



Matematikksenteret
Nasjonalt senter for matematikk i opplæringen

TEKNISK RAPPORT

NASJONAL PRØVE I REGNING 8. og 9. TRINN 2018

**Olav Dalsegg Tokle
Grethe Ravlo
Ole Harald Johansen
Stig Atle Myhre
Marte Thoresen**

MATEMATIKKSENTERET, NTNU 2019

Nasjonal prøve i regning for 8. (9.) trinn 2018

Olav Dalsegg Tokle, Grethe Ravlo, Stig Atle Myhre, Ole Harald Johansen og Marte Thoresen

Nasjonalt senter for matematikk i opplæringen (Matematikksenteret)

NTNU mars 2019

INNLEDNING	4
KAPITTEL 1: PREPILOT.....	5
KAPITTEL 2: 1. PILOT	6
KAPITTEL 3: 2. PILOT	8
HVORDAN 2. PILOTENE BLE SATT SAMMEN.....	8
GJENNOMFØRING AV 2. PILOTENE.....	9
KAPITTEL 4: SAMMENSETNING AV KOHORTPRØVE	11
INNLEDNING	11
PSYKOMETRI	11
VALIDITET	14
FORKLARING TIL NOEN OPPGAVER I ENDELIG PRØVE	16
OPPGAVER MED BIAS	16
OPPGAVER DER DISTRAKTORER HAR POSITIV POINT-BISERIAL.....	19
ANKEROPPGAVER	20
GENERELLE VURDERINGER AV PRØVEN SOM HELHET.....	21
KAPITTEL 5: ANALYSE AV NASJONAL PRØVE I REGNING 2018.....	22
INNLEDNING	22
PSYKOMETRISKE VERDIER	23
OVERSIKT OVER OPPGAVER NP8 (9) 2018	23
TESTINFORMASJON OG MÅLEUSIKKERHET	24
TEST RESPONSE FUNCTION	25
TOLKNINGER AV RESULTATENE.....	27
SAMMENLIGNING AV RESULTATER OVER TID.....	27
SAMMENLIGNING AV RESULTATER FOR 8. OG 9. TRINN	27
KJØNNSFORSKJELLER	29
UBESVARTE OPPGAVER	30
OPPGAVEANALYSE FOR ALLE OPPGAVENE.....	34
REFERANSER	37

Innledning

Denne rapporten beskriver prosessen for utvikling, gjennomføring og analyse av resultatene til nasjonal prøve i regning for 8. og 9. trinn 2018. Det tok to år å utvikle prøven, og oppgavene var med i tre piloteringsrunder før de ble godkjent til å være med i den nasjonale prøven. De tre første kapitlene beskriver prosessen rundt de tre utprøvingene som er gjennomført, kalt prepilot, 1. pilot og 2. pilot, og sier litt om formålet med og utfordringer i de ulike piloteringsfasene.

Kapittel 4 beskriver prosessen og begrunner sammensetningen av endelig prøve, mens kapittel 5 er en analyse og betraktninger rundt resultatene etter at nasjonal prøve for 2018 er gjennomført.

Gjennom hele utviklingsprosessen brukes klassisk test teori (CTT) og Item Response Theory (IRT) til å analysere resultatene. De utvalgte skolene som deltar i piloteringene, er trukket av Utdanningsdirektoratet og tilfredsstiller kravene til tilfeldige og representative utvalg.

For de ulike kapitlene er de viktigste analyseresultatene gjengitt i denne rapporten.

Kapittel 1: Prepilot

Første utprøving ble gjennomført i februar 2017 med elever på 8. trinn. Det var åtte prepiloter med fra 37 til 39 oppgaver i hver prøve, i alt 300 oppgaver som ble prøvd ut. Alle, unntatt 84 oppgaver som var rest fra tidligere utvikling, var nyutviklet spesielt for prøven til 2018. Medlemmer av prøveutviklingsgruppen var ute på skoler og observerte elever som løste oppgavene.

Oppgavene vi utvikler skal være regning i fag på fagenes premisser og måle grunnleggende ferdigheter slik dette er beskrevet i Rammeverket for grunnleggende ferdigheter (Utdanningsdirektoratet, 2017a), Rammeverk for nasjonale prøver (Utdanningsdirektoratet, 2017b) og i fagplanene til LK06 (Kunnskapsdepartementet & Utdanningsdirektoratet, 2006). Fagtilknytningene til en oppgave er derfor 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 Rammeverket for nasjonale prøver (Utdanningsdirektoratet, 2017b) 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 delprosessene 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, heretter kalt kohortprøven, ønsker vi en stor andel flervalgsoppgaver. Hovedhensikten med prepilotene er derfor å finne høyfrekvente gale elevsvar som viser en misoppfatning eller en typisk feiltenking. Disse feilsvarene blir distraktorer når en åpen oppgave endres til en flervalgsoppgave som skal prøves i neste piloteringsrunde.

Når vi velger oppgaver til prøvesett 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 kohortprøver, og som vi bruker til å sammenligne vanskelighetsgraden til prepilotene. I prepilotene til 2018-prøven er det til sammen ni slike prepilotanker. Åtte oppgaver er fordelt med én i hver pilot, mens én oppgave er felles i alle settene.

Det var til sammen 62 skoler og 2 247 elever som deltok i prepilot til 2018. Svarprosenten er høy, fra 93 til 96 prosent. Se tabell 1 for utvalg og resultater.

Prepiloter NP8 2018	Prepilot A	Prepilot B	Prepilot C	Prepilot D	Prepilot E	Prepilot F	Prepilot G	Prepilot H
Antall gutter	142	156	159	139	121	136	134	139
Antall jenter	141	155	148	131	144	135	126	141
Totalt antall elever	283	311	307	270	265	271	260	280
Gj.snitt Theta gutter	0,063	0,000	0,089	0,010	0,120	0,133	-0,018	-0,031
Gj.snitt Theta jenter	-0,022	-0,076	-0,029	0,038	-0,136	-0,092	0,047	-0,099
Gj.snitt Theta totalt	0,021	-0,038	0,032	0,024	-0,019	0,021	0,014	-0,065
Sign. forskjell J-G (Theta) $p < 0,05$	Nei	Nei	Nei	Nei	Ja	Ja	Nei	Nei
Standardavvik Theta	0,960	0,924	0,957	0,996	0,968	0,941	0,958	0,895
SE til gjennomsnitt Theta	0,057	0,052	0,055	0,061	0,059	0,057	0,059	0,053
Gj.snitt P-verdi gutter	56,7 %	58,0 %	65,3 %	58,0 %	59,2 %	57,8 %	53,3 %	55,7 %
Gj.snitt P-verdi jenter	55,7 %	56,7 %	63,1 %	58,9 %	54,0 %	53,9 %	54,8 %	54,9 %
Gj.snitt P-verdi totalt	56,2 %	57,4 %	64,2 %	58,4 %	56,4 %	55,9 %	54,0 %	55,3 %
Svarprosent	93,6 %	96,1 %	96,4 %	94,3 %	93,8 %	92,8 %	92,7 %	94,8 %
Cronbachs alpha	0,889	0,876	0,901	0,907	0,889	0,880	0,884	0,868
Antall item	38	38	39	37	37	37	37	37
Antall MC	8	7	12	8	7	9	6	7
Antall IA	3	5	2	4	3	2	1	4

Tabell 1 Resultater fra prepilot

Kapittel 2: 1. pilot

Andre utprøving, 1. pilot, ble gjennomført i april 2017. Det var til sammen fire prøvesett som hver inneholdt 50 oppgaver. 1. pilotene gjennomføres med elever på 7. trinn som på dette tidspunktet stemmer best i alder med den elevgruppen prøven er ment for. At elevene er fire måneder for unge, kan være en utfordring når det gjelder oppgavenes vanskelighetsgrad, men erfaring har vist at dette har mindre betydning enn vi hadde forventet.

En viktig jobb etter at analysene fra prepilotene foreligger, er å finne flest mulig åpne oppgaver som kan egne seg som flervalgsoppgaver. Vi går gjennom frekvenstabeller med svarprosenten på elevsvar til alle utprøvede oppgaver og leter etter distraktorer som viser misoppfatninger og feiltenkninger.

Figur 1 og 3 og tabell 2 viser arbeidsgangen med en åpen oppgave fra prepilot til den er blitt en flervalgsoppgave som skal prøves i 1. pilot. Figur 1 viser oppgaven som den var i prepilot.

Sonja har kveldsjobb i en klesbutikk og skal skrive på reklameplakater. Hun har skrevet førpris og nåpris på plakaten til høyre.

Sonja skal nå skrive hvor mange prosent prisen til buksa er satt ned.

Hvor mange prosent er prisen satt ned med?

Svar: %



Figur 1 Item 5131174 som åpen oppgave

Tabell 2 viser høyfrekvente svar på oppgaven og vår antagelse om hva slags tenkning som kan ligge bak feilsvarene. I noen tilfeller kan elevenes kladdemark fra prepilot hjelpe oss i denne fasen.

Elevsvar	Prosentandel	Mulig strategi
40	38,1	Riktig svar
60	17,7	Regner ut hvor stor del av førpris de må betale ($300 : 500 = 60\%$)
20	17,4	200 kr nedsatt blir til 20 %
15	4,5	Kan komme av at elevene bruker tallene i oppgaven og regner ut $5 \cdot 3$
25	4,2	
45	3,4	
Andre svar	10,8	
Ubesvart	4,2	

Tabell 2 Høyfrekvente svar item 5131174 som åpen oppgave

Figur 2 viser oppgaven som flervalg slik den ble prøvd i 1. pilot.

Sonja har kveldsjobb i en klesbutikk og skal skrive på reklameplakater. Hun har skrevet førpris og nåpris på plakaten til høyre.

Sonja skal nå skrive hvor mange prosent prisen til buksa er satt ned.

Hvor mange prosent er prisen satt ned?

☐ 15 %

☐ 20 %

☐ 40 %

☐ 60 %



Figur 2 Item 5131174 som flervalgsoppgave i 1. pilot

Et feilsvar betegnes som høyfrekvent når svarfrekvensen er 4–5 prosent eller høyere. I tillegg må riktig svar og distraktorene som velges, utgjøre en stor andel av alle elevsvarene til sammen for at en oppgave skal kunne endres fra åpen til flervalg. En annen viktig faktor er å få fram ulike feiltenkninger gjennom distraktorene. Det kan føre til at vi noen ganger prøver ut distraktorer i 1. pilot som har lavere frekvens enn ønskelig. Det er ikke alle oppgaver som kan gi gode elevsvar til distraktorer. Oppgaver med høy løsningsprosent og vanligvis også oppgaver med svært lav løsningsprosent gir sjelden gode elevsvar som kan brukes som distraktorer. Hvis en oppgave skal overleve som flervalgsoppgave med fire svaralternativer, må vi forvente en løsningsprosent høyere enn 25 prosent på oppgaven.

Nyproduksjon av oppgaver mellom prepilot og 1. pilot er alltid nødvendig for å erstatte oppgaver som for eksempel ikke har gode nok psykometriske verdier til å overleve etter prepiloten. Vi må hele tiden være sikre på at det finnes et bredt utvalg av oppgaver å velge mellom for alle mestringsnivåer, både når det gjelder relevans for de ulike prosessene gjenkjenne og beskrive, bruke og bearbeide og reflektere og vurdere, og når det gjelder matematiske områder og relevante kontekster. Primært er det oppgaver fra prepilot som er endret fra åpne til flervalg som blir pilotert i 1. pilotene, men også nyproduserte åpne oppgaver og oppgaver der tekst, tabeller/diagram og tall er blitt endret etter prepiloten, inngår i oppgavesettene.

Det var i alt 1659 elever fra 65 skoler som deltok i 1. pilot. Tabell 3 viser utvalg og resultater.

1. piloter NP8 2018	1. pilot A	1. pilot B	1. pilot C	1. pilot D
Antall gutter	208	218	188	205
Antall jenter	224	207	181	228
Totalt antall elever	432	425	369	433
Gj.snitt Theta gutter	0,040	0,034	0,091	0,078
Gj.snitt Theta jenter	-0,052	-0,031	-0,095	-0,062
Gj.snitt Theta totalt	-0,008	0,002	0,000	0,004
Sign.forskjell J - G (Theta) $p < 0,05$	Nei	Nei	Nei	Nei
Standardavvik Theta	0,965	0,969	0,962	0,910
SE til gjennomsnitt Theta	0,046	0,047	0,050	0,044
Gj.snitt P-verdi gutter	47,7 %	45,1 %	47,1 %	49,2 %
Gj.snitt P-verdi jenter	46,3 %	44,0 %	43,6 %	46,8 %
Gj.snitt P-verdi totalt	47,0 %	44,6 %	45,4 %	48,0 %
Svarprosent	93,1 %	94,6 %	92,0 %	94,0 %
Cronbachs alpha	0,891	0,904	0,911	0,882
Antall item	50	50	50	50
Antall MC	25	26	30	27
Antall IA	5	2	2	3

Tabell 3 Resultater til 1. pilot

Kapittel 3: 2. pilot

Hvordan 2. pilotene ble satt sammen

På bakgrunn av resultatene fra pre-pilotene og 1. pilotene setter vi sammen to ulike 2. piloter som hver for seg selv skal tilfredsstille kravene til en ferdig nasjonal prøve. Det betyr at oppgaverekkefølge, fordeling på mestringsnivå, fordeling på matematiske hovedtema og andel flervalgsoppgaver er så lik en kohortprøve som mulig.

Aktuelle oppgaver til 2. pilotene er oppgaver som fungerte som åpne oppgaver eller flervalgsoppgaver i 1. pilotene, og oppgaver som fungerte som åpne oppgaver i pre-pilotene. For at en flervalgsoppgave skal være aktuell for 2. pilotene, må den ha minimum to fungerende distraktorer i tillegg til det riktige svaret. Dette innebærer at distraktorene må være benyttet av minimum 4–5 prosent av elevene, og at gjennomsnittlig dyktighet til de som svarer riktig, skal være klart høyere enn for de som velger et annet alternativ.

I prosessen med å velge ut oppgaver tenker vi hele tiden på to parallelle prøver. Det betyr at vi gjennom hele prosessen plukker ut oppgaver som kan stå på lik plass i de to 2. pilotene. Erfaringen vår er at P-verdien til oppgavene øker noe fra 2. pilot til kohortprøve. Tilsvarende endring skjer også med oppgavens b-verdi, som er et annet mål på vanskelighetsgrad. Dette tar vi hensyn til når vi ut fra vanskelighetsgrad velger oppgaver til 2. piloten.

Oppgavesettene til 2. pilotene må i tillegg ta hensyn til ankeret til den fremtidige kohortprøven. Vi forsøker derfor å finne oppgaver som kan fungere som skyggeoppgaver, det vil si oppgaver så like som mulig når det gjelder format, vanskelighetsgrad, innhold og psykometriske verdier, til oppgavene som utgjør ankeret. Det gjør at ankeroppgavene legger føringer for hvilke oppgaver som skal plasseres på tilsvarende oppgavenummer når 2. pilotene settes sammen.

Vi satte sammen to 2. piloter, A og B, som hver bestod av 50 oppgaver. Tabell 4, 5 og 6 viser sammensetning før gjennomføring.

	2. pilot A	2. pilot B
Tall og algebra	26	23
Måling og geometri	18	18
Statistikk og sannsynlighet	6	9

Tabell 4 Fordeling av oppgaver innen områder

	2. pilot A	2. pilot B
Mestringsnivå 1	5	4
Mestringsnivå 2	11	10
Mestringsnivå 3	17	17
Mestringsnivå 4	9	9
Mestringsnivå 5	8	9

Tabell 5 Fordeling av oppgaver innen mestringsnivå før gjennomføring

Før gjennomføring	2.pilot A	2.pilot B
Gj.snitt P-verdi gutter	49,1 %	48,6 %
Gj.snitt P-verdi jenter	46,6 %	46,6 %
Gj.snitt P-verdi totalt	47,9 %	47,6 %
Svarprosent	93,9 %	92,3 %
Gj.snitt S-Rpbis	0,410	0,426
Gj.snitt T-Rpbis	0,444	0,458
Gj.snitt a-verdi	0,74	0,76
Gj.snitt b-verdi	0,055	0,205
Antall item	50	50
Antall MC	28	25
Antall IA	2	3

Tabell 6 Tekniske data før gjennomføring

Gjennomføring av 2. pilotene

2. pilotene for 2018 ble gjennomført av elever på 8. trinn høsten 2017. Dette skjedde et par uker før de samme elevene skulle gjennomføre årets nasjonale prøve. Det er viktig at oppgavene blir prøvd på riktig aldersgruppe. Gjennomføringstidspunktet for 2. pilotene må derfor ligge så nært som mulig opp til tidspunktet for gjennomføring av endelig prøve.

Det var 838 elever som gjennomførte 2. pilot A og 815 elever som gjennomførte 2. pilot B. Reliabiliteten til prøvene var henholdsvis 0,919 og 0,925. Verdiene tilfredsstiller anbefalingene i Rammeverket for nasjonale prøver (Utdanningsdirektoratet, 2017b). Tabell 7 viser utvalg og resultater fra gjennomføringen.

Etter gjennomføring	2.pilot A	2.pilot B
Antall gutter	445	427
Antall jenter	393	388
Totalt antall elever	838	815
Gj.snitt Theta gutter	0,082	0,066
Gj.snitt Theta jenter	-0,093	-0,070
Gj.snitt Theta totalt	0,000	0,001
Sign.forskjell J - G (Theta) $p < 0,05$	Ja	Ja
Standardavvik Theta	0,970	0,980
SE til gjennomsnitt Theta	0,034	0,034
Gj.snitt P-verdi gutter	48,0 %	47,6 %
Gj.snitt P-verdi jenter	45,0 %	45,5 %
Gj.snitt P-verdi totalt	46,6 %	46,6 %
Svarprosent	92,2 %	91,8 %
Cronbachs alpha	0,919	0,925

Tabell 7 Tekniske data etter gjennomføring

Tabell 7 viser at P-verdien i begge settene er noe lavere enn de estimerte verdiene før gjennomføring. Gjennomsnittlig b-verdi ved gjennomføring ble henholdsvis 0,139 og 0,142. Disse verdiene tyder også på at sett A var noe vanskeligere enn det vi trodde på forhånd, men ikke sett B. Ellers viser analysen av resultatene at 12 (8 i A og 4 i B) av de totalt 100 oppgavene har for lav a-verdi (diskriminering) i henhold til anbefalingene i Rammeverket (Utdanningsdirektoratet, 2017b) Disse oppgavene må vi ta en grundig vurdering av før de eventuelt blir tatt med i endelig prøve.

Tabell 8 viser fordelingen av oppgaver innen mestringsnivå etter gjennomføring av 2. pilotene.

	2. pilot A	2. pilot B
Mestringsnivå 1	3	2
Mestringsnivå 2	13	12
Mestringsnivå 3	13	18
Mestringsnivå 4	12	11
Mestringsnivå 5	9	7

Tabell 8 Fordeling av oppgaver innen mestringsnivå etter gjennomføring

Sammenligner vi tabell 5 og 8, ser vi at mange oppgaver har endret mestringsnivå fra før til etter gjennomføring av 2. pilotene. Noe av denne endringen kan forklares med at tallene som før-analysen er basert på, er hentet fra prepilot og 1. pilot, som igjen er basert på elever som har en annen alder ved gjennomføring enn hva 2. pilotene har. I tillegg er det foretatt en del endringer på oppgavene fram til 2. pilot som også påvirker resultatet.

Vanskelighetsgraden til oppgavene fra 2. pilot er utgangspunktet når vi setter sammen kohortprøven. Tabell 8 viser at vi for kohortprøven for 2018, har litt for få oppgaver å velge mellom til mestringsnivå 1.

Det at oppgaver endrer vanskelighetsgrad fra en utprøving til en annen, er en faktor vi prøver å analysere for å kunne forutse noe av denne endringen i arbeid med framtidige prøver.

Kapittel 4: Sammensetning av kohortprøve

Innledning

Høsten 2017 deltok til sammen 1653 elever fordelt på 74 skoler i 2. piloten til 2018-prøven. 2. pilot er den siste utprøvingen av oppgavene før den endelige gjennomføringen, og det er svarmønstrene disse elevene avga som danner grunnlag for tallene som blir løftet fram i dette kapitlet. I tillegg til at 2. pilotene er analysert, er det kjørt en IRT-analyse av kun de oppgavene som utgjør kohortprøven. 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 endelig prøve (tabell 9).

Nasjonal prøve for 8. og 9. trinn 2018 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. Det 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. Ankeret skiftes ut i løpet av en fem-årsperiode. Nye oppgaver må derfor piloteres i ankerprøven som «preanker». Det gjør at mer enn halvparten av oppgavene i ankerprøven må hemmeligholdes. De øvrige oppgavene i ankerprøven er de samme som i de tre andre versjonene. Ankerprøven er kommentert senere i kapitlet.

Psykometri

Andelen flervalgsoppgaver i kohortprøven er 52 prosent, og det er totalt 6 prosent interaktive oppgaver. Denne fordelingen er i tråd med tidligere prøver. Gjennomsnittlig b-verdi til oppgavene er -0,063, og gjennomsnittlig a-verdi er 0,76. Gjennomsnittlig P-verdi til oppgavene er 50,4 prosent, og ved utprøving er P-verdien til guttene tre prosentpoeng høyere enn jentenes (tabell 9). Det er seks oppgaver i kohortprøven 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. Dette blir omtalt senere i kapitlet.

Siden prøven er satt sammen av oppgaver fra to 2. piloter, som er besvart av to uavhengige elevgrupper, kan vi ikke beregne reliabiliteten til prøven direkte ved en IRT-analyse. Vi har derfor estimert reliabiliteten i tabell 9 ved bruke gjennomsnittlig måleusikkerhet¹ for prøven fra preanalysen (figur 5). I den estimerte reliabiliteten har vi beregnet gjennomsnittlig måleusikkerhet for elever fra theta -3 til 3. Dette tilsvarer ± 3 standardavvik som vil si at det 99,7 % av elevene skal være innenfor dette intervallet. Vi har også erfart at dette er typiske ytterpunkt i skalaen etter analyser av kohortprøver (tilsvarer 20–80 skalapoeng).

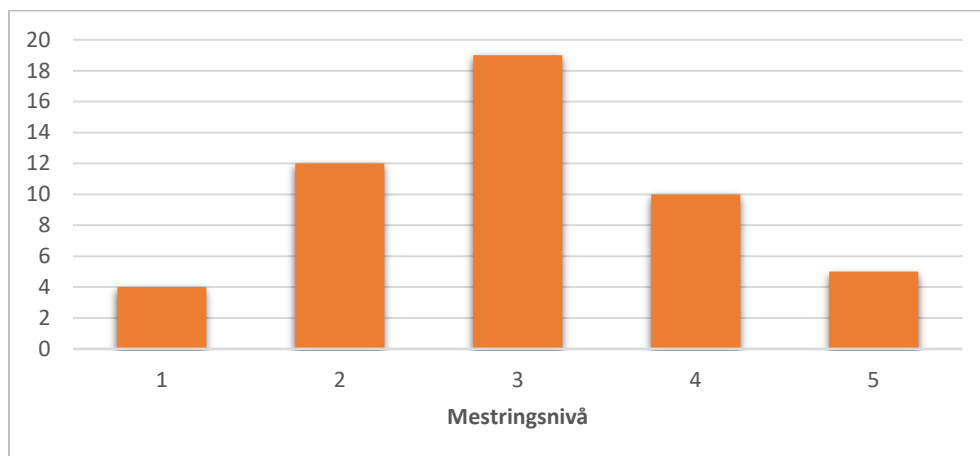
Analyse av tidligere kohortprøver har vist at reliabiliteten til prøvene har vært over 0,9. Vi antar derfor at kohortprøven for 2018 vil ha en slik reliabilitet, selv om estimatet er noe lavere.

¹ $1 - \overline{CSEM}^2$

NP8 2018	Kohortprøve	Ankerprøve ²
Gj.snitt Theta gutter	0,075	0,1
Gj.snitt Theta jenter	-0,084	-0,1
Gj.snitt Theta totalt	0,000	0
Sign.forskjell J - G (Theta) $p < 0,05$	Ja	Ja
Standardavvik Theta	0,961	1
SE til gjennomsnitt Theta (estimert)	0,004	0,02
Gj.snitt P-verdi gutter	51,8 %	53,5 %
Gj.snitt P-verdi jenter	49,0 %	50,7 %
Gj.snitt P-verdi totalt	50,4 %	52,1 %
Svarprosent	93,9 %	93,4 %
Cronbachs alpha (estimert)	0,877	> 0,9
Antall item	50	50
Antall MC	26	26
Antall IA	3	4

Tabell 9 Oppsummering forslag til endelig prøve. Tall fra preanalyse

Selv om gjennomføringstidspunktet for 2. pilotene er nært opp mot gjennomføringstidspunktet for kohortprøven, ser vi at gjennomsnittlig løsningsprosent øker, og at gjennomsnittlig 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 P-verdi økte i 2016 med to prosentpoeng og i 2017 med ett prosentpoeng fra siste utprøving til endelig prøve, mens gjennomsnittlig b-verdi endret seg med henholdsvis -0,08 og -0,05. Vi kan derfor anta at endelig prøve 2018 vil ha en gjennomsnittlig P-verdi rundt 0,50 og gjennomsnittlig b-verdi nær 0.

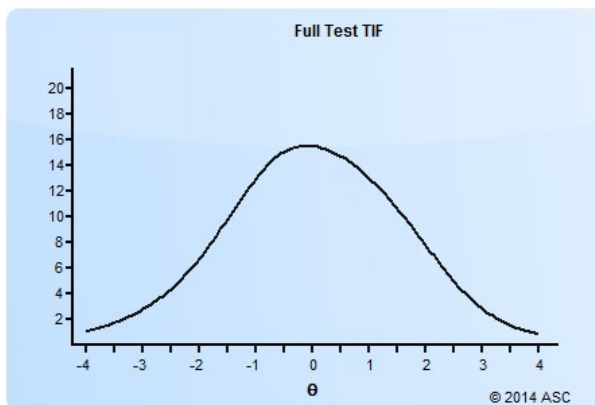


Figur 3 Fordeling av oppgaver etter mestringsnivå, basert på tall fra 2. pilot

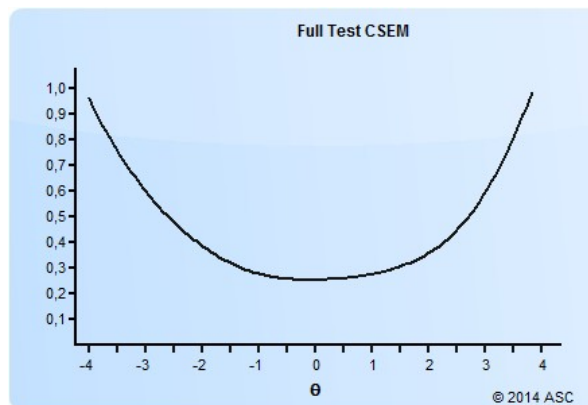
Figur 3 viser fordelingen av oppgaver etter mestringsnivå for 2018-prøven, med verdier fra utprøving. Siden oppgavens vanskelighetsgrad endrer seg noe fra siste utprøving til endelig gjennomføring, kan oppgaver endre mestringsnivå. Dette gjelder spesielt oppgaver som ligger nært opp til grensen mellom to nivåer. Figur 3 viser at kohortprøven inneholder tilstrekkelig antall oppgaver med lav og høy vanskelighetsgrad, samtidig som den inneholder flest oppgaver med middels vanskelighetsgrad.

² Tall fra IRT-analysen i denne tabellen er kun estimerte, siden vi ikke har en lenking i datasettet (2. pilot og ankersett) som gjør det mulig å beregne disse parameterne.

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



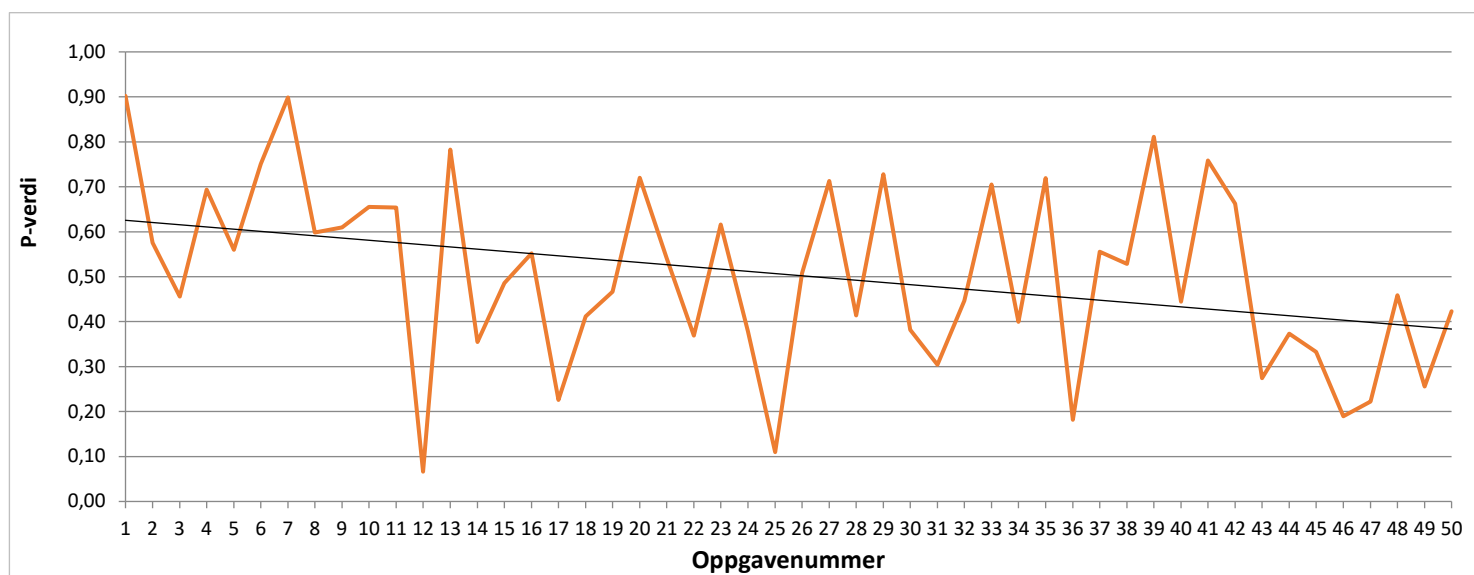
Figur 4 Testinformasjonen til prøven for 2018. Basert på tall fra preanalyse



Figur 5 Måleusikkerheten til prøven for 2018. Basert på tall fra preanalyse

En preanalyse av oppgavene som er plukket ut til kohortprøven, viser at testinformasjonen (figur 4) er størst for elever med $\theta = -0,010$. Det vil si at prøven gir mest informasjon om elever som presterer midt på ferdighetsskalaen. Det er også der vi har flest elever. Med fastsettelse av ny skala i 2014 ble det bestemt at gjennomsnittet skulle være $\theta = 0$ (50 skalapoeng) og standardavviket 1 (10 skalapoeng). Det betyr at 68 % 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 god informasjon om elever som presterer lavt og høyt på prøven. Figur 5 viser at måleusikkerheten er lavest der det er flest elever, mellom -1 og 1 på ferdighetsskalaen. Kurvene for TIF og CESM oppfyller anbefalingene i Rammeverket for nasjonale prøver (Utdanningsdirektoratet, 2017b)

Figur 6 viser P-verdi for hver enkelt oppgave i endelig prøve 2018, med tall fra siste utprøving. Dette oppsettet stemmer med tidligere prøver, der settet har synkende P-verdi utover i settet, samt at oppgaver med høy og lav P-verdi veksler gjennom hele prøven. Det er også et bevisst valg at det kommer en oppgave med noe høyere P-verdi rett etter en oppgave med svært lav P-verdi, for å unngå at elever får svært mange vanskelige oppgaver rett etter hverandre.



Figur 6 P-verdi til oppgavene etter oppgavennummer i kohortprøven 2018. Tall fra 2. pilot

Validitet

I Kunnskapsløftet (Kunnskapsdepartementet & 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. Rammeverket for nasjonale prøver (Utdanningsdirektoratet, 2017b)

Nasjonal prøve i regning for 8. og 9. trinn måler elevenes kompetanse etter 7. 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 8. trinn. Gruppen som utvikler nasjonale prøver i regning, har lang undervisningserfaring 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, 2017b), som handler om validitet, og opp mot beskrivelsene av den grunnleggende ferdigheten å kunne regne, slik den er beskrevet i Rammeverket for grunnleggende ferdigheter (Utdanningsdirektoratet, 2017a).

Oppgavene i prøven for 2018 knyttes til kompetansemål i fag som vist i tabell 10. En oppgave kan knyttes til kompetansemål i flere fag.

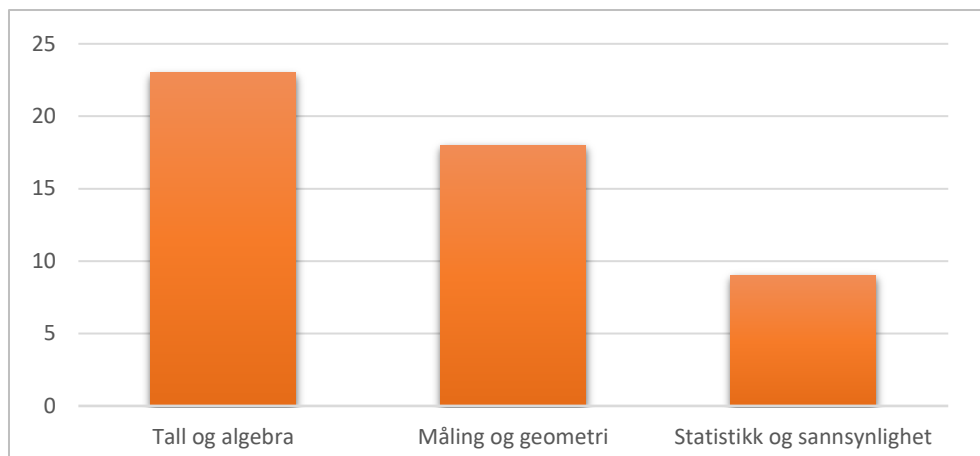
Fag	Antall oppgaver
Engelsk	12
KRLE	9
Kroppsøving	3
Kunst og håndverk	12
Matematikk	50
Mat og helse	14
Musikk	0
Naturfag	13
Norsk	9
Samfunnsfag	21

Tabell 10 Antall oppgaver som kan knyttes til kompetansemål i ulike fag etter LK06 7. trinn

Som tabell 10 viser, måler oppgavene i kohortprøven for 2018 ikke kompetansemål i alle fag. I år mangler vi oppgaver som har relevans for musikk. Faget musikk inneholder få kompetansemål etter 7. trinn som krever en form for regning, som kan måles i en prøvesituasjon. Dette gjelder blant annet kroppslige uttrykk som takt, tempo og rytme. Vi har imidlertid, som en konsekvens av at vi mangler oppgaver med relevans for musikk, satt ekstra fokus på dette faget i produksjonen av oppgaver til 2019-prøven. Vi håper vi kan få belyst musikkfaget i endelig prøve 2019.

Nasjonal prøve i regning for 8. og 9. trinn inneholder oppgaver innen områdene 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 Rammeverket for nasjonale prøver (Utdanningsdirektoratet, 2017b). Kohortprøven for 2018 består av 23 oppgaver i området tall og algebra, 18 innenfor måling og geometri og 9 statistikk- og sannsynlighetsoppgaver (figur 7). 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 (Kunnskapsdepartementet & Utdanningsdirektoratet, 2006).

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



Figur 7 Antall oppgaver i hver kategori

Opgavene i hver kategori er fordelt på fem mestringsnivåer (tabell 11), etter tall fra 2. pilot. Tabellen viser at det ikke er oppgaver på mestringsnivå 1 i S&S i endelig prøve. Det samme gjaldt prøven for 2017. De siste årene har vi hatt problemer med å utvikle oppgaver på dette nivået som har tilfredsstilt de psykmotriske kravene i Rammeverk for nasjonale prøver (Utdanningsdirektoratet, 2017b).

	Mestringsnivå 1	Mestringsnivå 2	Mestringsnivå 3	Mestringsnivå 4	Mestringsnivå 5
Tall og algebra	3	7	9	4	0
Måling og geometri	1	4	5	5	3
Statistikk og sannsynlighet	0	1	5	1	2

Tabell 11 Antall oppgaver på mestringsnivå i hver kategori. Tall fra 2. pilot

I tabellen ser vi at det ikke er oppgaver på mestringsnivå 5 i T&A. De fem andre oppgavene på mestringsnivå 5 i de to andre kategoriene inneholder mye tallbehandling. Hovedutfordringene i disse oppgavene tilhører kategoriene M&G og S&S.

I tillegg til å være knyttet til kategorier i læreplanen i matematikk, er oppgavene knyttet til de ulike prosessene i den helhetlige problemløsningsprosessen som er beskrevet i kravene i Rammeverket for grunnleggende ferdigheter (Utdanningsdirektoratet, 2017a). 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 12 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.

Prosess	Antall oppgaver
Gjenkjenne og beskrive	20
Bruke og bearbeide	28
Reflektere og vurdere	2

Tabell 12 Antall oppgaver knyttet til de ulike prosessene i den helhetlige problemløsningsprosessen

Det er få 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 vurdere og reflektere om det matematiske svaret de har funnet er et svar som er realistisk til konteksten oppgaven.

Forklaring til noen oppgaver i endelig prøve

Av de 50 oppgavene i kohortprøven er det 45 oppgaver som tilfredsstillende anbefalingene om at oppgavene skal ha en point-biserial-verdi $> 0,3$ og en a-verdi $> 0,5$. De fem som ikke tilfredsstillende begge disse kravene, er likevel tatt med i settet ut fra pedagogiske og didaktiske valg.

I IRT-analysen av 2. pilotene får seks oppgaver som er med i kohortprøven, bias mot kjønn, to mot gutter (oppgavene 24 og 28) og fire mot jenter (oppgavene 12, 18, 35 og 43). I tabell 13–18 er det satt inn tall for vanskelighetsgrad og diskriminering fra både CTT- og IRT-analysene. 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 $M-H > 1$, har guttene i gjennomsnitt over hele ferdighetsskalaen større odds for å løse oppgaven enn jentene. Dersom $M-H < 1$, har guttene mindre odds til å løse oppgaven enn jentene. Vi har også med p-verdi³ fra IRT-analysen som angir hvor sannsynlig det er at forskjellen mellom de to gruppene skyldes tilfeldigheter eller ikke. Vi har brukt 0,05 som signifikansmål, og oppgaver som har lavere p-verdi enn dette, kommer ut med bias.

Oppgaver med bias

Oppgave 35 (Bias mot jenter)

Sondre kjøper inn fire pakker med kjøttdeig. Hver pakke inneholder 400 g.

Hvor mange kilogram kjøttdeig kjøper Sondre?

☐ 1,6 kg

☐ 16 kg

☐ 160 kg

☐ 1600 kg



Itemnr.	T-Rpbis	S-Rpbis	P-verdi	M-H	p	a-verdi	b-verdi	Gjennomsnittlig theta			
								1	2	3	4
5160536	0,493	0,451	72,0	1,4423	0,0188	0,75	-1,064	0,48	-0,27	-0,20	-0,22

Tabell 13 Psykometriske verdier for oppgave 35, som har bias mot jenter. Tall fra 2. pilot

Oppgave 35 måler om elevene behersker omgjøring mellom prefikser i en kontekst som etter vår vurdering ikke favoriserer noen av kjønnene. Som tabell 13 viser er $p < 0,05$ og kommer derfor ut med bias og M-H er 1,4. Det betyr at gutter med lik ferdighet som jenter, i gjennomsnitt har 1,4 større odds for å løse oppgaven riktig.

I likhet med de siste årene har oppgaver med omgjøring mellom prefikser vært oppgaver som har kommet ut med bias mot jenter. Omgjøring av enheter er en essensiell ferdighet i den grunnleggende ferdigheten i å kunne regne, noe som over år har vært fokus i veiledningsmateriellet til prøvene.

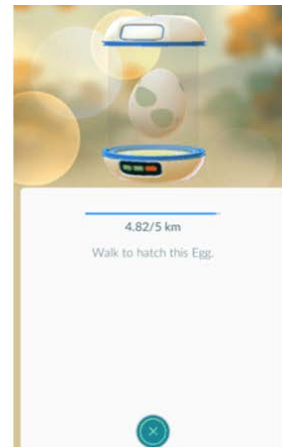
³ Merk at det er brukt P-verdi (stor P) for løsningsprosent og p-verdi (liten p) for mål for signifikans.

Oppgave 43 (Bias mot jenter)

Aleksander spiller Pokemon Go. Han har et egg som klekkes hvis han går 5 km. Aleksander har gått 4,82 km.

Hvor mange meter har Aleksander igjen å gå før egget klekkes?

Svar: m



Itemnr.	T-Rpbis	S-Rpbis	P-verdi	M-H	p	a-verdi	b-verdi	Gjennomsnittlig theta	
								0	1
5159918	0,534	0,495	27,5	2,096	0,0024	0,87	0,859	-0,31	0,82

Tabell 14 Psykometriske verdier for oppgave 43, som har bias mot jenter. Tall fra 2. pilot

Oppgave 43 handler også om omgjøring mellom prefikser, her er det kilometer og meter. Selv om Pokemon Go har vært svært populært blant både gutter og jenter, kan det være at denne oppgaven fenger gutter i større grad enn jenter. At M-H (tabell 14) er høyere og p lavere enn i oppgave 35, kan tyde på dette. Uansett er regneferdigheten som oppgaven måler, omgjøring mellom kilometer og meter, en grunnleggende ferdighet som er relevant for mange fag.

Oppgave 12 (Bias mot jenter)

Fredrik skal på sykkelturn med klassen.

Ifølge værvarslet skal vindstyrken bli 10 meter per sekund (m/s).

Fredrik lurer på hvor mange kilometer per time (km/h) dette tilsvarer.

Hvor mange kilometer per time tilsvarer 10 m/s?

Svar: km/h

Itemnr.	T-Rpbis	S-Rpbis	P-verdi	M-H	p	a-verdi	b-verdi	Gjennomsnittlig theta	
								0	1
5167499	0,405	0,359	6,6	2,8686	0,0407	1,16	2,071	0,15	1,58

Tabell 15 Psykometriske verdier for oppgave 12, som har bias mot jenter. Tall fra 2. pilot

Måling står også sentralt i oppgave 12, med omgjøring mellom enheter. Konteksten skal etter vår vurdering ikke favorisere noen av kjønnene, men vi ser at M-H-faktoren er høy (tabell 15). Imidlertid er p-verdien høyere enn i de andre oppgavene som har bias mot jenter.

Oppgave 18 (bias mot jenter)

Et idrettslag kjøper mobildeksel fra en fotobutikk for 40 kr per stykk.
Idrettslaget selger dekslene videre for 70 kr per stykk.

En dag solgte idrettslaget ti mobildeksel.

Hvor mye tjente idrettslaget denne dagen?

Svar: kr



Itemnr.	T-Rpbis	S-Rpbis	P-verdi	M-H	p	a-verdi	b-verdi	Gjennomsnittlig theta	
								0	1
5119906	0,569	0,531	41,2	1,4858	0,0395	0,89	0,34	-0,43	0,66

Tabell 16 Psykometriske verdier for oppgave 43, som har bias mot jenter. Tall fra 2. pilot

Oppgave 18 er den siste som har bias mot jenter (tabell 16). Det er en oppgave innen området tall og algebra der hovedutfordringen er å gjenkjenne og beskrive problemet. Analysene av resultatene viser at det er større prosentandel jenter enn gutter som svarer 700 kr. Det feilsvaret skyldes trolig at elevene regner ut inntekten og ikke hvor mye de tjente.

Konteksten er hentet fra dagliglivet og er en kontekst som vi antar elever i begynnelsen av ungdomsskolen kjenner seg igjen i. Konteksten og at regneferdigheten som måles er innen det som er definert som grunnleggende ferdighet i å kunne regne, er grunnen til at vi ønsker å inkludere oppgaven i prøvesettet.

Oppgave 24 (Bias mot gutter)

Kristine og to venninner skal bestille én ny turndrakt hver, og sender en samlet bestilling.

Hver turndrakt koster 630 kr. I tillegg må venninnene betale 60 kr i frakt for bestillingen. De blir enige om å dele fraktkostnaden likt.

Hvor mye skal Kristine betale til sammen?

Svar: kr

Itemnr.	T-Rpbis	S-Rpbis	P-verdi	H-M	p	a-verdi	b-verdi	Gjennomsnittlig theta	
								0	1
5159918	0,389	0,360	37,8	0,6417	0,0103	0,57	0,538	-0,28	0,51

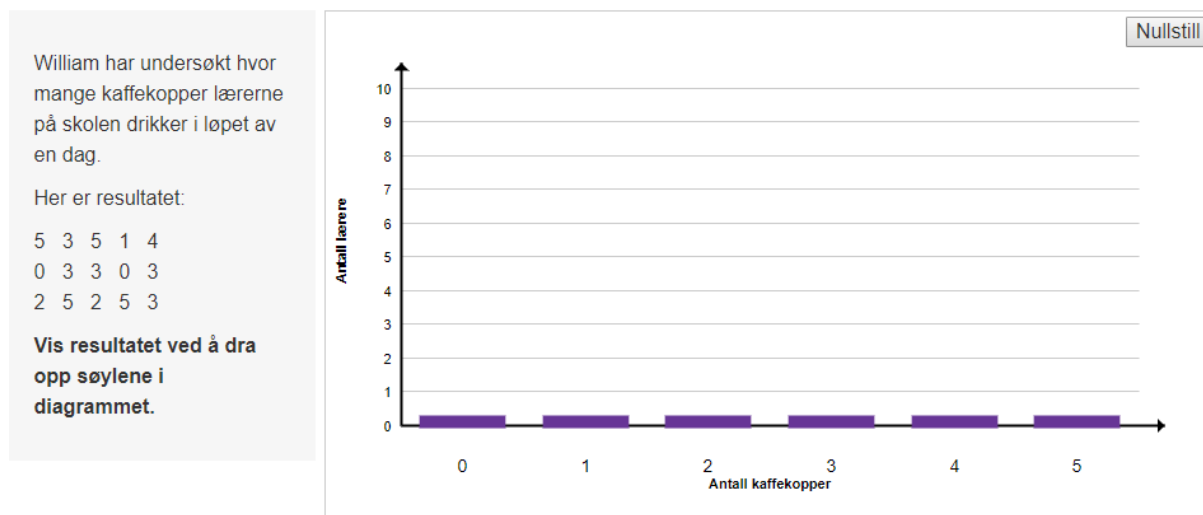
Tabell 17 Psykometriske verdier for oppgave 24, som har bias mot gutter. Tall fra 2. pilot

I oppgave 24 skal elevene gjenkjenne og beskrive et problem der tre jenter skal kjøpe hver sin turndrakt og dele på fraktkostnaden. Selv om konteksten handler om Kristine og to jenter som skal bestille turndrakt, noe som ved første øyekast kan virke som favoriserer jenter, er vår erfaring at valg av navn og for eksempel gjenstand som skal kjøpes, spiller liten rolle for hvordan oppgaven oppfører

seg psykometrisk (tabell 17). Vi har erfart, gjennom lignende oppgaver, at det denne oppgaven tester har vært oppgaver der jentene har gjort det bedre enn guttene.

Konteksten er hentet fra dagliglivet og er en kontekst som vi antar elever i begynnelsen av ungdomsskolen kjenner seg igjen i. Konteksten og at regneferdigheten som måles er innen det som er definert som grunnleggende ferdighet i å kunne regne, er grunnen til at vi ønsker å inkludere oppgaven i prøvesettet.

Oppgave 28 (Bias mot gutter)



Itemnr.	T-Rpbis	S-Rpbis	P-verdi	H-M	p	a-verdi	b-verdi	Gjennomsnittlig theta	
								0	1
5119720	0,430	0,401	41,4	0,6782	0,0221	0,60	0,42	-0,20	0,65

Tabell 18 Psykometriske verdier for oppgave 28, som har bias mot gutter. Tall fra 2. pilot

I oppgave 28 skal elevene sortere og systematisere et resultat fra en spørreundersøkelse, og framstille resultatet i et søylediagram. Konteksten er kjønnsnøytral, men oppgaven kommer ut med bias mot gutter (tabell 18). Samme oppgavetype fikk også bias mot gutter i fjor. Oppgaven krever nøyaktighetsarbeid, og det kan muligens være en årsak til at oppgaven får bias mot gutter.

Det å foreta undersøkelser og presentere resultatet i ulike diagram, er en kompetanse som kreves for å nå kompetansemål i mange fag. Vi ønsker derfor å inkludere oppgaven i prøvesettet.

Oppgaver der distraktorer har positiv point-biserial

I kohortprøven er det to flervalgsoppgaver der ett av svaralternativene fikk positiv point-biserial-verdi ved siste utprøving. Positive point-biserial-verdier for en distraktor oppstår gjerne når en gruppe elever som har høyt resultat på resten av prøvesettet, velger den distraktoren. Ofte kommer dette av at elevene har gjort en feil i prosessen gjenkjenne og beskrive, men bruker og bearbeider informasjon, slik de har tolket den, riktig.

Elevene viser dermed at de behersker å bruke og bearbeide i oppgaven, men finner et svar på et annet spørsmål enn det som spørres etter i oppgaven. Vi har her valgt å kommentere én av oppgavene for å vise dette.

Oppgave 50

Katja ønsker å kjøpe en ny genser som koster 150 kr.

Hun melder seg inn i en kundeklubb, og det gjør at hun får 30 % rabatt på genseren.

Hvor mye må Katja betale for genseren?

- ☐ 45 kr
- ☐ 50 kr
- ☐ 100 kr
- ☐ 105 kr
- ☐ 120 kr

Itemnr.	T-Rpbis	S-Rpbis	P-verdi	a-verdi	b-verdi	T-Rpbis correct for spuriousness				
						1	2	3	4	5
5160139	0,499	0,467	42,3	0,71	0,413	0,03	-0,14	-0,15	0,49	-0,30

Tabell 19 Oppgave der en distraktor korrelerer positivt

I oppgave 50 har distraktor 1 positiv verdi for point-biserial (tabell 19). Elevene som svarer dette, regner ut 30 % av 150 kr og finner rabatten, men ikke hvor mye Katja må betale. Disse elevene viser at de behersker å regne ut 30 % av 150 kr, noe som forklarer den positive verdien til distraktoren. Oppgaven har imidlertid høye diskrimineringsverdier, både ved CTT- og IRT-analysen (T-Rpbis, S-Rpbis og a-verdi) og er derfor tatt med i prøven.

Ankeroppgaver

Utgangspunktet for ankerdesignet i nasjonale prøver i regning for 8. og 9. trinn, var 20 utvalgte oppgaver fra den ordinære prøven for 2012. I 2014 ble disse, sammen med 30 oppgaver fra kohortprøven for 2014, satt sammen til en egen ankerprøve. I etterkant av gjennomføringen i 2014, ble ankeroppgavene kalibrert sammen med kohortprøven. Ankeroppgavene i 2014 ble også brukt i 2015, men da var vanskelighetsgraden til ankeroppgavene fastsatt ut fra gjennomføringen i 2014. I etterkant av gjennomføringen i 2015 ble det oppdaget et behov for å ha en modell som gjør det mulig å skifte ut ankeroppgaver. Et viktig poeng var da å utvikle et design som gjør det mulig å endre ankeret, uten at det går på bekostning av antall oppgaver som er ferdig kalibrert.

I 2016 ble det innført en modell der ankerprøven består av 20 ankeroppgaver, 4 framtidige ankeroppgaver og 26 oppgaver fra kohortprøven. De framtidige ankeroppgavene, kalt preankeroppgaver, er med i ankerprøven, men fungerer ikke som ankeroppgaver og behandles som oppgaver fra kohortprøven. I etterkant av gjennomføringen blir preankeroppgavene kalibrert sammen med kohortprøven, og ankeret, og går inn i ankeret som ankeroppgaver året etter. Det er derfor viktig at disse hemmeligholdes etter gjennomføring.

Modellen gjør det mulig å skifte ut opptil fire ankeroppgaver hvert år, som igjen betyr at hele ankeret kan fornyes i løpet av en femårsperiode. Ved hjelp av modellen kan dette skje uten å endre antall oppgaver som brukes til å kalibrere resultatene og måle utvikling over tid.

Ankeret i 2018 vil bestå av 20 oppgaver som er blitt kalibrert et tidligere år. 13 av disse oppgavene ble kalibrert i ankerprøven i 2014, mens tre ble kalibrert i 2016 og fire i 2017. I tillegg inneholder

ankerprøven fire preankeroppgaver. Preankeroppgavene er med i ankerprøven for å bli kalibrert i 2018, og skal etter planen gå inn i ankeret i 2019.

Da vi valgte ut preankeroppgaver, valgte vi gode pedagogiske og didaktiske oppgaver med høye diskrimineringsverdier fra 2. pilot 2018. Vi har også sett på ankeret som en helhet for å prøve å finne preankeroppgaver som gjør at ankeret i 2019 skal speile kohortprøven enda bedre enn tidligere år. Preankeroppgavene må ha en kontekst som vi tror vil være aktuell fem år fram i tid.

For å sikre mest mulig stabile tall, er det viktig at de nye ankeroppgavene står på samme plass som da de var preanker. Dette gjør at det er bevisste valg som ligger bak plasseringen av preankeroppgavene. Med denne modellen for utskiftning av ankeroppgaver, med fire nye ankeroppgaver hvert år, får vi også justert plasseringene til ankeroppgavene. Det gjør at ankeret for 2018 har flere ankeroppgaver som følger etter hverandre. Det kan være en fordel når vi skal sammenligne resultater fra år til år, da det er naturlig at oppgaven før påvirker elevenes resultat på oppgaven som kommer etter.

Generelle vurderinger av prøven som helhet

Gjennom utprøvinger har hver oppgave i endelig prøve vært prøvd ut minimum to ganger. 29 oppgaver har vært med i tre utprøvinger, mens 21 har vært med i to. De fleste oppgavene har små endringer i verdier fra 1. pilot til 2. pilot og holder seg stabile på mestringsnivå. Dette tyder på at oppgavene er relativt stabile.

Opgavene i prøven har stor bredde både når det gjelder vanskelighetsgrad, matematiske tema og oppgavens kontekster. Selv om ikke regning i alle fag er representert i prøven, har prøven relevans for mange fag. Det er en god blanding av oppgaver hvor elevene skal fylle inn svar og velge svaralternativer.

Kapittel 5: Analyse av nasjonal prøve i regning 2018

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 2018. Tallene det henvises til i dette kapitlet er basert på en analyse av elever som gjennomførte kohortprøven og ankerprøven på 8. trinn, der parameterne til ankeroppgaven er brukt for å kalibrere oppgavene i kohortprøven. Det var 59 125 elever på 8. trinn og 58 527 elever på 9. trinn som gjennomførte prøven. Rapporten handler i utgangspunktet om resultatene til 8. trinn, men vi sammenligner med resultatene til 9. trinn.

Nasjonal prøve i regning for 8. og 9. trinn måler elevenes kompetanse etter 7. 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 8. trinn. Verdien 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, 2017b). Endelig godkjenning av prøven ble gjort av Utdanningsdirektoratet og deres egne interne og eksterne kvalitetssikrere. Ingen oppgaver ble endret etter denne gjennomgangen.

Opgavene i kohortprøven er plassert på mestringsnivå ut fra vanskelighetsgrad, bestemt ved IRT-analyse. En oppgave med høy vanskelighetsgrad gir flere skalapoeng enn en oppgave med lav vanskelighetsgrad. Skalapoengene 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 22 viser de tekniske verdiene til kohortprøven og ankerprøven. I tillegg viser tabellen resultatene til 9. trinn for kohortprøven.

NP8 2018	Kohortprøve ⁴	Ankerprøve	9. trinn
Antall gutter	29 943	1870	29875
Antall jenter	29 182	1800	28952
Totalt antall elever	59 125	3670	58827
Gj.snitt Theta gutter	0,061	0,065	0,431
Gj.snitt Theta jenter	-0,063	-0,069	0,297
Gj.snitt Theta totalt	0,000	-0,001	0,365
Sign.forskjell J - G (Theta) $p < 0,05$	Ja	Ja	Ja
Standardavvik Theta	0,967	0,942	0,998
SE til gjennomsnitt Theta	0,004	0,004	0,004
Gj.snitt P-verdi gutter	54,1 %	53,5 %	62,3 %
Gj.snitt P-verdi jenter	51,8 %	50,8 %	60,0 %
Gj.snitt P-verdi totalt	53,0 %	52,2 %	61,2 %
Svarprosent	95,6 %	94,5 %	95,9 %
Cronbachs alpha	0,927	0,930	0,930
Antall item	50	50	50
Antall MC	26	26	26
Antall IA	3	4	3

Tabell 20 Psykometriske verdier kohortprøven og ankerprøven 2018, samt resultater for 9. trinn

Som tabellen viser er de psykometriske verdiene til kohortprøven på 8. trinn og ankerprøven tilnærmet identiske. Ankerprøven har noe høyere vanskelighetsgrad og lavere svarprosent enn kohortprøven, men forskjellene er marginale.

Oversikt over oppgaver NP8 (9) 2018

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, Målingsdivisjon	T&A	Åpen	Mat	90,6	-2,2	93,3	-1,7	0,10	-2,052	1
2	Tolke og lese tabell	S&S	Flervalg	Eng, Krle, Mat, Nor, Nat, Saf	53,9	2,9	62,6	3,0	0,18	-0,238	3
3	Negative tall, Temperatur	T&A	Flervalg	Mat, Nat	48,5	-2,6	59,2	-3,0	0,22	0,083	3
4	Tid, Omgjøring mellom timer, min og sek	M&G	Flervalg	Mat, Mhe	69,0	3,3	82,1	2,5	0,31	-0,931	2
5	Brøk, Sammenhengen brøk og desimaltall	T&A	Flervalg	Mat, Mhe	59,6	-7,8	73,4	-6,0	0,30	-0,341	3
6	Vurdere rimeligheten av svar	T&A	Flervalg	Mat	75,6	1,7	80,4	1,0	0,12	-1,047	2
7	Tid, Forståelse for enheter	M&G	Flervalg	Kro, Mat	91,6	-5,1	94,2	-3,7	0,10	-2,333	1
8	Prosent, Finne prosenten	T&A	Flervalg	Mat, Saf	58,3	-11,7	64,6	-12,3	0,13	-0,394	3
9	Tolke og lese tabell	S&S	Flervalg	Eng, Krle, Mat, Nor, Nat, Saf	63,2	-4,8	71,7	-3,8	0,18	-0,619	2
10	Brøk, Brøkdel av en hel	T&A	Flervalg	Mat	62,6	3,1	65,5	2,9	0,06	-0,655	2
11	Brøk, Regne med brøk	T&A	Åpen	Mat, Mhe, Nat, Nor, Saf	74,2	-3,9	79,6	-3,8	0,13	-0,871	2
12	Omgjøring mellom enheter	M&G	Åpen	Mat, Nat	7,5	-4,6	12,8	-7,7	0,18	2,063	5
13	Addisjon, Multiplikasjon, Desimaltall	T&A	Flervalg	Mat	77,7	0,9	82,8	1,0	0,13	-1,842	1
14	Areal, Forståelse av areal	M&G	Åpen	Khv, Mat	37,3	3,4	39,9	1,9	0,05	0,521	4
15	Overslag, Gjennomsnitt	S&S	Åpen	Mat, Saf	55,6	-3,8	63,0	-3,9	0,15	-0,239	3

⁴ Analysen er basert på alle elever som gjennomførte NPREG08, også de som tok ankerprøven som ikke tok alle oppgavene i kohortprøven

Teknisk rapport NP8 2018

16	Multiplikasjon, Hele tall	T&A	Åpen	Mat	57,7	7,9	64,0	7,5	0,13	-0,524	2
17	Forhold, Målestokk	M&G	Åpen	Khv, Kro, Mat, Saf	23,9	-6,7	33,7	-8,5	0,22	1,397	5
18	Subtraksjon, Multiplikasjon, Hele tall	T&A	Åpen	Mat	42,2	-8,8	58,9	-10,1	0,34	0,284	3
19	Forhold, Utvide/redusere matoppskrift	M&G	Flervalg	Mat, Mhe	44,2	-7,1	53,2	-6,5	0,18	0,221	3
20	Tolke og lese tabell, Multiplikasjon, Desimaltall	T&A	Flervalg	Eng, Krle, Mat, Nat, Nor, Saf	69,4	-10,6	73,5	-11,1	0,09	-1,060	2
21	Tid, Tidsintervall	M&G	Flervalg	Mat, Mhe	56,5	-2,6	66,1	-1,4	0,20	-0,354	3
22	Forhold, Blandingsforhold	M&G	Flervalg	Khv, Mat, Mhe, Nat	41,5	1,7	49,5	1,1	0,16	0,422	3
23	Algebraisk tenking, Sammensatt problem	T&A	Åpen	Mat	63,4	3,8	70,0	3,7	0,14	-0,640	2
24	Algebraisk tenking, Sammensatt problem	T&A	Åpen	Mat	37,2	3,4	44,3	3,7	0,14	0,654	4
25	Gjennomsnitt	S&S	Åpen	Mat, Saf	14,3	-2,9	21,3	-4,8	0,18	1,466	5
26	Tid, Stille analog klokke	M&G	Åpen	Mat, Mhe	49,8	4,8	57,7	5,0	0,16	0,009	3
27	Omgjøring mellom prefikser	M&G	Flervalg	Khv, Kro, Mat, Mhe, Nat, Saf	72,9	-11,0	79,9	-9,8	0,17	-0,959	2
28	Lage diagram, Systematisere data og framstille resultat	S&S	Åpen	Eng, Krle, Mat, Nat, Nor, Saf	49,0	11,7	62,2	12,6	0,27	0,051	3
29	Multiplikasjon, Hele tall	T&A	Åpen	Mat	72,3	5,9	76,8	3,0	0,10	-1,256	2
30	Areal, Forståelse av areal	M&G	Flervalg	Khv, Mat	43,4	0,8	52,3	-0,8	0,18	0,270	3
31	Tolke og lese tabell	S&S	Flervalg	Eng, Krle, Mat, Nat, Nor, Saf	58,9	1,5	67,3	3,3	0,17	-0,503	2
32	Forhold, Blandingsforhold	M&G	Åpen	Khv, Mat, Mhe, Nat	47,3	1,0	55,7	0,4	0,17	0,096	3
33	Overslag	T&A	Flervalg	Eng, Mat, Saf	70,9	-7,8	76,8	-8,0	0,13	-0,928	2
34	Brøk, Regne med brøk	T&A	Åpen	Khv, Mat	41,6	-8,3	48,1	-9,1	0,13	0,321	3
35	Omgjøring mellom prefikser	M&G	Flervalg	Khv, Mat, Mhe, Nat, Saf	69,5	-17,8	77,1	-14,0	0,17	-0,821	2
36	Volum, Forståelse av volum	M&G	Åpen	Khv, Mat	20,2	-4,1	29,3	-4,5	0,21	1,092	4
37	Multiplikasjon med 10, 100..	T&A	Flervalg	Mat	57,8	0,3	68,9	-0,2	0,23	-0,322	3
38	Tolke og lese diagram, Stolpediagram	S&S	Flervalg	Eng, Krle, Mat, Nat, Nor, Saf	53,2	4,5	64,8	3,7	0,24	-0,135	3
39	Omgjøring mellom prefikser, Forståelse av enheter	M&G	Flervalg	Khv, Mat	82,3	-3,0	85,9	-2,9	0,10	-1,285	2
40	Prosent, Finne prosenten	T&A	Flervalg	Mat, Saf	50,2	-10,9	60,9	-10,3	0,22	-0,010	3
41	Dobling, Desimaltall	T&A	Åpen	Mat	78,3	-3,1	84,4	-2,3	0,16	-1,095	2
42	Tolke og lese diagram, Stolpediagram	S&S	Flervalg	Eng, Krle, Mat, Nat, Nor, Saf	64,2	1,3	73,4	1,2	0,20	-0,637	2
43	Omgjøring mellom prefikser, Subtraksjon	M&G	Åpen	Khv, Mat, Mhe, Nat, Saf	35,4	-17,5	46,7	-18,4	0,23	0,547	4
44	Algebraisk tenking, Sammensatt problem	T&A	Åpen	Mat	48,4	4,8	60,7	5,0	0,25	0,058	3
45	Brøk, Forståelse av brøk	T&A	Åpen	Mat	33,5	-1,7	42,3	-2,4	0,18	0,595	4
46	Omgjøring mellom prefikser, Subtraksjon	M&G	Flervalg	Mat, Mhe Nat, Saf	28,3	-2,9	38,9	-2,9	0,23	0,778	4
47	Multiplikasjon, Divisjon, Hele tall	T&A	Åpen	Mat, Saf	24,6	-0,7	36,0	-0,3	0,25	0,960	4
48	Omgjøring mellom prefikser	M&G	Åpen	Khv, Mat, Mhe, Nat	47,1	-0,2	53,5	0,5	0,13	0,136	3
49	Tolke og lese diagram, Linjediagram	S&S	Flervalg	Eng, Krle, Mat, Nat, Nor, Saf	31,3	-0,7	41,3	-0,3	0,21	1,159	4
50	Prosent, Regne med prosent	T&A	Flervalg	Mat, Mhe, Saf	43,9	-4,5	57,3	-3,1	0,27	0,301	3
Gjennomsnitt					53,0	-2,2	61,2	-2,3		-0,172	

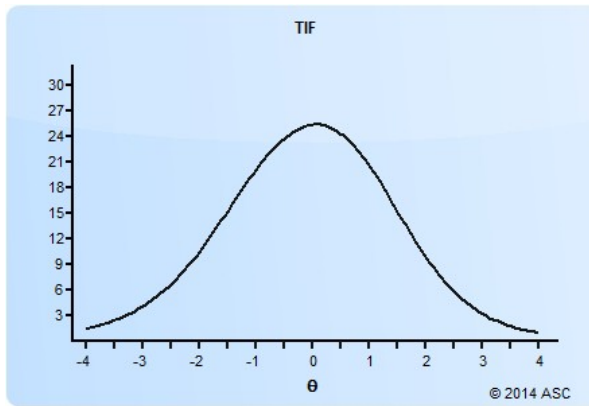
Tabell 21 Oppgaverekkefølge i versjon 1 av Nasjonal prøve i regning 2018. Resultater for 8. og 9. trinn

Tabell 23 viser rekkefølgen til oppgavene i versjon 1 av kohortprøven, hva oppgavene måler, fagområde, oppgaveformat, oppgavens tilknytning til fag, løsningsprosenter for 8. og 9. trinn, kjønnsforskjeller og forskjeller mellom trinn i effektstørrelse. I tillegg vises vanskelighetsgraden til hver enkelt oppgave (b-verdi) og oppgavens mestringsnivå.

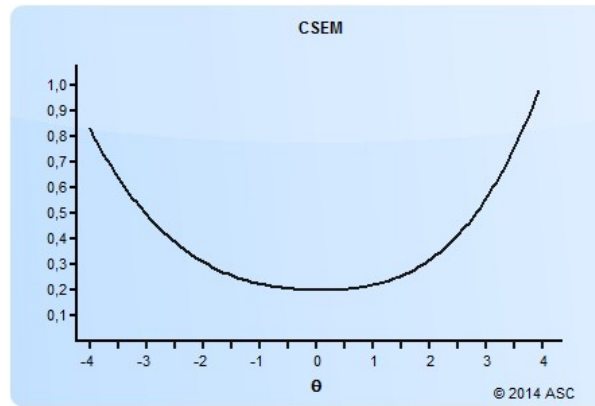
Testinformasjon og måleusikkerhet

Figur 8 og 9 viser testinformasjonen og måleusikkerheten til prøven for hvert ferdighetsnivå.

Testinformasjonen er høyest ved ferdighet 0,05, og måleusikkerheten er lavest for samme ferdighet.



Figur 8 Testinformasjonen til prøven for 2018



Figur 9 Måleusikkerheten til prøven for 2018

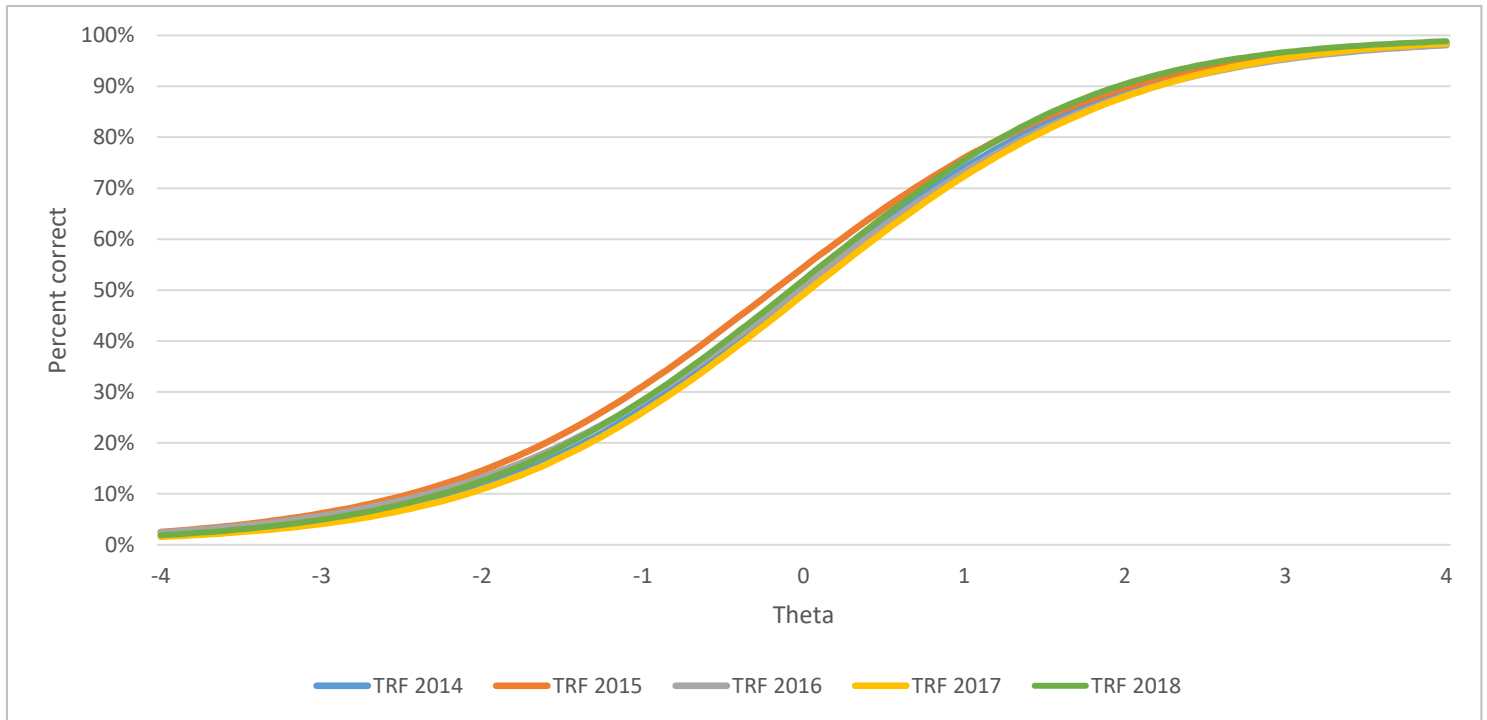
Prøven for 2018 hadde samsvarende kurver for testinformasjon og måleusikkerhet. Som vanlig har prøven høy testinformasjon og lav måleusikkerhet midt på ferdighetsskalaen. Dette er imidlertid en naturlig konsekvens av at det er flest oppgaver i prøven som ligger midt på ferdighetsskalaen (se figur 3, kapittel 4), som igjen er en konsekvens av hvordan elevene fordeler seg på denne skalaen.

Sammenlignet med prøven for 2017, hadde prøven for 2018 noe lavere testinformasjon, og dermed høyere måleusikkerhet, i hver ende av ferdighetsskalaen. Dette gjenspeiler at prøven for 2018 har færre oppgaver på mestringsnivå 1 og 5 enn hva som var tiltenkt på forhånd. Det er mange hensyn å ta når vi velger ut oppgaver til kohortprøven, der en god fordeling av oppgaver utover vanskelighetsgradsskalaen er ett av kriteriene. Figur 3 (kapittel 4) viser at til sammen ni oppgaver var på mestringsnivå 1 og 5 ut fra tall basert på 2. pilot, men etter gjennomføring var antallet redusert til seks oppgaver (figur 13). I forslaget til kohortprøven for 2019, har vi prøvd, i enda større grad enn tidligere, å finne oppgaver på mestringsnivå 1 der vanskelighetsgraden er et stykke unna grenseverdien til mestringsnivå 2. Dette for å prøve å hindre at oppgaver hopper mellom nivåer i særlig grad. Det samme gjelder for oppgaver på mestringsnivå 5.

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.

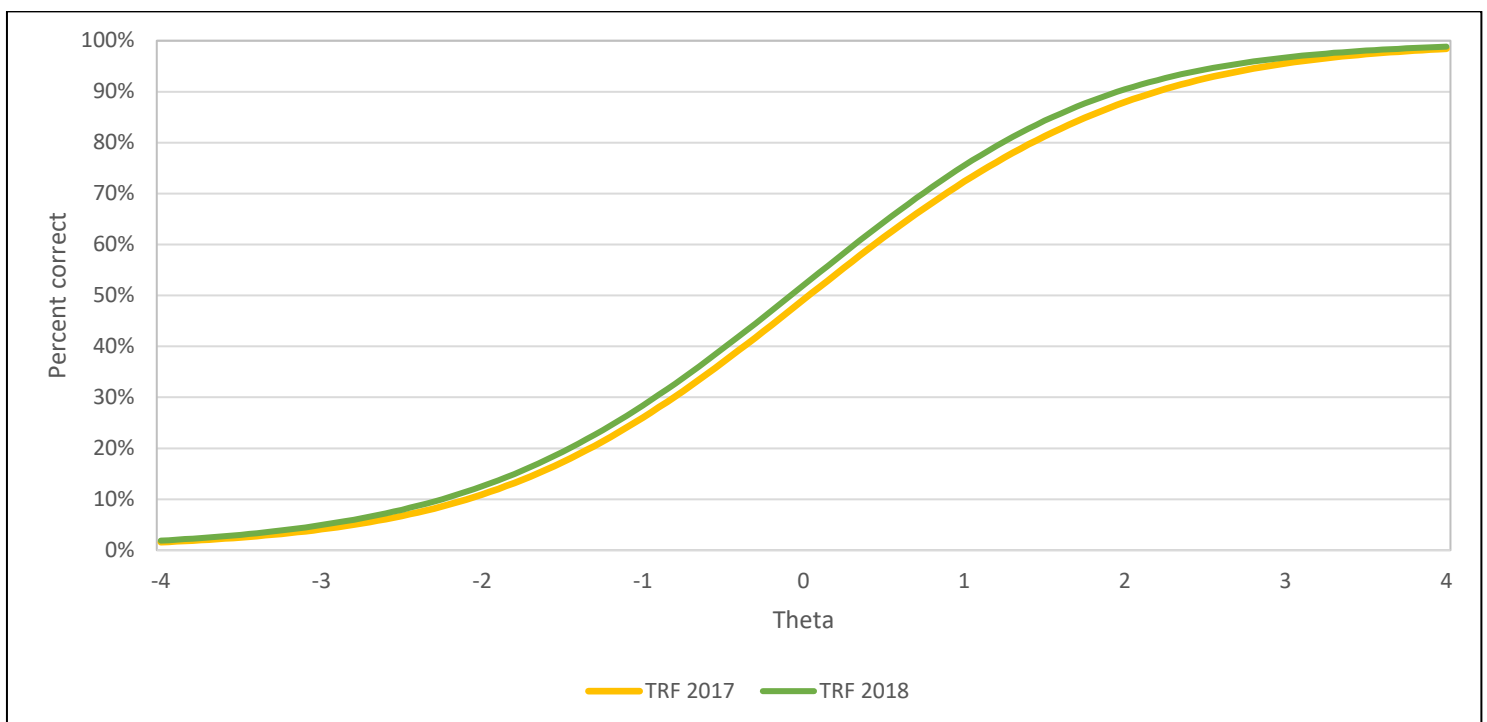
Figur 10 viser TRF for nasjonal prøve i regning fra og med 2014 til og med 2018, der resultatene for 2015–2018 er lenket til resultatene for 2014.



Figur 10 Test response function for prøvene i 2014–2018

Figur 10 viser at det er små forskjeller på prøvene, og at prøven for 2018 er lik i vanskelighetsgrad sammenlignet med de tidligere årene. Imidlertid ser vi at kurven for 2018 ligger marginalt høyere enn kurvene for årene 2014, 2016 og 2017. Det betyr at prøven for 2018 hadde noe lavere vanskelighetsgrad enn de andre. Forskjellen varierer noe på ulike plasser på ferdighetsskalaen sammenlignet med prøvene fra forskjellige år.

Dersom vi bare sammenligner TRF for 2017 og 2018 ser vi at prøven for 2018 har noe lavere vanskelighetsgrad enn prøven for 2017 over hele ferdighetsskalaen (figur 11).



Figur 11 Test response function for prøvene i 2017 og 2018

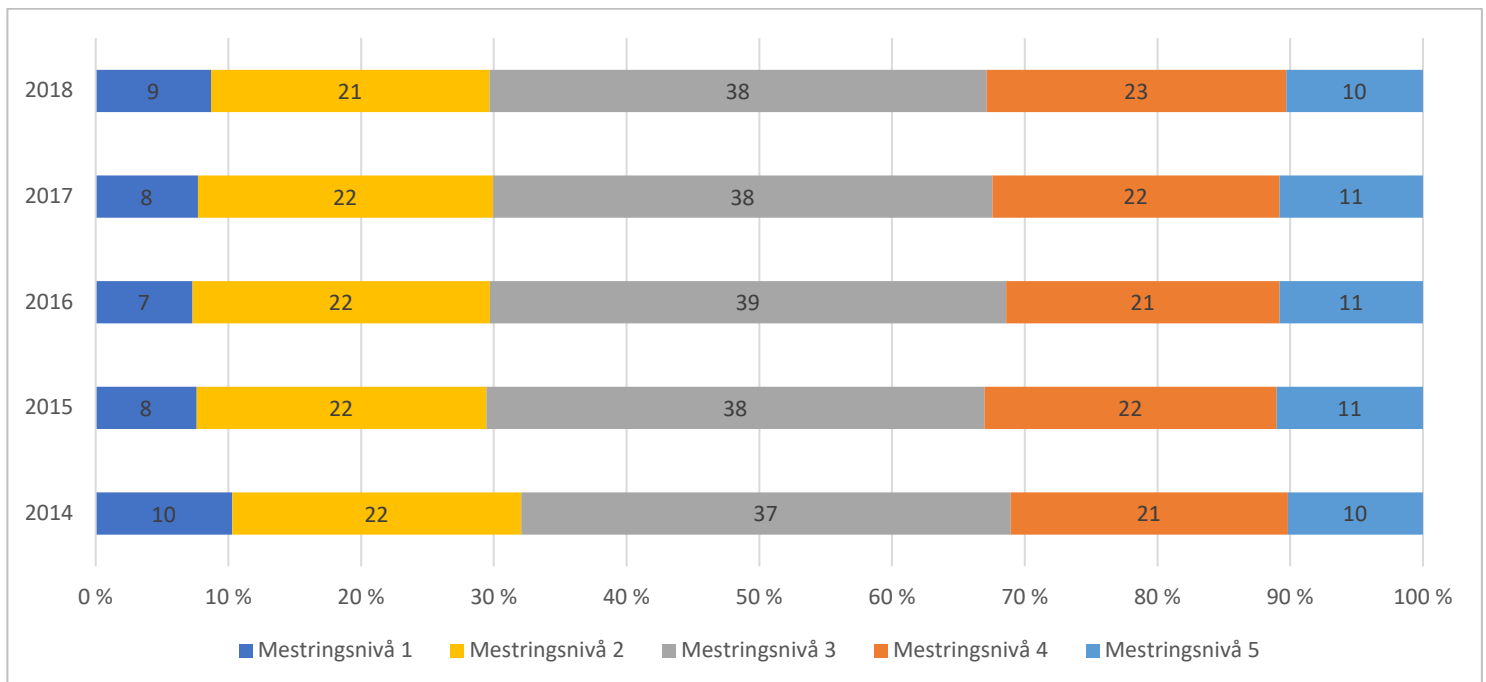
Tolkninger av resultatene

Sammenligning av resultater over tid

Innføring av ny vurderingsskala i 2014, gjør det mulig å måle utvikling over tid. Dette gjøres ved at omtrent seks prosent 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 vanskelighetsgrad 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. Fem år etter innføringen av ny skala er gjennomsnittet fortsatt 50 skalapoeng. Det betyr at elevenes resultater, i de grunnleggende ferdighetene i å kunne regne, har vært veldig stabil i løpet av denne perioden.

Selv om ankerdesignet i nasjonale prøver reduserer forskjellene i vanskelighetsgrad i kohortprøvene fra år til år, eksisterer det en viss lenkefeil (se figur 10 og 11). Ifølge Bjørnsson (2018), er lenkefeilen av den størrelsesorden, at for at en endring skal være statistisk signifikant, må endringen fra ett år til et annet være 0,2 standardavvik eller mer. Dette tilsvarer en endring på 2 skalapoeng fra ett år til et annet.



Figur 12 Prosentvis fordeling på mestringsnivå 2014–2018

Figur 12 viser prosentvis fordeling av elever på de ulike mestringsnivåer for prøvene 2014–2018. Som vi ser er resultatene på nasjonalt nivå stabile, også når det gjelder fordeling på ulike mestringsnivåer.

Sammenligning av resultater for 8. og 9. trinn

Prøven for 9. trinn blir analysert ved at oppgavens vanskelighetsgrad 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 2018 var gjennomsnittlig skalapoeng for 9. trinn 54. Det er samme resultat som alle år siden 2014.

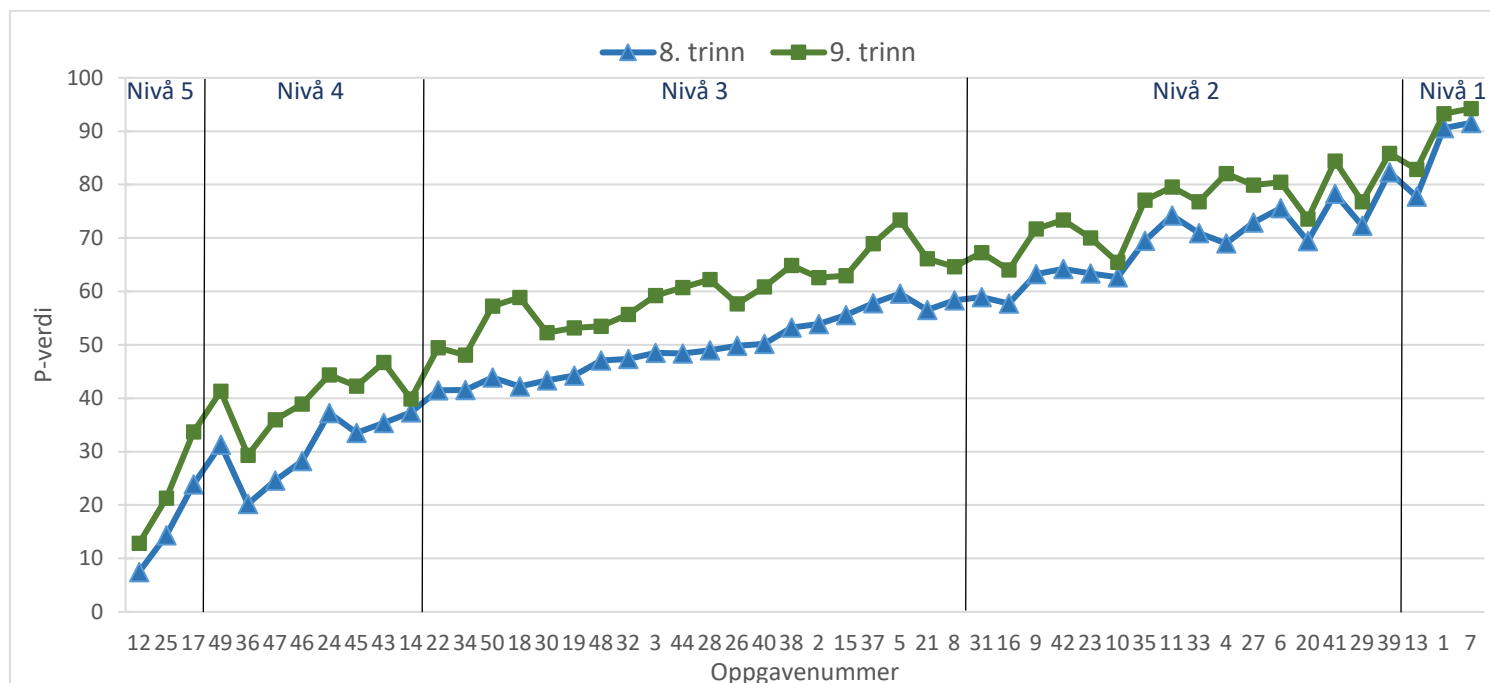
Sammenligner vi gjennomsnittlig P-verdi for 8. og 9. trinn, utgjør forskjellen mer enn 30 prosent av standardavviket ($e = 0,38$). Det betyr at det har middels betydning for resultatet til denne prøven om elevene som løser oppgavene, er fra 9. eller 8. trinn (tabell 24).

År	Effektstørrelse 8. / 9. trinn
2014	0,34
2015	0,36
2016	0,36
2017	0,38
2018	0,38

Tabell 22 Effektstørrelse 8. trinn sammenlignet med 9. trinn fra 2014–2018

Tabell 24 viser at forskjellen mellom trinnene er den samme som i 2017, som igjen er marginalt større sammenlignet med de øvrige årene. Effekten er omtrent lik det en skulle kunne forvente etter ett år lengre skolegang.

Figur 13 viser P-verdi 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 P-verdien til oppgaven. Vanskelighetsgraden defineres ut fra hvor mange skalapoeng en elev må ha for at sannsynligheten for å løse oppgaven skal være 50 prosent.



Figur 13 P-verdier for hver oppgave på 8. og 9. trinn. Oppgavene er sortert etter vanskelighetsgrad på 8. trinn

Analysen viser at oppgave 12 krevde høyest dyktighet og oppgave 7 krevde lavest dyktighet for at elevene skulle ha 50 prosent sannsynlighet for å løse oppgaven. I oppgave 12 skal elevene foreta en omgjøring mellom m/s og km/h, mens de i oppgave 7 skal reflektere og vurdere over tre alternativer knyttet til måling av tid.

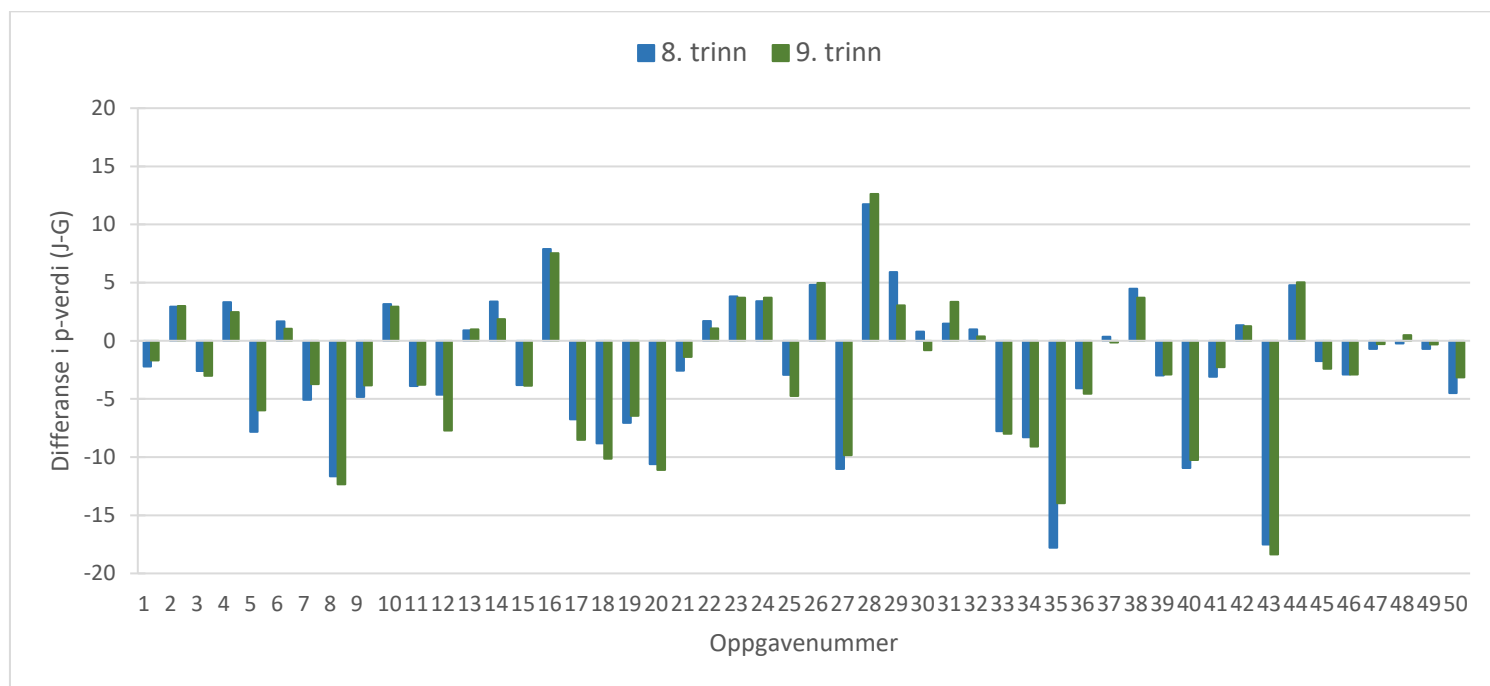
Oppgavene med størst prosentvis endring i P-verdi fra 8. trinn er oppgave 12, 25, 47, 36 og 17. P-verdien til disse oppgavene øker med over 40 % fra 8. trinn til 9. trinn. Det er naturlig at dette gjelder oppgaver med lav P-verdi på 8. trinn. En endring i P-verdi for disse oppgavene, utgjør større prosentvis endring enn samme endring i prosentpoeng for oppgaver med høyere P-verdi på 8. trinn. Det er vanskelig å plukke ut en fellesbetegnelse for hva de oppgavene med størst prosentvis endring måler, men flere av oppgavene handler om arbeid med prefikser og måleenheter. I tillegg utgjør algebraisk tekning, målestokk og prosent matematiske utfordringer i disse oppgavene.

Kjønnsforskjeller

Forskjellene mellom jenter og gutter på 8. trinn var i 2018 to skalapoeng. Guttene hadde i gjennomsnitt 51 skalapoeng, mens jentene hadde 49 skalapoeng. Denne forskjellen er den samme som i de to foregående årene. Tall fra Statistikkportalen til Utdanningsdirektoratet viser at guttene i gjennomsnitt har flere skalapoeng enn jentene i 16 av 18 fylker. Dersom vi ser på gjennomsnittlig løsningsprosent, er forskjellen i gjennomsnitt tilnærmet 2 prosentpoeng i guttenes favør, både for 8. trinn og 9. trinn. Denne forskjellen er lavere enn i prøven for 2017, men det er viktig å poengtere at forskjellen er en konsekvens av de oppgavene som er plukket ut til endelig prøve. Derfor kan det være like interessant å se på kjønnsforskjellen i ankeret. Den har vært stabil siden 2014, med tilnærmet 3 prosentpoeng i guttenes favør på 8. trinn.

Som i 2017, er det i 2018 størst prosentandel jenter på de tre laveste mestringsnivåene og flest gutter på de to høyeste nivåene. Guttene gjør det også best innenfor alle de tre matematiske områdene på de oppgavene som er med i endelig prøve.

Figur 14 viser forskjeller i P-verdier for jenter og gutter på oppgavenivå i prøven for 2018. Positive søyler viser oppgaver som jentene har løst i større grad enn guttene, og negative søyler viser oppgaver guttene har løst i større grad enn jentene. Det er til dels store forskjeller i P-verdi for jenter og gutter, og flere oppgaver som har høyere P-verdi for gutter enn jenter, enn motsatt.



Figur 14 Forskjell i P-verdi for jenter og gutter for hver oppgave, jente – gutt

I prøven for 2018 er det 13 oppgaver som skiller mer enn 5 prosentpoeng i løsningsprosent i guttefavør. Blant disse oppgavene finner vi seks oppgaver der elevene må gjøre om mellom ulike prefikser (35, 43, 27, 40, 17 og 7). Fire av disse oppgavene er blant de fem oppgavene med høyest forskjell i løsningsprosent mellom gutter og jenter. Andre temaer blant de oppgavene som skiller mest i guttefavør er posisjonssystemet, brøk og prosent.

Det er kun to oppgaver som skiller mer enn 5 prosentpoeng i løsningsprosent i jentefavør. I år, som i 2016 og 2017, er det oppgaven der elevene skal systematisere data og framstille resultatet i et diagram, som skiller mest i jentefavør. Oppgaven stiller høye krav til nøyaktighet.

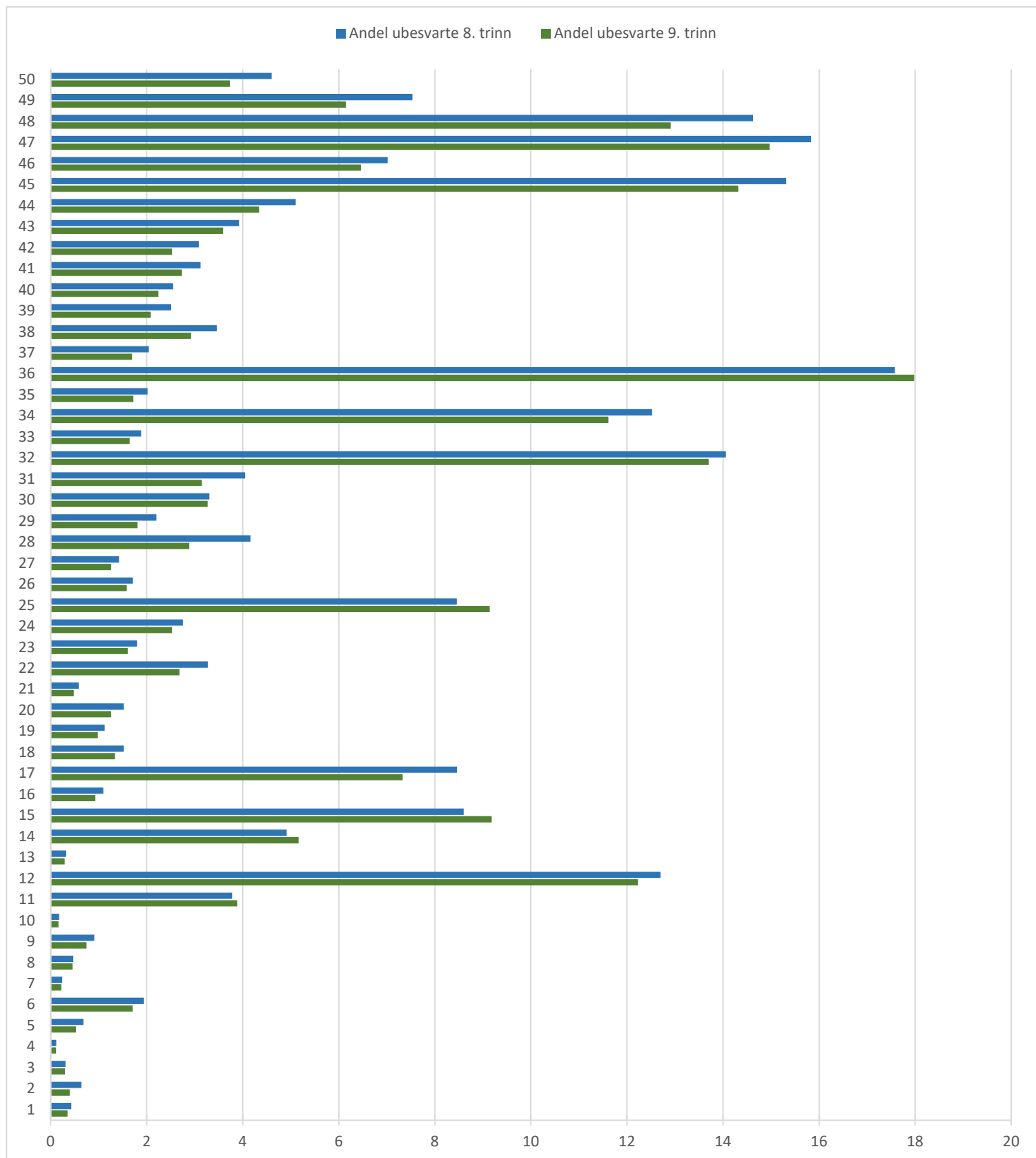
Ubesvarte oppgaver

Gjennomsnittlig andel ubesvart i 2018 var 4,5 prosent for 8. trinn og 4,1 prosent for 9. trinn. Det er omtrent det samme som for prøven i 2017, da de tilsvarende verdiene var 4,5 prosent for 8. trinn og 4,3 prosent for 9. trinn.

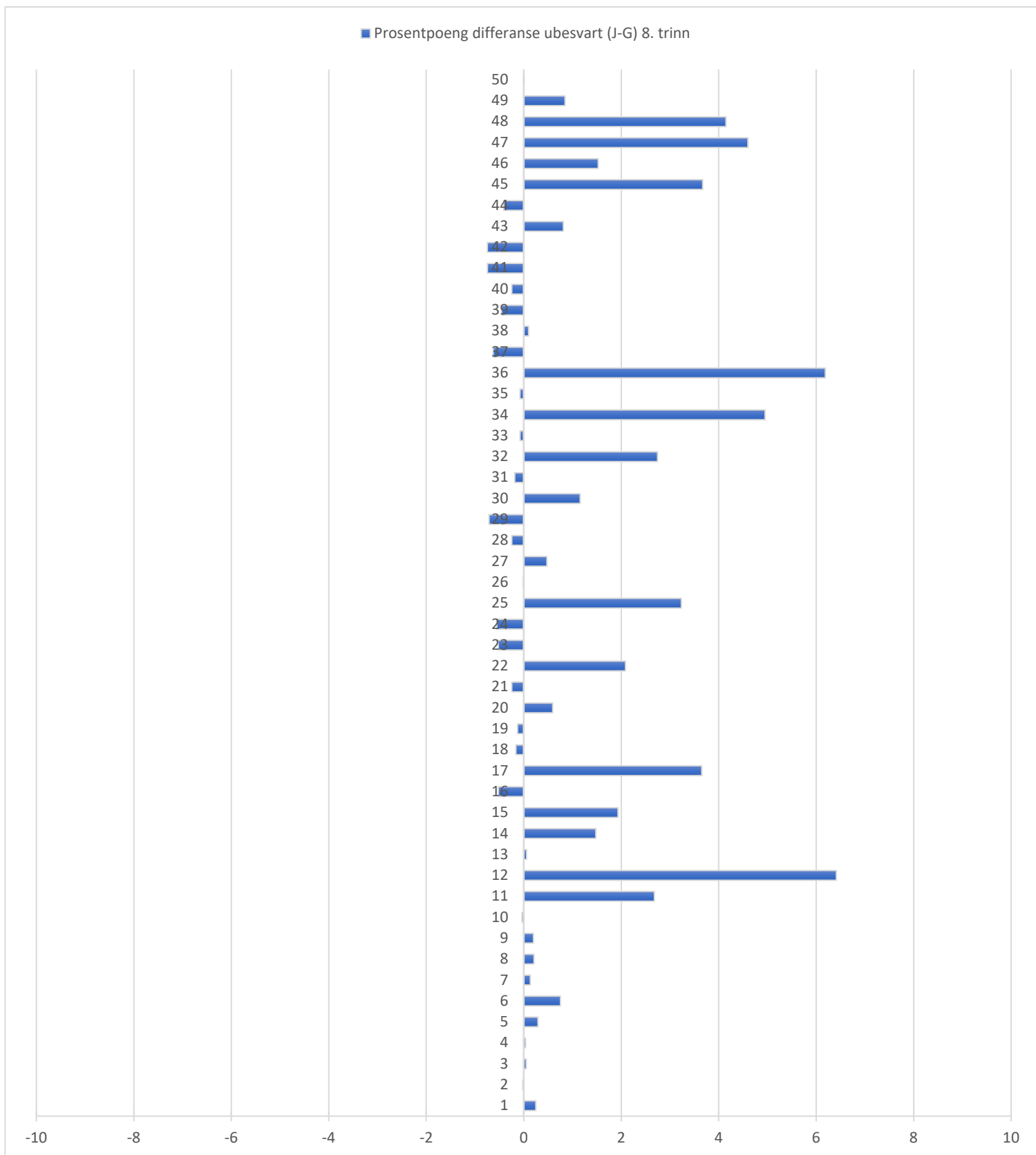
Figur 15 viser prosent ubesvart for hver oppgave for elever på 8. og 9. trinn. Det er ni oppgaver der andelen ubesvart er større enn 10 prosent.

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, kapitel 4), uten at oppgavene har en rekkefølge ut fra vanskelighetsgrad. Det 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.

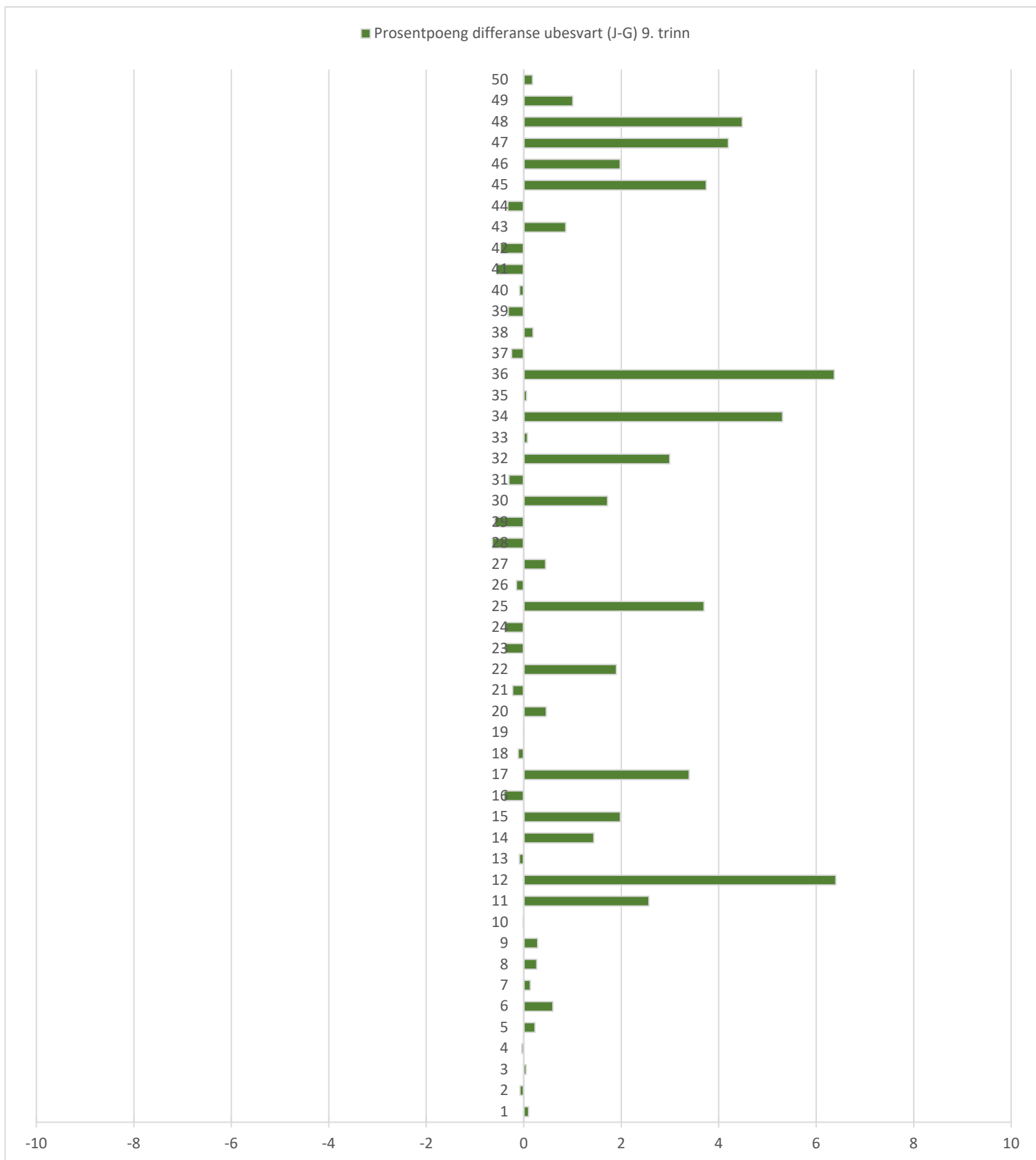
Figur 16 viser forskjellen i andel ubesvarte for jenter og gutter på 8. trinn, og figur 17 viser forskjellen i andel ubesvarte for jenter og gutter på 9. trinn.



Figur 15 Prosent ubesvart for hver oppgave for elever på 8. og 9. trinn



Figur 16 Forskjell i prosent ubesvart for jenter og gutter for hver oppgave på 8. trinn



Figur 17 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 24 og 25.

- 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 24 og 25: Oppgaveanalyse for alle oppgavene i kohortprøvene. Svarfordeling angitt i prosent. Gjennomsnittlig dyktighet (theta) 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 vanskelighetsgrad på en skala fra -4 til 4 ut fra IRT-analyse. Guttene har høyest løsningsprosent i de oppgavene hvor differansen (j-g) er negativ

Teknisk rapport NP8 2018

NP Regning 8. trinn 2018																					
Nr.	Svarfordeling i %							Gjennomsnittlig Theta						T- Rpbis	S- Rpbis	a- verdi	b- verdi	P- verdi	P - jente	P - gutt	Forskjell i P-verdi (J-G)
	F	1	2	3	4	5	Ube- svart	1	2	3	4	5	Ube- svart								
	Å	0	1					0	1												
1		9	91				0	-1,01	0,17				-1,14	0,36	0,33	0,864	-2,052	90,6	89,5	91,7	-2,2
2	1	14	54	23	9		1	-0,24	0,37	-0,46	0,00		-0,70	0,35	0,32	0,438	-0,238	53,9	55,4	52,4	2,9
3	1	22	20	9	48		0	-0,48	0,00	-0,54	0,44		-0,82	0,38	0,35	0,490	0,082	48,5	47,2	49,8	-2,6
4	1	14	69	17			0	-0,91	0,34	-0,28			-0,61	0,42	0,40	0,624	-0,931	69,0	70,7	67,4	3,3
5	1	24	60	10	6		1	-0,77	0,55	-0,57	-0,38		-0,49	0,60	0,57	1,062	-0,341	59,6	55,6	63,4	-7,8
6		22	76				2	-0,79	0,34				-0,99	0,50	0,47	0,904	-1,047	75,6	76,4	74,8	1,7
7	1	92	7	1			0	0,15	-0,93	-0,96			-0,68	0,31	0,29	0,754	-2,333	91,6	89,0	94,1	-5,1
8	1	4	27	58	10		0	-0,48	-0,66	0,43	0,08		-0,37	0,45	0,42	0,625	-0,394	58,3	52,4	64,1	-11,7
9	1	5	63	19	13		1	-0,12	0,40	-0,54	-0,63		-0,38	0,45	0,42	0,644	-0,619	63,2	60,8	65,6	-4,8
10	1	63	11	5	21		0	0,37	-0,33	-0,66	-0,48		-0,59	0,40	0,38	0,551	-0,655	62,6	64,2	61,1	3,1
11		22	74				4	-0,88	0,39				-0,93	0,58	0,54	1,175	-0,871	74,2	72,3	76,1	-3,9
12		80	7				13	0,02	1,37				-0,43	0,38	0,33	1,024	2,063	7,5	5,1	9,8	-4,6
13	1	5	12	78	4		0	-0,48	-0,35	0,22	-0,84		-0,85	0,30	0,28	0,447	-1,842	77,7	78,2	77,3	0,9
14		58	37				5	-0,30	0,71				-0,61	0,51	0,48	0,784	0,521	37,3	39,1	35,7	3,4
15		36	56				9	-0,46	0,49				-0,51	0,50	0,47	0,730	-0,239	55,6	53,6	57,5	-3,8
16		41	58				1	-0,28	0,32				-0,96	0,31	0,29	0,385	-0,524	57,7	61,7	53,8	7,9
17		68	24				8	-0,08	0,72				-0,68	0,38	0,34	0,585	1,397	23,9	20,4	27,2	-6,7
18		56	42				2	-0,41	0,73				-1,08	0,58	0,54	0,966	0,284	42,2	37,7	46,6	-8,8
19	1	26	14	15	44		1	-0,24	-0,77	-0,37	0,66		-0,64	0,54	0,51	0,851	0,221	44,2	40,7	47,7	-7,1
20	1	5	13	69	11		2	-0,27	-0,79	0,30	-0,24		-0,55	0,38	0,35	0,536	-1,060	69,4	64,1	74,7	-10,6
21	1	16	22	57	4		1	-0,22	-0,47	0,39	-0,36		-0,76	0,39	0,37	0,513	-0,354	56,5	55,2	57,8	-2,6
22	1	22	15	41	18		3	-0,16	-0,78	0,55	0,04		-0,55	0,42	0,39	0,577	0,422	41,5	42,3	40,6	1,7
23		35	63				2	-0,48	0,38				-0,94	0,44	0,41	0,624	-0,640	63,4	65,3	61,5	3,8
24		60	37				3	-0,22	0,58				-0,91	0,41	0,38	0,559	0,654	37,2	39,0	35,6	3,4
25		77	14				8	-0,09	1,23				-0,50	0,49	0,44	1,102	1,466	14,3	12,9	15,8	-2,9
26		48	50				2	-0,33	0,47				-0,85	0,42	0,40	0,566	0,009	49,8	52,2	47,4	4,8
27	1	6	15	73	5		1	-0,47	-0,73	0,35	-0,93		-0,75	0,49	0,46	0,831	-0,959	72,9	67,3	78,3	-11,0
28		47	49				4	-0,30	0,45				-0,54	0,40	0,37	0,519	0,051	49,0	54,9	43,2	11,7
29		25	72				2	-0,46	0,27				-0,90	0,36	0,34	0,525	-1,256	72,3	75,3	69,4	5,9
30	1	5	43	36	12		3	-0,76	0,64	-0,24	-0,60		-0,55	0,52	0,48	0,779	0,270	43,4	43,8	43,0	0,8
31	1	59	17	17	3		4	0,35	-0,34	-0,29	-0,58		-0,59	0,37	0,35	0,488	-0,502	58,9	59,7	58,2	1,5
32		39	47				14	-0,43	0,64				-0,56	0,58	0,54	0,941	0,096	47,3	47,8	46,9	1,0
33	1	9	13	71	5		2	-0,70	-0,45	0,34	-0,94		-0,72	0,47	0,44	0,740	-0,928	70,9	67,0	74,8	-7,8
34		46	42				13	-0,34	0,69				-0,57	0,55	0,52	0,881	0,321	41,6	37,4	45,7	-8,3
35	1	69	11	6	11		2	0,37	-0,57	-0,80	-0,62		-0,72	0,49	0,46	0,800	-0,821	69,5	60,5	78,3	-17,8
36		62	20				18	-0,15	1,14				-0,46	0,56	0,51	1,233	1,092	20,2	18,1	22,2	-4,1
37	1	11	17	58	13		2	-0,29	-0,57	0,48	-0,61		-0,70	0,51	0,48	0,774	-0,322	57,8	58,0	57,6	0,3
38	1	43	53				3	-0,46	0,53				-0,65	0,52	0,49	0,787	-0,134	53,2	55,5	51,0	4,5
39	1	6	82	10			3	-1,04	0,27	-0,93			-0,71	0,50	0,46	1,080	-1,285	82,3	80,8	83,8	-3,0
40	1	9	24	14	50		3	-0,21	-0,64	-0,48	0,63		-0,59	0,59	0,55	0,982	-0,010	50,2	44,6	55,6	-10,9
41		19	78				3	-0,89	0,32				-0,77	0,53	0,49	1,058	-1,095	78,3	76,7	79,8	-3,1
42	1	33	64				3	-0,52	0,39				-0,65	0,47	0,44	0,692	-0,637	64,2	64,9	63,5	1,3
43		61	35				4	-0,32	0,80				-0,74	0,56	0,52	0,950	0,547	35,4	26,5	44,0	-17,5
44		47	48				5	-0,44	0,63				-0,78	0,57	0,54	0,922	0,058	48,4	50,8	46,0	4,8
45		51	34				15	-0,25	0,86				-0,64	0,58	0,55	1,047	0,595	33,5	32,6	34,4	-1,7
46	1	65	28				7	-0,28	0,96				-0,47	0,59	0,55	1,134	0,778	28,3	26,8	29,7	-2,9
47		60	25				16	-0,19	0,99				-0,46	0,55	0,51	1,052	0,960	24,6	24,2	24,9	-0,7
48		38	47				15	-0,24	0,50				-0,58	0,45	0,42	0,615	0,136	47,1	46,9	47,2	-0,2
49	1	12	31	15	26	9	8	-0,23	0,52	-0,07	-0,16	0,10	-0,47	0,33	0,31	0,449	1,159	31,3	31,0	31,7	-0,7
50	1	11	9	17	44	15	5	0,08	-0,40	-0,26	0,51	-0,50	-0,44	0,42	0,39	0,570	0,301	43,9	41,6	46,1	-4,5
Gjennomsnitt							4,45							0,464	0,432	0,739	-0,177	53,0	51,9	54,1	-2,2

Tabell 23 Item-analyse for alle oppgaver NP8 2018

Teknisk rapport NP8 2018

NP Regning 9. trinn 2018																			
Nr.	Svarfordeling i %							Gjennomsnittlig Theta						T-Rpbis	S-Rpbis	P-verdi	P-jente	P-gutt	Forskjell i P-verdi (J-G)
	F	1	2	3	4	5	Ube-svart	1	2	3	4	5	Ube-svart						
	Å	0	1					0	1										
1		6	93				0	-1,03	0,14				-1,25	0,31	0,29	93,3	92,4	94,1	-1,7
2	1	12	63	17	8		0	-0,31	0,33	-0,58	-0,08		-0,80	0,35	0,33	62,6	64,1	61,1	3,0
3	1	17	16	8	59		0	-0,58	-0,15	-0,68	0,40		-0,90	0,42	0,39	59,2	57,7	60,7	-3,0
4	1	8	82	10			0	-1,15	0,25	-0,56			-0,64	0,42	0,40	82,1	83,3	80,9	2,5
5	1	15	73	8	3		1	-1,02	0,41	-0,79	-0,65		-0,66	0,59	0,57	73,4	70,4	76,3	-6,0
6		18	80				2	-0,86	0,29				-1,12	0,48	0,46	80,4	81,0	79,9	1,0
7	1	94	5	1			0	0,13	-1,07	-1,20			-0,83	0,30	0,28	94,2	92,4	96,1	-3,7
8	1	5	19	65	11		0	-0,62	-0,78	0,38	-0,03		-0,50	0,44	0,41	64,6	58,4	70,7	-12,3
9	1	5	72	14	9		1	-0,24	0,34	-0,67	-0,80		-0,61	0,46	0,44	71,7	69,7	73,6	-3,8
10	1	65	10	4	20		0	0,38	-0,43	-0,71	-0,57		-0,74	0,44	0,42	65,5	66,9	64,0	2,9
11		17	80				4	-0,94	0,32				-1,07	0,55	0,52	79,6	77,6	81,4	-3,8
12		75	13				12	-0,04	1,19				-0,52	0,44	0,37	12,8	8,9	16,6	-7,7
13	1	4	10	83	3		0	-0,58	-0,41	0,19	-0,95		-0,89	0,29	0,28	82,8	83,3	82,4	1,0
14		55	40				5	-0,32	0,68				-0,73	0,52	0,48	39,9	40,8	38,9	1,9
15		28	63				9	-0,50	0,41				-0,66	0,49	0,47	63,0	61,0	64,9	-3,9
16		35	64				1	-0,31	0,28				-1,03	0,30	0,28	64,0	67,9	60,3	7,5
17		59	34				7	-0,18	0,68				-0,82	0,45	0,40	33,7	29,4	37,9	-8,5
18		40	59				1	-0,62	0,54				-1,17	0,59	0,56	58,9	53,7	63,9	-10,1
19	1	21	11	13	53		1	-0,31	-0,91	-0,51	0,57		-0,73	0,56	0,53	53,2	49,9	56,4	-6,5
20	1	6	9	74	11		1	-0,27	-0,90	0,26	-0,27		-0,65	0,36	0,33	73,5	67,9	79,0	-11,1
21	1	13	17	66	4		0	-0,28	-0,56	0,32	-0,51		-0,85	0,38	0,36	66,1	65,4	66,8	-1,4
22	1	18	11	49	19		3	-0,39	-0,95	0,50	0,04		-0,71	0,44	0,41	49,5	50,0	48,9	1,1
23		28	70				2	-0,51	0,32				-1,10	0,41	0,39	70,0	71,9	68,2	3,7
24		53	44				3	-0,25	0,49				-1,06	0,40	0,37	44,3	46,2	42,5	3,7
25		70	21				9	-0,16	1,06				-0,60	0,53	0,47	21,3	18,9	23,6	-4,8
26		41	58				2	-0,38	0,40				-0,96	0,41	0,39	57,7	60,2	55,2	5,0
27	1	4	11	80	4		1	-0,64	-0,87	0,29	-1,08		-0,93	0,49	0,47	79,9	74,9	84,8	-9,8
28		35	62				3	-0,38	0,34				-0,74	0,38	0,36	62,2	68,6	56,0	12,6
29		21	77				2	-0,51	0,25				-1,04	0,36	0,34	76,8	78,3	75,3	3,0
30	1	3	52	32	9		3	-0,80	0,57	-0,37	-0,81		-0,63	0,55	0,52	52,3	51,9	52,7	-0,8
31	1	67	14	14	2		3	0,33	-0,47	-0,42	-0,77		-0,69	0,41	0,39	67,3	69,0	65,6	3,3
32		31	56				14	-0,51	0,56				-0,70	0,59	0,56	55,7	55,9	55,5	0,4
33	1	7	11	77	3		2	-0,88	-0,61	0,31	-1,06		-0,78	0,48	0,46	76,8	72,7	80,7	-8,0
34		40	48				12	-0,35	0,59				-0,72	0,54	0,50	48,1	43,4	52,5	-9,1
35	1	77	9	5	8		2	0,31	-0,69	-1,01	-0,76		-0,86	0,49	0,47	77,1	70,0	84,0	-14,0
36		53	29				18	-0,26	1,00				-0,55	0,62	0,56	29,3	27,0	31,6	-4,5
37	1	9	12	69	8		2	-0,30	-0,75	0,38	-0,82		-0,83	0,50	0,48	68,9	68,9	69,0	-0,2
38	1	32	65				3	-0,62	0,43				-0,79	0,53	0,51	64,8	66,7	63,0	3,7
39	1	4	86	8			2	-1,20	0,24	-1,02			-0,83	0,50	0,47	85,9	84,4	87,3	-2,9
40	1	9	15	13	61		2	-0,38	-0,78	-0,69	0,52		-0,74	0,60	0,57	60,9	55,6	65,9	-10,3
41		13	84				3	-1,03	0,26				-0,91	0,51	0,49	84,4	83,2	85,5	-2,3
42	1	24	73				3	-0,66	0,32				-0,80	0,47	0,45	73,4	74,0	72,8	1,2
43		50	47				4	-0,44	0,67				-0,91	0,59	0,55	46,7	37,4	55,7	-18,4
44		35	61				4	-0,60	0,51				-0,92	0,58	0,56	60,7	63,2	58,2	5,0
45		43	42				14	-0,34	0,74				-0,76	0,61	0,57	42,3	41,0	43,4	-2,4
46	1	55	39				6	-0,39	0,80				-0,61	0,61	0,57	38,9	37,5	40,4	-2,9
47		49	36				15	-0,28	0,81				-0,62	0,59	0,54	36,0	35,8	36,1	-0,3
48		34	54				13	-0,25	0,44				-0,73	0,45	0,42	53,5	53,7	53,2	0,5
49	1	9	41	14	21	9	6	-0,34	0,49	-0,16	-0,30	0,10	-0,60	0,38	0,35	41,3	41,1	41,4	-0,3
50	1	12	5	12	57	9	4	-0,11	-0,66	-0,47	0,44	-0,68	-0,59	0,46	0,43	57,3	55,7	58,8	-3,1
Gjennomsnitt							4,11							0,473	0,444	61,2	60,0	62,3	-2,3

Tabell 24 Item-analyse for alle oppgaver NP9 2018

REFERANSER

- Björnsson, J. K. (2018). Om lenkefeil og ekvivaleringsmetoder på nasjonale prøver: Evaluering av endring over tid. In (pp. 16): Acta Didactica Norge.
- Kunnskapsdepartementet, & Utdanningsdirektoratet. (2006). Læreplanverket for Kunnskapsløftet. Revidert utgave 2013. Retrieved from <https://www.udir.no/kl06/MAT1-04>
- Utdanningsdirektoratet. (2017a). Rammeverk for grunnleggende ferdigheter. Retrieved from https://www.udir.no/globalassets/upload/larerplaner/lareplangrupper/rammeverk_grf_2012.pdf
- Utdanningsdirektoratet. (2017b). Rammeverk for nasjonale prøver. Retrieved from <https://www.udir.no/eksamen-og-prover/prover/rammeverk-for-nasjonale-prover/>