



UNIVERSITETET I BERGEN

Nasjonale prøvar i engelsk 2018 5. trinn

Rapport februar 2019

Eli Moe og Hildegunn Lahlum Helness

Institutt for framandspråk

Innhold

OM RAPPORTEN	3
GENERELT OM NASJONAL PRØVE I ENGELSK FOR 5. TRINN	4
Nasjonal prøve i engelsk for 5. trinn og Læreplan i engelsk (ENG1-03)	4
Kva måler prøvane på 5. trinn	4
Om teksttypar og tema i prøvane på 5. trinn	5
Om kategorisering av item	6
Analyse av data	6
DEL 1 – PILOTERING AV NP ENGELSK 5. TRINN HAUSTEN 2017	7
1. Pilotering og forslag til prøvar 2018	7
1.1 Generell informasjon om pilotering til prøven i 2018	7
1.2 Forslag til prøve på 5. trinn 2018	8
1.3 Kva måler prøvane på 5. trinn i 2018	9
1.4 Om kjønnskilnader	10
DEL 2 – NASJONALE PRØVAR I ENGELSK 5. TRINN 2018	11
2. Nasjonal prøve i engelsk 5. trinn haust 2018 - hovudresultat	11
2.1 Gjennomføring 2018	11
2.2 Oppsummerande statistikk 5. trinn	11
2.3 Overordna IRT-analysar – theta, <i>b</i> -parameter og <i>a</i> -parameter	12
2.3.1 Theta	12
2.3.2 <i>b</i> -parameter	13
2.3.3 <i>a</i> -parameter	14
2.4 <i>p</i> -verdi	15
2.5 Skilnad i <i>p</i> -verdi pilotering og reell prøve 5. trinn	16
2.6 Polytome item	18
2.6.1 Gap-filling	18
2.6.2 Write the word	19
TIL SLUTT	25
Vedlegg 1: Piloteringsdesign hausten 2017	26
Vedlegg 2: Forslag til Nasjonal prøve i engelsk 5. trinn 2018 – oversikt og statistiske mål etter utprøving	28
Vedlegg 3: Hovudprøve 5. trinn 2018 – item for item-oversikt – reell prøve	31

OM RAPPORTEN

Denne rapporten beskriv grunnlaget for og utviklinga av Nasjonal prøve i engelsk for 5. trinn 2018. Samtidig dokumenterer rapporten at prøven er valid og reliabel. Det er ein samanheng mellom relevante kompetansemål i Læreplan i engelsk (ENG1-03) og innhaldet i prøven. Tema i prøven er til saman eit uttrykk for det prøven skal måle. Prøveresultatet skil mellom elevane på ulike ferdigheitsnivå på ein påliteleg måte.

Først i rapporten er ein del grunnleggjande informasjon om Nasjonal prøve i engelsk for 5. trinn oppsummert (til dømes samanhengen mellom læreplan i engelsk og prøve, kva prøven måler, itemformat og korleis tema i prøven blir kategorisert).

Rapporten har elles to hovuddelar: Del 1 som gjer greie for arbeidet før prøven vart gjennomført (utprøving og val av item til endeleg prøve), og Del 2 som dokumenterer korleis Nasjonal prøve i engelsk for 5. trinn 2018 fungerte for elevane på 5. trinn som tok prøven.

Sidan 2014 har det ved hjelp av dei nasjonale prøvane blitt undersøkt om ferdigheitene til elevane (i engelsk, lesing og rekning) endrar seg over tid. Dette blir gjort ved at eit utval elevane tar ein ankerprøve i engelsk, lesing og rekning, medan dei fleste tar hovudprøven som blir pilotert og sett saman året før.. Om lag halvparten av tema for prøvane i engelsk er felles for hovudprøven og ankerprøven. Resten av tema i ankerprøven er ankeritem som går igjen i ankerprøvar frå år til år. På den måten kan resultatane på nasjonale prøvar samanliknast frå år til år.

Denne rapporten fokuserer på Nasjonal prøve i engelsk for 5. trinn, hovudprøven. I samband med IRT-analysar er datamaterialet for både hovudprøve og ankerprøve lagt til grunn.

GENERELT OM NASJONAL PRØVE I ENGELSK FOR 5. TRINN

Nasjonal prøve i engelsk for 5. trinn og Læreplan i engelsk (ENG1-03)

Den nasjonale prøven i engelsk for 5.trinn måler nokre av kompetansemåla i engelsk etter 4. trinn. Når nokre av læreplanens kompetansemål blir målt i prøven (og ikkje alle), så er det fordi berre reseptive ferdigheiter kan målast og skårast automatisk i ein digital prøve (per i dag). Dei aktuelle kompetansemåla sorterer under «Skriftlig kommunikasjon» i læreplanen.

«Mål for opplæringen er at eleven skal kunne:

- bruke enkle lese- og skrivestrategier
- forstå sammenhengen mellom engelske språklyder og bokstaver og trekke lyder sammen til ord
- forstå betydningen av ord og uttrykk ut fra sammenhengen de er brukt i
- lese, forstå og skrive engelske ord og uttrykk knyttet til egne behov og følelser, dagligliv, fritid og interesser
- forstå hovedinnholdet i enkle tekster om kjente emner
- bruke noen vanlige småord og enkle stave- og setningsmønstre

Læreplanen uttrykkjer at elevane skal arbeide med engelsk med utgangspunkt i det kjende, for eksempel «dagligliv, fritid og interesser» og også «nære omgivelser og egne opplevelser». I tillegg understrekar læreplanen at elevane skal utforske engelsk i samband med «ulike levesett, tradisjoner og skikker i engelskspråklige land», og «engelskspråklig litteratur og barnekultur». Dette prøver vi sjølvsagt å ta omsyn til i tekstane som blir brukt i prøven. Det er likevel viktig å understreke at elevane ikkje treng spesifikk kunnskap for å svare på oppgåvene i prøven. Den nasjonale prøven i engelsk er ein språkprøve og ikkje ein kunnskapsprøve. Elevane skal kunne svare på spørsmåla ved å lese tekstane dei blir presentert for i prøven.

Hittil har leseferdigheiter blitt målt i prøven. Det kunne også vere aktuelt (og viktig) å måle elevane sine lytteferdigheiter i ein slik prøve. Så langt har ein vurdert det slik at det ikkje er teknisk forsvarleg å måle lytteferdigheiter i og med at ikkje alle skular har dataliner som kan takle at mange elevar skal laste lyttetekstar samtidig. Om nokre år ser ein kanskje annleis på dette.

Kva måler prøvane på 5. trinn

Dei nasjonale prøvane i engelsk skal måle generell leseforståing, vokabular og grammatikk. På 5. trinn betyr dette blant anna:

- finne informasjon i ulike enkle tekstar
- forstå enkeltord og hovudinnhald i ulike enkle tekstar

- forstå vanlege ord og uttrykk knytt til daglegliv og fritid
- forstå tydinga av ord og uttrykk ut frå samanhengen dei er brukt i
- bruke enkle stave- og setningsmønster
- bruke enkle setnings- og tekstmønster
- bruke ordforråd og omgrep

Det vi har funne ut i løpet av dei siste åra, er at konstruktet for prøvane eigentleg omfattar to hovuddimensjonar: 1) Nivået i teksten elevane må prosessere på for å løyse oppgåva og 2) Føremålet lesinga har (leseoperasjonane oppgåva krev)

Prosesseringsnivå: Eleven kan finne svaret på ei einskild oppgåve enten i eit ord/uttrykk, ein setning eller ved å kople informasjon frå ulike deler av teksten.

Føremålet lesinga har: Ei oppgåve kan krevje at eleven kan finne informasjon og forstå detaljar, forstå hovudinnhald, tolke informasjon (trekke slutningar), forstå samanheng i tekstar og oppfatte og bruke ord, stavemønster og grammatiske mønster.

Oppgåvene i prøven er utforma for å dekke alle desse aspekta. Det er likevel viktig å vere klar over at i mange tilfelle blir fleire ferdigheiter målt samtidig. I korte tekstar for dei yngste vil det til dømes ofte vere vanskeleg å skilje mellom å forstå hovudinnhald og detaljar. I slike tilfelle er det viktig å forstå detaljane (orda) for å forstå hovudinnhaldet.

Angela Hasselgreen, Craig Grocott og Torbjørn Torsheim har skrive om arbeidet vi har gjort dei siste åra om dette.

Angela Hasselgreen, Craig Grocott, Torbjørn Torsheim: Quality Assurance in the National Tests of English: Investigating What Makes Reading Difficult. In Cardona, M. D., R. Koppen and I. Lunde (eds.) *Remaining Relevant: Modern language Studies Today*, Bergen Language and Linguistics Studies, 7, Bergen: Norway. Lenke til artikkel: <https://bells.uib.no/index.php/bells/issue/view/247/showToc>

Om teksttypar og tema i prøvane på 5. trinn

Elevane på 5. trinn blir presentert for setningar og tekstar av varierende lengd og kompleksitet og frå ulike kjelder. Mange tekstar er korte, det vil seie at dei inneheld relativt få setningar. Mange av dei korte oppgåvene inneheld bilde til støtte for lesinga. Andre tekstar er mellom-lange, og to tekstar i prøven vil nok oppfattast som lange av elevane på 5. trinn. Dei såkalla lange tekstane finst i formata *Multiple choice long* og *Who could say*.

Elevar som har øvd på å lese med ulikt fokus og bruke ulike lesestrategiar, bør ha eit godt utgangspunkt for å svare på oppgåvene i dei nasjonale prøvane i engelsk. På same måte burde prøvane eignast seg godt i etterkant, for å øve på å lese tekstar av ulik lengd og kompleksitet.

Dersom ein kastar eit blikk nedover kolonna «Namn» i tabellane i Vedlegg 2 og 3, får ein eit inntrykk av ulike tema som finst i prøvane. Vi ser at tema ofte er knytt til kjente situasjonar og forhold som daglegliv, familie, dyr, fritidsaktivitetar, engelskspråklege land og liknande.

Elles møter elevane ulike teksttypar i prøvane som beskrivelsar, forteljingar, faktatekstar, dagboknotatar, brev, tabellar og liknande.

Om kategorisering av item

I rapporten vi skreiv for eitt år sidan, gjorde vi greie for arbeidet vi har gjort for å validere prøvekonstruktet først og fremst for dei nasjonale prøvane på 5. trinn. Planen er å sjå nærare på prøvane på 8. trinn også. Ein av dei viktige tinga som kom ut av arbeidet med prøvane på 5. trinn, er at vi kategoriserer itema ved hjelp av to dimensjonar: 1. Prosesseringsnivå og 2. Føremål med lesinga. Det vi ser når vi tar i bruk kategoriseringa, er at vi ynskjer å gjere nokre endringar. Til dømes såg vi at når vi skulle bruke denne kategoriseringa på prøvane på 8. trinn, så hadde vi behov for ein ny underkategori for føremål med lesinga, nemleg «å forstå avansert vokabular». Når vi går gjennom oppgåver og gjer endringar, kategoriserer vi spørsmåla samtidig. Og i samband med dette arbeidet, diskuterer vi stadig kategoriseringa og ynskje om endringar kan komme opp.

Analyse av data

Data frå piloteringa hausten 2017 har blitt analysert ved hjelp av to IRT-program: OPLM og Xcalibre. Norman Verhelst har hjelpt oss med OPLM-analysar. Hilde Olsen i Utdanningsdirektoratet har hjelpt oss med Xcalibre-analysar.

Data frå gjennomføringa av prøvane hausten 2018 er analysert ved hjelp av Xcalibre og SPSS. Xcalibre er brukt i analysar for heile datamaterialet for 5. trinn under eitt, det vil seie at data både frå hovudprøve og ankerprøve er inkludert. SPSS er brukt i ein del klassiske analysar med utgangspunkt i data frå hovudprøven på 5. trinn.

DEL 1 – PILOTING AV NP ENGELSK 5. TRINN HAUSTEN 2017

1. Pilotering og forslag til prøvar 2018

1.1 Generell informasjon om pilotering til prøven i 2018

Tidspunkt for pilotering:

September 2017

Tal på prøveversjonar som vart pilotert:

8 versjonar

Piloteringsdesign

Eit ufullstendig lenka design (incomplete linked design) som inkluderte 8 pilotprøvar for 5. trinn. Sjå vedlegg 1, s. 26.

Tal på item som vart pilotert:

174 item

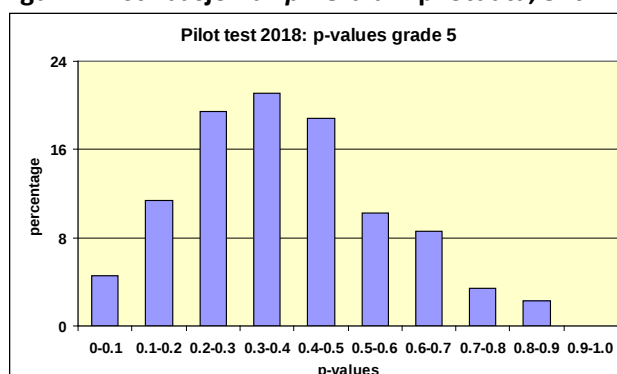
Tal på elevar som deltok:

1910

Samansetjing av 2018-prøve

Pilotdata vart analysert både ved hjelp av OPLM og Xcalibre. Gjennomsnittleg p -verdi for alle itema som vart pilotert på 5. trinn var 0,385. Vanskegraden på itema som vart piloterte varierte frå 0,046 (vanskelegaste item) til 0,898 (lettaste item). Dette betyr at det vart pilotert mange vanskeleg item hausten 2017. Sjølv om det er mange lette item blant dei som vart piloterte, vil vi dei kommande åra prøve å pilotere fleire lettare item. Figur 1 viser distribusjon av p -verdiar i pilotdata.

Figur 1: Distribusjon av p -verdiar i pilotdata, 5. trinn.



Etter dei innleiande analysane, vart åtte item tatt ut på grunn av svært dårleg diskriminering. I arbeidet med å setje saman eit forslag til prøve for 2018, tok vi utgangspunkt i 2017-prøven for 5. trinn når det gjaldt tal på item totalt og også innanfor kvart itemformat. Vi måtte også

balansere følgjande faktorar: itemformat, item totalt, tekstlengd, tema, vanskegrad og diskriminering.

1.2 Forslag til prøve på 5. trinn 2018

Tabell 1 gir ei oversikt over prøven vi foreslår for 5. trinn i 2018. Tabellen viser kor mange item som er valt ut innanfor kvart format, og formatrekkefølgja er den same som i 2017, versjon 1.

Tabell 1: Item per format i prøven for 5. trinn 2018

Format	Nasjonal prøve 2018 5. trinn
Click item	4
Click picture	5
Click and drag	4
Click text	4
What time is it?	2
Gap	2 (5)*
Mc short	5
Mc long	5
Who could say	6
Write the word (grammar/voc)	2 (6)*
Tal på item totalt	39 (46)

* Desse oppgåvene er polytome, det vil seie at eleven må gi fleire svar for å svare fullstendig på oppgåva. Dersom vi tar *Gap*-oppgåva som eit døme, så møter elevane der ein kort tekst med to eller tre luker. Via ein nedtrekksmeny må eleven velje eitt av fire alternativ for kvar av dei tre lukene.

Dei aller fleste oppgaveformata har item som blir skåra dikotomt (0/1). Eleven avgir eitt svar på desse oppgåvene. I dei polytome oppgåvene derimot, avgir eleven fleire svar.

Tabell 2 gir ei oversikt over vanskegrad (*p*- og *b*-verdiar) og diskrimineringsindeks (*a*-verdi) for prøven på 5. trinn. Desse verdiane er basert på piloteringsdata.

Tabell 2: Gjennomsnittleg p -, b - og a -verdi i prøven for 5. trinn 2018, basert på utprøvningsdata

	Nasjonal prøve 5. trinn
Tal på item	39 (46)
Gjennomsnittleg p -verdi	0,468
Gjennomsnittleg b -verdi, OPLM	-0,162
Gjennomsnittleg a -verdi OPLM*	4,13

*Den gjennomsnittlege a -verdien i analysen av alle piloterte item, vart sett til 3. Det vart difor lagt vekt på å velje item med høgre diskriminering til prøven, der dette var mogeleg.

1.3 Kva måler prøvane på 5. trinn i 2018

Tabell 3 gir ei oversikt over kva hovudprøven 5. trinn måler.

Tabell 3 : Oversikt over kva itema i hovudprøven på 5. trinn 2018 måler

		Finne info og forstå detaljar	Forstå hovudinnhald	Tolke	Bruke ord, stave- og gram. mønst.
Prosesseringsnivå	Vokabular	5		1	
	Setning	15		7	4
	Ulike deler av teksten	3	2	2	

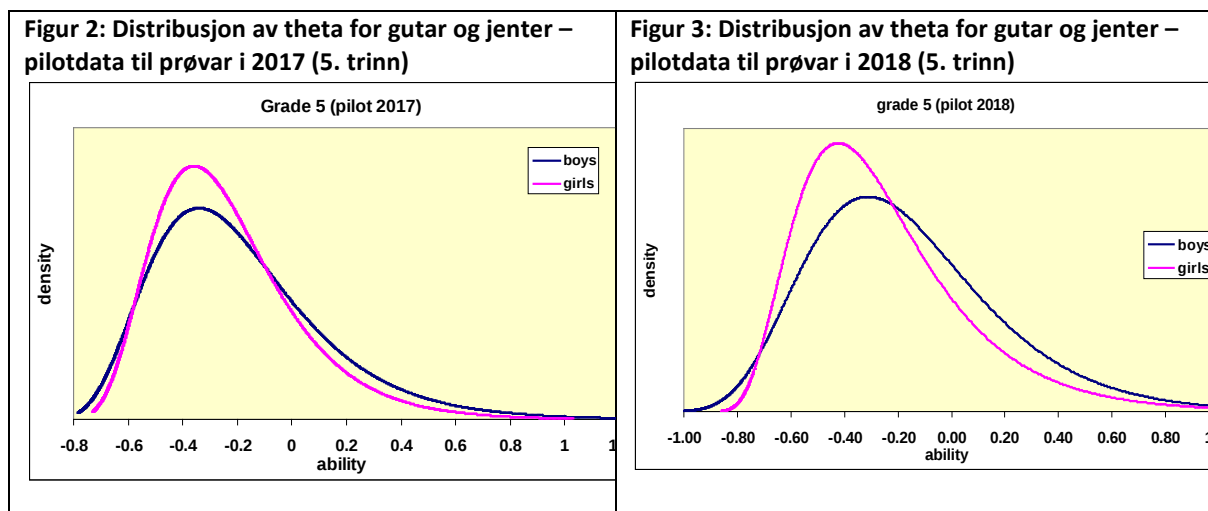
Tabell 3 må forståast slik:

Hovudprøven på 5. trinn inneheld

- 23 item der eleven skal finne informasjon. I fem tilfelle er svaret å finne i eit ord, i 15 tilfelle i ein setning og i tre tilfelle ved å kople informasjon frå ulike delar av teksten.
- to item der eleven skal forstå hovudinnhald. I begge tilfella kan dette gjerast ved å kople informasjon frå ulike deler av teksten.
- 10 item der eleven må tolke innhaldet for å finne svaret. I eitt tilfelle finst svaret i eit ord, i 7 tilfelle finst svaret i ein setning og i to tilfelle må eleven kople informasjon frå ulike stader i teksten.
- fire item der eleven skal bruke ord, stavemønster eller grammatiske mønster. For å gjere dette, må eleven forstå setninga dette skal brukast i.

1.4 Om kjønnskilnader

Norman Verhelst viser i sin rapport at pilotdata indikerer at skilnaden mellom gutar og jenter på 5. trinn har auka dei siste åra i gutane sin favør. Dette er med utgangspunkt i pilotdata. Figurane 2 og 3, som er henta frå Verhelst sin rapport, indikerer at distribusjonen av theta for gutar og jenter er til dels forskjellig, med utgangspunkt i piloteringsdata.



Vi ser at i 2018 er distribusjonen av theta for jentene meir asymmetrisk enn den for gutane. Det er ein større del av jentene med theta i den lågare enden av skalaen..

Tabell 4: (Verhelsts tabell): Effektstørrelse for kjønn – basert på piloteringsdata

År:	2016	2017	2018
5. trinn	-0.031	-0.117	-0.269

Dersom effektstørrelsen for kjønn er rundt 0, betyr det at forskjellen mellom kjønnene er minimal. Sjølv om kjønnskilnaden ikkje er stor på 5. trinn, er han større enn før. Sidan 2014 har gutane prestert best på dei nasjonale prøvane i engelsk. Denne skilnaden ser altså ut til å auke. Når prøvane er gjennomført til hausten (2018), vil vi sjå nærare på om skilnaden mellom gutar og jenter framleis viser samme tendens.

DEL 2 – NASJONALE PRØVAR I ENGELSK 5. TRINN 2018

2. Nasjonal prøve i engelsk 5. trinn haust 2018 - hovudresultat

2.1 Gjennomføring 2018

Totalt var det 60 003 elevar som tok Nasjonal prøve i engelsk for 8. trinn hausten 2018. 56 463 elevar tok hovudprøven som vart sett saman etter piloteringa hausten 2017. Resten av elevane, 3540 elevar, tok ankerprøven.

Hovudprøven vart gjennomført i tre versjonar som inneheldt dei same itema, men rekkefølga på itema varierte frå versjon til versjon.

2.2 Oppsummerande statistikk 5. trinn

Vi har analysert data frå dei nasjonale prøvane både ved hjelp av klassiske analysar (Excel og SPSS) og IRT-analysar (Xcalibre).

Dei aller fleste oppgaveformata har item som blir skåra dikotomt (0/1). To av formata inneheld item som blir skåra polytomt (eleven svarar på fleire deloppgåver og kan dermed få fleire rette svar): *Gap filling* og *Write the word*. Kvar av *Write the word*-oppgåvene til dømes inneheld tre luker der eleven skal skrive inn eitt ord i kvar luke. På desse oppgåvene kan elevane få 0, 1, 2 og 3 rette svar.

Tabell 5 viser statistikk for nasjonale prøvar i engelsk 5. trinn. Tabellen viser data både for kvar av dei tre versjonane av hovudprøven og for alle elevane på trinnet (samla resultat for dei tre hovudprøvane).

Tabell 5: Oppsummerande statistikk NP-engelsk 5. trinn 2018 – tre versjonar og samla resultat*

Hovud-prøve	Tal på item	Opne item	Lukka item	Reliabilitet	Gj.sn p-verdi	Std-avvik	Diskrimi-nering	Tal på elevar
NP5-1	39(46)	17 (19)	22 (27)	0,921	0,572	10,588	0,468	18288
NP5-2	39(46)	17 (19)	22 (27)	0,922	0,576	10,652	0,471	18530
NP5-3	39(46)	17 (19)	22 (27)	0,922	0,570	10,666	0,472	18645
NP5-alle	39(46)	17 (19)	22 (27)	0,922	0,573	10,635	0,470	56463

*b-verdien er frå analyse av data frå alle kalibrerte item (med ankring).

NP5-1 = nasjonal prøve i engelsk for 5. trinn, versjon 1

NP5-2 = nasjonal prøve i engelsk for 5. trinn, versjon 2

NP5-3 = nasjonal prøve i engelsk for 5. trinn, versjon 3

Hovudprøven på 5. trinn er reliabel (0,922) og diskriminerer på ein fornuftig måte mellom elevane (0,470). Prøven har ein gjennomsnittleg p -verdi på 0,57, som er noko høgare enn det vi siktar mot. Vi ser at dei tre hovudprøvene (versjon 1, 2 og 3) oppfører seg så og seie identisk. Dette er forventa. Den gjennomsnittlege p -verdien etter piloteringa var 0,468 (sjå tabell 2), og det betyr at prøven har blitt vel 10,5 prosentpoeng lettare enn piloteringsdata indikerte.

2.3 Overordna IRT-analysar – theta, b -parameter og a -parameter

Til grunn for *Item respons teori* (IRT-teori) ligg det synet at ein underliggjande ferdigheit (evne, kompetanse), kan representerast som ein uavhengig kontinuerleg variabel. Det betyr at variabelen kan ha ein skala frå $-\infty$ til $+\infty$ (fra minus uendelig til pluss uendelig). Denne variabelen, for eksempel leseferdigheiter i engelsk, kan ikkje observerast direkte, difor kallar vi det ein underliggjande variabel eller eit underliggjande trekk. Her refererer vi til variabelen som ein ferdighet, og ein ser for seg at ferdighetene ein person har, kan målast og plasserast på den latente variabelen. Vi kan også seie at den latente variabelen korresponderer til ei line, eller ein skala, og at leseferdigheitene i engelsk som individuelle personar har, kan plasserast på denne skalaen. Å måle leseferdigheiter i engelsk blir då å finne ut, så nøyaktig som råd, kor på skalaen personar skal plasserast.

I IRT-modellar er den greske bokstaven *theta* (θ) eit uttrykk for ferdigheit/evne/kompetanse (ability). Minusverdiar viser til lågare ferdigheit, mens plussverdiar viser til høgare. Gjennomsnittet i ein populasjon vil ligge rundt 0. Di lågare evne/ferdigheit ein person har, jo lågare verdi vil θ ha. Og motsett, di høgare evne/ferdigheit ein person har, di høgare verdi vil θ -verdien ha.

Personar sine ferdigheiter (θ) og vanskegrad på item kan plasserast på same skala. Når vi refererer til vanskegrad på item, refererer vi til b -verdiar. På same måte som for θ , får lette item minusverdiar og vanskelege item plussverdiar. Eit anna mål vi forholdt oss til, er a -verdiar som refererer til kor godt item skil mellom personar med høge/låge ferdigheiter.

Når det er presisert under at tala viser til ankra analysar, så betyr det at analysane for 2018 kan relaterast til analysar for perioden 2014 – 2017 i og med at eit sett ankeritem går igjen frå år til år. a - og b -verdiar for ankeritema frå tidlegare år blir brukt i analysane.

Estimata i analysane under viser til item som vart brukt både i hovudprøven og ankerprøven på 5. trinn.

2.3.1 Theta

Tabell 7 og Figur 4 oppsummerer θ -mål i 5. trinns populasjonen. Alle Xcalibre-figurar og -tabellar er frå analysar med ankring, der ankeritema er inkludert. Vi ser i Tabell 6 at distribusjonen av θ (ability) i 5. trinnsgruppa varierer frå -2,884 og +2,498.

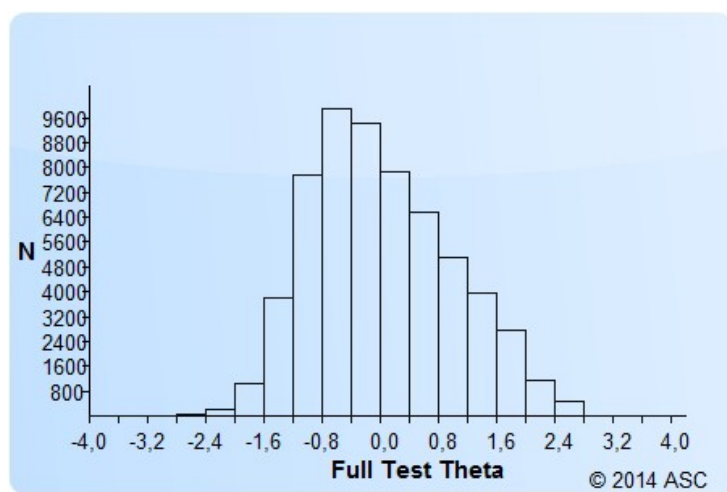
Tabell 6: Oppsummerende statistikk for theta (Xcalibre med ankring)

Elevnr	Gjennomsnitt	Standardavvik	Minimum	Median	Max
60003	0,026	0,968	-2,884	-0,097	2,498

Variasjonen er framstilt grafisk i figur 4 og meir detaljert i Tabell 7.

Vi ser at distribusjonen av theta i 5.trinnspopulasjonen er tilnærma normalfordelt. Data frå prøven viser altså at det finst elevnr som har svært låge og låge ferdigheiter i engelsk, og samtidig finst det dei med høge og svært høge ferdigheiter i engelsk. Dei aller fleste elevane har ferdigheiter ein eller annan stad mellom desse ytterpunkta.

Figur 4: Distribusjon av theta i populasjonen på 5. trinn



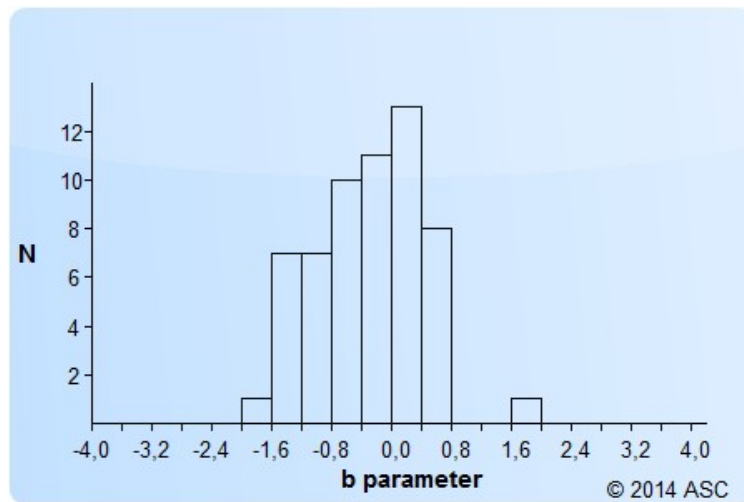
Tabell 7: Fordeling av theta

Frå	Til	Tal på elevnr
Under -3,2		0
-3,2	-2,8	16
-2,8	-2,4	42
-2,4	-2,0	201
-2,0	-1,6	1045
-1,6	-1,2	3787
-1,2	-0,8	7730
-0,8	-0,4	9892
-0,4	0	9429
0	0,4	7850
0,4	0,8	6536
0,8	1,2	5115
1,2	1,6	3971
1,6	2,0	2765
2,0	2,4	1163
2,4	2,8	461
Over 2,8		0

2.3.2 *b*-parameter

Om vi i neste omgang ser på distribusjon av *b*-parameter (vanskegrad på itema), så ser vi at denne også er nokolunde normalfordelt. Figur 5 og Tabell 8 viser korleis *b*-parametra for alle dikotome item som vart administrert i 5. trinnsprøvene i 2018, fordeler seg på thetaskalaen.

Figur 5: Fordeling av item etter b -parameter i prøven (dikotome item)



Tabell 8: Fordeling av item etter b -parameter (dikotome item)

Frå	Til	Tal på item
Under -2,0		0
-2,0	-1,6	1
-1,6	-1,2	7
-1,2	-0,8	7
-0,8	-0,4	10
-0,4	0	11
0	0,4	13
0,4	0,8	8
0,8	1,2	0
1,2	1,6	0
1,6	2,0	1
Over 2,0		0

Vanskegraden for dei dikotome itema i hovudprøven på 5. trinn varierer frå -1,66 (lettaste item) til +1,62 (vanskelegaste item). Vi ser også at gjennomsnittleg er det fleire relativt lettare item (med minus verdier) enn vanskelege (med +-verdier) av dei som gjekk i prøven på 5. trinn.

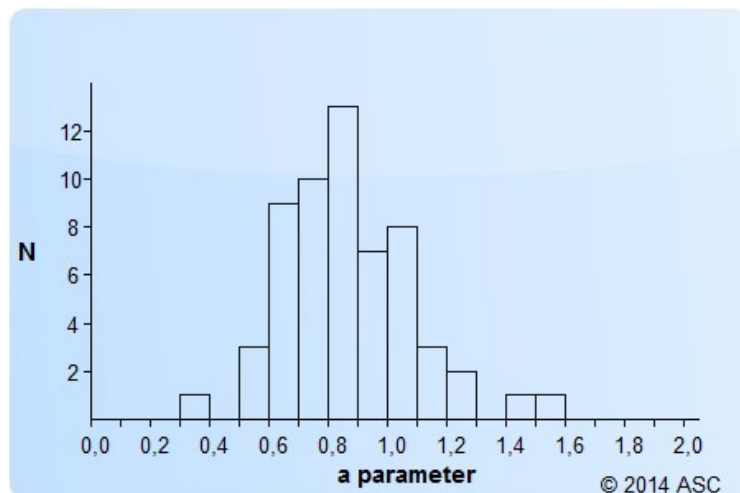
I hovudprøve og ankerprøve for 5. trinn finst det til saman 6 polytome oppgåver, oppgåver der elevane svarar på fleire item og får ein samleskår for kvar av desse oppgåvene. Det er item i dei polytome oppgåvene som både er lettare (-1,87) og vanskelegare (+2,42) enn nokon av dei dikotome oppgåvene.

2.3.3 α -parameter

α -parametra viser som sagt til kor godt item diskriminerer. Xcalibre-skalaen for α -parameter går frå 0 og oppover. Akseptabel diskriminering med utgangspunkt i Xcalibre-analysar, bør vere over 0,5. Målet er sjølvsagt å velje item med høg diskriminering. Dersom vi vel item med lågare diskriminering (dvs. item som skil dårlegare mellom elevar på ulike ferdigheitsnivå), så blir prøven totalt sett mindre reliabel (påliteleg). Set vi grensa for diskriminering høgre, til dømes ved 0,7 eller 0,8, blir prøven meir reliabel. Samstundes vil vi ha færre item å velje mellom når vi set saman den endelege prøven, og det kan gå ut over balansen mellom alle faktorane vi skal ta omsyn til, slik som dei aspekta som prøven skal måle, ulike itemformat og så vidare. Når tabell 5 viser at reliabilitetsestimatet for 5. trinnsprøven er på ca 0,92, noko som vi er godt nøgde med, så støttar dette at ei grensa for α -verdi på 0,5, verkar fornuftig. Figur 6 og tabell 9 viser korleis α -parametra til itema som vart

administrert i prøvane på 5. trinn, er distribuert. Oversikta viser til dikotome item som vart administrert i 5. trinnsprøvene i 2018.

Figur 6: Distribusjon av item etter α -parameter (diskriminering) i prøven på 5. trinn (dikotome item)



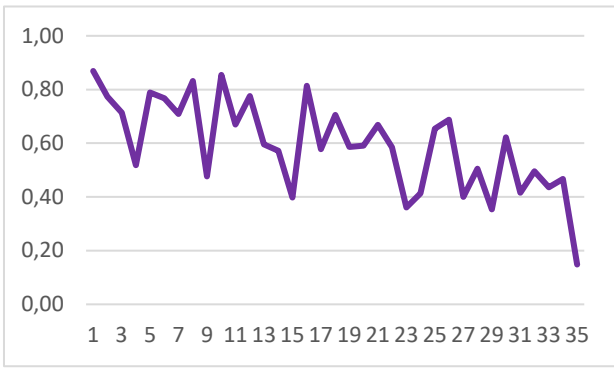
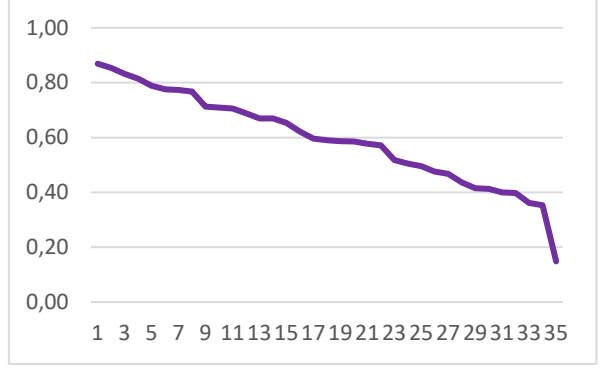
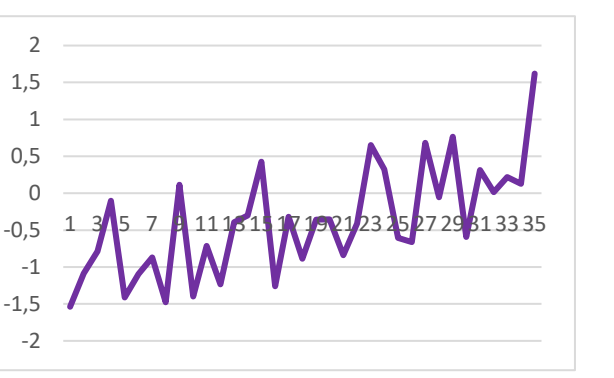
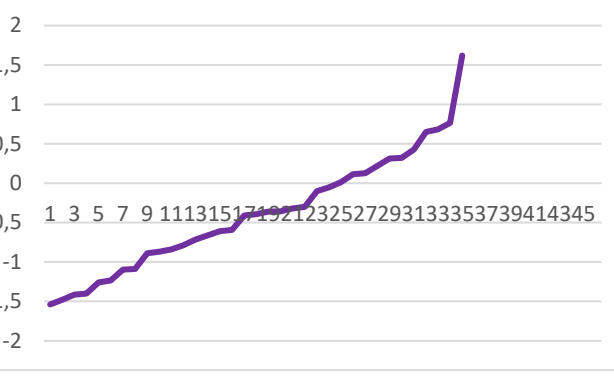
Tabell 9: Fordeling av α -parameter (dikotome item)

Frå	Til	Tal på item
Under 0,5		0
0,3	0,4	1
0,4	0,5	0
0,5	0,6	3
0,6	0,7	9
0,7	0,8	10
0,8	0,9	13
0,9	1,0	7
1,0	1,1	8
1,1	1,2	3
1,2	1,3	2
1,3	1,4	0
1,4	1,5	1
1,5	1,6	1
Over 1,6		0

Av dei dikotome itema som vart brukt i prøvane på 5. trinn i 2018, er det eitt item som har α -verdi under 0,5. Fleire item har ein diskrimineringsindeks på over 1. Den gjennomsnittlege α -verdien på alle dei dikotome itema som var med i prøvane på 5. trinn, er 0,866, mens for dei polytome itema (seks item) var den gjennomsnittlege α -verdien 0,946. Det itemet med høgast diskriminering er eit *Click text*-item med α -verdi på 1,501.

2.4 p -verdi

Figurane 7 - 10 viser til vanskegrad på 5. trinns prøven. Figurane 7 og 8 viser p -verdi i prøvane. Figur 7 viser vanskegraden på itema etter rekkefølga itema opptrer i prøven, og deretter er dei same itema sortert frå det lettaste til det vanskelegaste i figur 8. Figurane 9 og 10 viser vanskegrad i b -verdi, først i rekkefølga itema opptrer i prøven (figur 9), og deretter sortert frå det lettaste til det vanskelegaste itemet (figur 10). Dei polytome itema er ikkje tatt med i figurane under. Til saman er det 4 item som ikkje er med i den grafiske framstillinga i figurane 7 – 10.

Figur 7: Vanskegrad i p-verdi 5. trinn 2018 – rekkefølge som i prøven	Figur 8: Item sortert etter vanskegrad (p-verdi) – 5. trinn 2018
	
Figur 9: Vanskegrad i b-verdi, 5. trinn 2018 – rekkefølge som i prøven	Figur 10: Item sortert etter b-verdi – 5. trinn 2018
	

Dersom vi skal samanlikne dei to måla i figurane 7 og 9 eller 8 og 10, må vi speglvende den eine figuren fordi låge verdiar viser til vanskelege item når vi har med p -verdiar å gjere. I samband med b -verdiar er det omvendt, minusverdiar viser til lette item og +-verdiar til vanskelege.

Vanskegraden i prøven for 5. trinn varierer mellom ein p -verdi på 0,98 (lettaste) og 0,15 (vanskelegaste). Vanskegraden i b -verdi varierer mellom -1,538 (lettaste) og +1,62 (vanskelegaste).

Figurane viser at prøven på 5. trinn inneheld item som passar for elevar på ulike ferdigheitsnivå. Vanskegraden på itema varierer frå svært lette til svært vanskelege. Dette betyr at prøven inneheld oppgåver som burde passe for elevar på ulike ferdigheitsnivå.

2.5 Skilnad i p -verdi pilotering og reell prøve 5. trinn

Det er ei kjent sak at vanskegraden på item fell frå pilotering til reell prøve. Motivasjonen til å gjere sitt beste er større i samband med ein reell prøve enn ved pilotering. Dette gjer at det

ikkje er heilt lett å vere treffsikker på vanskegraden dei reelle prøvane helst bør ha. I piloteringa hausten 2017 var gjennomsnittleg p -verdi på prøven vi sette saman for elevar på 5. trinn 0,468. Tabell 5 i denne rapporten viser at gjennomsnittleg p -verdi for den reelle prøven var 0,573. Den same prøven vart altså 10,5 prosentpoeng lettare frå pilot til prøve.

Då vi starta med dei nasjonale prøvane tidleg på 2000-talet vart det sagt at prøven samla sett burde ha ein p -verdi på cirka 0,5.

I tabell 10 samanliknar vi p -verdi ved pilotering og reelle prøvar på 5. trinn. I tabellen er det brukt forkortingar for format på denne måten:

ci = Click item

cp = Click picture

c&d = Click and drag

ct = Click text

wt = What time?

Gap = Gap filling

MCs = MC short

MCl = MC long

who = Who could say

wtw = Write the word

Tabell 10: Skilnad i p -verdi mellom itema i prøven basert på pilotering (2017) og reell prøve 2018, 5. trinn

Rekke-følgje	Format	item no	p pilot	p prøve	diff	Rekke-følgje	Format	item	p pilot	p prøve	diff
1	ci	5112848	0,801	0,87	-0,07	22	Mcs	5167542	0,48	0,59	-0,11
2	ci	5112169	0,66	0,77	-0,11	23	Mcs	5161541	0,58	0,67	-0,09
3	ci	5145368	0,63	0,71	-0,08	24	Mcs	5166980	0,52	0,59	-0,07
4	ci	5148366	0,42	0,52	-0,10	25	Mcs	5166882	0,25	0,36	-0,11
5	cp	5162160	0,65	0,79	-0,14	26	Mcs	5166886	0,32	0,41	-0,09
6	cp	5162182	0,66	0,77	-0,11	27	Mcl	5167558	0,49	0,65	-0,16
7	cp	5162112	0,58	0,71	-0,13	28	Mcl	5167567	0,53	0,69	-0,16
8	cp	5166926	0,75	0,83	-0,08	29	Mcl	5167570	0,31	0,40	-0,10
9	cp	5159857	0,38	0,48	-0,10	30	Mcl	5167561	0,38	0,50	-0,12
10	c&d	5118186	0,78	0,85	-0,07	31	Mcl	5167564	0,23	0,35	-0,12
11	c&d	5074782	0,57	0,67	-0,10	32	Who	5167800	0,45	0,62	-0,17
12	c&d	5110828	0,63	0,78	-0,15	33	Who	5167783	0,24	0,42	-0,18
13	c&d	5135148	0,47	0,60	-0,13	34	Who	5167777	0,38	0,50	-0,12
14	wt	5167801	0,39	0,57	-0,18	35	Who	5167785	0,26	0,44	-0,18
15	wt	5167997	0,35	0,40	-0,05	36	Who	5167788	0,30	0,47	-0,16
16	ct	5167475	0,66	0,81	-0,15	37	Who	5167768	0,07	0,15	-0,08
17	ct	5118313	0,50	0,58	-0,08	38	wtw*	5167840	0,14	0,28	-0,14
18	ct	5162769	0,62	0,71	-0,09	39	wtw*	5162617	0,08	0,17	-0,09
19	ct	5167724	0,46	0,59	-0,13						
20	Gap*	5167587	0,23	0,29	-0,06						
21	Gap*	5161294	0,16	0,29	-0,13						

*Desse itema er polytome. Eleven avgir meir enn eitt svar (som oftast 3), og p -verdien som er oppgitt, viser til elevar med maks skår. Svært mange av elevane klarar deler av desse oppgåvene. Sjå vedlegg 2 og 3.

Den lilla kolonna i tabellen viser differansen i p -verdi mellom pilotering og nasjonale prøvar. Cellene med den mørkaste lillafargen viser til item der differansen er større enn 10 prosentpoeng. Det er 26 item (av 39) i 5. trinnsprøven der auken i p -verdi er større enn 10 prosentpoeng.

Ein annan faktor er piloteringstidspunkt. Oppgåvene blir pilotert i overgangen mellom august og september. Prøven vart gjennomført i slutten av oktober/begynnelsen av november i 2018. Det betyr at elevane har nesten to månader å lære engelsk på og dermed førebu seg til prøven på etter piloteringa på 5. trinn. Dette kan sjølvsagt ha noko å seie for resultatet. I 2019 vil dei nasjonale prøvane bli avvikla i september (2. – 27. september). Piloteringa skjer 19. – 30. august, altså rett før. Vi vil følgje med på kva dette vil ha å seie for kor vanskeleg prøven blir for elevane.

2.6 Polytome item

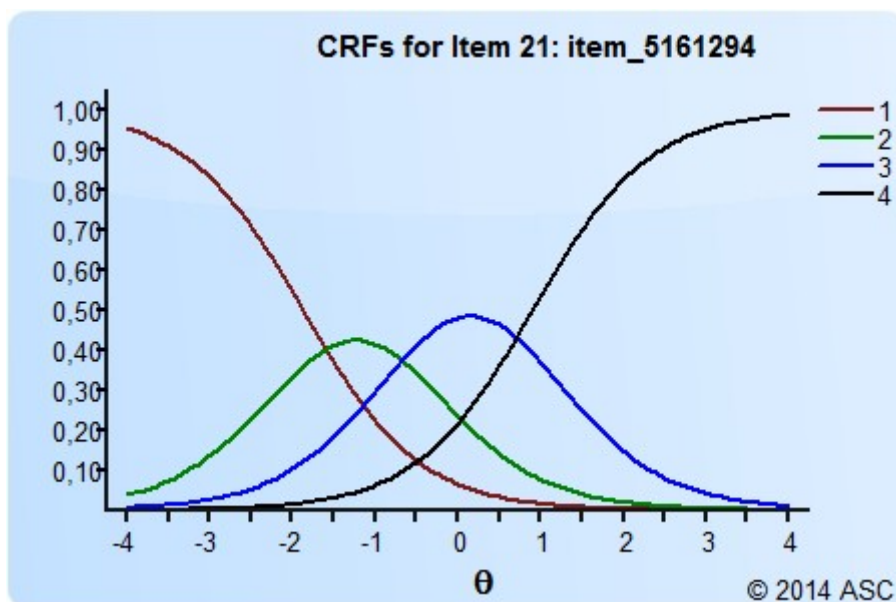
Dei første åra vart alle item i dei Nasjonale prøvane i engelsk skåra binært (0, 1). Dette betydde at når vi til dømes laga *Gap filling*-oppgåver, så piloterte vi kanskje ei oppgåve med fem luker. Etter pilotering kunne kanskje analysane vise at to eller tre av desse ikkje fungerte, enten fordi dei var alt for vanskelege/lette, eller fordi dei hadde svært låg diskriminering. Dersom vi likevel ville bruke itema som fungerte bra i ein prøve, ville vi ta ut dei itema som ikkje fungerte, før oppgåva vart presentert i ein prøve. Dette kunne føre til at det vart ei ujamn fordeling av luker reint grafisk i den «ferdige» oppgåva.

Dei siste åra er det innført polytom skåring på to itemformat: *Gap filling* og *Write the word* på femte trinn. Dette ser ut til å fungere bra. Under gir vi nokre eksempel på oppgåver frå 2018-prøvane med polytom skåring.

2.6.1 Gap-filling

Oppgåva som blir vist grafisk i figur 11, er ei *Gap filling*-oppgåve med tre luker. Det er mogeleg å få 0, 1, 2 og 3 rette svar på denne oppgåva.

Figur 11: Oppgave 21 i hovudprøven på 5. trinn



Figur 10 viser at oppgåva skil godt mellom elevar på ulike ferdigheitsnivå. Elevar som ikkje hadde nokon rette, hadde gjennomsnittleg theta på -0,965. Dei som hadde eitt rett svar, hadde ein theta på -0,643, medan dei som hadde to rette svar, hadde ein theta på 0,130. I gjennomsnitt hadde elevane med tre rette svar på denne oppgåva ein theta på 0,290. Tabell 11 viser samanhengen mellom rette svar, prosentvis fordeling av elevar og thetamål.

Tabell 11: Oppgave 21 i 5. trinnsprøven – rette svar, fordeling i prosent av elevar og thetamål

Rette svar	Elevar i prosent	Gjennomsnittleg theta
0	11,4	-0,965
1	23,4	-0,598
2	36,1	0,154
3	29,1	0,756

Statistiske analysar viser at begge *Gap filling*-oppgåvene i 5. trinnsprøven skil klart mellom elevar på ulike ferdigheitsnivå.

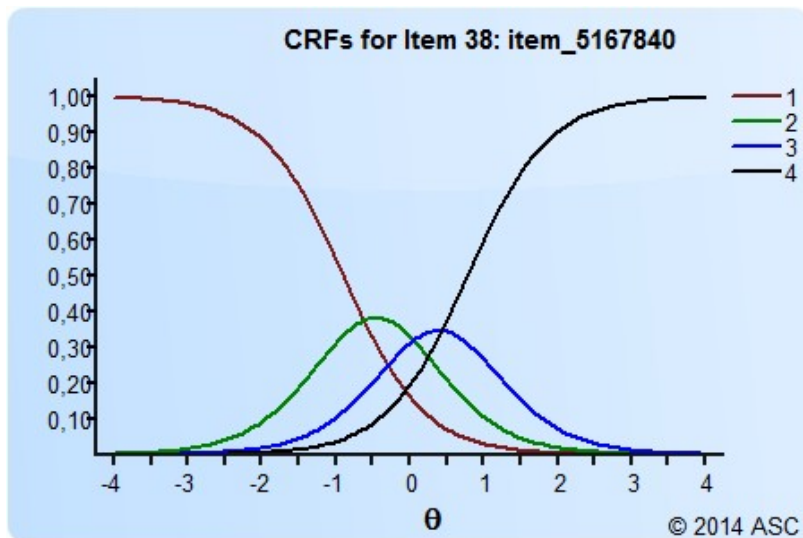
Både på 5. og 8. trinn fungerer dette oppgåveformatet godt.

2.6.2 Write the word

I formatet *Write the word* møter eleven ein tekst med tre luker. Oppgåva er å skrive inn svar sjølv i lukene. Elevane på femte trinn får to slike oppgåver. I den første oppgåva (oppgåvenummer 38) skal dei hente dei orda som passar i lukene frå ein tabell. Den andre oppgåva har ingen tabell (oppgåvenummer 39), så der må elevane finne det rette svaret sjølve.

Det kjem klart fram av analysane at dette oppgåveformatet skil godt mellom elevar på ulike ferdigheitsnivå. Figur 12 og 13 viser oppgåve nummer 38 og 39 i prøven.

Figur 12: Oppgåva *Loud music* – med tabell (oppgåve 38)



Elevar som ikkje svarte rett på noko, hadde ein gjennomsnittleg theta på -0,847, medan dei som hadde alt rett, hadde gjennomsnittleg theta på (+)1,429. Tabell 12 viser samanhengen mellom rette svar, prosentvis fordeling av elevar og thetamål.

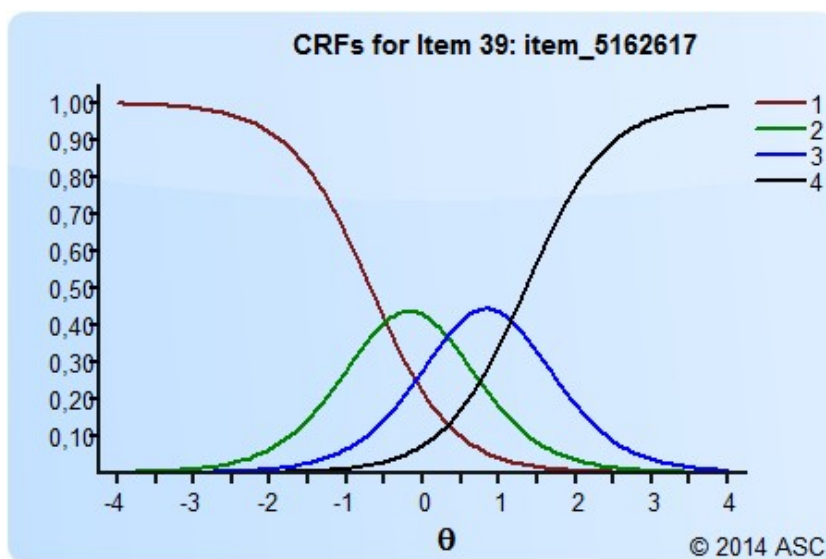
Tabell 12: Oppgåve 38 i 5. trinnsprøven – rette svar, fordeling i prosent av elevar og thetamål

Rette svar	Elevar i prosent	Gjennomsnittleg theta
0	27,2	-0,781
1	24,5	-0,356
2	19,8	0,170
3	28,4	1,008

Den prosentvise fordelinga av elevar i oversikta over viser at om lag ein fjerdedel av elevane klarar å få eitt rett svar og at vel ein fjerdedel klarar tre rette svar. Tabell 12 viser også at det er dei svakaste elevane som ikkje får nokon rette svar, den gjennomsnittlege thetaverdien i denne gruppa er -0,781. På same måte viser tabellen at den gjennomsnittlege thetaverdien i gruppa som svarar rett på alle tre lukene, er mykje høgare 1,008.

I figur 13 ser vi korleis oppgåve 39 i prøven for 5. trinn fungerte. Når elevane svarte på denne oppgåva, måtte dei sjølv finne orda dei skulle skrive inn i oppgåvelukene. Vi ser at det er litt færre elevar enn i oppgåve 38 som svarar rett på alle lukene. Dette er som forventa.

Figur 13: Oppgåve *Embarrassing hat story* – utan tabell (oppgåve 39)



Tabell 13: Oppgåve 39 i 5. trinnspørven – rette svar, fordeling i prosent av elevar og thetamål

Rette svar	Elevar i prosent	Gjennomsnittleg theta
0	32,5	-0,742
1	28,4	-0,169
2	22	0,439
3	17	1,283

Vi ser den same tendensen i data frå oppgåve 39 som vi såg i oppgåve 38. Gruppa som ikkje svarar rett på nokon av lukene, har ein gjennomsnittleg theta på -0,742, mens gruppa som svarar rett på alle itema, har ein gjennomsnittleg theta på 1,283.

Oppgåveformatet fungerer godt både på 5. og 8. trinn.

2.7 Kjønnskilnader 5. trinn

I statistisk samanheng betyr ordet signifikant «ikkje tilfeldig». Ein t-test vil avsløre om skilnadene mellom gutar og jenter er signifikante. At ein skilnad er signifikant, vil seie at det er liten sjanse for at resultatet er tilfeldig. Dette blir uttrykt som ein p -verdi som kan tolkast som kor sannsynleg det er at resultatet kan oppstå ved eit reint tilfelle. Ein liten verdi er difor eit uttrykk for at skilnadene er store (t.d. $>0,01$), dette kan også tolkast som ein prosent der ein multipliserer p -verdien med 100. Tradisjonelt brukar ein grenser på 5% (0,05) og 1% (0,01) som viktige markeringar.

Når vi har analysert data frå nasjonale prøvar i engelsk tidlegare år, har det i det store og heile ikkje vore særlege skilnader mellom gutane og jentene sine resultat. Ofte har jentene gjort det litt betre enn gutane, men skilnadene har vore små. Når vi baserer analysane på store utval (til dømes fleire tusen elevar), er det lett at trivielle skilnader blir signifikante. Desse analysane baserer seg på data frå om lag 60000 elevar for kvar av prøvane på 5. og 8. trinn. Her vil vi meir eller mindre forvente at trivielle skilnader blir signifikante om vi samanliknar korleis gutar og jenter presterer. Pilotdata indikerte skilnader mellom jenter og gutar (sjå figur 2 og tabell 4).

Gjennomsnittleg theta for 2018-avviklinga viser at skilnaden mellom gutar og jenter på 5. trinn er ikkje så stor, men likevel signifikant. Data frå dei fem siste åra viser at gutane gjer det betre enn jentene på den nasjonale prøven i engelsk.

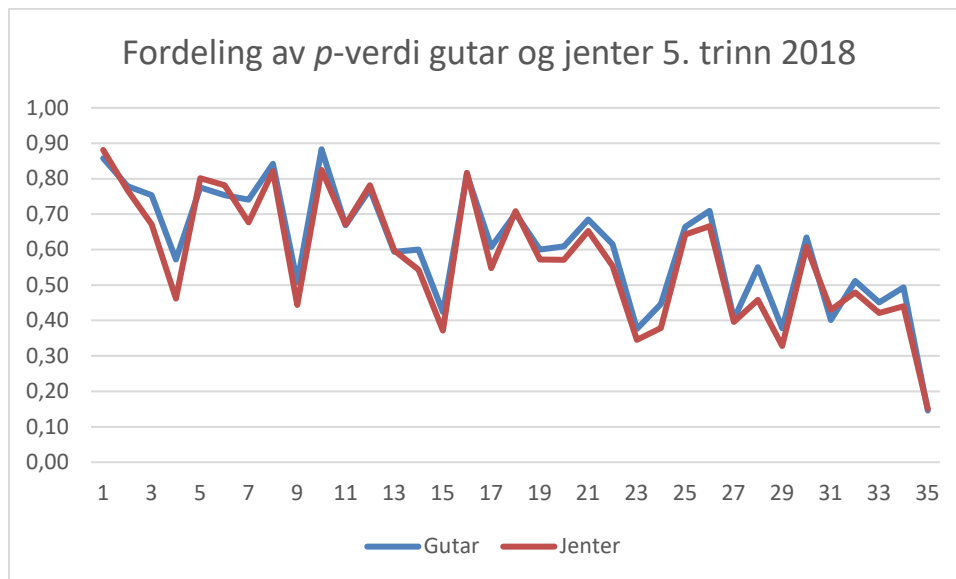
Tabell 14 viser skilnadene mellom gutar og jenter basert på ein t-test. Vi ser at mens gjennomsnittleg theta for gutane er (+)0,0985, er theta for jentene -0,0489. Dette indikerer altså at gutane har betre leseferdigheiter på engelsk enn jentene, skilnaden i theta er på 0.1474 i gutane sin favør. Denne skilnaden er større enn forventet. Standardavviket er nesten likt i dei to gruppene, noko som seier at variasjonen i jentegruppa og gutegruppa er omtrent lik.

Tabell 14: Skilnader mellom jenter og gutar på 5. trinn – t-test

Prøve	Kjønn	Gjennomsnittleg theta	<i>n</i>	Standardavvik	<i>t</i>	<i>Sig</i>
NP-5	Gutar	0,0985	30303	0,9809	18,718	0,000
	Jenter	-0,0489	29700	0,9484		

Data viser at skilnaden mellom jenter og gutar er signifikant på 0,01 nivå. Det betyr altså at skilnaden ikkje er tilfeldig. Som nemnt tidlegare vil ein meir eller mindre forvente signifikante skilnader når ein baserer seg på eit så stort datamateriale som vi gjer i samband med dei nasjonale prøvane. Det kan likevel vere viktig å følgje med på utviklinga dei neste åra.

Figur 14: *p*-verdiar gitar og jenter 5. trinn 2018



Den raude lina viser *p*-verdiar for jenter og den blå for gitar. Vi ser at prøven i all hovudsak oppfører seg likt i dei to gruppene. Sjølv om skilnaden mellom gitar og jenter på 5. trinn er større enn før, så er ikkje skilnaden stor på dei fleste enkeltitem. Gjennomsnittleg råskår for gitar er 27,17 (av max 46) og for jenter 25,48. Dette vil seie at gutane svarar rett på vel eitt og eit halvt fleire item enn jentene i gjennomsnitt. I vedlegg 3 blir *p*-verdiane på enkeltitem oppgitt, både totalt og for jenter og gitar separat. Gutane gjer det best på 29 av 39 item. Jentene skårar i gjennomsnitt høgare enn gutane på 6 av 39 item. I dei fleste tilfelle er skilnaden svært liten, men i nokre tilfelle er skilnaden ganske stor. Dette kan vi også sjå i figur 13.

Resultata på 5. trinn viser ein signifikant skilnad mellom jenter og gitar, der gutane skårar i gjennomsnitt høgare enn jentene

Ein rask kikk i arkivet av rapportar frå dei siste åra med fokus på jenter og gitar, er oppsummerte i tabell 15.

Tabell 15: Kjønnsskilnader 2010 – 2018.

	5. trinn	
	Kven skårar høgst?	Signifikant?
2010	Jenter	Ja
2011	*	
2012	Jenter	Ja
2013	Jenter	Nei
2014	Gutar	Ja
2015	Gutar	Ja
2016	Gutar	Ja
2017	Gutar	Ja
2018	Gutar	Ja

* Nasjonale prøvar vart avlyst på 5. trinn i 2011 på grunn av tekniske problem.

Oversikta i tabell 15 viser gutane har gjort det best på dei nasjonale prøvane i engelsk for 5. trinn sidan 2014. Skilnaden mellom kjønna har vore størst i perioden 2016 - 2018.

Dette betyr at det kan vere ei endring på gang når det gjeld kjønn og engelskferdigheiter på 5. trinn. Spørsmålet er kvifor ei slik endring oppstår. Er det oppgåvene vi lagar som favoriserer gutar? Er gutar meir på datamaskin og les og snakkar meir engelsk i samband med slike aktivitetar? Dette bør ein prøve å sjå nærare på dei neste åra.

TIL SLUTT

Den nasjonale prøven i engelsk på 5. trinn fungerer i all hovudsak som forventa. Prøven er reliabel, inneheld item med varierende vanskegrad, og diskriminerer på ein fornuftig måte mellom elevane. Prøven er litt lettare enn det vi siktar mot.

Prøvane skal måle nokre av kompetansemåla i læreplanen, dette gjeld kompetansemål som omhandlar lesing på engelsk, vokabular og grammatikk. Vi prøver å inkludere eit variert sett med lesestekstar. Tekstane representerer ulike sjangrar, omhandlar ulike tema og har ulik vanskegrad. Vi prøver også å finne tekstar som kan vere interessante for jenter og gutar på 5. og 8. trinn.

Dei siste åra har vi sett at gutane på 5. trinn gjer det betre enn jentene. Det neste året bør vi sjå nærare på mogelege årsaker til dette. Inneheld prøven tema og oppgåver som appellerer meir til gutar enn til jenter? Brukar gutar engelsk meir enn jenter utanfor skulen? Andre årsaker?

Sidan prøvane er digitale og blir skåra automatisk, kan vi ikkje måle produktive ferdigheiter (munnlege og skriftlege ferdigheiter), i alle fall ikkje ennå. Prøven måler heller ikkje lytting på engelsk. Lytting er ein ferdigheit det kunne vere viktig å inkludere i nasjonale prøvar i engelsk dersom og når det kan forsvarest teknisk. Lytteitem kan også, på same måte som leseitem, skårast automatisk.

Til slutt vil vi nemne at det er skrive lærarrettleiingar som fylgjer prøvane kvart år. Lærarrettleiingane informerer om kva prøvane måler, viser døme på item som representerer ulike meistringsnivå og kjem med idear om korleis lærarar kan arbeide med elevar på ulike meistringsnivå.

Vedlegg 1: Piloteringsdesign hausten 2017

Pakke	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
5-1	x	x														
5-2			x	x												
5-3					x	x										
5-4							x	x								
5-5	x							x								
5-6		x	x													
5-7				x	x											
5-8						x	x									
5-9	x		x													
5-10		x		x												
5-11					x		x									
5-12						x		x								
5-13																
5-14	x						x									
5-15		x						x								
5-16			x		x											
5-17				x		x										
58-1	x			x					x	x						
58-2		x					x								x	x
58-3			x			x					x	x				
58-4					x			x					x	x		
58-5	x					x								x	x	
58-6		x			x					x	x					
58-7			x					x	x							x
58-8				x			x					x	x			
8-1									x	x						
8-2											x	x				
8-3													x	x		
8-4															x	x
8-5									x							x
8-6										x	x					
8-7												x	x			
8-8														x	x	
8-9									x		x					
8-10										x		x				
8-11													x		x	
8-12														x		x
8-13									x						x	
8-14										x						x
8-15											x		x			
8-16												x		x		

Forklaring til designet:

Oversikta viser at det vart pilotert 8 pilotprøvar for 5. trinn (1-8) og åtte for 8. trinn (9-16). Vi piloterer vanlegvis åtte pilotprøvar både for 5. og 8. trinn i eit delvis lenka design (incomplete linked design). Dette vart også gjort hausten 2017. Slik designet er konstruert, er alle item og itempakkar knytt til alle andre enten direkte eller indirekte. Dette gjer at vi kan analysere data frå alle pilotprøvene saman ved hjelp av IRT-analyse.

Ei itempakke omfattar mellom 6 og 10 – 12 item, vanlegvis cirka 9 item. Nokre itempakkar er mynta spesifikt på elevar på 5. trinn (5-1, 5-2, 5-3 osv.) og andre på elevar på 8. trinn (8-1, 8-2 osv.). Nokre få pakker blir prøvd ut både på 5. og 8. trinn (58-1, 58-2 osv.). Vi ser til dømes at itempakke 5-1 finst i pilotprøvene 1 og 2, mens itempakke 5-5 er i pilotprøvene 1 og 8.

Dersom vi ser på pilotprøve 1, er denne knytt til alle dei andre pilotprøvene på 5. trinn via ulike itempakkar. I tillegg er pilotprøve 1 knytt til pilotprøvene 9, 10, 14 og 15 på 8. trinn via itempakkane 58-1 og 58-5. Slik designet er konstruert, er alle item og itempakkar knytt til alle andre enten direkte eller indirekte. Dette gjer at vi kan analysere data frå alle pilotprøvene saman ved hjelp av IRT-analyse.

Vedlegg 2: Forslag til Nasjonal prøve i engelsk 5. trinn 2018 – oversikt og statistiske mål etter utprøving

Rekkefølge*	Format	Item	Namn	Teksttype	p-value	Maks skåre	α (OPLM)**	b
1	Click item	5112848	Airport line 2	Beskrivelse	0,801	1	5	-0,631
2	Click item	5112169	Nature museum 4-lynx	Beskrivelse	0,661	1	5	-0,444
3	Click item	5145368	Museum-Simon	Beskrivelse	0,633	1	4	-0,417
4	Click item	5148366	Boy'sRoom2Aeroplane	Beskrivelse	0,423	1	6	-0,165
5	Click picture	5162160	Flags 2-New Z	Beskrivelse	0,655	1	3	-0,485
6	Click picture	5162182	Dinosaurs 4	Beskrivelse	0,657	1	4	-0,450
7	Click picture	5162112	Sam's dinner spaghetti	Beskrivelse	0,578	1	3	-0,360
8	Click picture	5166926	Atlas holding up the world	Beskrivelse	0,753	1	4	-0,589
9	Click picture	5159857	Campground1	Beskrivelse	0,384	1	5	-0,113
10	Click and drag	5118186	LivingroomChair2	Instruksjon	0,782	1	6	-0,566
11	Click and drag	5074782	Bathroom-towel rack	Instruksjon	0,572	1	4	-0,316
12	Click and drag	5110828	Family dining out 3	Beskrivelse	0,634	1	4	-0,398
13	Click and drag	5135148	Market 1-seller	Beskrivelse	0,469	1	4	-0,204
14	What time is it	5167801	Laura 1-03:00	Forteljing	0,393	1	5	-0,099
15	What time is it	5167997	Holiday1-04:15	Forteljing	0,351	1	5	-0,043
16	Click text	5167475	Hats2	Beskrivelse	0,657	1	5	-0,439
17	Click text	5118313	WeatherMeteor	Faktatekst	0,501	1	4	-0,218
18	Click text	5162769	Strange facts 4	Faktatekst	0,623	1	4	-0,384

Vedlegg 2

19	Click text	5167724	Favourite books1	Forteljing	0,464	1	4	-0,173
20	Gap*	5167587	Soup	Beskrivelse	0 = 0,40 1 = 0,37 2 = 0,23	2	4	-0,135
21	Gap*	5161294	Letter from San Frans	Korrespondanse	0 = 0,26 1 = 0,26 2 = 0,33 3 = 0,16	3	3	-0,200
22	Mc-short	5167542	Drenched	Forteljing	0,479	1	4	-0,197
23	Mc-short	5161541	Ariel diary	Forteljing	0,583	1	4	-0,322
24	Mc-short	5166980	Fishing with Grandad1	Forteljing	0,521	1	4	-0,256
25	Mc-short	5166882	Samuel shoes story 1	Forteljing	0,247	1	3	0,234
26	Mc-short	5166886	Mole book	Anmeldelse	0,322	1	5	0,026
27	Mc-story	5167558	Why bears have short tails 2	Forteljing	0,493	1	4	-0,217
28	Mc-story	5167567	Why bears have short tails 5	Forteljing	0,533	1	6	-0,273
29	Mc-story	5167570	Why bears have short tails 6	Forteljing	0,305	1	2	0,244
30	Mc-story	5167561	Why bears have short tails 3	Forteljing	0,382	1	5	-0,084
31	Mc-story	5167564	Why bears have short tails 4	Forteljing	0,232	1	3	0,271
32	Who	5167800	BenReubHoliday15-Benjamin	Forteljing	0,452	1	4	-0,147
33	Who	5167783	BenReubHoliday8-mum/Paula	Forteljing	0,238	1	4	0,194
34	Who	5167777	BenReubHoliday5-Benjamin	Forteljing	0,378	1	4	-0,051
35	Who	5167785	BenReubHoliday9-Reuben	Forteljing	0,264	1	4	0,141
36	Who	5167788	BenReubHoliday10-Benjamin	Forteljing	0,308	1	4	0,069
37	Who	5167768	BenReubHoliday1-Mrs Banks	Forteljing	0,067	1	4	0,707

38	Write the word*	5167840	Loud music	Forteljning	0 = 0,49 1 = 0,22 2 = 0,14 3 = 0,14	3	3	0,015
39	Write the word*	5162617	Embarrassing hat-story	Forteljning	0 = 0,54 1 = 0,23 2 = 0,15 3 = 0,08	3	3	0,150

*Oppgåva er polytom. p -verdien viser til proporsjonen av elevar som har 0, 1, 2, eller 3 rette svar.

**Gjennomsnittleg a -verdi for alle piloterte item i OPLM-analysen av allepiloterte item, vart sett til 3. Det er difor lagt vekt på å velje item med høgre diskriminering i forslaget til prøve – der dette er mogeleg.

Vedlegg 3: Hovudprøve 5. trinn 2018 – item for item-oversikt – reell prøve

Rekke- følge	Format	Item	Namn	p	p gutar	p jenter	diff j-g	a	b
1	Click item	5112848	Airport line 2	0,87	0,86	0,88	-0,02	1,042	-1,538
2	Click item	5112169	Nature museum 4-lynx	0,77	0,78	0,77	0,01	0,959	-1,086
3	Click item	5145368	Museum-Simon	0,71	0,75	0,67	0,08	1,057	-0,789
4	Click item	5148366	Boy'sRoom2Aeroplane	0,52	0,57	0,46	0,11	1,408	-0,101
5	Click picture	5162160	Flags 2-New Z	0,79	0,78	0,8	-0,03	0,671	-1,412
6	Click picture	5162182	Dinosaurs 4	0,77	0,75	0,78	-0,03	0,882	-1,096
7	Click picture	5162112	Sam's dinner spaghetti	0,71	0,74	0,68	0,06	0,820	-0,868
8	Click picture	5166926	Atlas holding up the world	0,83	0,84	0,82	0,02	0,849	-1,476
9	Click picture	5159857	Campground1	0,48	0,51	0,44	0,06	0,573	0,114
10	Click and drag	5118186	LivingroomChair2	0,85	0,88	0,82	0,06	1,122	-1,400
11	Click and drag	5074782	Bathroom-towel rack	0,67	0,67	0,67	0	0,782	-0,714
12	Click and drag	5110828	Family dining out 3	0,78	0,77	0,78	-0,01	0,761	-1,235
13	Click and drag	5135148	Market 1-seller	0,6	0,6	0,6	0	0,841	-0,393
14	What time?	5167801	Laura 1-03:00	0,57	0,6	0,54	0,06	0,844	-0,300
15	What time?	5167997	Holiday1-04:15	0,4	0,42	0,37	0,05	0,734	0,426
16	Click text	5167475	Hats2	0,81	0,81	0,81	0	1,001	-1,260
17	Click text	5118313	WeatherMeteor	0,58	0,61	0,55	0,06	0,771	-0,320
18	Click text	5162769	Strange facts 4	0,71	0,7	0,71	-0,01	0,769	-0,888

19	Click text	5167724	Favourite books1	0,59	0,6	0,57	0,03	0,823	-0,359
20	Gap filling*	5167587	Soup	0 = 0,31 1 = 0,40 2 = 0,28	0,33	0,25	0,08	0,944	-0,73, 0,83
21	Gap filling*	5161294	Letter from San Frans	0 = 0,11 1 = 0,23 2 = 0,36 3 = 0,29	0,32	0,27	0,05	0,834	-1,87, -0,59, 0,91
22	MC short	5167542	Drenched	0,59	0,61	0,57	0,04	0,900	-0,356
23	MC short	5161541	Ariel diary	0,67	0,69	0,65	0,03	0,587	-0,839
24	MC short	5166980	Fishing with Grandad1	0,59	0,62	0,55	0,06	0,616	-0,409
25	MC short	5166882	Samuel shoes story 1	0,36	0,38	0,35	0,03	0,647	0,650
26	MC short	5166886	Mole book	0,41	0,45	0,38	0,07	0,855	0,322
27	MC long	5167558	Why bears have short tails 2	0,65	0,66	0,64	0,02	0,903	-0,607
28	MC long	5167567	Why bears have short tails 5	0,69	0,71	0,67	0,04	1,225	-0,663
29	MC long	5167570	Why bears have short tails 6	0,4	0,4	0,4	0,01	0,389	0,681
30	MC long	5167561	Why bears have short tails 3	0,5	0,55	0,46	0,09	1,225	-0,054
31	MC long	5167564	Why bears have short tails	0,35	0,38	0,33	0,05	0,556	0,764
32	Who	5167800	BenReubHoliday15-Benj	0,62	0,63	0,61	0,03	0,605	-0,592
33	Who	5167783	BenReubHoliday8-mum/P	0,42	0,4	0,43	-0,03	0,852	0,314
34	Who	5167777	BenReubHoliday5-Benjamin	0,5	0,51	0,48	0,03	0,641	0,013
35	Who	5167785	BenReubHoliday9-Reuben	0,44	0,45	0,42	0,03	0,919	0,218
36	Who	5167788	BenReubHoliday10-Benj	0,47	0,49	0,44	0,05	0,718	0,127

37	Who	5167768	BenReubHoliday1-Mrs B	0,15	0,15	0,15	0	0,898	1,620
38	Write the word*	5167840	Loud music	0 = 0,27 1 = 0,25 2 = 0,20 3 = 0,28	0,32	0,27	0,05	1,107	-0,85, 0,01, 0,78
39	Write the word*	5162617	Embarrassing hat-story	0 = 0,33 1 = 0,28 2 = 0,22 3 = 0,17	0,31	0,26	0,05	1,134	-0,63, 0,34, 1,33

*Oppgåva er polytom. *p*-verdien viser til proporsjonen av elevar som har 0, 1, 2, eller 3 rette svar.