



UNIVERSITETET I BERGEN

Nasjonale prøvar i engelsk 2018 8. trinn

Rapport februar 2019

Eli Moe og Hildegunn Lahlum Helness

Institutt for framandspråk

Innhold

OM RAPPORTEN	3
GENERELT OM NASJONAL PRØVE I ENGELSK FOR 8. TRINN	4
Nasjonal prøve i engelsk for 8. trinn og Læreplan i engelsk (ENG1-03)	4
Kva måler prøvane på 8. trinn	4
Om teksttypar og tema i prøvane	5
Kategorisering av item.....	6
Analyse av data.....	6
DEL 1 – PILOTERING AV NP ENGELSK 8. 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 8. trinn 2018	8
1.3 Kva måler prøvane på 8. trinn i 2018	9
1.4 Om kjønnskilnader	10
DEL 2 – NASJONALE PRØVAR I ENGELSK 8. TRINN 2018	11
2. Nasjonal prøve i engelsk 8. trinn haust 2018 - hovudresultat	11
2.1 Gjennomføring 2018	11
2.2 Oppsummerande statistikk 8. 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 8. trinn.....	16
2.6 Polytome item	18
2.6.1 Gap filling.....	19
2.6.2 Write the word	19
2.6.3 Move paragraph	20
2.7 Kjønnskilnader 8. trinn.....	21
Til slutt	25
Vedlegg 1: Piloteringsdesign hausten 2017	26
Vedlegg 2: Forslag til Nasjonal prøve i engelsk 8. trinn 2018 – oversikt og statistiske mål etter utprøving	28
Vedlegg 3: Hovudprøve 8. trinn 2018 – item for item-oversikt	31

OM RAPPORTEN

Denne rapporten beskriv grunnlaget for og utviklinga av Nasjonal prøve i engelsk for 8. 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 8. 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 8. trinn 2018 fungerte for elevane på 8. 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 8. trinn, hovudprøven. I samband med IRT-analysar er datamaterialet for både hovudprøve og ankerprøve lagt til grunn.

GENERELT OM NASJONAL PRØVE I ENGELSK FOR 8. TRINN

Nasjonal prøve i engelsk for 8. trinn og Læreplan i engelsk (ENG1-03)

Dei nasjonale prøvane i engelsk for 8. trinn måler nokre av kompetansemåla i engelsk etter 7. 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 lese- og skrivestrategier
- forstå og bruke et ordforråd knyttet til kjente emner
- forstå hovedinnholdet i (selvvalgte) tekster
- lese og forstå ulike typer tekster av varierende omfang fra forskjellige kilder
- bruke grunnleggende mønstre for rettskriving, ordbøying, setnings- og tekstbygging i produksjon av tekst
- bruke digitale verktøy og andre hjelpemidler for å finne relevant informasjon og lage ulike typer tekster»

Læreplanen uttrykkjer fleire stader at elevane skal arbeide med engelsk med utgangspunkt i «kjente emner». Med kjende emne forstår vi emne som vi kan forvente at elevar på same alder har nokon kjennskap til, til dømes forhold som gjeld heim, fritid, hobbiar, skule, samfunn, dyr og plantar osv. Læreplanen understrekar også nokre tema som relevante (under *Kultur, samfunn og litteratur*) t.d. «personer, stader og begivenheter fra engelskspråklige land», «omgangsformer i ulike kulturer i engelskspråklige land», «engelskspråklige litterære tekster (fra forskjellige kilder), film, nettkultur, bilder og musikk». Dette prøver vi sjølvsagt å ta omsyn til i tekstane som blir brukt 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. 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.

Kva måler prøvane på 8. trinn

Dei nasjonale prøvane i engelsk skal måle generell leseforståing, vokabular og grammatikk.

Elevar på 8. trinn skal kunne

- finne, forstå og bruke informasjon
- forstå ord, omgrep og hovudinnhald i tekstar av varierende omfang frå ulike kjelder
- forstå og bruke ord og uttrykk som dekkjer daglegdagse situasjonar
- forstå tydinga av ord og uttrykk ut frå samanhengen dei er brukt i
- forstå og bruke grunnleggjande mønster for rettskriving, ordbøying, setnings- og tekstoppygging
- reflektere over innhaldet i tekstar

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

Tekstane elevane skal lese på i 8. trinnsprøven, er jamt over noko lengre og meir komplekse enn dei i prøven for 5. trinn. Nokre av oppgåvene i 8. trinnsprøven har også bilde, men færre enn i prøven for 5. trinn. Prøven inneheld tre relativt lange tekstar i formata *Multiple choice long*, *Who could say* og *Move paragraph*.

Tabellane som viser oversikter over prøvane på 8. trinn i Vedlegg 2 og 3, viser at prøvane inneheld fleire faktatekstar og relativt færre beskrivingar og forteljingar enn på 5. trinn. Dette er også naturleg, og speglar nok tekstutvalet dei arbeider med i skulesamanheng. Mange av tekstane har emne frå natur, historie og samfunn. Tekstane har i større grad eit samfunnsperspektiv, sjølv om det også finst tekstar i prøven som omhandlar meir nære og personlege forhold.

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 pilotinga 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. Xcalibre er brukt i analysar for heile datamaterialet for 8. trinn under eitt, det vil seie at data både frå hovudprøve og ankerprøve er inkludert. I tillegg har vi brukt SPSS til ein del klassiske analysar med utgangspunkt i data frå hovudprøven på 8. trinn.

DEL 1 – PILOTERING AV NP ENGELSK 8. 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:

192 item

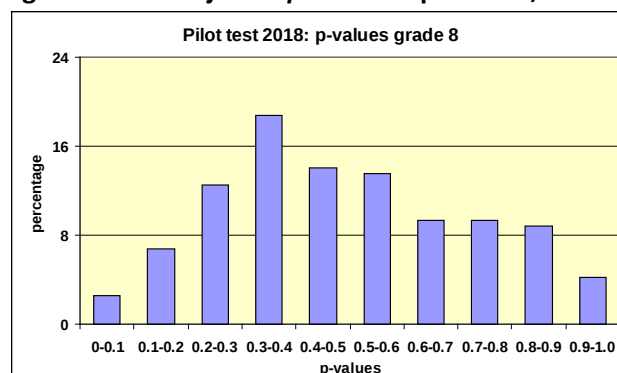
Tal på elevar som deltok:

1902

Arbeid med pilotdata

Pilotdata vart analysert både ved hjelp av OPLM og Xcalibre. Gjennomsnittleg p -verdi for alle itema som vart pilotert på 8. trinn var 0,493. Vanskegraden på itema som vart piloterte varierte frå 0,048 (vanskelegaste item) til 0,968 (lettaste item). Figur 1 viser distribusjon av p -verdiar i pilotdata.

Figur 1: Distribusjon av p -verdiar i pilotdata, 8. 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 8. 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 8. trinn 2018

Tabell 1 gir ei oversikt over prøven vi foreslår for 8. trinn i 2018. Tabellen viser kor mange item som er valt ut innanfor kvart format, og formatrekkefølga er den same som i 2017, versjon 1.

Tabell 1: Item per format i prøven for 8. trinn 2018

Format	Nasjonal prøve 2018 8. trinn
Click picture	3
Click and drag	3
Click text	4
Gap	1 (3)*
Mc short	6
Mc long	5
Who could say	6
Click word	5
Complete sentence (grammar)	6
Write the word (grammar/voc)	2 (6)*
Move paragraph	1 (4)*
Tal på item totalt	42 (51)*

* Disse 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 tre luker. Via ein nedtrekksmeny må eleven velje eitt av fire alternativ for kvar av dei tre lukene.

Tabell 2 gir ei oversikt over gjennomsnittleg vanskegrad (*p*- og *b*-verdiar) og diskrimineringsindeks (*a*-verdi) for prøven på 8. trinn. Desse verdiane er basert på pilotdata.

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

	Nasjonal prøve 8. trinn
Tal på item	42 (51)
Gjennomsnittleg <i>p</i> -verdi	0,467
Gjennomsnittleg <i>b</i> -verdi, OPLM	0,045
Gjennomsnittleg <i>a</i> -verdi OPLM	4,43

*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 8. trinn i 2018

Tabell 3 gir ei oversikt over kva hovudprøven på 8. trinn måler.

Tabell 3 : Oversikt over kva itema i hovudprøven på 8. trinn 2018 måler

		Føremål med lesinga					
		Finne info og forstå detaljar	Forstå hovudinnhald	Tolke	Bruke ord, stave- og gram. mønst.	Oppfatte samanheng i tekstar	Forstå avansert vokabular
Prosesseringsnivå	Vokabular	3					
	Setning	9		7	9		3
	Ulike deler av teksten	3	5	2		1	

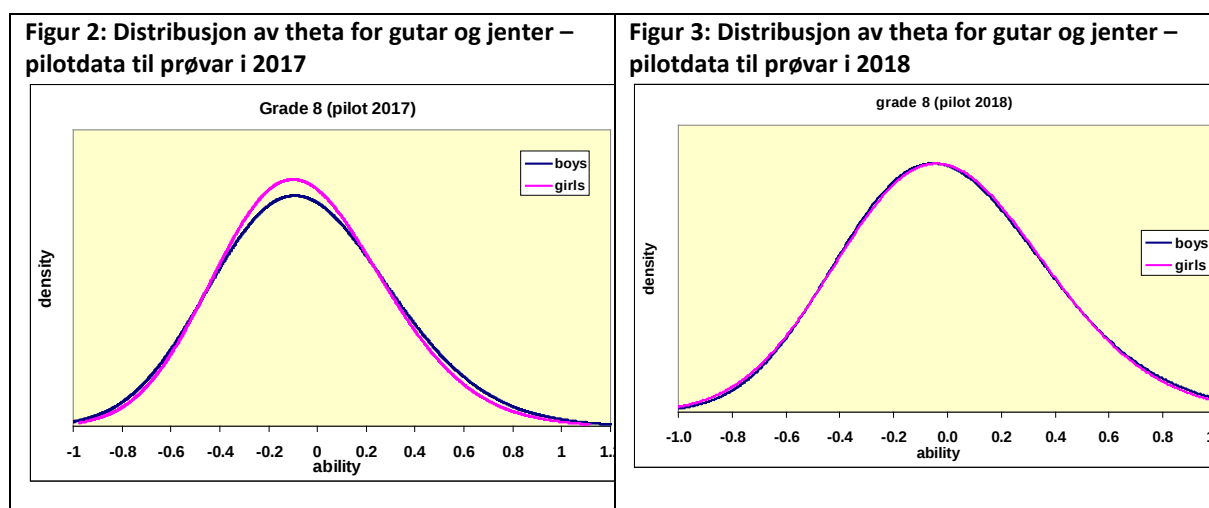
Tabell 3 må forståast slik:

Hovudprøven på 8. trinn inneheld

- 15 item der eleven skal finne informasjon. I tre tilfelle er svaret å finne i eit ord, i ni tilfelle i ei setning og i tre tilfelle ved å kople informasjon frå ulike delar av teksten.
- fem item der eleven skal forstå hovudinnhald. I alle tilfella kan dette gjerast ved å kople informasjon frå ulike deler av teksten.
- ni item der eleven må tolke innhaldet for å finne svaret. I sju tilfelle finst svaret i ei setning og i to tilfelle må eleven kople informasjon frå ulike stader i teksten.
- ni oppgåver av meir formell karakter der eleven må bruke ord og mønster for staving og grammatikk. I nokre tilfelle skal dei velje eitt av fire alternative svar. I andre tilfelle skal dei skrive ord, eller formulere deler av setningar sjølv.
- eitt item der eleven må vise at han forstår samanhengen i eit tekst. Dette er ei *Move paragraph*-oppgåve der eleven kan svare rett på fire item.
- tre oppgåver der eleven er avhengig av å forstå relativt avansert vokabular for å finne svaret. I slike oppgåver kan eleven forstå kva ordet betyr ut frå samanhengen i teksten.

1.4 Om kjønnskilnader

Norman Verhelst viser i sin rapport at pilotdata indikerer at skilnaden mellom gutar og jenter på 8. trinn ikkje har endra seg særleg dei siste åra. Dette er med utgangspunkt i pilotdata. På 8. trinn ser det ikkje det ut til å vere særleg skilnad mellom gutar og jenter. Figurane 2 og 3, som er henta frå Verhelst sin rapport, viser at distribusjonen av theta for gutar og jenter ser ut til å vere relativt lik, med utgangspunkt i piloteringsdata.



Vi ser at i 2018 er distribusjonen av theta for gutar og jenter meir eller mindre samanfallande.

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

År:	2016	2017	2018
8. trinn	+0.064	-0.024	-0.007

Dersom effektstørrelsen for kjønn er rundt 0, betyr det at forskjellen mellom kjønnene er minimal. På 8. trinn ser ikkje skilnaden ut til å vere stor. Når vi tidlegare år har signifikantesta resultata for gutar og jenter, har det variert litt frå år til år om det er gutar eller jenter som har gjort det best på prøvane på 8. trinn. Med data frå mellom 50 000 og 60 000 elevar, har vi nesten alltid fått signifikante resultat (som forventa), men skilnadene har alltid vore små. Etter prøven er gjennomført til hausten (2018), vil vi sjå nærare på om skilnaden mellom gutar og jenter framleis er liten.

DEL 2 – NASJONALE PRØVAR I ENGELSK 8. TRINN 2018

2. Nasjonal prøve i engelsk 8. trinn haust 2018 - hovudresultat

2.1 Gjennomføring 2018

Totalt var det 58 878 elevar som tok Nasjonal prøve i engelsk for 8. trinn hausten 2018. 55 387 elevar tok hovudprøven som vart sett saman etter piloteringa hausten 2017. Resten av elevane, 3491 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 8. 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). Tre av formata inneheld item som blir skåra polytomt (eleven svarar på fleire deloppgåver og kan dermed få fleire rette svar): *Gap filling*, *Move Paragraph* 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 8. trinn. Tabellen viser estimat både for kvar av dei tre versjonane av hovudprøven og for alle elevane på trinnet (samla resultat for dei tre versjonane).

Tabell 5: Oppsummerande statistikk NP-engelsk 8. 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
NP8-1	42(51)	22 (27)	20 (24)	0,935	0,524	11,946	0,509	18864
NP8-2	42(51)	22 (27)	20 (24)	0,934	0,523	11,887	0,506	18355
NP8-3	42(51)	22 (27)	20 (24)	0,934	0,527	11,892	0,505	18158
NP8-alle	42(51)	22 (27)	20 (24)	0,935	0,525	11,909	0,507	55387

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

NP8-1 = nasjonal prøve i engelsk for 8. trinn, versjon 1

NP8-2 = nasjonal prøve i engelsk for 8. trinn, versjon 2

NP8-3 = nasjonal prøve i engelsk for 8. trinn, versjon 3

Prøven på 8. trinn er reliabel (0,935) og diskriminerer på ein fornuftig måte mellom elevane. Prøven har ein gjennomsnittleg p-verdi på 0,525, som er litt lettare enn det vi siktar mot når vi

konstruerer prøvane. Vi ser at dei tre hovudprøvane oppfører seg så og seie identisk. Dette er forventa. Den gjennomsnittlege p -verdien etter pilotinga var 0,467 (sjå tabell 2), og det betyr at prøven har blitt vel 6 prosentpoeng lettare enn piloteringsdata indikerte.

2.3 Overordna IRT-analysar – θ , 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 $+\infty$ (frå 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 θ (theta) 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 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 Xcalibre-analysane viser til analysar av alle itema som var brukt på Nasjonal prøve i engelsk for 8. trinn i 2018, både i hovudprøven og ankerprøven.

2.3.1 Theta

Tabell 6 og Figur 4 oppsummerer thetamål i 8. 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 8. trinnsgruppa varierer frå -2,889 og +2,911.

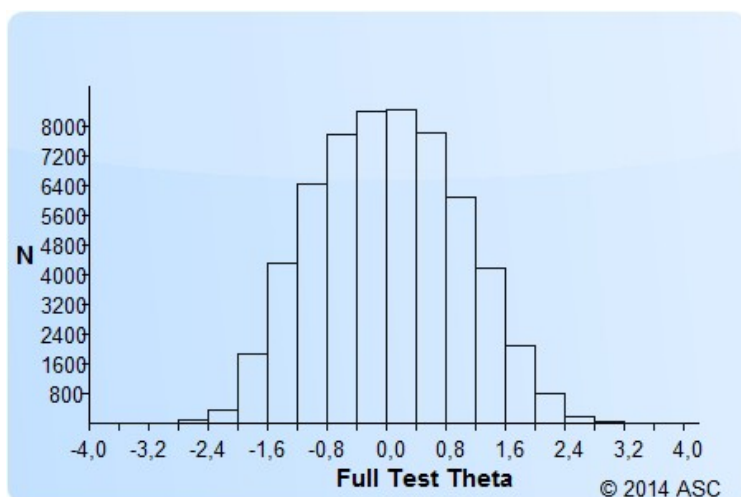
Tabell 6: Oppsummerende statistikk for theta (Xcalibre med ankring)

Elevlar	Gjennom- snitt	Standard- avvik	Minimum	Median	Max
58878	0,022	0,970	-2,889	0,009	2,911

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

Vi ser at distribusjonen av theta i 8. trinnspopulasjonen er tilnærma normalfordelt. Data frå prøven viser altså at det finst elevlar 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 8. trinn



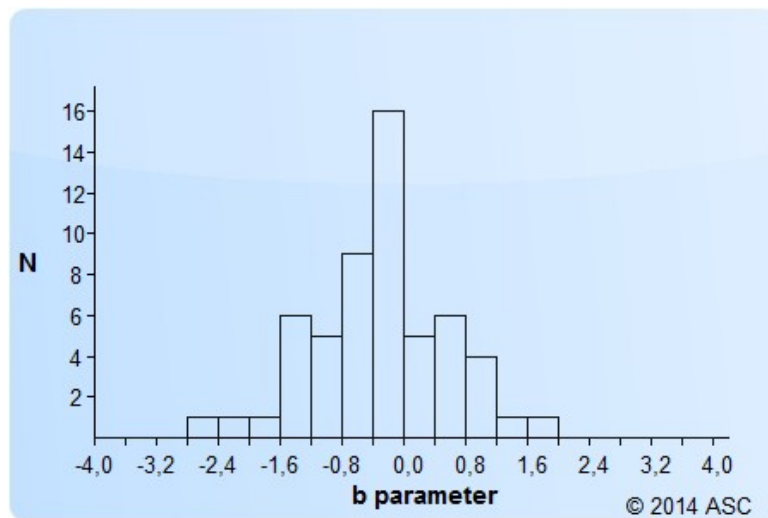
Tabell 7: Fordeling av theta

Frå	Til	Tal på elevlar
Under -3,2		0
-3,2	-2,8	17
-2,8	-2,4	68
-2,4	-2,0	367
-2,0	-1,6	1853
-1,6	-1,2	4325
-1,2	-0,8	6427
-0,8	-0,4	7781
-0,4	0	8402
0	0,4	8439
0,4	0,8	7815
0,8	1,2	6074
1,2	1,6	4169
1,6	2,0	2065
2,0	2,4	815
2,4	2,8	196
2,8	3,2	65
Over 3,2		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 8. trinnsprøvene i 2018, fordeler seg på thetaskalaen.

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



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

Frå	Til	Tal på item
Under -2,8		0
-2,8	-2,4	1
-2,4	-2,0	1
-2,0	-1,6	1
-1,6	-1,2	6
-1,2	-0,8	5
-0,8	-0,4	9
-0,4	0	16
0	0,4	5
0,4	0,8	6
0,8	1,2	4
1,2	1,6	1
1,6	2,0	1
Over 2,0		0

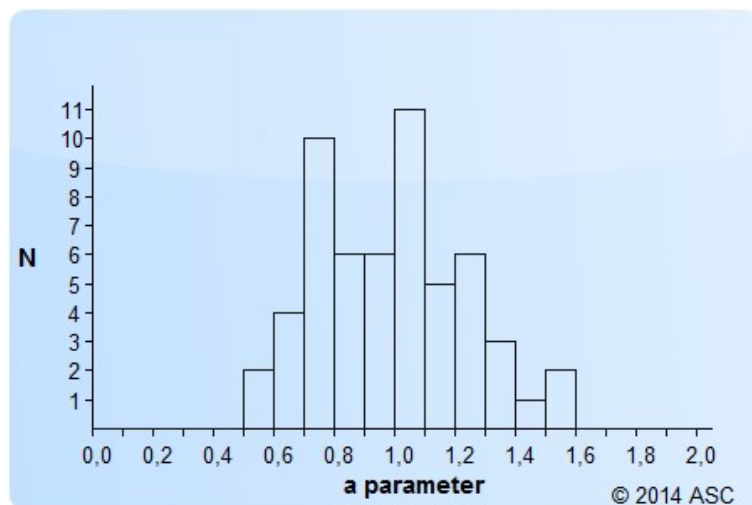
Vanskegraden for dei dikotome itema i hovudprøven på 8. trinn varierer frå -2,69 (lettaste item) til +1,80 (vanskelegaste item). Vi ser også at gjennomsnittleg er det fleire relativt lettare item (med minus verdiar) enn vanskelege (med +-verdiar) av dei som gjekk i prøvane på 8. trinn.

I hovudprøve og ankerprøve for 8. trinn finst det til saman 7 polytome oppgåver, oppgåver der elevane svarar på fleire item og får ein samleskår for kvar av desse oppgåvene. Fleire av dei polytome oppgåvene har ein høgre b -verdi (totalt) enn dei dikotome oppgåvene. Det finst såleis polytome oppgåver med b -verdi for 3 rette svar både på 2,17, 2,54 og 3,07. (Dette gjeld altså når vi ser på b -verdien for alle rette svar. b -verdi for eitt eller to rette svar blir sjølvsagt lågare i slike oppgåver.)

2.3.3 a -parameter

a -parametra viser som sagt til kor godt item diskriminerer. Xcalibre-skalaen for a -parameter går frå 0 og oppover. Akseptabel diskriminering med utgangspunkt i Xcalibre-analysar bør vere over 0,5. Figur 6 og tabell 9 viser korleis a -parametra til itema som vart administrert i prøvane på 8. trinn, er distribuert. Oversikta viser til dikotome item.

Figur 6: Distribusjon av item etter a-parameter (diskriminering) i prøven på 8. trinn



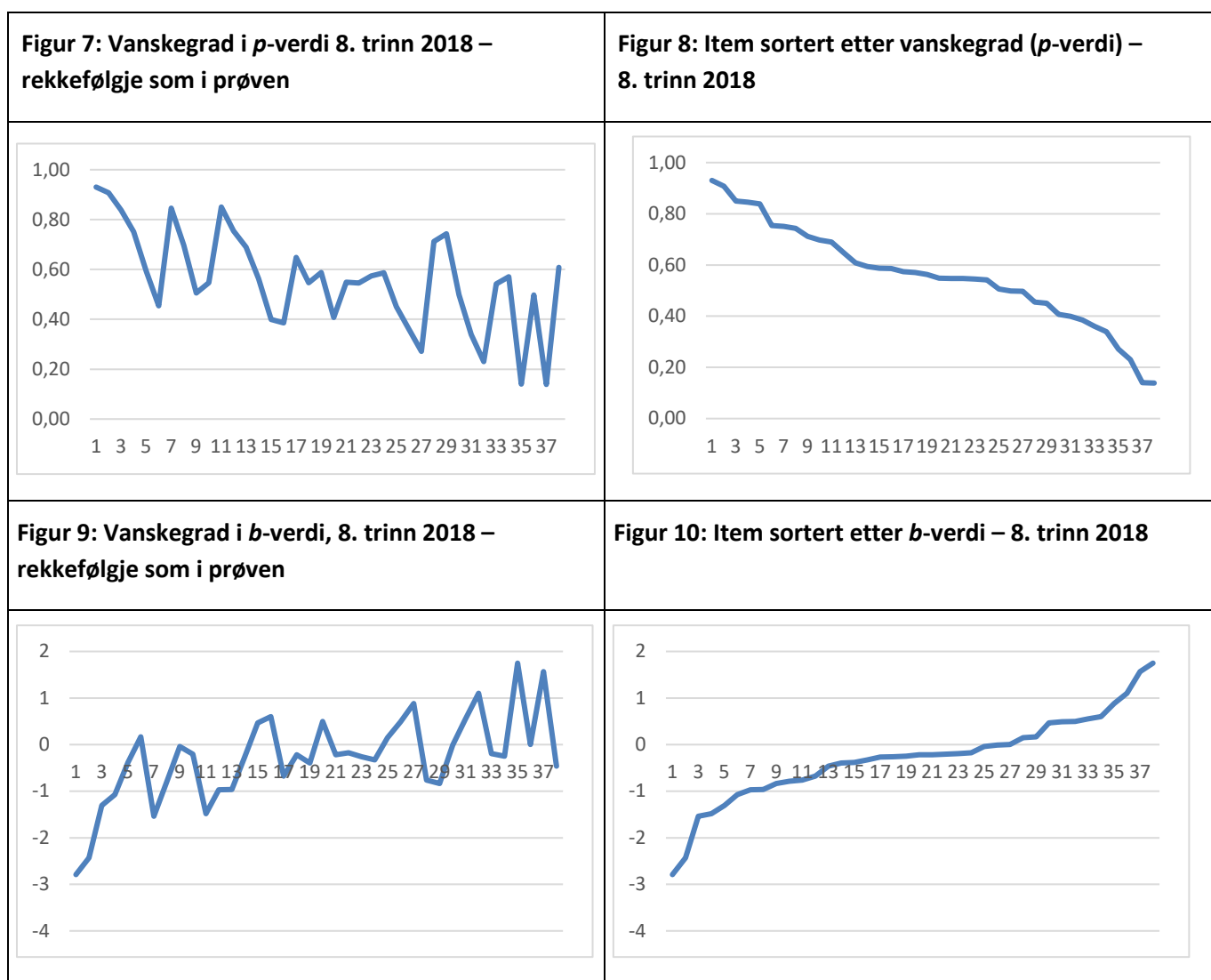
Tabell 9: Fordeling av a-parameter (dikotome item)

Frå	Til	Tal på item
Under 0,5		0
0,5	0,6	2
0,6	0,7	3
0,7	0,8	10
0,8	0,9	6
0,9	1,0	6
1,0	1,1	11
1,1	1,2	5
1,2	1,3	6
1,3	1,4	3
1,4	1,5	1
1,5	1,6	2
Over 1,6		0

Av dei dikotome itema som vart brukt i prøvane på 8. trinn i 2018, er det ingen item som har a -verdi under 0,5. Det itemet med høgast diskriminering er eit *Click word*-item med a -verdi på 1,595.

2.4 p -verdi

Figurane 7 - 10 viser til vanskegrad på 8. trinnsprøven. Figurane 7 og 8 viser p -verdi i prøvane. Figur 7 viser vanskegraden på itema etter rekkefølgja 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 rekkefølgja 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.



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 8. trinn varierer mellom ein p -verdi på 0,87 (lettaste) og 0,15 (vanskelegaste). Vanskegraden i b -verdi varierer mellom -2,69 (lettaste) og +1,80 (vanskelegaste). Figurane viser at prøven på 8. 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 8. trinn

Det er ei kjent sak at vanskegraden på item fell frå pilotering til reell prøve. 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 8.

trinn 0,464. Tabell 5 i denne rapporten viser at gjennomsnittleg p -verdi for den reelle prøven var 0,525. Den same prøven vart altså 6 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 åttande trinn. I tabellen er det brukt forkortingar for format på denne måten:

cp = Click picture

c&d = Click and drag

ct = Click text

Gap = Gap filling

MCs = MC short

MCl = MC long

who = Who could say

cw = Click word

cs = Complete sentence

wtw = Write the word

Mp = Move paragraph

Tabell 10: Skilnad i p -verdi mellom itema i prøven basert på pilotering (2017) og reell prøve 2018, 8. trinn

Rekkefølge	Format	Item	p -pilot	p -prøve	Diff	Rekkefølge	Format	Item	p -pilot	p -prøve	Diff
1	cp	5162160	0,90	0,93	-0,03	23	who	5167346	0,46	0,55	-0,08
2	cp	5166934	0,86	0,91	-0,05	24	who	5167291	0,45	0,57	-0,13
3	cp	5162071	0,83	0,84	-0,01	25	who	5167324	0,45	0,59	-0,13
4	c&d	5167918	0,76	0,75	0,00	26	who	5167330	0,30	0,45	-0,16
5	c&d	5120316	0,55	0,59	-0,05	27	who	5167349	0,19	0,36	-0,18
6	c&d	5167936	0,37	0,45	-0,08	28	Who	5167273	0,23	0,27	-0,04
7	ct	5167724	0,83	0,85	-0,01	29	cw	5162218	0,58	0,71	-0,13
8	ct	5160141	0,66	0,70	-0,03	30	cw	5162222	0,57	0,74	-0,17
9	ct	5167737	0,45	0,51	-0,06	31	cw	5167160	0,34	0,50	-0,16
10	ct	5167954	0,49	0,55	-0,05	32	cw	5168005	0,26	0,34	-0,08
11	gap*	5161306	0,27	0,33	-0,06	33	cw	5167194	0,20	0,23	-0,04
12	mcs	5161580	0,76	0,85	-0,10	34	cs	5167655	0,42	0,54	-0,13
13	mcs	5161594	0,71	0,75	-0,04	35	cs	5167657	0,30	0,57	-0,27
14	mcs	5160649	0,63	0,69	-0,06	36	cs	5167645	0,09	0,14	-0,05
15	mcs	5162238	0,48	0,56	-0,08	37	cs	5167643	0,28	0,50	-0,21
16	mcs	5161667	0,37	0,40	-0,03	38	cs	5167605	0,13	0,14	-0,01
17	mcs	5167431	0,34	0,39	-0,05	39	cs	5162693	0,44	0,61	-0,17
18	mcl	5161703	0,57	0,65	-0,08	40	wtw*	5162609	0,02	0,05	-0,03
19	mcl	5161707	0,52	0,55	-0,02	41	wtw*	5162598	0,04	0,07	-0,03
20	mcl	5161719	0,46	0,59	-0,13	42	mp*	5167208	0,11	0,13	-0,02
21	mcl	5161717	0,39	0,41	-0,02						
22	mcl	5161709	0,46	0,55	-0,09						

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

Den grå kolonna i tabellen viser differansen i p -verdi mellom pilotering og nasjonale prøvar. Cellene med den mørkaste gråfargen viser til item der differansen er større enn 10 prosentpoeng. Det er 11 item (av 42) i 8. trinnsprøven der auken i p -verdi er større enn 10 prosentpoeng.

2.6 Polytome item

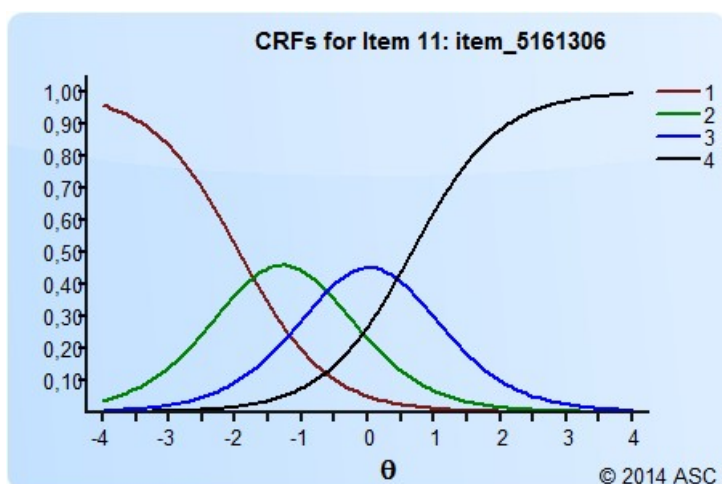
Dei første åra vart alle item i dei Nasjonale prøvane i engelsk skåra binært (rett/galt eller 1/0). Dette betydde at når vi til dømes laga *Gap filling*-oppgåver, så piloterte vi kanskje ei oppgåve med fem luker. Etter pilotering kunne kanskje analysane vise at to eller tre av desse ikkje fungerte, enten 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 tre itemformat: *Gap filling*, *Move paragraph* og *Write the word* på 8. trinn. Dette ser ut til å fungere bra. Under gir vi nokre eksempel på oppgåver frå 2018-prøvene 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: Oppgåve 11 i hovudprøven på 8. trinn – Gap filling



Figur 11 viser at oppgåva skil godt mellom elevar på ulike ferdigheitsnivå. Elevar som ikkje hadde nokon rette, hadde gjennomsnittleg theta på -0,1059. Dei som fekk tre rette svar, hadde i gjennomsnitt theta på 0,777. Tabell 11 viser samanhengen mellom rette svar, prosentvis fordeling av elevar og thetamål.

Tabell 11: Oppgåve 11 i 8. trinnsprøven – rette svar, fordeling i prosent av elevar og thetamål

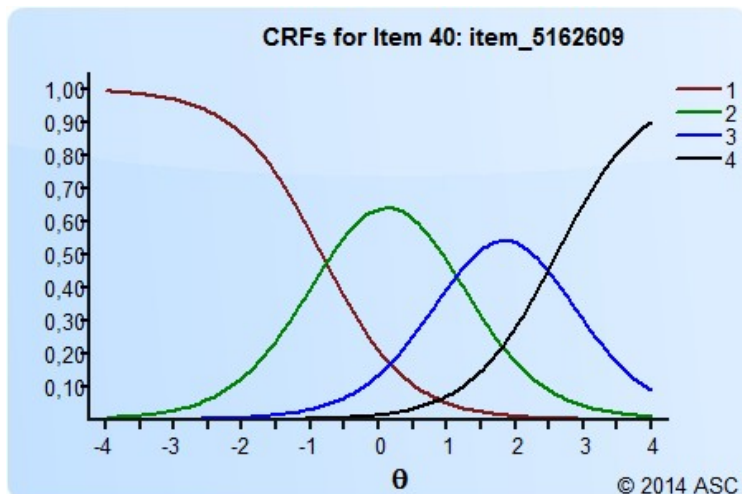
Rette svar	Elevar i prosent	Gjennomsnittleg theta
0	10,6	-1,013
1	24,1	-0,616
2	31,9	0,041
3	33,4	0,775

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 8. trinn får to slike oppgåver. Det kjem klart fram av analysane at dette oppgåveformatet skil godt mellom elevar på ulike ferdigheitsnivå. Figur 12 viser oppgåve 40 i prøven.

Figur 12: Oppgåve 40 - Record rocking chair



Figur 12 viser at oppgåva skil godt mellom elevar på ulike ferdigheitsnivå. 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 40 i 8. trinnspørven – rette svar, fordeling i prosent av elevar og thetamål

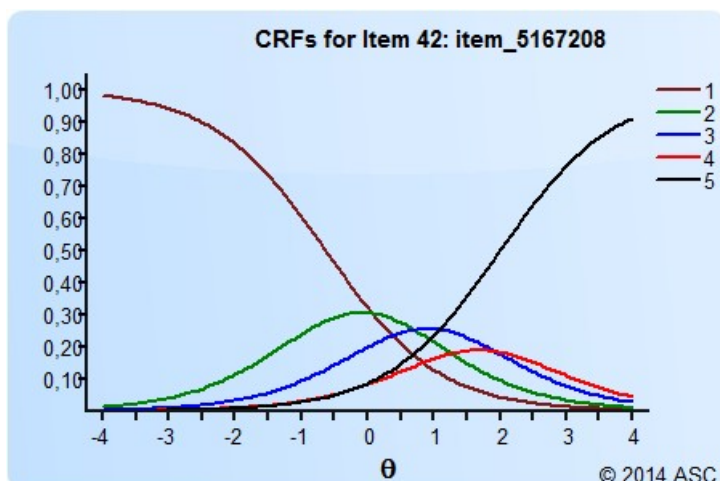
Rette svar	Elevar i prosent	Gjennomsnittleg theta
0	29,0	-0,847
1	47,3	0,122
2	19,3	0,731
3	4,5	1,429

Den prosentvise fordelinga av elevar i oversikta over viser at om lag halvparten av elevane klarar å få eitt rett svar, og at relativt få klarar to rette og ikkje minst tre rette svar. Oppgåveformatet fungerer godt både på 5. og 8. trinn.

2.6.3 Move paragraph

I formatet *Move paragraph* skal elevane flytte fem avsnitt slik at resultatet blir ein samanhengande tekst. Det er mogeleg å ende opp med 0, 1, 2, 3 eller 4 rette svar. Figur 13 viser oppgåve 42 i prøven.

Figur 13: Oppgave 42 i hovudprøven på 8. trinn – *Move paragraph*



Både figur 13 og tabell 13 (under) viser at dette formatet skil godt mellom dei som har 0, 1, 2/3 og 4 rette svar. Det kan sjå ut som om formatet ikkje skil like klart mellom dei som har to og tre rette svar. Det kan enten vere fordi formatet i seg sjølv ikkje fangar opp skilnaden mellom nokre av ferdigheitsnivåa, det kan ha med måten formatet er skåra på, og det kan vere fordi IRT-analyse ikkje passar like godt til alle polytome format. Dette vil vi sjå nærare på framover.

Tabell 13: Oppgave 42 i 8. trinnsprøven – rette svar, fordeling i prosent av elevar og thetamål

Rette svar	Elevar i prosent	Gjennomsnittleg theta
0	37	-0,515
1	24,1	-0,209
2	16,9	0,428
3	8,9	0,409
4	13,1	1,141

2.7 Kjønnskilnader 8. 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), blir lett trivielle skilnader signifikante. Desse analysane baserer seg på data frå mellom 50000 og 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 viste ikkje skilnader mellom jenter og gutar på 8. trinn (sjå figur 3 og tabell 4).

Gjennomsnittleg theta for 2018-avviklinga viser at skilnaden mellom gutar og jenter på 8. trinn ikkje er så stor, men kanskje større enn forventa. Skilnaden er signifikant. Som i 2017 gjer gutane det betre enn jentene. Tidlegare år har jentene gjort det litt betre enn gutane.

Tabell 14 viser skilnadene mellom gutar og jenter basert på ein t-test. Vi ser at mens gjennomsnittleg theta for gutane er (+)0,0711, er theta for jentene -0,0284. Dette indikerer altså at gutane har betre leseferdigheiter på engelsk enn jentene, skilnaden i theta er på 0.0995 i gutane sin favør. Denne skilnaden er større enn forventa. Standardavviket er nesten likt i dei to gruppene, noko som seier at variasjonen i jentegruppa og gutegruppa er omtrent lik.

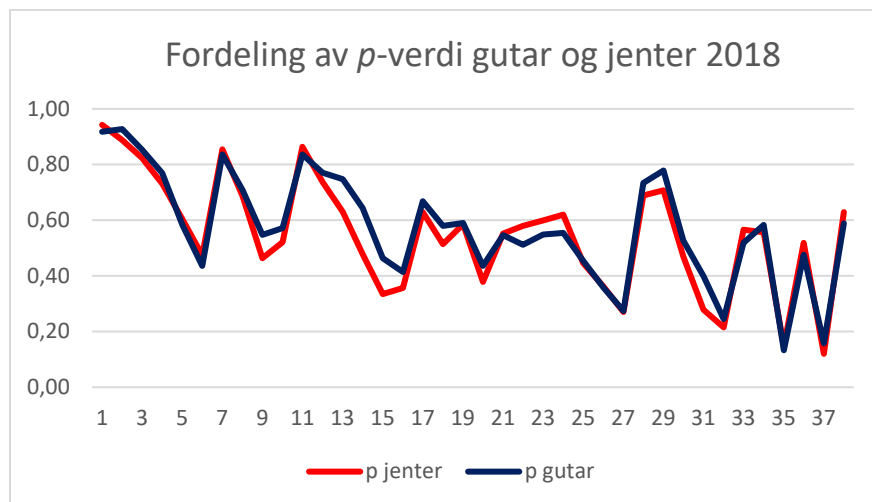
Tabell 14: Skilnader mellom jenter og gutar på 8. trinn – t-test

Prøve	Kjønn	Gjennomsnittleg theta	<i>n</i>	Standardavvik	<i>t</i>	<i>Sig</i>
NP-8	Gutar	0,0711	29741	0,9823	12,684	0,000
	Jenter	-0,0284	29137	0,9546		

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 undersøke dette nærare i tida framover.

Figur 14 viser p -verdi knytt til kjønn.

Figur 14: p -verdiar gutar og jenter 8. trinn 2018



Den raude lina viser p -verdiar for jenter og den blå for gutar. Vi ser at prøven i all hovudsak oppfører seg likt i dei to gruppene. Sjølv om det er ein skilnad mellom gutar og jenter på 8. trinn, så er ikkje skilnaden stor på dei fleste enkeltitem. Gjennomsnittleg råskår for gutar er 27,46 (av maks 51) og for jenter 26,08. Dette vil seie at gutane skårar nesten eitt og eit halvt poeng høgare enn jentene i gjennomsnitt. I vedlegg 3 blir p -verdiane på enkeltitem oppgitt, både totalt og for jenter og gutar separat. Gutane gjer det best på 27 av 42 item. I dei fleste tilfelle er skilnaden svært liten, men i nokre tilfelle er skilnaden ganske stor. Dette kan vi også sjå i figur 14.

Fram til og med 2016 viste resultatet på dei nasjonale prøvane i engelsk at skilnaden mellom gutar og jenter var svært liten. Ofte gjorde jentene det litt betre enn gutane, men det var heile tida snakk om svært små skilnader, sjølv om resultata var signifikante. På 5. trinn har tendensen vore at gutane har gjort det betre enn jentene dei siste åra. Både i 2017 og 2018 er det gutane som gjer det betre enn jentene på 8. trinn. Dei neste åra vil vise om denne tendensen vil halde seg, og om vi vil sjå det same i resultata for 8. trinn som vi har sett nokre år på 5. trinn.

Ein rask kikk i arkivet av rapportar frå dei siste åra med fokus på jenter og gutar, er summert opp i Tabell 15.

Tabell 15: Kjønnsskilnader 2010 – 2018.

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

Oversikta i tabell 15 viser at gutane har gjort det best på dei nasjonale prøvane i engelsk for 8. trinn dei to siste åra. Dette betyr at det kan vere ei endring på gang når det gjeld kjønn og engelskferdigheiter på 8. 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 8. trinn fungerer i all hovudsak som forventa. Prøven er reliabel, har ein passande vanskegrad, inneheld item med varierende vanskegrad, og diskriminerer på ein fornuftig måte mellom elevane.

Prøvene 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 lesetekstar. 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 8. trinn.

Dei to siste åra har vi sett at gutane på 8. 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 forsvarast teknisk. Lytteitem kan også, på same måte som leseitem, skårast automatisk.

Til slutt vil vi nemne at det er skrive lærarretteiingar som fylgjer prøvane kvart år. Lærarretteiingane 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

	Pilotprøvar															
Item-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 8. trinn 2018 – oversikt og statistiske mål etter utprøving

Rekke-følgje*	Format	item no	Task name	Teksttype	<i>p</i>	Maks skår	<i>a (OPLM)**</i>	<i>b</i>
1	Click picture	5162160	Flags 2-New Z-fact	Beskrivelse	0,90	1	4	-0,739
2	Click picture	5166934	Greek god of war	Faktatekst	0,86	1	3	-0,742
3	Click picture	5162071	New animal on the farm	Beskrivelse	0,83	1	6	-0,451
4	Click and drag	5167918	Gym1-bag2	Beskrivelse	0,76	1	5	-0,349
5	Click and drag	5120316	LockerroomComb	Beskrivelse	0,55	1	5	-0,095
6	Click and drag	5167936	Outerspace-water	Beskrivelse	0,37	1	4	0,152
7	Click text	5167724	Favourite books1-Amy	Forteljning	0,83	1	4	-0,537
8	Click text	5160141	Smartest animals1	Faktatekst	0,66	1	4	-0,261
9	Click text	5167737	Holidays3-Martha	Forteljning	0,45	1	3	0,079
10	Click text	5167954	Funny animal stories2cockatiel	Faktatekst	0,49	1	3	0,006
11	Gap*	5161306	Sinkhole-bbc	Faktatekst	0 = 0,13 1 = 0,28 2 = 0,32 3 = 0,27	3	2	-0,160
12	Mc-short	5161580	Boxing training1	Faktatekst	0,76	1	5	-0,371
13	Mc-short	5161594	Johnny's prize	Forteljning	0,71	1	5	-0,266
14	MC-short	5160649	Coffee car2	Faktatekst	0,63	1	4	-0,187
15	Mc-short	5162238	Balloon record 3-bbc	Faktatekst	0,48	1	4	-0,009
16	MC-short	5161667	GolfFan-fact	Faktatekst	0,37	1	3	0,201
17	MC-short	5167431	Experiment bending water	Faktatekst	0,34	1	3	0,305
18	Mc-long	5161703	Cities in northern England1	Faktatekst	0,57	1	4	-0,111

Vedlegg 2

19	Mc-long	5161707	Cities in northern England3	Faktatekst	0,52	1	4	-0,055
20	Mc-long	5161719	Cities in northern England7	Faktatekst	0,46	1	5	0,024
21	Mc-long	5161717	Cities in northern England6	Faktatekst	0,39	1	4	0,132
22	Mc-long	5161709	Cities in northern England4	Faktatekst	0,46	1	4	0,028
23	Who	5167346	The life of Henry VIII Jane S	Faktatekst	0,46	1	5	0,041
24	Who	5167291	The life of Henry VIII Cath P	Faktatekst	0,45	1	6	0,058
25	Who	5167324	The life of Henry VIII Elizab	Faktatekst	0,45	1	5	0,020
26	Who	5167330	The life of Henry VIII Anne of	Faktatekst	0,30	1	7	0,203
27	Who	5167349	The life of Henry VIII Catherine Howard2	Faktatekst	0,19	1	5	0,433
28	Who	5167273	The life of Henry VIII Anne of	Faktatekst	0,23	1	5	0,348
29	Click words	5162218	Bronze Age village 2	Faktatekst	0,58	1	7	-0,105
30	Click words	5162222	Courtroom 1-joke-	Faktatekst	0,57	1	6	-0,089
31	Click words	5167160	Jungle explorer2	Faktatekst	0,34	1	6	0,164
32	Click words	5168005	Kilimanjaro effort	Faktatekst	0,26	1	6	0,263
33	Click words	5167194	Popular reality TV5	Faktatekst	0,20	1	4	0,550
34	Complete sentence	5167655	Comp-younger	Vokabularsetn.	0,42	1	4	0,154
35	Complete sentence	5167657	Comp-less	Vokabularsetn.	0,30	1	5	0,248
36	Complete sentence	5167645	Dir/Indir-she could not	Grammatikksetn.	0,09	1	4	0,831
37	Complete sentence	5167643	Dir/Indir-he is	Grammatikksetn.	0,28	1	5	0,257
38	Complete sentence	5167605	Neg.-does not	Grammatikksetn.	0,13	1	5	0,626
39	Complete sentence	5162693	It-there 4-it is	Grammatikksetn.	0,44	1	5	0,046

Vedlegg 2

40	Write the word*	5162609	Record rocking chair-fact	Faktatekst	0 = 0,43 1 = 0,36 2 = 0,17 3 = 0,05	3	3	0,565
41	Write the word*	5162598	Crazy horse-fact	Faktatekst	0 = 0,45 1 = 0,15 2 = 0,33 3 = 0,07	3	3	0,275
42	Move paragraph*	5167208	The Lego brand	Faktatekst	0 = 0,50 1 = 0,21 2 = 0,12 3 = 0,05 4 = 0,11	4	3	0,415

*Oppgåva er polytom. *p*-verdien viser til proporsjonen av elevar som har 0, 1, 2, 3 eller 4 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 mogele

Vedlegg 3: Hovudprøve 8. trinn 2018 – item for item-oversikt

Rekke-følgje	Format	Item	Namn	<i>p</i>	<i>p</i> jenter	<i>p</i> gutar	diff j-g	a	b
1	Click picture	5162160	Flags 2-New Z-fact	0,93	0,94	0,92	0,02	0,674	-2,689
2	Click picture	5166934	Greek god of war	0,91	0,89	0,93	-0,04	0,699	-2,337
3	Click picture	5162071	New animal on the farm	0,84	0,82	0,85	-0,03	1,321	-1,246
4	Click and drag	5167918	Gym1-bag2	0,75	0,73	0,77	-0,04	0,888	-1,020
5	Click and drag	5120316	LockerroomComb	0,59	0,60	0,58	0,02	0,888	-0,354
6	Click and drag	5167936	Outerspace-water	0,45	0,47	0,44	0,04	0,917	0,17
7	Click text	5167724	Favourite books1-Amy	0,85	0,85	0,84	0,02	0,965	-1,472
8	Click text	5160141	Smartest animals1	0,70	0,69	0,71	-0,02	0,979	-0,743
9	Click text	5167737	Holidays3-Martha	0,51	0,46	0,55	-0,08	0,752	-0,015
10	Click text	5167954	Funny animal stories	0,55	0,52	0,57	-0,05	0,722	-0,178
11	Gap*	5161306	Sinkhole-bbc	0 = 0,11 1 = 0,24 2 = 0,32 3 = 0,33	0,26	0,40	-0,14	0,915	-1,86, -0,58, 0,67
12	Mc-short	5161580	Boxing training1	0,85	0,86	0,84	0,03	1,085	-1,419
13	MC-short	5161594	Johnny's prize	0,75	0,74	0,77	-0,03	1,162	-0,920
14	Mc-short	5160649	Coffee car2	0,69	0,63	0,75	-0,12	0,629	-0,914
15	Mc-short	5162238	Balloon record 3-bbc	0,56	0,48	0,64	-0,17	0,813	-0,242
16	MC-short	5161667	GolfFan-fact	0,40	0,33	0,46	-0,13	0,651	0,474

17	MC-short	5167431	Experiment bending water	0,39	0,36	0,41	-0,06	0,546	0,606
18	Mc-long	5161703	Cities in northern England1	0,65	0,63	0,67	-0,04	0,726	-0,639
19	Mc-long	5161707	Cities in northern England3	0,55	0,51	0,58	-0,06	0,730	-0,191
20	Mc-long	5161719	Cities in northern England7	0,59	0,59	0,59	-0,01	0,747	-0,362
21	Mc-long	5161717	Cities in northern England6	0,41	0,38	0,44	-0,06	0,531	0,500
22	Mc-long	5161709	Cities in northern England4	0,55	0,55	0,55	0,01	0,762	-0,195
23	Who	5167346	The life of Henry VIII	0,55	0,58	0,51	0,07	1,248	-0,149
24	Who	5167291	The life of Henry VIII	0,57	0,60	0,55	0,05	1,523	-0,230
25	Who	5167324	The life of Henry VIII	0,59	0,62	0,55	0,07	1,065	-0,298
26	Who	5167330	The life of Henry VIII	0,45	0,45	0,45	-0,01	1,427	0,164
27	Who	5167349	The life of Henry VIII	0,36	0,36	0,36	0,00	1,118	0,498
28	Who	5167273	The life of Henry VIII	0,27	0,27	0,27	0,00	1,035	0,874
29	Click word	5162218	Bronze Age village 2	0,71	0,69	0,73	-0,04	1,309	-0,718
30	Click word	5162222	Courtroom 1-joke-	0,74	0,71	0,78	-0,07	1,548	-0,790
31	Click word	5167160	Jungle explorer2	0,50	0,47	0,53	-0,06	1,209	-0,009
32	Click word	5168005	Kilimanjaro effort	0,34	0,28	0,40	-0,12	1,195	0,559
33	Click word	5167194	Popular reality TV5	0,23	0,22	0,24	-0,03	0,997	1,086
34	Complete sentence	5167655	Comp-younger	0,54	0,57	0,52	0,05	0,772	-0,165
35	Complete sentence	5167657	Comp-less	0,57	0,56	0,58	-0,03	1,393	-0,223
36	Complete sentence	5167645	Dir/Indir-she could not	0,14	0,15	0,13	0,01	0,856	1,712
37	Complete sentence	5167643	Dir/Indir-he is	0,50	0,52	0,48	0,04	0,772	0,020
38	Complete sentence	5167605	Neg.-does not	0,14	0,12	0,16	-0,04	1,061	1,537
39	Complete sentence	5162693	It-there 4-it is	0,61	0,63	0,59	0,04	0,833	-0,428

40	Write the word*	5162609	Record rocking chair-fact	0 = 0,29 1 = 0,47 2 = 0,19 3 = 0,05	0,03	0,06	-0,03	0,965	-0,79, 1,07, 2,54
41	Write the word*	5162598	Crazy horse-fact	0 = 0,28 1 = 0,23 2 = 0,41 3 = 0,07	0,06	0,08	-0,02	1,128	-0,76, 0,08, 2,03
42	Move paragraph*	5167208	The Lego brand	0 = 0,37 1 = 0,24 2 = 0,17 3 = 0,09 4 = 0,13	0,14	0,13	0,01	0,711	-0,59, 0,46, 1,33, 1,96

*Oppgåva er polytom. *p*-verdien viser til proporsjonen av elevar som har 0, 1, 2, 3 eller 4 rette svar.