved AEA-conference Arnhem - Nijmegen, Nederland.