



Matematikksenteret
Nasjonalt senter for matematikk i opplæringen

TEKNISK RAPPORT

NASJONAL PRØVE I REGNING
8. og 9. TRINN 2021

Olav Dalsegg Tokle
Bård Vinje
Stig Atle Myhre

MATEMATIKKSENTERET, NTNU 2022

Nasjonal prøve i regning for 8. (9.) trinn 2021

Olav Dalsegg Tokle, Bård Vinje og Stig Atle Myhre

Nasjonalt senter for matematikk i opplæringen (Matematikksenteret)

NTNU, mars 2022

Innholdsfortegnelse

Innledning.....	4
Kapittel 1: Prepilot.....	5
Kapittel 2: 1. pilot	6
Kapittel 3: 2. pilot	7
HVORDAN 2. PILOTENE BLE SATT SAMMEN.....	7
GJENNOMFØRING AV 2. PILOTENE.....	8
Kapittel 4: Sammensetting av kohortprøve	10
INNLEDNING	10
PSYKOMETRI	10
FORKLARING TIL NOEN OPPGAVER I KOHORTPRØVEN	14
Oppgaver med lave diskrimineringsverdier	15
Oppgaver med bias.....	17
VALIDITET	17
GENERELLE VURDERINGER AV PRØVEN SOM HELHET.....	20
Kapittel 5: Analyse av nasjonal prøve i regning.....	21
INNLEDNING	21
PSYKOMETRISKE VERDIER.....	21
Oversikt over oppgaver NP8 (9) 2021.....	22
Testinformasjon og målesikkerhet.....	23
Test response function	24
TOLKNINGER AV RESULTATENE	25
Sammenligning av resultater over tid	25
Sammenligning av resultater for 8. og 9. trinn	26
Kjønnforskjeller	27
Ubesvarte oppgaver.....	28
OPPGAVEANALYSE FOR ALLE OPPGAVENE	33
OPPSUMMERING	36
Referanser	37

Innledning

Denne rapporten beskriver prosessen for utvikling, gjennomføring og analyse av resultatene til nasjonal prøve i regning for 8. og 9. trinn 2021. Det tok to år å utvikle prøven. I et normalår er oppgavene til nasjonale prøver med i tre utprøvingsrunder, før de blir godkjent til å være med i den nasjonale prøven. På grunn av koronapandemien våren 2020, ble én av utprøvingene avlyst. Oppgavene i prøven for 2021 er derfor blitt prøvd ut to ganger.

De første kapitlene i denne tekniske rapporten beskriver prosessen rundt utprøvingene, og sier litt om formålet med og utfordringer i de ulike utprøvingsfasene. Kapittel 4 beskriver prosessen og begrunner sammensetningen av forslag til endelig prøve, mens kapittel 5 er en analyse og betraktninger rundt resultatene etter at nasjonal prøve for 2021 er gjennomført.

I tillegg er det et konfidensielt vedlegg til denne rapporten som beskriver sammensetningen av ankerprøven.

Gjennom hele utviklingsprosessen brukes klassisk test teori (CTT) og Item Response Theory (IRT) til å analysere resultatene. IRT-analysene er foretatt med Xcalibre (Assessment Systems Corporation, 2014), 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 piloteringene, er trukket av Utdanningsdirektoratet.

Kapittel 1: Prepilot

Første utprøving ble gjennomført i februar 2020 med elever på 8. trinn. Det var åtte oppgavesett med 31 oppgaver i ett sett og 32 oppgaver i resten, i alt 255 oppgaver.

Opgavene vi utvikler skal være regning i fag på fagenes premisser og måle grunnleggende ferdigheter slik dette er beskrevet i Rammeverk for nasjonale prøver (Utdanningsdirektoratet, 2017) og i fagplanene til LK06 (Utdanningsdirektoratet, 2006). En utfordring med å utvikle oppgaver til 2021-prøven er at når den blir gjennomført, vil elevene ha hatt undervisning etter ny læreplan (LK20) i ett år. 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.

Fagtilknytningene til en oppgave er viktig allerede når oppgavene blir produsert. Oppgavene skal være i kontekster som elevene kjenner seg igjen i, og favne alle vanskelighetsgrader innen de matematiske områdene som Rammeverk for nasjonale prøver (Utdanningsdirektoratet, 2017) beskriver. Slik vil oppgavene kunne plasseres på fem mestringsnivåer når endelig prøve er ferdig utviklet. I tillegg skal den enkelte oppgave spisses for å utfordre elevene i hver av delferdighetene i den helhetlige problemløsningsprosessen. Viktige elementer i utviklingen av oppgaver er også «vask» av språk, diskusjoner om bildebruk og kontekst, og om tallene som er valgt, er realistiske og hensiktsmessige ut ifra hva en oppgave skal måle.

I første fase av utviklingen er de fleste oppgavene åpne oppgaver, uten svaralternativer. I den endelige prøven skal en stor andel av oppgavene være flervalgsoppgaver. Hovedhensikten med prepiloten er derfor å finne høyfrekvente elevsvar som viser en typisk feiltenkning. Disse svarene kan være tegn på misoppfatninger. De høyfrekvente feilsvarene blir distraktorer når en åpen oppgave endres til en flervalgsoppgave som skal prøves i neste piloteringsrunde. Når vi velger oppgaver til prøvesettene i en pilot, ønsker vi å gjøre settene så like som mulig når det gjelder innhold, antatt vanskelighetsgrad, arbeidsmengde og oppgaveformat. Fra 2016 har prepilotene inneholdt oppgaver som vi kaller prepilotanker. Dette er oppgaver som har tilfredsstillende psykometriske verdier fra tidligere prøver, og som vi bruker til å sammenligne vanskelighetsgraden til prepilotene. I prepiloten til 2020-prøven er det til sammen ni slike prepilotanker. Åtte oppgaver er fordelt med én i hvert prøvesett, mens én oppgave er felles i alle settene.

Det var til sammen 3275 elever som deltok i prepilot til 2021. Svarprosenten i prøvene varierer fra 86 prosent i prepilot E til 93 prosent i prepilot F. En slik variasjon har vi også sett tidligere år. Ser vi på gjennomsnittet av oppgavenes vanskegrad er ikke prepilot E oppgavesettet med høyest vanskegrad, men det er grunn til å tro at noen utfordrende enkeltoppgaver i settet kan ha påvirket svarprosenten. Se tabell 1 for utvalg og resultater.

Prepiloter NP8 2020	Prepilot A	Prepilot B	Prepilot C	Prepilot D	Prepilot E	Prepilot F	Prepilot G	Prepilot H
Antall gutter	247	193	188	218	189	195	212	210
Antall jenter	207	228	210	202	182	196	193	205
Totalt antall elever	454	421	398	420	371	391	405	415
Gj.snitt Theta gutter	0,037	0,129	-0,076	0,029	-0,055	0,096	0,062	0,228
Gj.snitt Theta jenter	-0,041	-0,095	-0,002	-0,062	-0,050	-0,004	-0,017	-0,093
Gj.snitt Theta totalt	-0,002	0,008	-0,037	-0,015	-0,052	0,046	0,024	0,070
Sign.forskjell J-G (Theta) $p < 0,05$	Nei	Ja	Nei	Nei	Nei	Nei	Nei	Ja
Standardavvik Theta	0,566	0,960	0,923	0,957	0,960	0,942	0,099	0,961
SE til gjennomsnitt Theta	0,027	0,047	0,046	0,047	0,050	0,048	0,005	0,047
Gj.snitt løsningsprosent gutter	44,2 %	47,4 %	49,7 %	44,2 %	38,1 %	47,6 %	46,9 %	41,7 %
Gj.snitt løsningsprosent jenter	41,1 %	42,3 %	51,4 %	41,9 %	38,2 %	45,9 %	44,6 %	35,3 %
Gj.snitt løsningsprosent totalt	42,8 %	44,7 %	50,6 %	43,1 %	38,2 %	46,8 %	45,8 %	38,5 %
Svarprosent	89,6 %	89,9 %	89,5 %	91,3 %	85,5 %	92,8 %	90,5 %	89,8 %
Cronbachs alpha	0,886	0,889	0,870	0,881	0,882	0,881	0,901	0,880
Antall oppgaver	32	31	32	32	32	32	32	32
Antall flervalgsoppgaver (MC)	5	4	8	4	6	8	6	9
Antall interaktive oppgaver (IA)	1	0	2	2	2	2	1	1

TABELL 1 RESULTATER FRA PREPILOT FOR 2021

Kapittel 2: 1. pilot

Vår normale andre utprøving, 1. pilot, skulle egentlig vært gjennomført i april 2020. På grunn av koronapandemien i landet ble det derimot ikke gjennomført 1. pilot.

Kapittel 3: 2. pilot

Hvordan 2. pilotene ble satt sammen

Siden 1. pilot for 2021-prøven ikke ble gjennomført, kunne vi ikke gjennomføre 2. piloten på samme måte som det vi har gjort tidligere.

Normalt setter vi, på bakgrunn av resultatene fra prepilotene og 1. pilotene, sammen to ulike oppgavesett til 2. pilot som hver for seg selv skal tilfredsstillere kravene til en ferdig nasjonal prøve. Det betyr at oppgaverekkefølge, fordeling på mestringsnivå, fordeling på matematisk hovedtema og andel flervalgsoppgaver må være så lik en endelig prøve som mulig. Denne gangen hadde vi bare resultatene fra prepilot som grunnlag, og vi hadde behov for å prøve ut et større antall oppgaver enn normalt, for å sikre at vi ville få nok oppgaver å velge mellom i sammensetning av endelig prøve. Dette resulterte i at vi til denne 2. piloten satte sammen tre ulike oppgavesett.

Aktuelle oppgaver til 2. piloten var oppgaver som fungerte som åpne oppgaver i prepiloten, oppgaver som ble gjort om til flervalgsoppgaver etter prepiloten og enkelte nyutviklede oppgaver. I tillegg hadde vi med oss noen ubrukte, men tidligere utprøvde oppgaver fra tidligere år.

I prosessen med å velge ut oppgaver plukket vi denne gangen først ut ti oppgaver som skulle være felles for alle elevene. Hensikten med dette var å sikre lenking mellom de ulike oppgavesettene. Fellesoppgavene var oppgaver vi håpet skulle fungere godt, og de ble plassert på samme plass i de tre settene. Deretter valgte vi ut 40 oppgaver til hver av prøvene. Da var vi blant annet bevisst på å sikre en relativt lik fordeling av emneområder og flervalgsoppgaver. For enkelte oppgaver hadde vi ulike varianter, og da ble de plassert på samme plass i ulike prøver.

Vi satte sammen tre prøvesett, A, B og C, som hver bestod av 50 oppgaver. Til sammen inneholdt prøvesettene 130 ulike oppgaver. Tabell 4, Tabell 5 og Tabell 6 viser sammensetningen før gjennomføring. Antall oppgaver innenfor hvert område i tabell 4 avhenger av i hvilken grad grunnleggende ferdigheter i å kunne regne og de ulike områdene inngår i kompetansemål for fag i LK 06 (Utdanningsdirektoratet, 2006). Det er færre kompetansemål relatert til området statistikk og sannsynlighet, og dette området er derfor representert med færrest oppgaver i prøvene.

	2. pilot A	2. pilot B	2. pilot C
Tall og algebra	22	19	21
Måling og geometri	19	21	18
Statistikk og sannsynlighet	9	10	11

TABELL 2 FORDELING AV OPPGAVER INNEN OMRÅDER

For mestringsnivåene er ønsket fordeling av antall oppgaver til endelig prøve 5–10–20–10–5. På grunn av manglende 1. pilot var det større usikkerhet rundt analyseverdier før 2. pilot enn tidligere, og det var også enkelte oppgaver vi ikke hadde verdier på. Av tabell 5 ser vi uansett at oppgavene i de tre ulike prøvene hadde en grei fordeling på mestringsnivå. Oppgavene fordeles på mestringsnivå ut fra b-verdier (vanskelighetsgrad IRT) etter siste utprøving.

	2. pilot A	2. pilot B	2. pilot C
Mestringsnivå 1	6 (13 %)	6 (13 %)	6 (13 %)
Mestringsnivå 2	11 (24 %)	8 (17 %)	14 (30 %)
Mestringsnivå 3	17 (38 %)	18 (38 %)	15 (32 %)
Mestringsnivå 4	4 (9 %)	7 (15 %)	3 (6 %)
Mestringsnivå 5	7 (16 %)	9 (19 %)	9 (19 %)

TABELL 3 FORDELING AV OPPGAVER INNEN MESTRINGSNIVÅER

Før gjennomføring	2. pilot A	2. pilot B	2. pilot C
Gj.snitt løsningsprosent totalt	50,8 %	46,9 %	51,7 %
Gj.snitt løsningsprosent jenter	49,4 %	45,1 %	50,1 %
Gj.snitt løsningsprosent gutter	52,2 %	48,8 %	53,3 %
Gj.snitt S-Rpbis	0,424	0,424	0,419
Gj.snitt T-Rpbis	0,476	0,476	0,471
Gj.snitt a-verdi	0,817	0,810	0,781
Gj.snitt b-verdi	-0,002	0,151	-0,086
Antall oppgaver	50 (45)	50 (48)	50 (47)
Antall flervalgsoppgaver (MC)	24	23	23
Antall interaktive oppgaver (IA)	5	3	4

TABELL 4 TEKNISKE DATA FØR GJENNOMFØRING AV 2. PILOT

Det er interaktive oppgaver både blant de åpne oppgavene og blant flervalgsoppgavene. Interaktiviteten består i å flytte på objekter eller å dra opp søyler.

Gjennomføring av 2. pilotene

2. piloten for 2021 ble gjennomført av elever på 8. trinn høsten 2020. Dette skjedde i ukene før de samme elevene skulle gjennomføre årets nasjonale prøve. Det er viktig at oppgavene blir prøvd på riktig aldersgruppe. Gjennomføringstidspunktet for 2. piloten må derfor ligge så nært som mulig opp til tidspunktet for gjennomføring av endelig prøve.

Det var 1418 elever som gjennomførte 2. pilot A, 1437 elever som gjennomførte 2. pilot B og 1397 elever som gjennomførte 2. pilot C. Reliabiliteten til prøvene var henholdsvis 0,907, 0,906 og 0,911. Verdiene tilfredsstillt kravene i Rammeverk for nasjonale prøver (Utdanningsdirektoratet, 2017). Tabell 5 viser utvalg og resultater fra gjennomføringen.

2. pilot NP8 2020	2. pilot A	2. pilot B	2. pilot C
Antall gutter	732	751	698
Antall jenter	686	686	699
Totalt antall elever	1418	1437	1397
Gj.snitt Theta gutter	0,141	0,136	0,064
Gj.snitt Theta jenter	-0,101	-0,157	-0,092
Gj.snitt Theta totalt	0,024	-0,004	-0,014
Sign.forskjell J - G (Theta) $p < 0,05$	Ja	Ja	Ja
Standardavvik Theta	0,885	0,958	0,966
SE til gjennomsnitt Theta	0,024	0,025	0,026
Gj.snitt løsningsprosent gutter	50,1 %	49,0 %	46,7 %
Gj.snitt løsningsprosent jenter	45,5 %	43,3 %	43,3 %
Gj.snitt løsningsprosent totalt	47,9 %	46,3 %	45,0 %
Svarprosent	90,7 %	90,3 %	90,0 %
Cronbachs Alpha	0,907	0,906	0,911
Antall item	50	50	50
Antall flervalgsoppgaver (MC)	24	23	23
Antall interaktive oppgaver (IA)	5	3	4

TABELL 5 TEKNISKE DATA ETTER GJENNOMFØRING AV 2. PILOT

Tabell 5 viser at løsningsprosenten i begge settene er noe lavere enn de estimerte verdiene før gjennomføring. Gjennomsnittlig vanskegrad (b-verdi) ved gjennomføring ble henholdsvis 0,125, 0,214 og 0,211. Disse verdiene tyder på at settene er like i vanskelighetsgrad totalt sett. Ellers viser analysen av resultatene at 29 av de totalt 130 oppgavene har for lav a-verdi (diskriminering) i henhold til kravene i Rammeverk for nasjonale prøver (Utdanningsdirektoratet, 2017). Disse oppgavene må vi vurdere grundig før de eventuelt blir tatt med i endelig prøve.

Kapittel 4: Sammensetting av kohortprøve

Innledning

Høsten 2020 deltok til sammen 4252 elever på 2. piloten til 2021-prøven. 2. pilot er den siste utprøvingen av oppgavene før den endelige prøven, og det er svarmønstrene disse elevene avga 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. Tabeller eller diagram som henviser til disse er kommentert med «tall fra preanalyse». Det vil si at den er basert på elevbesvarelser fra 2. pilot, men at det er foretatt en analyse på kun de 50 oppgavene som utgjør forslag til endelig prøve (tabell 8).

Designet som er brukt på 2. pilotene 2021, sikrer en overlapp av oppgaver i alle tre prøvesettene. Dette gjør at 2. pilotene er analyserte med en concurrent calibration-analyse som sikrer at oppgavene er plassert på samme skala. I 2. pilotene var det ti oppgavene oppgaver som var felles for alle tre prøvesettene, der seks av disse er med i forslag til endelig prøve. Selv om ankringen mellom oppgavene da er noe svakere enn først tiltenkt, er også preanalysen foretatt med en concurrent calibration.

Vi har tidligere erfart liten endring fra preanalyse til analyse av endelig gjennomføring. Det at utprøvingen av oppgaver til 2021-prøven avviker noe fra tidligere, som kan gjøre at det er større usikkerhet til forslag til prøven for 2021 sammenlignet med en normalsituasjon. Gitt at vilkårene for gjennomføring av nasjonal prøve i regning 2021 vil være normale, forventer vi imidlertid ikke betydelig større endringer enn det vi har vært vitne til de tidligere år. Vi forventer derfor at preanalysen også i år gir en god pekepinn på hvordan resultatene etter gjennomføring blir.

Nasjonal prøve for 8. og 9. trinn 2021 består av 50 oppgaver. Prøven foreligger i fire versjoner. Alle versjonene inneholder 50 oppgaver. Oppgavene i tre av versjonene er de samme, men rekkefølgen til de ni første oppgavene er ulik i de tre versjonene. Dette er gjort for å redusere sannsynligheten for at to elever som sitter nært hverandre, arbeider med samme oppgave samtidig. Versjonen som er ulik de andre tre, er en ankerprøve som brukes til å måle utvikling over tid. Omtrent halvparten av oppgavene i ankerprøven har vært med tidligere og utgjør det hemmelige ankeret. De øvrige oppgavene i ankerprøven er de samme som i de tre andre versjonene.

Psykometri

Andelen flervalgsoppgaver i prøven er 50 prosent (25 oppgaver), og det er totalt 6 prosent (3 oppgaver) interaktive oppgaver. Denne fordelinga er i tråd med tidligere prøver, men i sammensetningen av prøvesettet i år var det ekstra utfordrende å sørge for at andelen flervalgsoppgaver ble høy nok. Hovedårsaken til dette er at vi, på grunn av koronasituasjonen, ikke fikk gjennomført 1. pilot våren 2020, der vi vanligvis tester alle oppgavene som vi gjør om fra kortsvarsformat til flervalgsformat. Dette resulterte i at vi prøvde våre flervalgsoppgaver for første gang i 2. pilot, der mange viste seg å fungere dårligere enn hva vi hadde håpet på.

Gjennomsnittlig vanskegrad (b-verdi) til oppgavene i kohortprøven er 0,033, og gjennomsnittlig diskriminering (a-verdi) er 0,690. Gjennomsnittlig løsningsprosent til oppgavene er 49,0 prosent, og ved utprøving er løsningsprosenten til guttene fem prosentpoeng høyere enn jentenes (tabell 6). Det er ni oppgaver i prøven med bias mot kjønn, og det er å anta at det kan bli en signifikant forskjell mellom kjønnene i klassisk test-teori (CTT) ved endelig gjennomføring. Disse oppgavene blir omtalt senere i kapitlet.

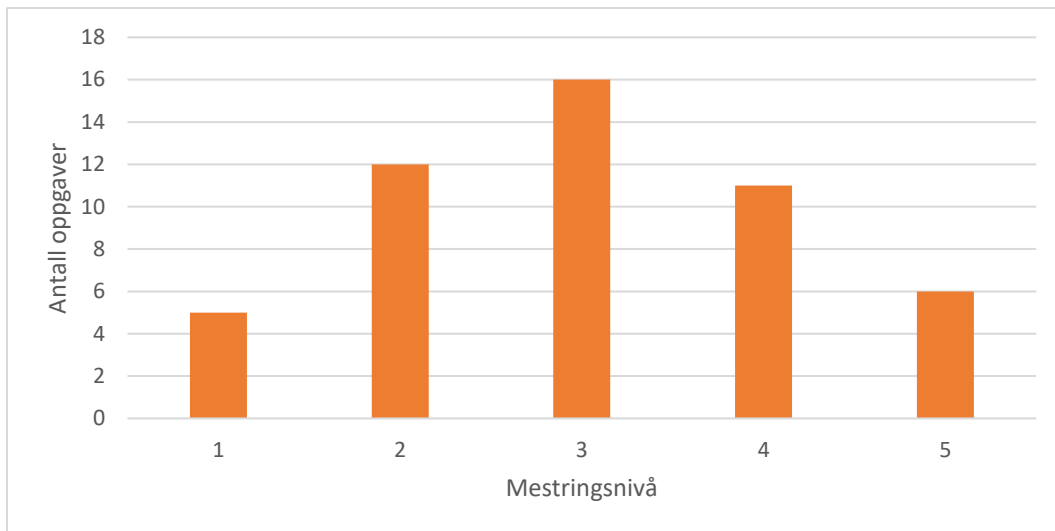
Siden prøven er satt sammen av oppgaver fra tre 2. piloter, som med noen unntak er besvart av to uavhengige elevgrupper, er det ikke mulig å beregne reliabilitet for den endelige prøven. Hver av de tre 2. pilotene har $\alpha \approx 0,9$. Basert på tidligere erfaringer med å sette sammen prøver etter uavhengige piloter, kan vi anta at reliabiliteten til den endelige prøven kommer til å bli 0,85–0,90. Dette tilfredsstillende kravet i Rammeverk for nasjonale prøver (Utdanningsdirektoratet, 2017).

Forslag til prøve NP8 2021	Kohortprøve	Ankerprøve	Ankeret ¹
Gj.snitt Theta gutter	0,102	X	0,003
Gj.snitt Theta jenter	-0,104		-0,224
Gj.snitt Theta totalt	0,019		-0,107
Sign.forskjell J - G (Theta) $p < 0,05$	Ja		Ja
Standardavvik Theta	0,944		0,935
SE til gj.snitt Theta	0,014		0,016
Gj.snitt løsningsprosent gutter	51,5 %	51,0 %	50,6 %
Gj.snitt løsningsprosent jenter	46,4 %	45,4 %	44,6 %
Gj.snitt løsningsprosent totalt	49,0 %	48,3 %	47,7 %
Svarprosent	91,2 %	91,9 %	93,4%
Cronbachs Alpha			0,863
Gj.snitt S-Rpbis	0,379	0,416	0,471
Gj.snitt T-Rpbis	0,464	0,450	0,501
Gj.snitt a-verdi	0,690	0,721	0,811
Gj.snitt b-verdi	0,033	0,064	0,087
Antall items	50	50	20
Antall MC	25	26	10
Antall IA	3	3	1

TABELL 6 OPPSUMMERING FØR GJENNOMFØRING

Selv om gjennomføringstidspunktet for 2. pilotene er nært opp mot gjennomføringstidspunktet for den endelige prøven, ser vi at gjennomsnittlig løsningsprosent øker, og at gjennomsnittlig vanskegrad (b-verdi) synker fra siste utprøving til endelig gjennomføring. Det kan muligens forklares med at en endelig prøve tas mer seriøst av både lærere og elever enn en utprøving, og at elevene er bedre forberedt til den endelige prøven enn ved en utprøving. Endringen er imidlertid liten. Gjennomsnittlig løsningsprosent økte i både 2019 og i 2020 med ett prosentpoeng fra siste utprøving til endelig prøve, mens gjennomsnittlig vanskegrad endret seg med henholdsvis -0,024 og -0,060. Vi kan derfor anta at endelig prøve 2021 vil ha en gjennomsnittlig p-verdi rundt 0,50 og gjennomsnittlig b-verdi nær 0.

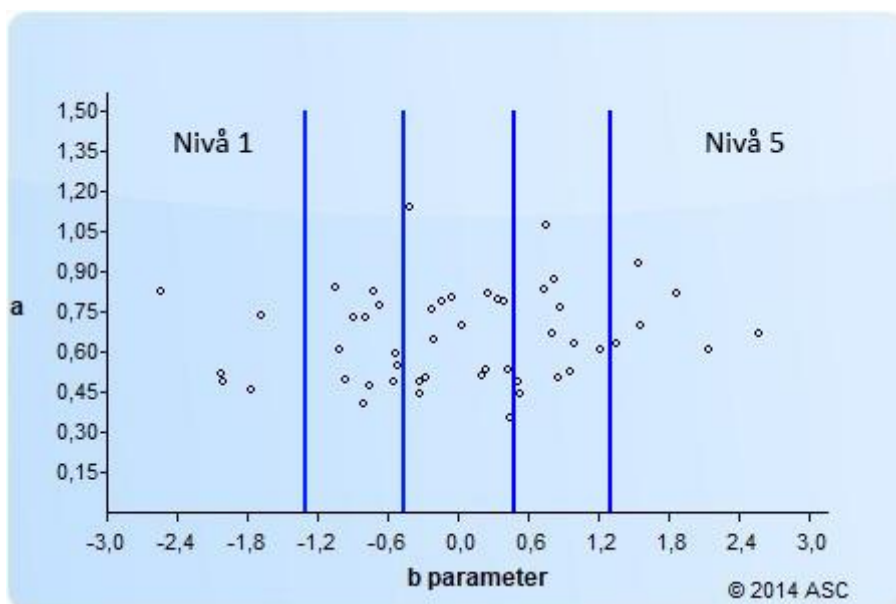
¹ Preanalysen av ankeret er basert på en analyse fra ankerprøven, der de 20 oppgavene som er med i ankeret i 2021 inngår. Parametrene er kalibrert gjennom en concurrent calibration med alle oppgavene i ankerprøven 2020, men tallene som er gjengitt i tabellen er kun knyttet til de 20 oppgavene som utgjør ankeret i 2021.



FIGUR 1 FORDELING AV OPPGAVER ETTER MESTRINGSNIVÅ. BASERT PÅ TALL FRA PREANALYSE

Figur 1 viser fordelingen av oppgaver etter mestringsnivå for 2021-prøven, med verdier fra preanalyse. Siden oppgavens vanskelighetsgrad endrer seg noe fra siste utprøving til endelig gjennomføring, kan oppgaver endre mestringsnivå. Dette gjelder spesielt oppgaver som ligger nært opp til grensen mellom to nivåer.

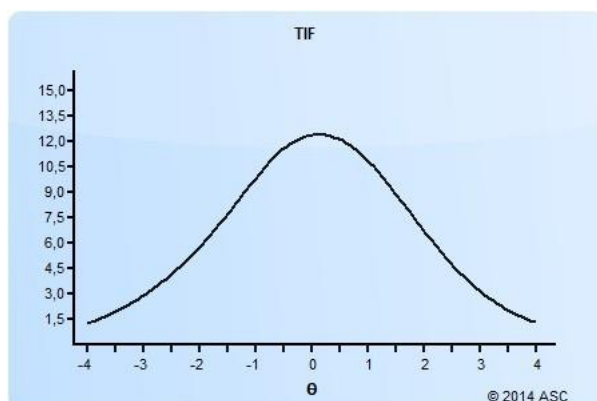
Figur 2 er et plottdiagram som viser oppgavens vanskegrad (b-verdi) og diskriminering (a-verdi), ut fra preanalysen. I figuren er det satt på vertikale linjestykker som illustrerer skiller mellom de ulike mestringsnivåene.



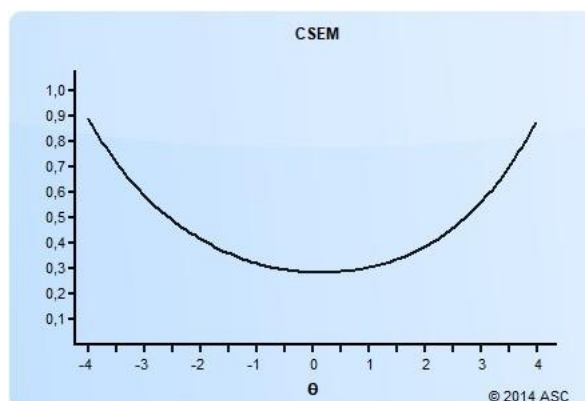
FIGUR 2 PLOTTDIAGRAM SOM VISER OPPGAVENES VANSKEGRAD OG DISKRIMINERING. BASERT PÅ TALL FRA PREANALYSE

Figur 1 og figur 2 viser at prøven inneholder tilstrekkelig antall oppgaver med lav og høy vanskelighetsgrad, og at oppgavene har høy diskrimineringsverdi. Samtidig viser figurene at prøven inneholder flest oppgaver med middels vanskelighetsgrad. Oppgaver med lav diskrimineringsverdi er nevnt senere i dette kapitlet. Figur 2 viser at mange oppgaver, ifølge preanalysen, har en vanskegrad som er tett opp til grensen til et annet mestringsnivå. Det er derfor naturlig å anta at flere oppgaver vil endre mestringsnivå etter den endelige gjennomføringen.

Opggavene er godt fordelt på mestringsnivåene, og gir mulighet til gode målinger på alle nivå og i begge ender av skalaen.



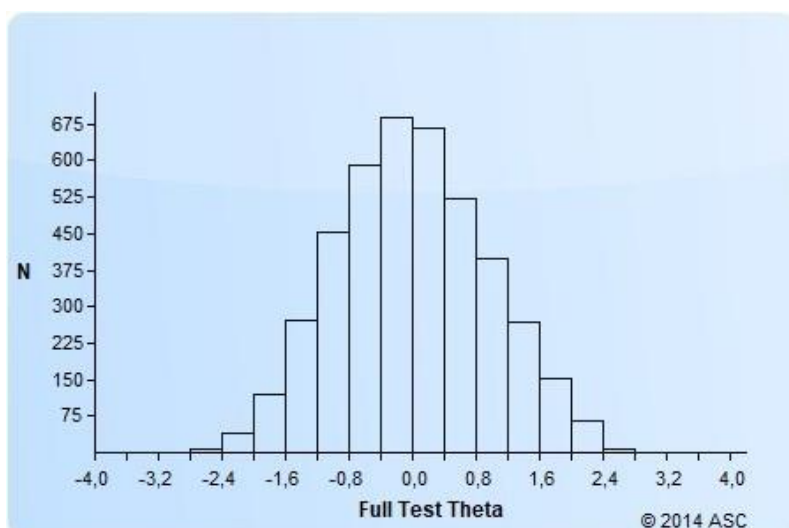
FIGUR 3 TESTINFORMASJONEN TIL PRØVEN FOR 2021.
BASERT PÅ TALL FRA PREANALYSE



FIGUR 4 MÅLEUSIKKERHETEN TIL PRØVEN FOR 2021.
BASERT PÅ TALL FRA PREANALYSE

En preanalyse av oppgavene til kohortprøven, viser at testinformasjonen (figur 4) er størst for elever med theta lik 0,150. Det vil si at prøven gir mest informasjon om elever som presterer litt over middels på ferdighetsskalaen. Med fastsettelse av ny skala i 2014 ble det bestemt at gjennomsnittet skulle være 0 (50 skalapoeng) og standardavviket 1 (10 skalapoeng). Det betyr at 68 prosent av elevene skal ligge mellom -1 og 1 på ferdighetsskalaen, og være i intervallet 40 til 60 skalapoeng. I tillegg ser vi at prøven gir bra med informasjon om elever som presterer lavt og høyt på prøven, selv om måleusikkerhet her er noe høyere enn midt på ferdighetsskalaen. Figur 5 viser at måleusikkerheten er lavest der det er flest elever, mellom -1 og 1 på ferdighetsskalaen. Kurvene for TIF og CSEM oppfyller kravene i Rammeverk for nasjonale prøver (Utdanningsdirektoratet, 2017).

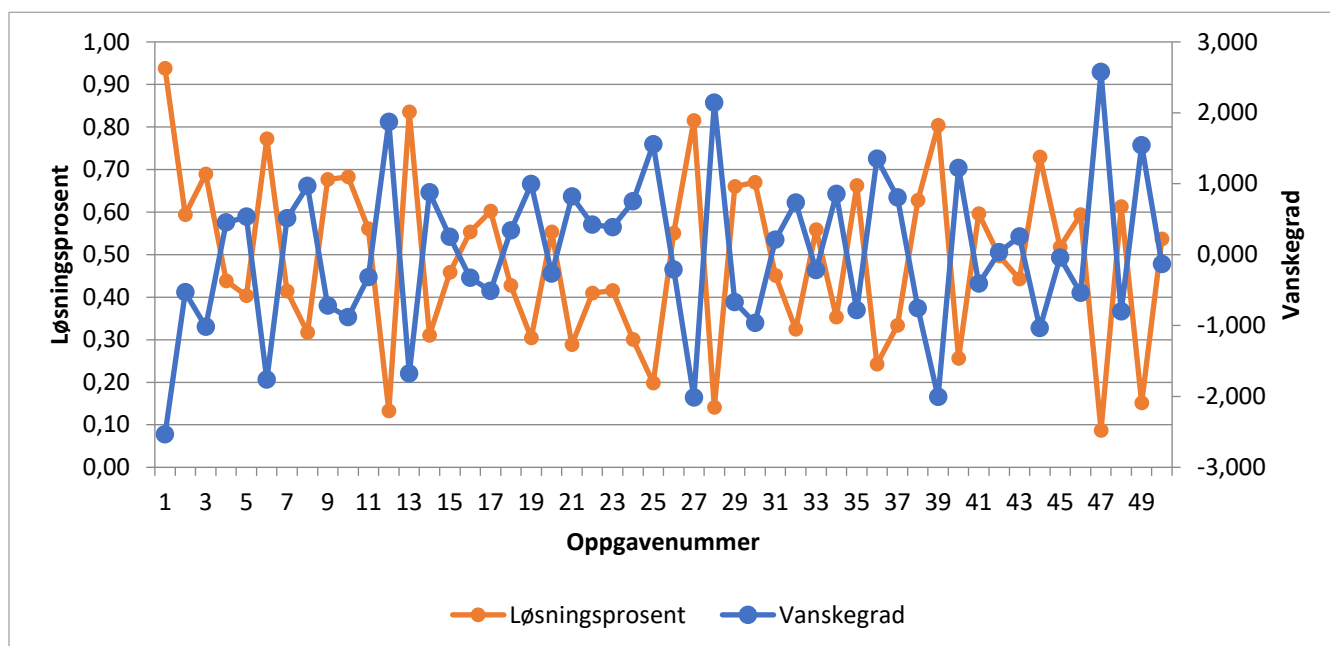
Figur 5 viser hvordan elevene som deltok på siste utprøving, er fordelt på samme skala som oppgavene. Sammenligner vi Figur 3 og figur 4 ser vi at testinformasjonen er høyest der det er flest elever, samtidig som prøven gir god informasjon om elever med høy og lav ferdighet.



FIGUR 5 FORDELING AV ELEVER UT FRA DYKTIGHET. BASERT PÅ TALL FRA PREANALYSE.

Figur 6 viser løsningsprosent og vanskegrad (b-verdi) for hver enkelt oppgave i kohortprøven for 2021, med tall fra preanalysen og siste utprøving. Dette oppsettet stemmer med tidligere prøver, der settet tenderer til å ha økende vanskegrad utover i settet, samt at vi finner oppgaver med høy og lav vanskegrad gjennom hele prøven. Det er også et bevisst valg at det kommer en oppgave med noe

lavere vanskegrad rett etter en oppgave med svært høy vanskegrad, for å unngå at elever får mange vanskelige oppgaver rett etter hverandre.



FIGUR 6 LØSNINGSPROSENT OG VANSKEGRAD TIL OPPGAVENE ETTER OPPGAVENUMMER I KOHORTPRØVEN FOR NP 2021. LØSNINGSPROSENT FRA 2. PILOT, VANSKEGRAD FRA PREANALYSE.

Forklaring til noen oppgaver i kohortprøven

Av de 50 oppgavene i forslag til endelig prøve er det 39 oppgaver som tilfredsstillende anbefalingene i Rammeverk for nasjonale prøver (Utdanningsdirektoratet, 2017) om at oppgavene skal ha en point-biserial-verdi $> 0,3$ ved klassisk testteori og en a -verdi $> 0,5$ ved item response theory. Dette er noe flere enn som har vært tilfelle de tidligere årene. En mulig betraktning av dette, er gjort til slutt i dette delkapittelet. Av disse elleve er det én oppgave som ikke tilfredsstillende noen av anbefalingene om diskrimineringsverdier, oppgave 4. Oppgaven er beskrevet nedenfor.

De øvrige fem oppgavene (1, 6, 27, 39 og 47) som har lav point-biserial-verdi, har a -verdi høyere enn 0,5. Disse oppgavene har enten veldig høy eller lav vanskegrad, og befinner seg dermed på et sted på ferdighetsskalaen der det er få elever, og er trolig årsaken til de lave diskrimineringsverdiene i klassisk testteori. Det er ingen av disse oppgavene som har lavere point-biserial-verdi når vi sammenligner svarmønster i oppgaver med elevenes theta ($T-R_{pbis}$). Disse oppgavene er ikke kommentert videre i dette kapittelet.

Oppgavene som har lavere a -verdi enn 0,5 (oppgave 5, 11, 16 og 48), har a -verdi fra 0,45 til og med 0,49. Oppgaven med lavest a -verdi, oppgave 48 er beskrevet nærmere nedenfor.

Oppgavene med noe lave diskrimineringsverdier er inkludert i forslaget til prøven for 2021 ut fra pedagogiske og didaktiske valg, i tillegg til at flere av disse er flervalgsoppgaver. Å sørge for at prøvesettet inneholdt stor nok andel flervalgsoppgaver var som tidligere påpekt en utfordring i sammensetningen av nasjonal prøve i regning for 8. trinn 2021. Oppgavene med diskrimineringsverdier marginalt under anbefalingene, er ikke kommentert nærmere.

I forslag til endelig prøve er det én oppgaver som i analysene etter 2. pilot, blir markert med som bias. Oppgave 46 favoriserer jenter sammenlignet med gutter med lik dyktighet. I tabell 7–tabell 9 er det satt inn tall for vanskelighetsgrad og diskriminering fra både CTT- og IRT-analysene for disse

oppgavene. I tillegg er gjennomsnittlig dyktighet (Theta) for elevene som har valgt de ulike alternativene 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 $M-H > 1$, så har guttene i gjennomsnitt over hele ferdighetsskalaen større odds for å løse oppgaven enn jentene. Dersom $M-H < 1$ så har guttene mindre odds til å løse oppgaven enn jentene. Vi har også med p-verdi fra IRT-analysen som angir hvor sannsynlig det er at forskjellen mellom de to gruppene skyldes tilfeldigheter eller ikke. Vi har satt 0,05 som signifikansmål Xcalibre (Assessment Systems Corporation, 2014) er, og oppgaver som har lavere p-verdi enn dette, kommer ut med bias. Analysen er basert på å sammenligne odds ratio for sju deler av ferdighet

Oppgaver med lave diskrimineringsverdier

Oppgave 4

Jon er på Island. Far sier at Jon kan regne om fra islandske kroner (ISK) til norske kroner (NOK) ved å dele beløpet på 10 og deretter trekke fra $\frac{1}{4}$.

Jon kjøper en pose lakris som koster 320 ISK.

Hvor mange norske kroner koster posen med lakris?

- 8 NOK
- 24 NOK
- 32 NOK
- 80 NOK

Itemnr.	T-Rpbis	S-Rpbis	Løsningsprosent	M-H	p	a-verdi	b-verdi	Gjennomsnittlig Theta			
								1	2	3	4
5196927	0,307	0,284	0,44	0,88	0,157	0,379	0,424	0,38	0,45	-0,54	-0,30

TABELL 7 PSYKOMETRISKE VERDIER FOR OPPGAVE 4, SOM HAR BIAS MOT GUTTER. TALL FRA 2. PILOT.

I denne oppgaven er det mest sannsynlig svaralternativ 1 som gjør at a-verdien er under 0,5 og point-biserial-verdien er under 0,3. Vi ser av analysene at dette svaralternativet diskriminerer positivt. Elevene som har valgt dette svaralternativet har delt 320 ISK på 10 og deretter funnet riktig brøkdel. De har derimot ikke trukket fra denne brøkdelen for å finne riktig pris i norske kroner. Svaralternativ 1 er dermed knyttet til utfordringer med delferdigheten *gjenkjenne og beskrive* som handler om å matematisere situasjonen korrekt. Av de gjennomsnittlige Theta-verdiene ser vi at elever som har valgt alternativ 2 i snitt er dyktigere enn de som har valgt alternativ 1, og vi har derfor valgt å ta med oppgaven, selv om den har diskrimineringsverdier under grenseverdiene.

Oppgave 48

Emilie har tatt bilde av en flokk med fugler. Hun har delt bildet i like deler.

Omtrent hvor mange fugler er det i bildet?

- Under 200 fugler
- Mellom 200 og 400 fugler
- Mellom 400 og 500 fugler
- Over 500 fugler



Itemnr.	T-Rpbis	S-Rpbis	Løsningsprosent	M-H	p	a-verdi	b-verdi	Gjennomsnittlig Theta			
								1	2	3	4
5196671	0,337	0,300	0,61	1,08	0,478	0,453	-0,689	-0,18	0,30	-0,04	-0,37

TABELL 8 PSYKOMETRISKE VERDIER FOR OPPGAVE 48, SOM HAR BIAS MOT GUTTER. TALL FRA 2. PILOT.

Denne oppgaven har godkjente verdier for point-biserial, men a-verdien er under 0,5. Av de gjennomsnittlige Theta-verdiene kan det se ut til at svaralternativ 3 er det som gjør at diskrimineringen er lavere enn ønsket. En forklaring på dette kan være at elevene som velger dette alternativet har forstått hva de skal gjøre, men velger kanskje å ta utgangspunkt i en rute som ikke inneholder et representativt antall fugler. Det kan også være mulig at elevene har gjort en noe unøyaktig optelling. Av analysene ser vi derimot at svaralternativ 3 diskriminerer negativt, slik en distraktor skal gjøre, og vi har derfor valgt å ta med oppgaven. I tillegg er dette en uvanlig oppgave sammenlignet med de andre oppgavene, og det kan gjøre at den har mindre til felles med resten av settet. Vi vurderer at det er interessant å prøve ut denne typen oppgaver, selv om det kan oppleves som noe nytt for elevene.

Oppgaver med bias

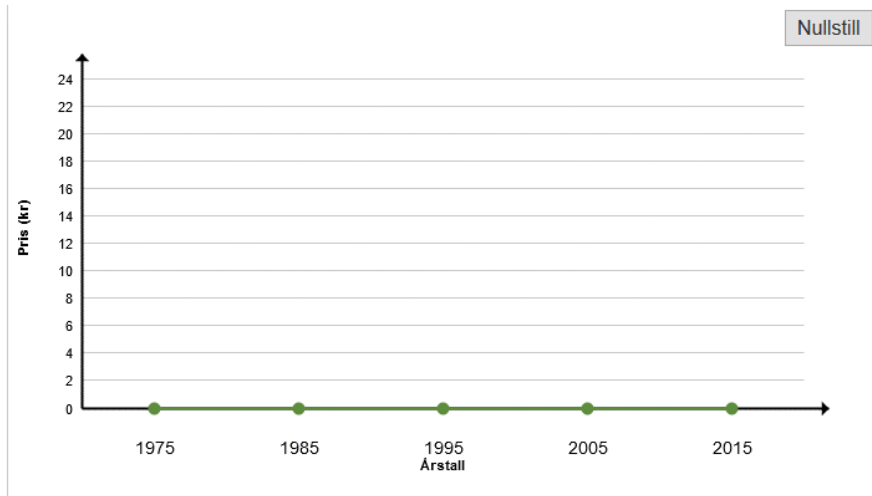
Oppgave 46 (Bias mot gutter)

Krone-isen ble lansert i 1953, og siden den gangen har prisen utviklet seg mye.

Tabellen viser hvordan prisen utviklet seg i perioden 1975–2015.

Årstall	Pris (kr)
1975	2,00
1980	4,00
1985	5,00
1990	8,00
1995	11,00
2000	14,00
2005	16,00
2010	20,00
2015	24,00

Fremstill utviklingen i et linjediagram.



Itemnr.	T-Rpbis	S-Rpbis	Løsningsprosent	M-H	p	a-verdi	b-verdi	Gjennomsnittlig Theta	
								0	1
5197593	0,411	0,370	0,59	0,67	0,018	0,549	-0,490	-0,24	0,38

TABELL 9 PSYKOMETRISKE VERDIER FOR OPPGAVE 46, SOM HAR BIAS MOT JENTER. TALL FRA 2. PILOT.

Oppgaven måler om elevene behersker å lage diagram med utgangspunkt i en tabell, der konteksten er prisen på Krone-is. Denne konteksten tenker vi i utgangspunktet er kjønnsnøytral. Vår erfaring er som tidligere nevnt, at denne typen oppgaver setter krav til elevenes nøyaktighet, og vi ser at jenter som regel løser slike oppgaver i større grad enn gutter. I denne oppgaven stilles det også store krav til å koble tekst/tabell med diagrammet, siden det ikke er alle årstall som skal være med i diagrammet.

Validitet

Høsten 2020 ble nye læreplaner innført i Norge. På grunn av at vi bruker to år på å utvikle en nasjonal prøve, var oppgavene som danner grunnlaget prøven i 2021 ferdig utviklet, og prøvd ut, før og samtidig som læreplanene ble innført. Nasjonal prøve i regning 2021 tar derfor utgangspunkt i Kunnskapsløftet, LK06. I LK06 (Utdanningsdirektoratet, 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. Rammeverk for nasjonale prøver (Utdanningsdirektoratet, 2017).

Nasjonal prøve i regning for 8. og 9. trinn måler elevenes kompetanse etter 7. trinn i den grunnleggende ferdigheten å kunne regne. Derfor må valideringen skje ut fra det en kan forvente av elever som nettopp har begynt i 8. trinn. Gruppen som utvikler nasjonale prøver i regning, har lang undervisnings erfaring fra grunnskolen og kjenner både læreplanen og den aktuelle aldersgruppen godt.

I tillegg har en lærergruppe med fire personer gjennomført ekstern validering av prøvesettet. Det vil si at de har vurdert hver enkelt oppgave og prøvesettet som helhet opp mot kravene i Rammeverket for nasjonale prøver (Utdanningsdirektoratet, 2017), som handler om validitet, og opp mot

beskrivelsene av den grunnleggende ferdigheten å kunne regne, slik den er beskrevet i samme dokument og i læreplaner for fag i Kunnskapsløftet (Utdanningsdirektoratet, 2006).

Nr.	Innhold ¹	Område ²	Kommentarer	Hovedutfordring i løsningsprosess ³	Relevans til fag ⁴	Mestringsnivå	Underpunkt mestringsnivå ⁵	Fasit	P-verdi 2. pilot
1	Halvering	Tall og algebra		BB	Mat, saf	1	1.1	525	0,93
2	Omgjøring mellom enheter	Måling og geometri	Var innom underpunkt 3.7 og 3.12	RV	Mat, Khv	3	1.2, 1.4	mm (alt. 4)	0,54
3	Subtraksjon hele tall	Tall og algebra		BB	Mat, saf, kro, nat	2	2.1	628	0,69
4	Divisjon, subtraksjon og brøk	Tall og algebra		BB	Mat, saf	3	2.1, 2.2, 3.2, 3.5	24 NOK (alt. 2)	0,44
5	Prosent.	Tall og algebra		BB (RV)	Mat, saf	4	3.3, 3.5, 3.12	1470 kr (alt. 4)	0,34
6	Omgjøring mellom enheter.	Måling og geometri		BB (RV)	Mat, Khv	1	1.4, 3.7	5 m (alt. 2)	0,72

FIGUR 7 UTSNITT FRA ARBEIDET TIL LÆRERNE SOM GJENNOMFØRTE EKSTERN VALIDERING AV NP8 OG NP9 2021

Den eksterne valideringen går blant annet ut på at lærerne fyller ut en tabell med deres validering av oppgavene i kohortprøven. Figur 7 viser et utsnitt av denne tabellen. Lærerne diskuterer oppgavens innhold (hva oppgaven måler), hvilket matematisk hovedområde oppgaven er knyttet til, eventuelle kommentarer, hvilken del av den helhetlige problemløsningsprosessen som er hovedutfordringen, hvilke fag oppgaven har relevans for og hvor de finner dekning for oppgaven i mestringsbeskrivelsene for nasjonale prøver i regning 8. og 9. trinn. Etter at lærerne har validert alle oppgavene, diskuterer lærerne valideringen med en representant fra Matematikksenteret. I tillegg har prosjektgruppa som jobber med nasjonale prøver gått igjennom lærernes validering i ettertid og vurdert hvert enkelt innspill. Eksempel på endringer som har skjedd på bakgrunn av den eksterne valideringen er utskiftning av oppgaver, satt inn/tatt ut bilder, tekstendringer og justeringer av mestringsbeskrivelsene.

Opgavene i prøven for 2021 knyttes til kompetansemål i fag som vist i tabell 10. En oppgave kan knyttes til kompetansemål i flere fag.

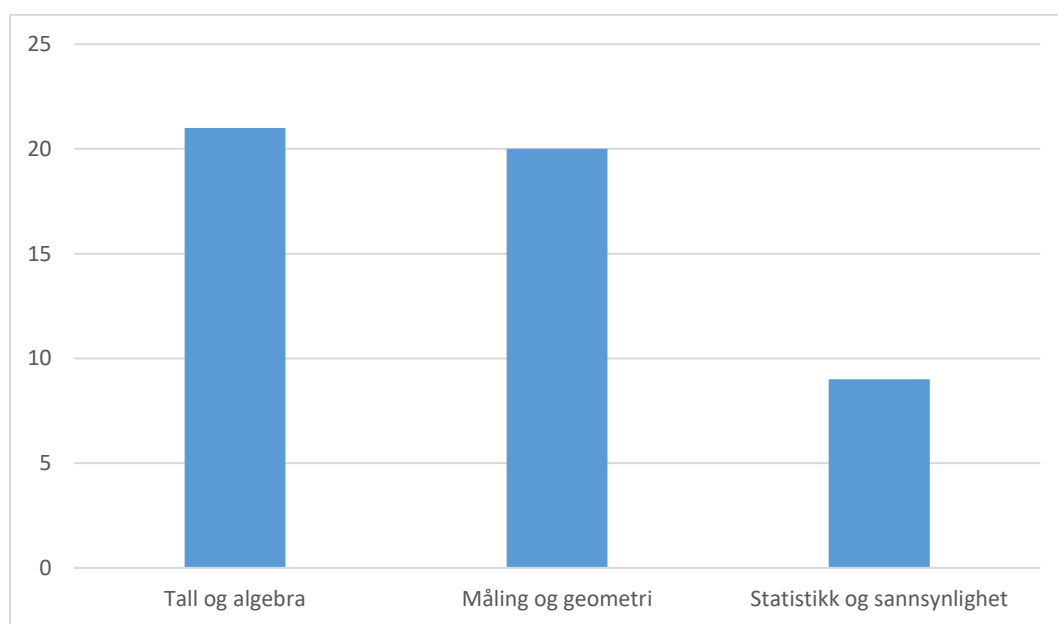
Fag	Antall oppgaver
Engelsk	7
KRLE	7
Kroppsøving	9
Kunst og håndverk	7
Matematikk	50
Mat og helse	8
Musikk	0
Naturfag	21
Norsk	7
Samfunnsfag	19

TABELL 10 ANTALL OPPGAVER SOM KNYTTES TIL KOMPETANSEMÅL I ULIKE FAG ETTER LK06 7. TRINN

Selv om gjeldende læreplan sier at regning er en grunnleggende ferdighet i alle fag (Utdanningsdirektoratet, 2006), er det ulikt antall kompetansemål som kan knyttes til regning i de ulike fagene. I noen fag er kompetansemålene, der elevens regneferdighet har betydning for om eleven når kompetansemålet, tydelig knyttet til regning, mens i andre er regneferdigheten mer skjult. Det er også en del kompetansemål i praktiske fag, der den regneferdigheten som kreves for å nå kompetansemålet ikke kan måles i det formatet som nasjonale prøver har. Som tabell 10 viser, kan ikke oppgavene i forslag til kohortprøven 2020 relateres til kompetansemål i alle fag. I år, som ved flere tidligere år, har vi ingen oppgaver som kan knyttes til kompetansemål i musikk. Dette er et fag som inneholder få kompetansemål etter 7. trinn som krever en form for regning som kan måles i en formatet til nasjonale prøver.

Nasjonal prøve i regning for 8. og 9. trinn inneholder oppgaver innen hovedområdene i læreplanen i matematikkfaget (Utdanningsdirektoratet, 2013); tall og algebra (T&A), måling og geometri (M&G), og statistikk og sannsynlighet (S&S). Oppgavene er knyttet til områdene i tråd med beskrivelsene i Rammeverk for nasjonale prøver (Utdanningsdirektoratet, 2017). Prøven består av 21 oppgaver i området tall og algebra, 20 innenfor måling og geometri og 9 statistikk- og sannsynlighetsoppgaver (figur 8). Denne fordelingen med omtrent like mange oppgaver innen tall og algebra og måling og geometri, samt halvparten av dette antallet i statistikk og sannsynlighet, gjenspeiler etter vår vurdering hvordan regning er synliggjort i kompetansemålene i Kunnskapsløftet (Utdanningsdirektoratet, 2006).

Det er imidlertid viktig å påpeke at mange av oppgavene krever en kompetanse fra flere områder. For eksempel krever ofte oppgaver innen statistikk og sannsynlighet også at elevene bearbeider tall eller måleenheter. Vi har imidlertid forsøkt å kategorisere oppgavene etter hvilken kategori hovedutfordringen ligger.



FIGUR 8 ANTALL OPPGAVER I HVERT OMRÅDE

Opgavene i hver kategori er fordelt på fem mestringsnivåer (tabell 11), etter tall fra preanalysen. Tabellen viser at forslag til prøven for 2021 har oppgaver innenfor alle mestringsnivåer for de ulike områdene. Selv om vi erfarer at oppgavene endrer vanskegrad noe fra siste utprøving til endelig prøve, er det til dels store endringer som må til for at vi ikke skal ha oppgaver på alle nivåer for hvert område også etter gjennomføringen høsten 2021.

	Mestringsnivå 1	Mestringsnivå 2	Mestringsnivå 3	Mestringsnivå 4	Mestringsnivå 5	Totalt
Tall og algebra	2	4	8	6	1	21
Måling og geometri	2	6	5	3	4	20
Statistikk og sannsynlighet	1	2	3	2	1	9
Totalt	5	12	16	11	6	

TABELL 11 ANTALL OPPGAVER PÅ HVERT MESTRINGSNIVÅ INNENFOR HVERT OMRÅDE. TALL FRA PREANALYSE

I tillegg til å være knyttet til kategorier i læreplanen i matematikk, er oppgavene knyttet til de ulike delferdighetene i den helhetlige problemløsningsprosessen som er beskrevet i Rammeverk for nasjonale prøver (Utdanningsdirektoratet, 2017). Noen oppgaver har hovedutfordring i å formulere

et regneuttrykk (gjenkjenne og beskrive), andre i å foreta beregningen (bruke og bearbeide) eller å avmatematisere innholdet i oppgaven (reflektere og vurdere).

Tabell 17 viser en oversikt over fordeling av oppgaver knyttet til disse prosessene. 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 bruke og bearbeide i samme problem. Derfor har vi ikke sett på en fordeling av de matematiske områdene T&A, M&G og S&S opp mot prosessene i den helhetlige problemløsningsprosessen, siden områdene og prosessene er to kategoriseringer med høy unøyaktighet i seg selv.

Prosess	Antall oppgaver
Gjenkjenne og beskrive	23
Bruke og bearbeide	23
Reflektere og vurdere	4

TABELL 12 ANTALL OPPGAVER KNYTTET TIL DE ULIKE PROSESSENE I DEN HELHETLIGE PROBLEMLØSNINGSPROSESSEN

Det er færrest oppgaver der hovedutfordringen er reflektere og vurdere (tabell 12). Det betyr ikke at dette er en prosess som i liten grad måles i prøven. Elevene bør i alle oppgaver reflektere over og vurdere om det matematiske svaret de har funnet er et svar som er realistisk til konteksten oppgaven, men det er få oppgaver der vi anser dette å være hovedutfordringen i oppgaven.

Generelle vurderinger av prøven som helhet

Koronapandemien har påvirket prosessen med å utvikle nasjonale prøver. Det har blant annet ført til at vi i 2021 har en prøve som har vært gjennom en utprøvingsrunde færre enn planlagt. I et normalår har vi følt vi har hatt god kontroll over oppgavene idet vi har satt sammen 2. pilotene, mens for 2021-prøven var situasjonen mer uoversiktlig siden vi hadde mindre data om oppgavene tilgjengelig. Dette medførte at vi endret design på 2. pilotene, for å øke sjansene for at vi ville få et tilstrekkelig antall gode nok oppgave å velge mellom når vi skulle sette sammen forslag til endelig prøve. På grunn av at vi hadde liten informasjon om oppgavene før 2. pilot, var det en betydelig større andel av oppgavene som fungerte dårlig i siste utprøving enn normalt. Det er derfor knyttet noe større usikkerhet til hvordan oppgavene vil fungere ved endelig gjennomføring sammenlignet med tidligere år.

Gjennom utprøvinger har hver oppgave i forslag til endelig prøve vært prøvd ut minimum én gang. 35 oppgaver har vært med i to utprøvinger, mens 15 har vært med i bare hovedutprøvingen.

På tross av at prosessen med utvikling av nasjonal prøve 2021 har avviket fra en normal utviklingsprosess vurderer vi at prøven, både oppgave og oppgave og som helhet, er valid og reliabel, selv om det er ting som tyder på at prøven for 2021 er mindre reliabel enn hva tidligere prøver har vært.

Kapittel 5: Analyse av nasjonal prøve i regning

Innledning

Dette kapitlet er en analyse av resultatene til elevene som gjennomførte nasjonal prøve i regning på bokmål eller nynorsk for 8. og 9. trinn høsten 2021. Tallene det henvises til i dette kapitlet er basert på en analyse av elever som gjennomførte kohortprøven og ankerprøven på 8. trinn. I analysene er parameterne til ankeroppgavene fastsatt på forhånd, for å sikre at kohortoppgavene kommer på samme skala som oppgaver fra tidligere prøver. Dette gjør det mulig å sammenlignbare resultater over tid. I analysene av prøven for 9. trinn er parameterne fra analysen av prøven for 8. trinn fastsatt på forhånd, for å sikre at også resultatene for 9. trinn kommer på samme skala. Det var 61 697 elever på 8. trinn og 59 271 elever på 9. trinn som gjennomførte prøven. Rapporten omhandler i utgangspunktet resultatene til 8. trinn, men vi sammenligner med resultatene til 9. trinn.

Nasjonal prøve i regning for 8. og 9. trinn måler elevenes grunnleggende ferdigheter i å kunne regne med utgangspunkt i kompetansemålene i alle fag etter 7. trinn. Derfor må valideringen skje ut fra det en kan forvente av elever som nettopp har begynt i 8. trinn. Verdiene for 9. trinn er tatt med til sammenligning.

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 Rammeverk for nasjonale prøver (Utdanningsdirektoratet, 2017). Endelig godkjenning av prøven ble gjort av Utdanningsdirektoratet og deres egne interne og eksterne kvalitetssikrere.

Opgavene i kohortprøven er plassert på mestringsnivå ut fra vanskegrad, bestemt ved IRT-analyse. Oppgavens vanskegrad påvirker elevenes estimerte ferdighetsnivå. I praksis betyr det at elever som besvarer et likt antall oppgaver riktig, kan komme ut med ulikt antall skalapoeng. Antall oppnådde skalapoeng avgjør mestringsnivået til elevene.

For 8. og 9. trinn er det fem mestringsnivåer, nivå 1 til 5, der 5 er det høyeste nivået. Grensene til nivåene er faste fra år til år, og følgende grenser gjelder for nasjonal prøve i regning for 8. og 9. trinn:

Mestringsnivå 1:	36 eller færre skalapoeng
Mestringsnivå 2:	37–44 skalapoeng
Mestringsnivå 3:	45–54 skalapoeng
Mestringsnivå 4:	55–62 skalapoeng
Mestringsnivå 5:	63 eller flere skalapoeng

Psykometriske verdier

Tabell 13 viser de psykometriske verdiene til kohortprøven og ankerprøven. I tillegg viser tabellen resultatene til 9. trinn for kohortprøven.

NP8 2021	Kohortprøve ²	Ankerprøve ³	9. trinn
Antall jenter	30 248	1807	29 080
Antall gutter	31 449	1892	30 191
Totalt antall elever	61 697	3 699	59 271
Gj.snitt a-verdi	0,723	0,809	X
Gj.snitt b-verdi	-0,047	-0,052	
Gj.snitt Theta jenter	-0,171	-0,252	0,177
Gj.snitt Theta gutter	0,119	-0,003	0,448
Gj.snitt Theta totalt	-0,023	-0,125	0,315
Sign.forskjell J - G (Theta) p<0,05	Ja	Ja	Ja
Standardavvik Theta	0,964	0,946	1,01
SE til gjennomsnitt Theta	0,004	0,016	0,004
Gj.snitt løsningsprosent gutter	53,50 %	50,80 %	60,40 %
Gj.snitt løsningsprosent jenter	47,40 %	44,00 %	55,00 %
Gj.snitt løsningsprosent totalt	50,50 %	47,40 %	57,70 %
Svarprosent	93,60 %	93,40 %	94,40 %
Cronbachs alpha	0,922	0,868	0,926
Antall item	50	20	X
Antall MC	25	9	
Antall IA	3	0	

TABELL 13 PSYKOMETRISKE VERDIER KOHORTPRØVEN OG ANKERPRØVEN 2021, SAMT RESULTATER FOR 9. TRINN

Som tabellen viser, er de psykometriske verdiene til kohortprøven og ankerprøven på 8. trinn veldig like. Vi ser også at forskjellen mellom gjennomsnittlig løsningsprosent for gutter og jenter er rundt 6 prosentpoeng i begge prøvene, og at det er signifikant forskjell mellom kjønnene når det gjelder dyktighet.

Oversikt over oppgaver NP8 (9) 2021

Nr.	Innhold	Område	Format	Relevans til fag	P-verdi 8.trinn	Forskjell P-verdi (J-G) 8. trinn	P-verdi 9. trinn	Forskjell P-verdi (J-G) 9. trinn	Effektstørrelse fra 8. til 9. trinn	b-verdi	Mest-rings-nivå
1	Divisjon. Hele tall.	T&A	Fyll inn tall	Mat, Mhe	0,92	-0,03	0,95	0,0	0,07	-2,1348	1
2	Prefikser.	M&G	Flervalg	Khv, Kro, Mat, Nat	0,58	-0,08	0,70	-0,1	0,17	-0,6146	2
3	Subtraksjon. Hele tall.	T&A	Fyll inn tall	Kro, Mat	0,71	0,01	0,77	0,0	0,14	-0,9928	2
4	Divisjon. Subtraksjon.	T&A	Flervalg	Mat	0,44	0,01	0,52	0,0	0,16	0,3751	3
5	Prosent. Regne med prosent.	T&A	Flervalg	Mat, Saf	0,39	-0,07	0,52	-0,1	0,18	0,3816	3
6	Forståelse av enheter.	M&G	Flervalg	Khv, Mat, Nat	0,74	-0,10	0,82	-0,1	0,08	-1,6654	1
7	Regneark i kontekst. Lag formel/uttrykk.	T&A	Flervalg	Mat, Saf	0,46	0,02	0,55	0,0	0,21	0,2616	3
8	Areal	M&G	Flervalg	Khv, Mat	0,32	0,00	0,42	0,0	0,19	0,9294	4
9	Forhold. Multiplikasjon. Divisjon. Omgjøring mellom enheter.	M&G	Fyll inn tall	Mat, Mhe	0,61	-0,05	0,72	0,0	0,19	-0,5350	2
10	Tolke og lese diagram.	S&S	Fyll inn tall	Eng, Krle, Mat, Nat, No, Saf	0,68	0,02	0,73	0,0	0,12	-0,8199	2
11	Prefikser.	M&G	Flervalg	Mat, Mhe, Nat	0,50	-0,12	0,64	-0,1	0,16	-0,3340	3
12	Tolke og lese av diagram.	S&S	Fyll inn tall	Eng, Krle, Mat, Nat, No, Saf	0,14	-0,02	0,21	0,0	0,16	1,6824	5
13	Divisjon	T&A	Fyll inn tall	Mat	0,82	-0,05	0,88	0,0	0,11	-1,5835	1
14	Forhold	T&A	Flervalg	Mat, Nat, Saf	0,26	-0,11	0,38	-0,1	0,14	0,8308	4
15	Tid. Beregne tid, timer og minutter.	M&G	Flervalg	Kro, Mat, Nat, Saf	0,41	-0,11	0,53	-0,1	0,13	0,1635	3
16	Multiplikasjon. Hele tall. Algebraisk tenkning.	T&A	Fyll inn tall	Mat	0,58	0,03	0,63	0,0	0,14	-0,3588	3
17	Tid. Stille analog klokke.	M&G	Klokke	Mat, Mhe, Nat, Saf	0,60	-0,01	0,68	0,0	0,16	-0,5212	2
18	Forhold.	M&G	Fyll inn tall	Khv, Mat	0,36	-0,10	0,48	-0,1	0,15	0,3220	3
19	Forhold.	M&G	Fyll inn tall	Khv, Mat	0,33	-0,01	0,43	0,0	0,20	0,7059	4
20	Multiplikasjon. Brøk.	T&A	Flervalg	Mat, Mhe	0,45	-0,17	0,63	-0,2	0,20	-0,1865	3

² Analysen er basert på elever som gjennomførte versjon 1, 2 eller 3 av NPREG08

³ Ankerprøven består av omtrent halvparten av oppgavene fra kohortprøven og halvparten ankeroppgaver.

21	Multiplikasjon. Algebraisk tenking.	T&A	Fyll inn tall	Mat	0,28	-0,07	0,42	-0,1	0,22	0,6467	4
22	Lese av diagram. Subtraksjon.	T&A	Flervalg	Eng, Krle, Mat, Nat, No, Saf	0,46	0,01	0,52	0,0	0,14	0,1992	3
23	Gjennomsnitt	S&S	Fyll inn tall	Krle, Mat, Nat, Saf	0,36	-0,05	0,54	-0,1	0,31	0,4026	3
24	Multiplikasjon. Innsetting i formel.	T&A	Fyll inn tall	Mat, Nat	0,34	-0,06	0,49	0,0	0,24	0,4246	3
25	Tid. Beregne tid, sekunder og hundredeler.	M&G	Fyll inn tall	Kro, Mat, Nat	0,19	-0,10	0,28	-0,1	0,08	1,0832	4
26	Lese av diagram.	S&S	Fyll inn tall	Eng, Krle, Mat, Nat, No, Saf	0,54	-0,10	0,67	-0,1	0,15	-0,4421	3
27	Geometriske figurer. Systematisere data og framstille resultat.	M&G	Fyll inn tall	Khv, Mat, Nat, Saf	0,84	0,06	0,85	0,1	0,10	-1,8011	1
28	Prosent. Forhold.	T&A	Flervalg	Mat, Saf	0,08	-0,10	0,19	-0,1	0,16	1,8937	5
29	Divisjon/multiplikasjon.	T&A	Flervalg	Mat	0,66	-0,03	0,74	0,0	0,15	-0,7240	2
30	Addisjon. Negative tall.	T&A	Flervalg	Mat, Nat, saf	0,66	-0,01	0,73	0,0	0,13	-0,8764	2
31	Prosent. Regne med prosent.	T&A	Flervalg	Mat, Saf	0,41	-0,10	0,53	-0,1	0,14	0,1869	3
32	Tidsintervall.	M&G	Fyll inn tall	Krle, Mat, Saf	0,27	-0,10	0,41	-0,1	0,17	0,6742	4
33	Omgjøring mellom prefikser	M&G	Flervalg	Mat, Mhe, Nat	0,38	-0,20	0,56	-0,2	0,16	0,0496	3
34	Tolke og lese av diagram.	S&S	Fyll inn tall	Eng, Krle, Mat, Nat, No, Saf	0,25	-0,06	0,30	-0,1	0,05	1,1747	4
35	Omgjøring mellom enheter.	M&G	Fyll inn tall	Mat, Mhe	0,62	-0,08	0,74	-0,1	0,18	-0,7104	2
36	Blandingsforhold.	M&G	Fyll inn tall	Khv, Mat, Nat	0,26	-0,10	0,39	-0,1	0,16	0,7556	4
37	Forhold. Multiplikasjon.	T&A	Flervalg	Mat	0,28	-0,11	0,40	-0,1	0,12	0,7482	4
38	Vei, fart, tid.	M&G	Flervalg	Kro, Mat	0,57	-0,11	0,68	-0,1	0,11	-0,7044	2
39	Lage diagram. Sektordiagram.	S&S	Dra og slipp	Eng, Krle, Mat, Nat, No, Saf	0,76	0,01	0,80	0,0	0,11	-1,3023	1
40	Sannsynlighet	S&S	Flervalg	Mat	0,32	0,00	0,41	0,0	0,19	0,8322	4
41	Subtraksjon. Brøk.	T&A	Flervalg	Mat, Mhe	0,55	-0,10	0,68	-0,1	0,16	-0,3701	3
42	Divisjon.	T&A	Flervalg	Mat, Saf	0,56	-0,08	0,68	0,0	0,18	-0,4170	3
43	Lese av tabell. Algebraisk tenking	S&S	Fyll inn tall	Mat, Saf	0,48	-0,07	0,61	-0,1	0,19	-0,0932	3
44	Valuta.	M&G	Fyll inn tall	Mat	0,75	-0,08	0,83	-0,1	0,10	-1,1348	2
45	Subtraksjon. Desimaltall.	T&A	Flervalg	Mat, Nat, Saf	0,54	-0,15	0,69	-0,1	0,15	-0,4361	3
46	Lage diagram.	S&S	Søyler	Eng, Krle, Mat, Nat, No, Saf	0,68	0,04	0,73	0,1	0,16	-0,7842	2
47	Vei, fart, tid.	M&G	Fyll inn tall	Mat, Nat	0,07	-0,05	0,13	-0,1	0,10	2,3809	5
48	Overslag.	T&A	Flervalg	Mat, Nat	0,64	-0,03	0,68	0,0	0,06	-1,0628	2
49	Volum. Anvende formler.	M&G	Fyll inn tall	Mat, Nat	0,15	-0,01	0,31	0,0	0,37	1,4722	5
50	Omgjøring mellom enheter. Addisjon.	M&G	Flervalg	Mat, Nat, Saf	0,46	-0,24	0,65	-0,2	0,15	-0,3386	3

TABELL 14 OPPGAVEREKKEFØLGE I VERSJON 1 AV NASJONAL PRØVE I REGNING 2021. RESULTATER FOR 8. OG 9. TRINN

Tabell 14 viser rekkefølgen til oppgavene i versjon 1 av kohortprøven, hva oppgavene måler, fagområde, oppgaveformat, oppgavens tilknytning til fag, løsningsprosent for 8. og 9. trinn, kjønnsforskjeller og forskjeller mellom trinn i effektstørrelse. I tillegg vises vanskegraden til hver enkelt oppgave (b-verdi) og oppgavens mestringsnivå.

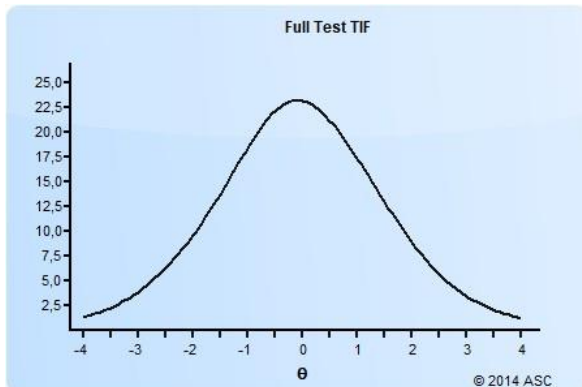
Testinformasjon og måleusikkerhet

Figur 9 og figur 10 viser testinformasjonen og måleusikkerheten til prøven for hvert ferdighetsnivå. Testinformasjonen er høyest ved ferdighet -0,15, og måleusikkerheten er lavest for samme ferdighet.

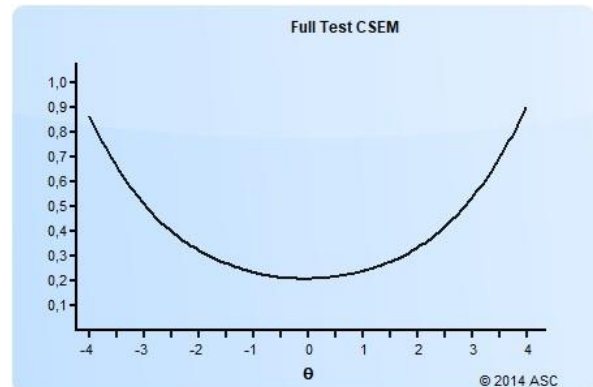
Prøven for 2021 har kurver for testinformasjon og måleusikkerhet som samsvarer med hverandre. Som vanlig har prøven høy testinformasjon og lav måleusikkerhet midt på ferdighetsskalaen. Dette er imidlertid en naturlig konsekvens av at det er flest oppgaver i prøven som ligger midt på ferdighetsskalaen (se figur 1, kapittel 4), som igjen er en konsekvens av hvordan elevene fordeler seg på denne skalaen.

Sammenlignet med prøvene for 2020, har prøven for 2021 litt lavere testinformasjon, og dermed høyere måleusikkerhet, for elever med høy dyktighet. For elever med gjennomsnittlig og lav dyktighet gir prøven for 2021 marginalt mer informasjon enn prøven for 2020. Det er mange hensyn å ta når vi velger ut oppgaver til kohortprøven, der en god fordeling av oppgaver utover vanskegradsskalaen er ett av kriteriene. Figur 1 (kapittel 4) viser at til sammen elleve oppgaver var på

mestringsnivå 1 og 5 ut fra tall basert på 2. pilot. Etter gjennomføring har dette endret seg for to oppgaver. I preanalysen var det 6 oppgaver på nivå 1 og 5 oppgaver på nivå 5, mens det etter gjennomføring var 5 oppgaver på nivå 1 og 4 oppgaver på nivå 5 (se figur 14).



FIGUR 9 TESTINFORMASJON - PRØVEN 2021

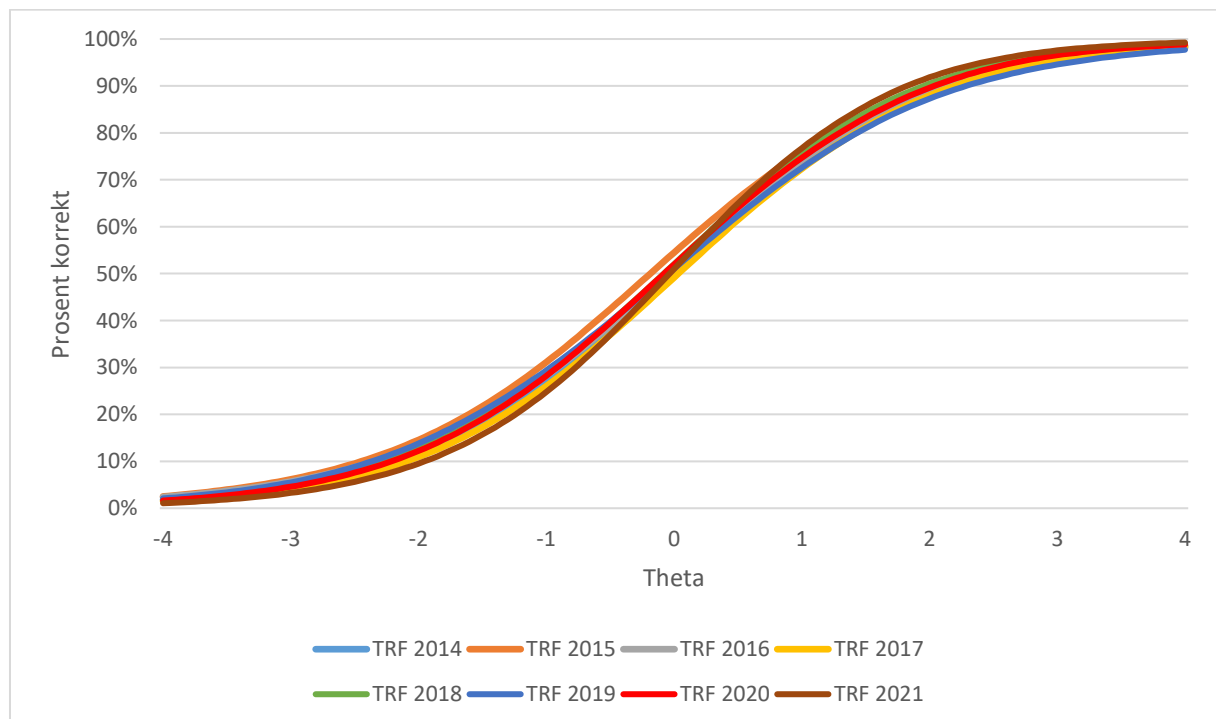


FIGUR 10 MÅLEUSIKKERHET - PRØVEN 2021

Test response function

Ved hjelp av IRT-analyse får vi en Test Response Function (TRF) som viser sammenhengen mellom dyktigheten til elevene og antall riktige besvarte oppgaver på prøven.

Figur 11 viser TRF for nasjonal prøve i regning fra og med 2014 til og med 2021 der resultatene for 2015–2021 er lenket til resultatene for 2014. Figuren viser at det er små forskjeller på prøvene, og at prøven for 2021 er lik i vanskelighetsgrad sammenlignet med de tidligere årene. Vi ser at kurven for 2021 er marginalt lavere for elever med lav theta, og marginalt høyere for elever med høy theta sammenlignet med tidligere prøver. Det tyder på at prøven for 2021 har litt færre oppgaver med veldig lav og veldig høy vanskegrad, enn flere av de foregående prøvene.



FIGUR 11 TEST RESPONSE FUNCTION FOR PRØVENE I 2014–2020

Dersom vi bare sammenligner TRF for 2020 og 2021 ser vi forskjellene mellom disse prøvene tydeligere (figur 12). For elever med lavere theta enn 0, ligger kurven for 2021 prøven lavere enn

prøven for 2020, noe som betyr at elever med denne ferdigheten i 2021 hadde lavere gjennomsnittlig løsningsprosent enn elever i 2020. For elever med høyere theta enn 0, ser vi at kurven for prøven i 2021 ligger høyere enn kurven for prøven for 2020.



FIGUR 12 TEST RESPONSE FUNCTION FOR PRØVENE I 2020–2021

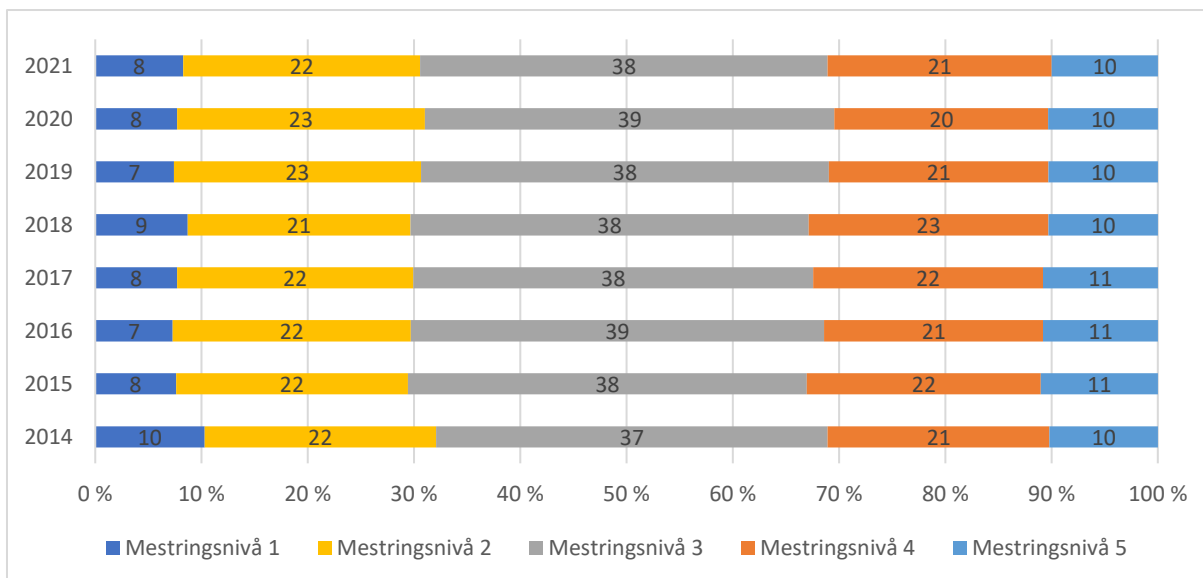
Tolkninger av resultatene

Sammenligning av resultater over tid

Innføring av ankerdesign i nasjonale prøver i 2014, gjør det mulig å måle utvikling over tid. Dette gjøres ved at omtrent 6 % av elevene gjennomfører en ankerprøve hvert år, der omtrent halvparten av oppgavene er ankeroppgaver og resten er oppgaver fra kohortprøven.

2014 er år null for vurdering av utvikling over tid. Gjennomsnittlig vanskegrad til prøven i 2014 ble satt til 50 skalapoeng med standardavvik lik 10 skalapoeng. Ved hjelp av ankeroppgavene er det mulig å sammenligne resultatene fra og med 2014 og vurdere om elevenes dyktighet har endret seg. Seks år etter innføringen av ny skala er gjennomsnittet fortsatt 50 skalapoeng. Det betyr at elevenes resultater, i den grunnleggende ferdigheten å kunne regne, har vært veldig stabil i løpet av denne perioden.

Selv om ankerdesignet i nasjonale prøver reduserer forskjellene i vanskegrad i kohortprøvene fra år til år, eksisterer det en viss lenkefeil (se figur 13 og 14). Ifølge Björnsson (2018), var lenkefeilen i 2018 på 0,2 standardavvik. Det betyr at for at en endring i resultater fra ett år til et annet skal være statistisk signifikant (med 68 prosent sikkerhet), må endringen i skalapoeng være 2 eller mer. Lenkefeilen for nasjonale prøve i regning for 8. trinn 2021 er beregnet til 0,14 standardavvik, eller litt over ett skalapoeng, etter en modell beskrevet av Björnsson.



FIGUR 13 PROSENTVIS FORDELING PÅ MESTRINGSNIVÅ 2014–2021

Figur 13 viser prosentvis fordeling av elever på de ulike mestringsnivåene for prøvene 2014–2021. Som vi ser er resultatene på nasjonalt nivå stabile, også når det gjelder fordeling på mestringsnivåer. De endringene vi ser er at 2014 skiller seg noe ut fra de øvrige årene, noe som kan gjenspeile at prøven for 2014 inneholdt åtte flere oppgaver enn prøvene fra og med 2015 har gjort.

Sammenligning av resultater for 8. og 9. trinn

Prøven for 9. trinn blir analysert ved at oppgavens vanskegrad er fastsatt ut fra analysen av prøven for 8. trinn. Dette gjør det mulig å sette resultatene for 9. trinn på samme skala som 8. trinn. I 2021 var gjennomsnittlig skalapoeng for 9. trinn 53.

Sammenligner vi gjennomsnittlig løsningsprosent (råskåre) for 8. og 9. trinn, utgjør forskjellen mer enn 30 % av standardavviket ($e = 0,37$). Ifølge Cohen, Manion og Morrison (2018) tilsvarer en effektstørrelse mellom 0,2 og 0,5 liten effekt, mens en effektstørrelse mellom 0,5 og 0,8 tilsvarer middels effekt. Det betyr at det har liten effekt for resultatet til denne prøven om elevene som løser oppgavene, er fra 9. eller 8. trinn (tabell 15).

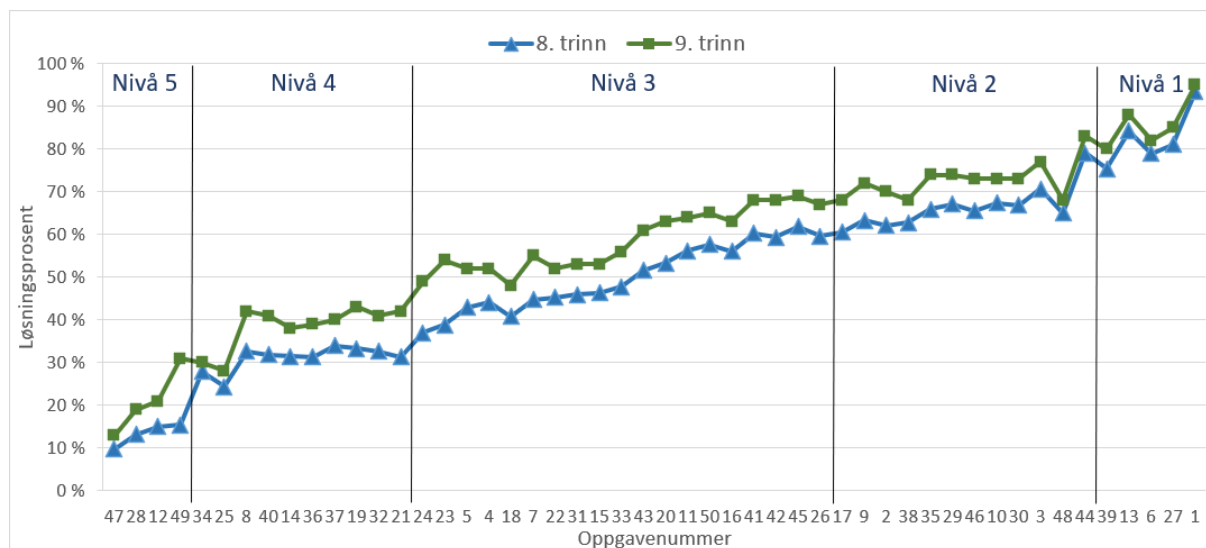
År	Effektstørrelse 8. / 9. trinn
2014	0,34
2015	0,36
2016	0,36
2017	0,38
2018	0,38
2019	0,37
2020	0,41
2021	0,37

TABELL 15 EFFEKTSTØRRELSE 8. TRINN SAMMENLIGNET MED 9. TRINN I PERIODEN 2014–2021

Tabell 15 viser at forskjellen mellom trinnene har vært veldig stabil siden 2014. Effekten er omtrent lik det en skulle kunne forvente etter ett år lengre skolegang (Ræder, Tokle & Olsen, 2019).

Figur 14 viser gjennomsnittlig løsningsprosent for hver oppgave på 8. og 9. trinn. Oppgavene er sortert etter vanskelighetsgraden oppgavene fikk på 8. trinn, og figuren viser at det ikke alltid er samsvar mellom vanskelighetsgraden og løsningsprosenten til oppgaven. Vanskelighetsgraden

defineres ut fra dyktigheten en elev må ha for at sannsynligheten for å løse oppgaven skal være 50 prosent.



FIGUR 14 P-VERDIER FOR HVER OPPGAVE PÅ 8. OG 9. TRINN. OPPGAVENE ER SORTERT ETTER VANSKEGRAD PÅ 8. TRINN.

Analysen viser at oppgave 47 krevde høyest dyktighet og oppgave 1 krevde lavest dyktighet, for at elevene skulle ha 50 % sannsynlighet for å løse oppgaven korrekt. Oppgave 47 stiller store krav til elevene når det gjelder å gjenkjenne og beskrive den matematiske utfordringen i oppgaven og velge en hensiktsmessig strategi for å gjøre beregninger. I oppgave 1 skal elevene foreta en halvering i en lett gjenbar kontekst.

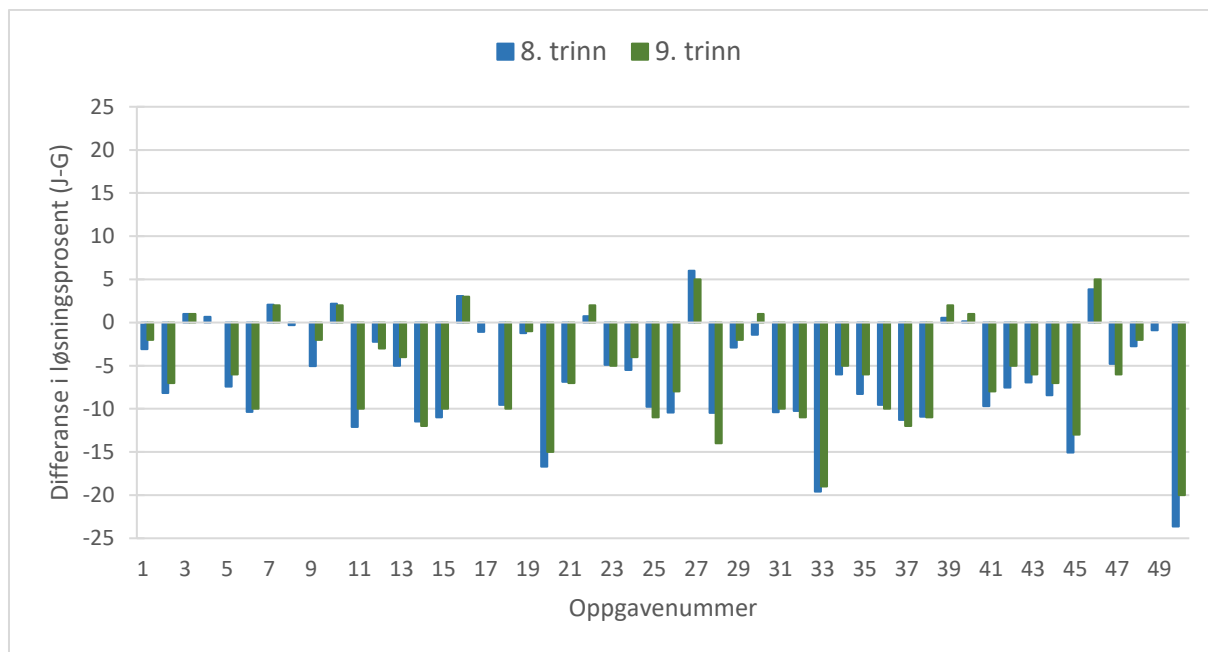
Opgavene med størst prosentvis endring i løsningsprosent fra 8. trinn til 9. trinn er oppgave 49, 28 og 12. Løsningsprosenten til disse oppgavene øker med over 40 %, noe som gjør at det er oppgaver med relativt lav løsningsprosent på 8. trinn. En endring i løsningsprosent for disse oppgavene, utgjør større prosentvis endring enn samme endring i prosentpoeng for oppgaver med høyere løsningsprosent på 8. trinn. Alle oppgavene med størst prosentvis endring har en vanskegrad på nivå 5. Det er vanskelig å plukke ut en fellesbetegnelse for hva de oppgavene med størst prosentvis endring måler, men oppgavene krever god begrepsforståelse og at elevene fleksible når det gjelder valg av strategier.

Kjønnsforskjeller

Forskjellene mellom jenter og gutter på 8. trinn var i 2021 tre skalapoeng. Guttene hadde i gjennomsnitt 51 skalapoeng, mens jentene hadde 48 skalapoeng. Forskjellen mellom jenter og gutter har variert mellom to og tre skalapoeng helt siden innføring av ny skala for nasjonale prøver i 2014. Tall fra statistikkportalen til Utdanningsdirektoratet (Utdanningsdirektoratet, 2021) viser at guttene i gjennomsnitt har flere skalapoeng enn jentene i alle fylker. Dersom vi ser på gjennomsnittlig løsningsprosent, er forskjellen i gjennomsnitt over seks prosentpoeng i guttenes favør, både for 8. trinn og 9. trinn. Denne forskjellen er høyere enn de foregående årene, men det er viktig å poengtere at forskjellen alltid vil være en konsekvens av de oppgavene som er plukket ut til endelig prøve. Derfor kan det være like interessant å se på kjønnsforskjellen i ankeret, selv om det også der er utskiftning av oppgaver fra år til år. Kjønnsforskjellen i ankeret lå lenge på tre prosentpoeng, men var i 2019 og 2020 på fem prosentpoeng. I 2021 var forskjellen på sju prosentpoeng.

Som i de foregående årene er det størst prosentandel jenter på de tre laveste mestringsnivåene og flest gutter på de to høyeste nivåene. I 2021 var det over dobbelt så stor andel gutter på mestringsnivå 5 som jenter.

Figur 15 viser forskjeller i løsningsprosent for jenter og gutter på oppgavenivå i prøven for 2021. Søylor med positiv verdi viser oppgaver som jentene har løst i større grad enn guttene, og søylor med negativ verdi viser oppgaver guttene har løst i større grad enn jentene. Det er til dels store forskjeller i løsningsprosent for jenter og gutter, og langt flere oppgaver som har høyere løsningsprosent for gutter enn jenter, enn motsatt. Vi ser også at forskjellene i guttefavør er mye større enn forskjellene i jentefavør. Fordelingen av oppgaver i guttefavør og jentefavør var den samme i preanalysen, så resultatene har ikke endret seg mye fra 2. pilot til endelig gjennomføring.



FIGUR 15 FORSKJELL I LØSNINGSPROSENT FOR JENTER OG GUTTER FOR HVER OPPGAVE, JENTE - GUTT

I prøven for 2021 er det 29 oppgaver som skiller mer enn fem prosentpoeng i løsningsprosent i guttefavør på 8. trinn. Blant de oppgavene som skiller mest omhandler de fleste måleenheter, både rene omgjøring mellom prefikser og mer sammensatte oppgaver med måleenheter. Oppgaver som omhandler måleenheter, har vært en gjenganger blant oppgaver som guttene i større grad har løst riktig enn jentene. Etter vår vurdering virker dette å være dette uavhengig av kontekst i oppgavene.

Det er én oppgave som skiller mer enn fem prosentpoeng i løsningsprosent i jentefavør på 8. trinn i 2021. Det er oppgave 46, den eneste oppgaven som i preanalysen kommer ut med bias (se kapittel 4). I oppgave 46 der elevene skal plassere punkter som danner et linjediagram slik at diagrammet viser samme resultat som en tabell. Oppgaven krever en god del nøyaktighet, noe som har vært en gjenganger for oppgaver som jentene i størst grad løser riktig sammenlignet med guttene i nasjonale prøver.

Ubesvarte oppgaver

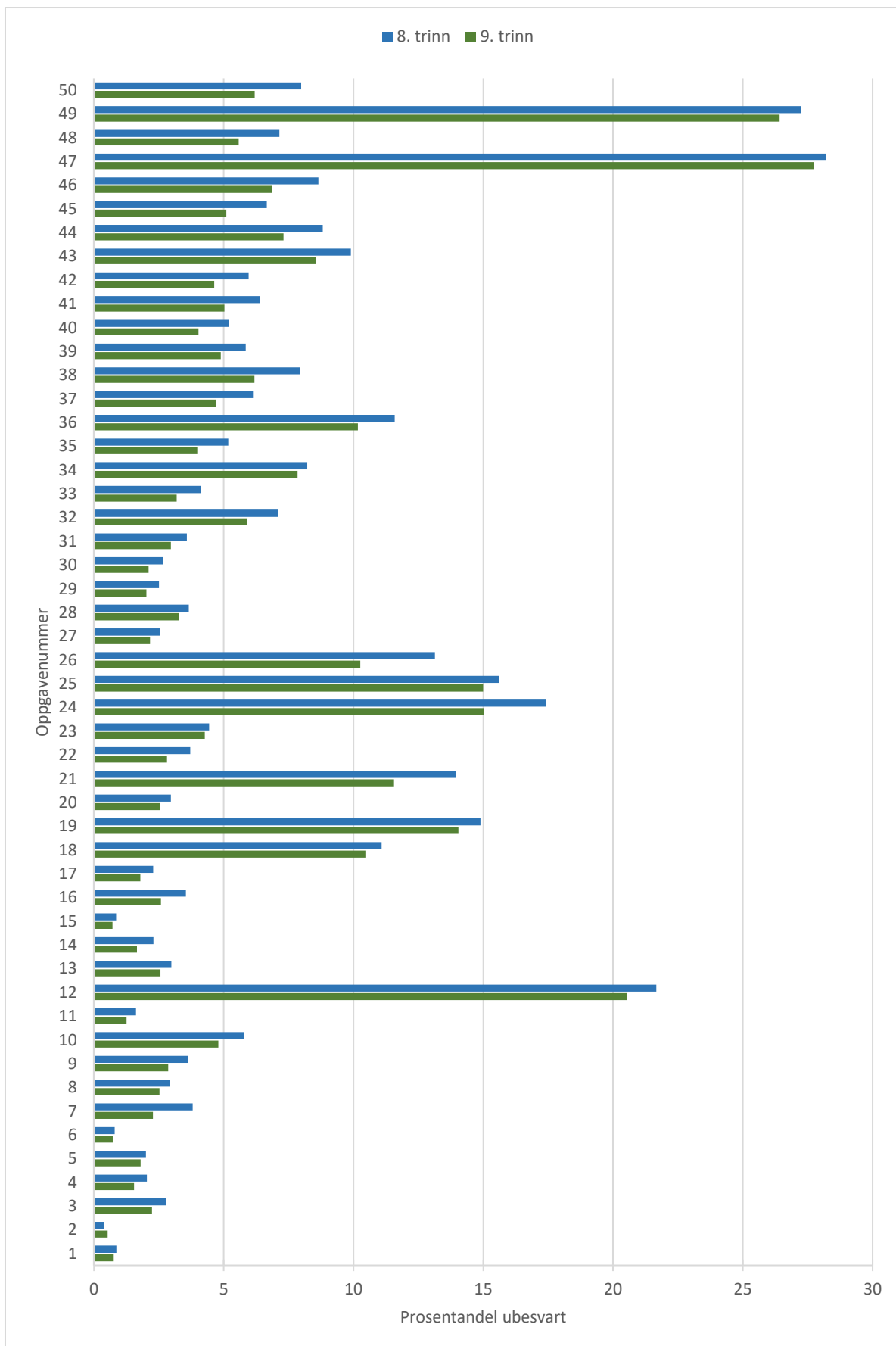
Gjennomsnittlig andel ubesvart i 2021 var 7,0 prosent for 8. trinn og 6,1 prosent for 9. trinn. Dette er noe høyere enn tilsvarende tall for tidligere prøver. Trenden med økende ubesvart fra år til år, er én av årsakene til at Matematikksenteret har foreslått at prøven fra og med 2022 reduseres i omfang.

Figur 16 viser prosent ubesvart for hver oppgave for elever på 8. og 9. trinn. Det er ti oppgaver på 8. trinn der andelen ubesvart er større enn 10 %. Dette er omtrent samme antall som for tidligere prøver.

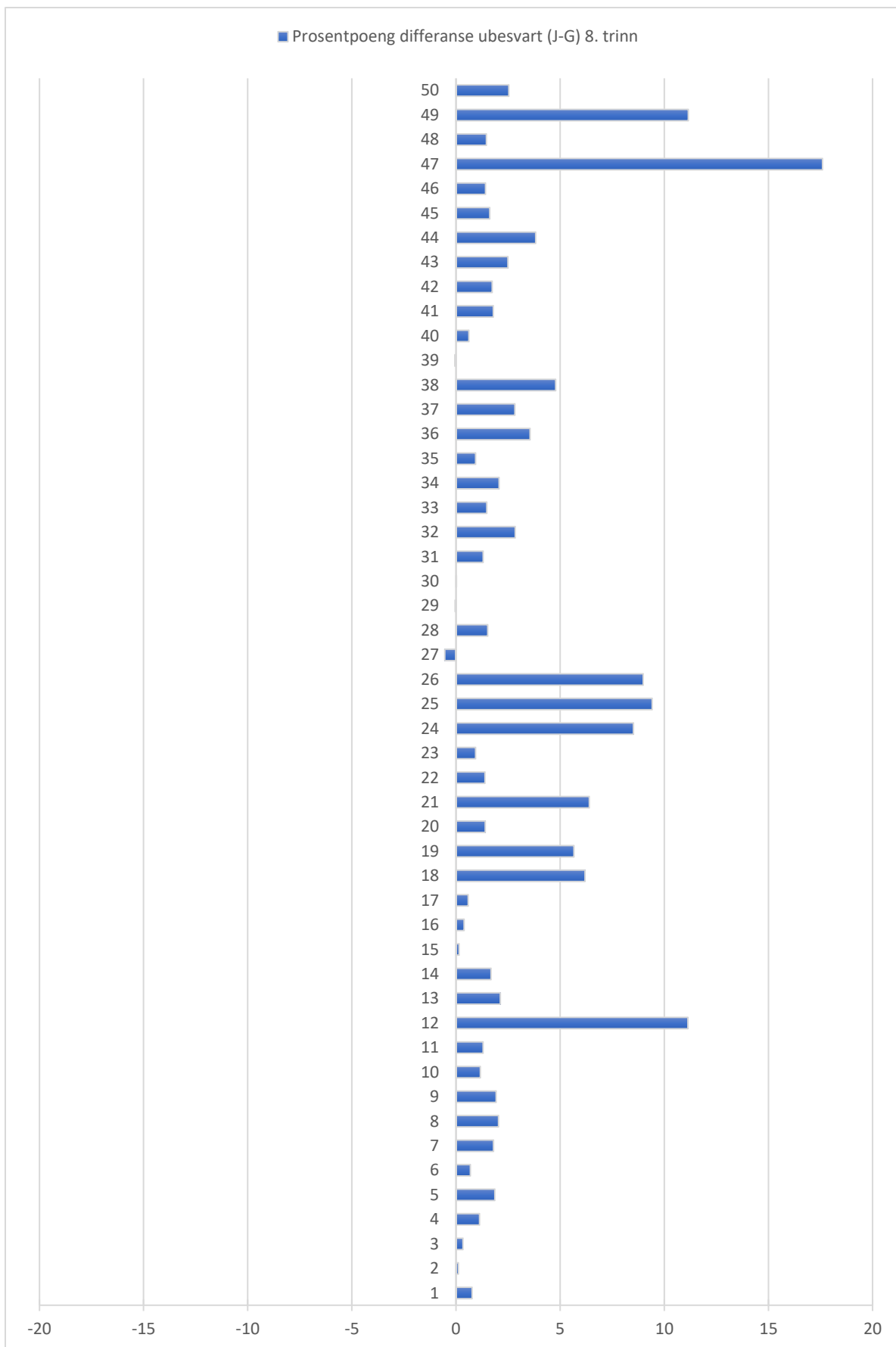
Det er en tendens til at andelen ubesvarte oppgaver øker mot slutten av prøven. Prøven er bygd opp slik at det er flere vanskelige oppgaver mot slutten enn i begynnelsen (Figur 6, kapittel 4), selv om

oppgavene ikke er sortert etter vanskegrad. Dette kan være en medvirkende årsak til økende tendens for ubesvart utover i prøven. Det er også en tendens til at andelen jenter som ikke svarer, øker mer i siste del av prøven enn andel gutter som ikke svarer.

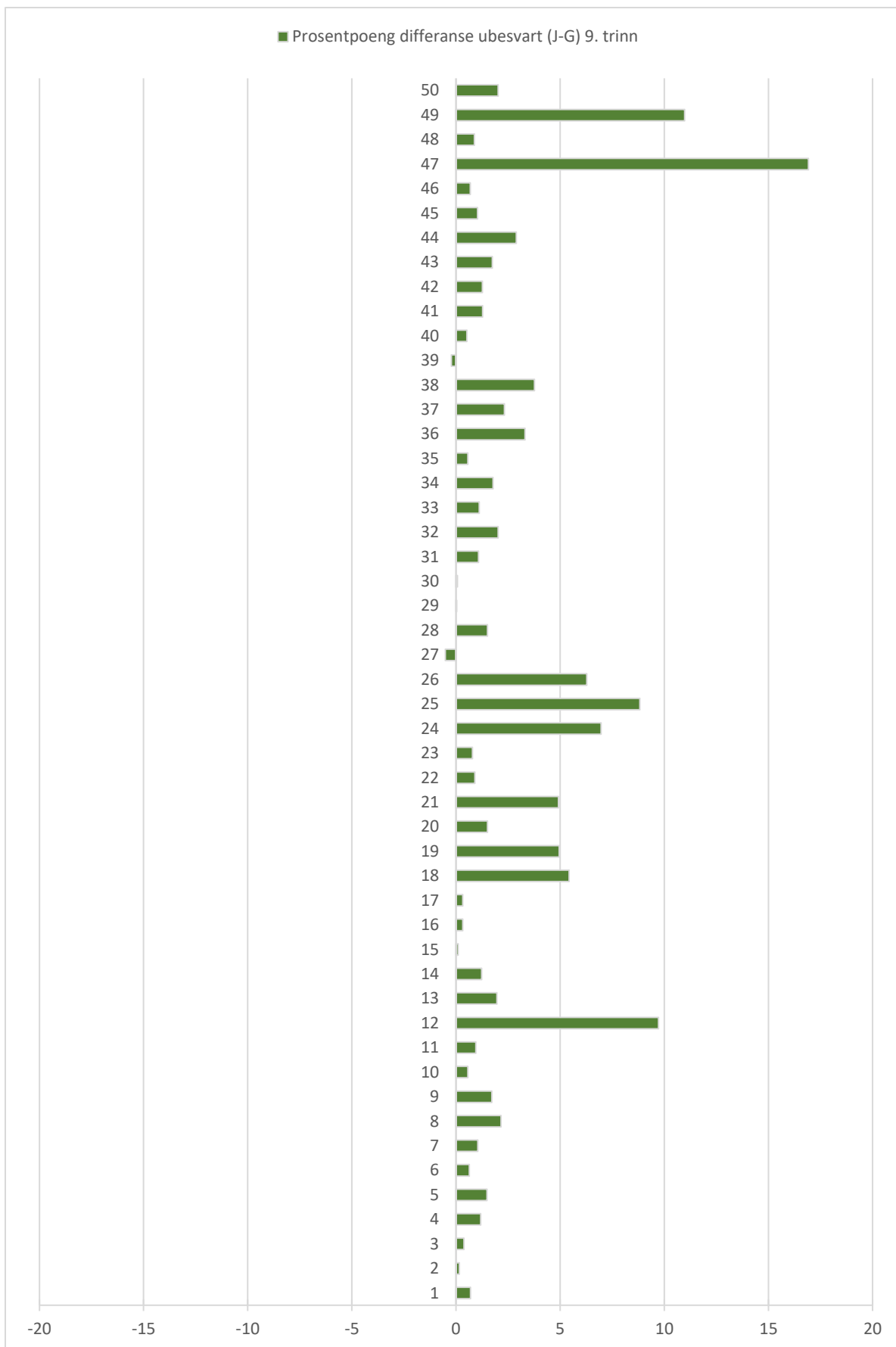
Figur 17 viser forskjellen i andel ubesvarte for jenter og gutter på 8. trinn, og figur 18 viser forskjellen i andel ubesvarte for jenter og gutter på 9. trinn. Det er stort samsvar mellom figur 17 og figur 18, og vi ser at det nesten ikke er oppgaver, der andelen ubesvart er større blant guttene enn blant jentene (søylar med negativ). I forberedelsene til de nasjonale prøvene bør derfor lærere være bevisst på å motivere elevene til å besvare oppgaver hvor de også er litt usikre på hva som er riktig svar.



FIGUR 16 PROSENT UBESVART FOR HVER OPPGAVE FOR ELEVER PÅ 8. OG 9. TRINN



FIGUR 17 FORSKJELL I PROSENT UBESVART FOR JENTER OG GUTTER FOR HVER OPPGAVE PÅ 8. TRINN



FIGUR 18 FORSKJELL I PROSENT UBESVART FOR JENTER OG GUTTER FOR HVER OPPGAVE PÅ 9. TRINN

Oppgaveanalyse for alle oppgavene

Oppgaveanalyse for alle oppgavene er vist i tabell 16 og tabell 17.

- I kolonnen forskjell j-g prosentpoeng, betyr positive verdier at jentene har høyere løsningsverdi enn guttene, og negative verdier at guttene har høyere løsningsprosent enn jentene
- F betyr flervalgsoppgave. 1, 2, 3, 4 og 5 er svaralternativene
- Å betyr åpen oppgave, 0 betyr galt svar og 1 betyr rett svar

Tabell 27 og 28: Oppgaveanalyse for alle oppgavene i kohortprøvene. Svarfordeling angitt i prosent. Gjennomsnittlig dyktighet (θ) for hvert svaralternativ og for rett eller galt i de åpne oppgavene. Point-biserial betyr oppgavens korrelasjon med summen av poeng. a-verdi betyr oppgavens diskriminering i IRT-analyse, og b-verdien er oppgavens vanskegrad på en skala fra -3 til 3 ut fra IRT-analyse. Guttene har høyest løsningsprosent i de oppgavene hvor differansen (j-g) er negativ. IRT-analysen for 9. trinn er foretatt som en uavhengig analyse. Dette betyr at oppgaveparametere fra 8. trinn ikke er brukt i analysene, slik de er ved fastsettelse av skalapoeng for elever på 9. trinn.

NP Regning 9. trinn 2021																					
Nr.	F Å	Svarfordeling i %						Ube- svart	Gjennomsnittlig Theta						T- Rpbis	S- Rpbis	Løsningsprosent				
		1	2	3	4	6	6		1	2	3	4	5	6			Alle	J	G	J-G	
1	Å	4	95					1	-1,28	0,18					-1,05	0,330	0,301	95,2	94,1	96,3	-2,2
2	F	20	9	1	70			0	-0,57	-0,33	-0,79	0,38			-0,46	0,415	0,389	69,7	65,9	73,2	-7,3
3	Å	21	77					2	-0,59	0,33					-0,87	0,414	0,388	76,6	77,1	76,2	0,9
4	F	26	52	16	5			1	0,17	0,42	-0,80	-0,33			-0,52	0,329	0,304	51,9	52,4	51,5	0,8
5	F	12	5	13	52	15		2	0,08	-0,44	-0,33	0,49	-0,55		-0,25	0,400	0,374	52,3	49,2	55,3	-6,1
6	F	5	82	10	3			1	-0,32	0,26	-0,62	-1,03			-0,57	0,346	0,324	82,4	77,4	87,3	-9,9
7	F	55	22	21				2	0,45	-0,40	-0,17				-0,47	0,383	0,356	55,1	55,9	54,2	1,8
8	F	28	16	42	12			2	-0,46	-0,01	0,60	-0,01			-0,30	0,425	0,391	41,9	41,6	42,2	-0,5
9	Å	25	72					3	-0,70	0,43					-0,75	0,524	0,498	72,0	70,6	73,3	-2,7
10	Å	22	73					5	-0,52	0,36					-0,85	0,443	0,419	73,2	74,2	72,2	1,9
11	F	3	21	64	11			1	-0,71	-0,53	0,40	-0,08			-0,39	0,409	0,384	64,1	58,5	69,4	-10,9
12	Å	58	21					21	-0,04	0,94					-0,33	0,446	0,401	21,4	20,2	22,5	-2,3
13	Å	9	88					3	-0,91	0,25					-0,82	0,410	0,382	87,8	85,8	89,6	-3,8
14	F	38	35	25				2	0,72	-0,15	-0,40				-0,52	0,481	0,446	38,0	32,2	43,6	-11,4
15	F	10	53	10	6	18	3	0	-0,17	0,49	-0,26	-0,68	-0,25	-0,70	-0,52	0,420	0,390	53,0	48,2	57,6	-9,5
16	Å	23	64					3	-0,29	0,37					-0,91	0,356	0,328	63,5	64,9	62,1	2,8
17	Å	29	69					2	-0,49	0,39					-0,70	0,431	0,404	68,5	68,5	68,5	0,0
18	Å	41	48					11	-0,41	0,68					-0,41	0,568	0,532	47,8	42,8	52,7	-9,9
19	Å	43	43					14	-0,26	0,68					-0,48	0,512	0,478	42,8	42,1	43,5	-1,4
20	F	14	63	9	12			2	-0,14	0,47	-0,72	-0,76			-0,41	0,492	0,467	62,6	55,0	70,0	-14,9
21	Å	45	43					12	-0,32	0,77					-0,61	0,589	0,547	42,5	39,3	45,6	-6,3
22	F	5	52	23	10	7		3	-0,40	0,54	-0,23	-0,61	-0,36		-0,52	0,469	0,438	52,4	53,4	51,5	1,9
23	Å	42	54					4	-0,47	0,62					-0,55	0,567	0,535	53,7	51,2	56,0	-4,7
24	Å	36	49					15	-0,44	0,71					-0,49	0,611	0,572	48,7	46,7	50,7	-4,0
25	Å	57	28					15	-0,09	0,84					-0,52	0,479	0,434	28,3	22,7	33,7	-11,0
26	Å	23	67					10	-0,44	0,42					-0,67	0,486	0,459	67,1	62,7	71,3	-8,6
27	Å	13	85					2	-0,57	0,24					-0,74	0,346	0,325	84,7	87,2	82,3	4,9
28	F	26	36	16	19			3	-0,32	-0,12	0,36	1,00			-0,26	0,439	0,390	19,1	12,1	25,9	-13,7
29	F	9	74	10	5			2	-0,47	0,39	-0,83	-0,87			-0,55	0,507	0,483	74,1	73,3	74,9	-1,6
30	F	73	17	7				3	0,32	-0,34	-0,76				-0,57	0,392	0,366	73,4	73,7	73,1	0,5
31	F	27	53	8	9			3	-0,28	0,51	-0,39	-0,48			-0,32	0,446	0,417	53,1	48,3	57,8	-9,4
32	Å	53	41					6	-0,26	0,73					-0,82	0,537	0,499	40,5	35,0	45,9	-10,9
33	F	8	32	57				3	-0,52	-0,46	0,56				-0,48	0,547	0,515	56,5	47,1	65,5	-18,5
34	Å	62	30					8	-0,07	0,64					-0,49	0,369	0,335	30,2	27,7	32,6	-4,9
35	Å	22	74					4	-0,59	0,36					-0,74	0,478	0,452	74,3	71,2	77,3	-6,1
36	Å	51	39					10	-0,22	0,72					-0,57	0,518	0,479	38,9	33,5	44,0	-10,5
37	F	24	29	40	3			4	-0,24	-0,17	0,63	-0,58			-0,48	0,454	0,417	40,1	34,3	45,7	-11,4
38	F	5	68	19	2			6	-0,45	0,35	-0,36	-0,63			-0,39	0,409	0,383	68,2	62,7	73,6	-11,0
39	F	15	80					5	-0,44	0,26					-0,63	0,380	0,353	80,0	81,0	79,0	2,0
40	F	7	10	11	27	41		4	-0,41	-0,30	-0,29	-0,16	0,65		-0,45	0,472	0,437	40,5	40,7	40,4	0,4
41	F	68	11	14	3			4	0,51	-0,66	-0,87	-0,89			-0,45	0,639	0,613	68,0	64,1	71,8	-7,7
42	F	3	9	16	68			4	-0,63	-0,57	-0,64	0,43			-0,42	0,534	0,507	68,3	65,6	70,9	-5,3
43	Å	30	61					9	-0,47	0,49					-0,56	0,537	0,505	60,9	58,2	63,5	-5,3
44	Å	10	83					7	-0,91	0,29					-0,64	0,514	0,484	83,3	79,9	86,6	-6,8
45	F	69	16	10				5	0,47	-0,73	-0,74				-0,45	0,604	0,575	68,6	62,1	74,9	-12,8
46	Å	20	73					7	-0,44	0,32					-0,52	0,426	0,400	73,3	75,7	71,0	4,7
47	Å	59	13					28	0,05	1,03					-0,20	0,378	0,329	13,3	10,0	16,4	-6,4
48	F	9	68	14	3			6	-0,19	0,27	-0,18	-0,33			-0,34	0,296	0,273	68,0	67,1	68,9	-1,8
49	Å	43	31					26	-0,21	0,92					-0,32	0,577	0,531	30,8	30,9	30,7	0,2
50	F	65	18	6	4			7	0,45	-0,39	-0,82	-0,94			-0,38	0,530	0,499	65,2	54,8	75,3	-20,5
								6,1							0,461	0,430	57,7	55,0	60,4	-5,4	

TABELL 17 ITEM-ANALYSE FOR ALLE OPPGAVER NP9 2021

Oppsummering

Resultatene fra nasjonale prøver i regning 2021 for 8. trinn viser ingen endringer når det gjelder gjennomsnittlig antall skalapoeng for hele elevgruppa. Ser vi på fordeling innen de ulike mestringsnivåene, er det også snakk om svært marginale endringer, og ikke større endringer enn mellom andre år. Koronasituasjonen har ført til annerledes opplæringstilbud for mange elever, men det er ingen ting fra analysene fra nasjonale prøver i regning for 8. trinn som tyder på at hverken 8. trinns elever i 2020 eller 2021 viser lavere måloppnåelse i den grunnleggende ferdigheten i å kunne regne, sammenlignet med tidligere elevkull.

Som denne rapporten antyder kan det virke som om forskjellene i resultater for gutter og jenter har økt de siste årene. Siden vi gjennom vår lange erfaring med å utvikle og analysere oppgaver, vet at noen matematiske områder vil med veldig stor sannsynlighet gi store forskjeller i løsningsprosent, er dette en vanskelig avveining vi må ta når vi setter sammen prøver. Matematikksenteret kan utvikle prøver der forskjellene mellom gutter og jenter blir betydelig lavere, men etter vår mening vil ikke disse prøvene være valide og gjenspeile læreplanen på en god måte. For å studere kjønnsforskjeller nærmere, er Matematikksenteret i et samarbeid med ILS og miljøet som har ansvaret for TIMSS (Trends in Mathematics and Science) i Norge. Formålet med det samarbeidet er å undersøke om vi finner noen forskjeller, eller likheter, mellom kjønnsforskjeller i de to prøvene.

En prøve i den grunnleggende ferdigheten å kunne regne krever oppgaver i kontekst. Selv om vi jobber mye med tekst, og prøver å gjøre den så lettlest og presis som mulig, kan vi ikke garantere at leseferdigheten ikke påvirker enkeltelevers prestasjoner i nasjonale prøver i regning. Vi håper derfor at lesestøtte for elever som har nytte av det blir prioritert i det nye prøvegjennomføringssystemet som skal tas i bruk for prøvene fra og med høsten 2022, slik at vi kan redusere forstyrrende faktorer i testingen av elevenes regneferdigheter.

Referanser

- Assessment Systems Corporation. (2014). Xcalibre 4.2 IRT Item parameter calibration (Versjon 4.2.2.0).
- Björnsson, J. K. (2018). Om lenkefeil og ekvivaleringsmetoder på nasjonale prøver: Evaluering av endring over tid. *Acta Didactica Norge*, 12(4), 1-24.
- Cohen, L., Manion, L. & Morrison, K. (2018). *Research methods in education* (8. utg.). New York: Routledge.
- Ræder, H. G., Tokle, O. D. & Olsen, R. V. (2019). Utvikling av nasjonale prøver – rapport 1 og 2a. Utdanningsdirektoratet. (2006). Læreplanverket. Hentet 24.10 2019 fra <https://www.udir.no/laring-og-trivsel/lareplanverket/>
- Utdanningsdirektoratet. (2013). Læreplan i matematikk fellesfag (MAT1-04). Hentet 19.02.19 fra <https://www.udir.no/kl06/MAT1-04>
- Utdanningsdirektoratet. (2017). *Rammeverk for nasjonale prøver*. www.udir.no. Hentet fra <https://www.udir.no/eksamen-og-prover/prover/rammeverk-for-nasjonale-prover/>
- Utdanningsdirektoratet. (2021, 14.02.2021). Statistikk. I. Hentet fra <https://www.udir.no/tall-og-forskning/statistikk/statistikk-grunnskole/nasjonale-prover-8.-og-9.-trinn/>