



**BØRNE- OG
UNDERVISNINGS-
MINISTERIET**
STYRELSEN
FOR IT OG LÆRING

Evaluering af de statistiske aspekter ved de nationale test

Januar 2020

Indhold

Forord	4
De nationale test	4
Opgavebeskrivelse	5
Udarbejdet materiale	6
1 Notat 1 Algoritmen i testsystemet og beregning af elevdygtigheden	7
Sammenfatning	7
Indledning	8
Rasch-modellen og beregning af elevdygtighed	8
Run-in perioden	9
Efter run-in perioden	10
Valg af opgaver i den adaptive algoritme	10
Eksempler på valg af opgave i testforløb	11
Tjek af beregnet dygtighed og statistisk usikkerhed	17
2 Notat 2 De nationale tests måleegenskaber	24
Sammenfatning	24
Indledning	25
Sammenhæng mellem testresultater og prøvekarakter	25
Sammenhæng mellem testresultater og standpunktskarakter	29
Sammenhæng mellem testresultater og PISA test	31
3 Notat 3 Den statistiske usikkerhed og testenens reliabilitet	32
Sammenfatning	32
Indledning	33
Den statistiske usikkerhed	33
Den statistiske usikkerhed og testenens stopkriterie	34
Den statistiske usikkerhed i de enkelte profilområder	36
Den statistiske usikkerhed og antal opgaver	37
Den statistiske usikkerhed og elevdygtigheden	39
Den statistiske usikkerhed og sikkerhedsintervaller	40
Reliabiliteten	42
Forslag til forbedringer af den statistiske sikkerhed	46
Antal opgaver	46
Polytome opgaver	46

Justering af algoritmen	47
Flere svære opgaver	47
4 Notat 4 Opgavebanken og opgavernes sværhedsgrad	48
Sammenfatning	48
Indledning	49
Opgaveafprøvningen.....	49
Opgavebanken	51
Stabiliteten af opgavernes sværhedsgrad over tid	53
Forskellen i opgavernes sværhedsgrad baseret på lineær og adaptiv testadministration	54
Ændring over tid	56
Opgavernes alder	58
Opgavernes sværhedsgrad	59
Forskel i beregnet elevdygtighed	60
5 Notat 5 Samling af testresultater fra flere profilområder	63
Sammenfatning	63
Indledning	64
Korrelationen mellem elevresultater.....	64
Antal besvarede opgaver og SEM i profilområderne.....	65
Analyse af mulighed for anvendelse af én Rasch-model pr. test ..	65
Samlet elevdygtighed	66
Bilag 1.1 Anvendte skalaer til præsentation af elevernes beregnede dygtigheder	68
Bilag 1.2 Opgavebanken i dansk læsning 8. klasse – sprogforståelse ...	70
Bilag 2.1 Sammenhæng mellem testresultater og karakterer	81
Bilag 3.1 Statistisk usikkerhed på elevdygtighederne	84
Bilag 3.2 Reliabilitet	92
Bilag 4.1 Opgaveafprøvningsperioder	95
Bilag 4.2 Skærmdumps fra RUMM	98
Dansk læsning 2. klasse – sprogforståelse	99
Dansk læsning 2. klasse – afkodning	103
Dansk læsning 2. klasse – tekstforståelse	106
Dansk læsning 4. klasse – sprogforståelse	111

Matematik 3. klasse – geometri og måling	115
Matematik 6. klasse – tal og algebra	120
Bilag 4.3 Opgavebankens sammensætning i forhold til opgavernes sværhedsgrad	124
Bilag 4.4 Sammenhæng mellem elevernes dygtighed og opgavernes sværhedsgrad	125
Bilag 4.5 Undersøgelse af link-opgavernes ændrede sværhedsgrad ..	136
Bilag 4.6 Forskel i opgavernes sværhedsgrad	141

Forord

De nationale test

Der er ti obligatoriske nationale test i folkeskolen (Figur 1), hvor hver test består af tre faglige profilområder¹. En test kan gennemføres på 45 minutter.

Figur 1 Frivillige og obligatoriske nationale test

Fag og klasstrin	1.	2.	3.	4.	5.	6.	7.	8.	9.
Dansk, læsning	■	■	■						
			■	■	■				
					■	■	■		
Matematik		■	■	■					
					■	■	■		
							■	■	■
Engelsk			■	■	■				
						■	■	■	
								■	■
Fysik/kemi						■	■	■	
Biologi							■	■	■
Geografi							■	■	■
Dansk som andetsprog				■	■	■			
							■	■	■

■ Obligatorisk test målrettet klasstrinnet
 ■ Frivillig test målrettet klasstrinnet
 ■ Frivillig test målrettet klasstrinnet over eller under

Kilde: www.uvm.dk/folkeskolen/elevplaner-nationale-test--trivselsmaaling-og-sprogproever/nationale-test

De nationale test er it-baserede, selvscorende og adaptive. At testene er adaptive betyder, at opgaverne i et testforløb udvælges så de bedst muligt passer til elevens dygtighedsniveau undervejs i forløbet. Dygtige elever får de sværeste opgaver, mens elever med større faglige udfordringer får de lettere opgaver.

De første obligatoriske nationale test blev gennemført i folkeskolen i skoleåret 2009/2010.

¹ <https://www.uvm.dk/folkeskolen/elevplaner-nationale-test--trivselsmaaling-og-sprogproever/nationale-test/klasstrin-fag-og-profilomraader>

Børne- og undervisningsministeren igangsatte en evaluering af de nationale test i 2018. I den forbindelse blev der nedsat en rådgivningsgruppe bestående af forskere og praktikere, der blandt andet skulle rådgive i forhold til indholdet af evalueringen af de nationale test.

Evalueringsspørgsmålene er udarbejdet på baggrund af anbefalinger fra rådgivningsgruppen.

Opgavebeskrivelse

Evalueringen af de nationale test består af to dele:

1. Validering af teknisk beregning
2. Undersøgelse af betydningen og brugen af de nationale test

Under delopgave 1 skal følgende evalueringsspørgsmål besvares:

1. *Regner de nationale test rigtigt?*

Ved besvarelse af spørgsmålet skal den kritik, der rejses af den statistiske sikkerhed og reliabiliteten i de nationale test adresseres. Herunder skal det klarlægges om:

- a. opgavernes sværhedsgrader stadig er korrekte?
 - b. opgaverne fortsat passer til Rasch-modellen?
 - c. det er muligt at forbedre den adaptive algoritme med henblik på at reducere den statistiske usikkerhed?
2. Det skal afdækkes, om sikkerheden i målingerne af elevernes færdigheder kan forbedres ved at kombinere resultater fra forskellige profilområder? Herunder sigter spørgsmålet på at klarlægge følgende:
- a. Kan det påvises, at profilområderne måler forskellige aspekter af den samme bagvedliggende færdighed?
 - b. Som følge af spørgsmål a: Kan testresultaterne fra profilområderne slås sammen og dermed forbedre sikkerheden i testene?

Styrelsen for It og Læring (STIL) under Børne- og undervisningsministeriet har tidligere undersøgt mange af evalueringens temaer og formidlet disse på www.uvm.dk². Børne- og undervisningsministeriet har derfor bedt STIL om at foretage analyserne knyttet til delopgave 1. Evalueringen af de statistiske aspekter ved de nationale test (delopgave 1) samler de tidligere gennemførte analyser og supplerer disse med opdaterede data og nye analyser.

² <https://www.uvm.dk/folkeskolen/elevplaner-nationale-test--trivselsmaaling-og-sprogproever/nationale-test/om-de-nationale-test>

Udarbejdet materiale

Til besvarelse af evalueringens delopgave 1 er der udarbejdet 5 notater med tilhørende bilag:

- Notat 1. Algoritmen i testsystemet og beregning af elevdygtigheden
 - Bilag 1.1. Anvendte skalaer til præsentation af elevernes beregnede dygtigheder
 - Bilag 1.2. Opgavebanken i dansk læsning 8. klasse - sprogforståelse
- Notat 2. De nationale tests måleegenskaber
 - Bilag 2.1. Sammenhæng mellem testresultater og karakterer
- Notat 3. Den statistiske usikkerhed og testenes reliabilitet
 - Bilag 3.1. Statistisk usikkerhed på elevdygtighederne
 - Bilag 3.2. Reliabilitet
- Notat 4. Opgavebanken og opgavernes sværhedsgrad
 - Bilag 4.1. Opgaveafprøvningsperioder
 - Bilag 4.2. Skærmdumps fra RUMM
 - Bilag 4.3. Opgavebankens sammensætning i forhold til opgavernes sværhedsgrad
 - Bilag 4.4. Sammenhæng mellem elevernes dygtighed og opgavernes sværhedsgrad
 - Bilag 4.5. Undersøgelse af link-opgavernes ændrede sværhedsgrad
 - Bilag 4.6. Forskel i opgavernes sværhedsgrad
- Notat 5. Samling af testresultater fra flere profilområder

De 5 notater inklusiv bilag er samlet i denne rapport.

1 Notat 1 Algoritmen i testsystemet og beregning af elevdygtigheden

Sammenfatning

- Den adaptive algoritme i testsystemet fungerer efter hensigten både i forhold til valg af opgaver fra opgavebanken og i forhold til beregning af elevdygtigheden og den statistiske usikkerhed
- I testsystemet vælges opgaverne til elevernes testforløb, så de passer til elevens dygtighed. Eksempler illustrerer, hvordan algoritmen virker
- De beregnede elevdygtigheder og tilhørende statistiske usikkerheder i testsystemet i dansk læsning 8. klasse og matematik 6. klasse fra de obligatoriske nationale test i 2018 er sammenlignet med tilsvarende beregnede elevdygtigheder og usikkerheder ved anvendelse af det kommercielle software program RUMM
- Der er ingen statistisk signifikant forskel mellem elevdygtighederne beregnet i testsystemet og beregnet i RUMM. Den gennemsnitlige forskel er på 0,02 logit
- Beregningerne viser overensstemmelse indenfor $\pm 0,1$ logit mellem elevdygtighederne beregnet i testsystemet og i RUMM for over 99,2 % af alle elevforløb og overensstemmelse indenfor $\pm 0,2$ logit mellem elevdygtighederne beregnet i testsystemet og i RUMM for over 99,7 % af alle elevforløb

Indledning

I evalueringen af de nationale test skal følgende spørgsmål besvares:

Regner de nationale test rigtigt?

For at svare på spørgsmålet er det først og fremmest vigtigt at vurdere, om algoritmen i testsystemet fungerer efter hensigten.

I notatet beskrives, hvordan opgaverne fra opgavebanken vælges til elevernes testforløb, og hvordan elevernes dygtighed beregnes. De elevdygtigheder, der beregnes i testsystemet, sammenholdes med elevdygtighederne beregnet i et kommercielt software program.

Formålet er således, at vurdere om elevernes dygtighed beregnes rigtigt i testsystemet.

Notatet indeholder følgende afsnit:

- Rasch-modellen og beregning af elevdygtigheden
- Valg af opgaver i den adaptive algoritme
- Sammenligning af beregnet elevdygtighed og statistisk usikkerhed i test-system med tilsvarende beregnet i kommercielt software

Rasch-modellen og beregning af elevdygtighed

De nationale test er baseret på Rasch-modellen^{3,4}. Rasch-modellen er en sandsynlighedsmodel, der i den simpleste udgave, kaldet det dikotome tilfælde, giver sandsynligheden for, at en elev nummer n med dygtighedsparameteren θ_n svarer rigtigt (svarende til scoringen $X_{ni}=1$) på item⁵ nummer i med item sværhedsparameteren β_i :

$$P\{X_{ni} = x\} = \frac{e^{x(\theta_n - \beta_i)}}{1 + e^{(\theta_n - \beta_i)}}$$

Sandsynligheden for, at en elev svarer rigtigt på et item, afhænger således kun af elevens dygtighed θ (theta) og opgavens sværhed β (beta).

I Rasch-modellen optræder itemsværheder og elevdygtigheder på samme skala. Skalaen kaldes logit-skalaen⁶. Anvendte skalaer i de nationale test ses i Bilag 1.1.

³ Rasch, G.: Probabilistic Models for Some Intelligence and Attainment Tests. Danish National Institute for Educational Research, Copenhagen 1960.

⁴ Karl Bang Christensen, Svend Kreiner, Mounir Mesbah (edt): Rasch Models in Health. Wiley 2013.

⁵ I Rasch-modellen anvendes begrebet item i stedet for opgave.

⁶ Logits er en transformation med den naturlige logaritme af odds, $p/(1-p)$, hvor p er sandsynligheden for at svare rigtigt på et item.

Rasch-modellen kan udvides til at inkludere items med flere subitems, således at scoringen x kan antage højere heltalsværdier end 1 svarende til, at flere subitems er besvaret korrekt. Items med flere subitems kaldes polytome items.

Det polytome tilfælde resulterer i en lidt mere kompliceret matematisk model med flere parametre, de såkaldte tærskelværdier τ , og en størrelse m_i , der angiver den maksimale scoring for det polytome item:

$$P\{X_{ni} = x\} = \frac{e^{-\tau_{1i}-\tau_{2i}\dots-\tau_{xi}+x(\theta_n-\beta_i)}}{\sum_{z=0}^{m_i} e^{-\tau_{1i}-\tau_{2i}\dots-\tau_{zi}+z(\theta_n-\beta_i)}} \quad \{1\}$$

Det ses, at det dikotome tilfælde er indeholdt i modellen med $m_i=1$ og $\tau_{1i}=0$.

I de nationale test anvendes en blanding af dikotome og polytome items.

Tærskelværdierne for de enkelte items i opgavebanken er beregnet på baggrund af opgaveafprøvningerne.

Opgavernes sværhedsgrad, location, beregnes som gennemsnittet af de ikke-centraliserede tærskelværdier.

Ved at anvende de estimerede tærskelværdier for opgaverne i opgavebanken kan elevernes dygtighed beregnes. Elevernes dygtighed beregnes som maximum likelihood estimatet af person parameteren, θ . Tilsvarende kan den statistiske usikkerhed, Standard Error of Measurement (SEM), på elevdygtigheden beregnes.

I testsystemets algoritme håndteres elevforløb uden for intervallet ± 7 logit ikke. Dvs. alle elevdygtigheder under -7 logit sættes lig med -7 og alle elevdygtigheder over $+7$ logit sættes lig med $+7$.

Enkelte elevforløb er såkaldte ekstreme forløb. Et ekstremt forløb i denne sammenhæng er, når eleven enten svarer forkert eller rigtigt på alle opgaver. Dette giver problemer i beregningerne, da værdier, der anvendes, kan nærme sig uendeligt. Beregningen af elevdygtigheden i ekstreme forløb håndteres ved at ændre elevens samlede score med $+0,22$ ved ekstremt lav score og $-0,22$ ved ekstremt høj score.

Selve algoritmen til valg af opgaver fra opgavebanken og beregning af elevernes dygtighed i testsystemet er programmeret i C#, og beregningen af elevdygtigheden foregår iterativt ved anvendelse af bisektion metoden.

I testsystemets algoritme skelnes mellem 'run-in perioden' og 'perioden efter run-in'.

Run-in perioden

Run-in perioden defineres ved de tre første opgaver i hvert profilområde. Eleven starter med en dygtighed, der passer til en middelsvær opgave i det

pågældende profilområde. Startværdien på logit-skalaen for elevens dygtighed afhænger således af det enkelte profilområde.

Svarer eleven forkert, nedsættes elevens dygtighed med en fast step-værdi. Svarer eleven rigtigt, forhøjes elevens dygtighed med step-værdien. Step-værdien afhænger af spredningen på opgavernes sværhedsgrad i det enkelte profilområde. Step-værdien er på 0,25 eller 0,50 logit afhængigt af profilområde.

Efter elevens besvarelse på anden opgave gentages ovenstående fastsættelse af elevdygtigheden.

Efter besvarelse af tredje opgave beregnes elevens dygtighed for første gang – baseret på elevens besvarelser på de tre opgaver.

Efter run-in perioden

Elevens dygtighed beregnes efter hver besvarelse efter run-in perioden ifølge modellen {1}. Beregningen baseres på elevens besvarelser af alle foregående opgaver samt på opgavernes sværhedsgrad.

Valg af opgaver i den adaptive algoritme

Valg af opgaver fra opgavebanken til elevernes testforløb foregår efter det adaptive princip. Hvis eleven svarer rigtigt på en opgave, da vil næste opgave være med en større sværhedsgrad. Hvis eleven svarer forkert, vil næste opgave være med en mindre sværhedsgrad.

Valg af den næste opgave i en elevs testforløb i de nationale test foregår ved, at vælge den opgave, der har en sværhedsgrad, der er lig med eller ligger så tæt som muligt på elevens dygtighed. Derved opnås den højeste statistiske informationsværdi og den mindst mulige statistiske usikkerhed på beregningen af elevens dygtighed.

I den adaptive algoritme søges efter opgaver i et interval omkring elevens aktuelt beregnede dygtighed i følgende rækkefølge:

1. Elevens beregnede dygtighed $\pm 0,1$ logit
2. Elevens beregnede dygtighed $\pm 0,2$ logit
3. Elevens beregnede dygtighed $\pm 0,5$ logit
4. Elevens beregnede dygtighed $\pm 1,0$ logit
5. Elevens beregnede dygtighed ligger tættest muligt på opgavens sværhedsgrad

Blandt de opgaver, der ligger i intervallet, vælges den aktuelle opgave simpelt tilfældigt. Der kan ikke vælges opgaver, eleven allerede har fået i samme testforløb.

Søgning efter den næste opgave starter således i et lille interval ($\pm 0,1$ logit) omkring elevens dygtighed. Findes ingen opgaver her, eller har eleven allerede fået alle de opgaver, der lå i intervallet, da udvides intervallet til $\pm 0,2$ logit. Findes heller ingen opgaver her, da udvides søgeintervallet til $\pm 0,5$ logit og endelig til $\pm 1,0$ logit. Kan der stadig ikke findes en opgave, som eleven ikke har besvaret tidligere i testforløbet, da vælges den opgave, der har en sværhedsgrad tættest på elevens dygtighed.

I testsystemet skiftes mellem opgaver i de tre profilområder, testen består af. Eleven får således stillet opgaverne i rækkefølgen 1, 2, 3, 1, 2, 3, 1, ..., hvor 1, 2 og 3 står for de tre profilområder. Når den statistiske usikkerhed på elevens dygtighed er kommet under 0,55, eller eleven har besvaret 29 opgaver i et profilområde fortsættes udelukkende med opgaver i de øvrige profilområder.

Eksempler på valg af opgave i testforløb

Nedenstående tre eksempler viser hvilke opgaver, der vælges for den enkelte elev i ét profilområde. Der er valgt tre eksempler fra de obligatoriske test i dansk læsning 8. klasse i foråret 2018. Alle tre eksempler stammer fra profilområdet sprogforståelse.

Algoritmens valg af opgave fungerer fuldstændigt ens for alle test og profilområder. Eksemplerne er således repræsentative for algoritmens valg af opgaver uanset test, klassetrin og profilområde.

Startværdien for elevens dygtighed i dansk læsning 8. klasse sprogforståelse er sat til 0,50 og step-værdien i run-in perioden er på 0,50.

Bilag 1.2 indeholder alle opgaver i opgavebanken i dansk læsning 8. klasse sprogforståelse med tilhørende sværhedsgrad. Opgaverne er sorteret efter stigende sværhedsgrad.

Nedenstående tabeller, der viser tre elevs testforløb, indeholder:

- **Svartidspunkt:** Dato og tid for elevens besvarelse af den pågældende opgave
- **Theta:** Elevens dygtighed på logit skalaen når opgavens stilles
- **Opgavenummer:** Opgavens id-nummer i opgavebanken
- **Location:** Opgavens sværhedsgrad på logit skalaen
- **Afstand:** Forskellen på logit skalaen mellem elevens dygtighed og opgavens sværhedsgrad
- **Score:** Elevens antal rigtige i opgaven
- **Ny theta:** Elevens beregnede dygtighed efter besvarelse af opgaven
- **SEM:** Standard Error of Measurement. Den statistiske usikkerhed på elevens beregnede dygtighed

Eksempel 1

Tabel 1.1 Elevforløb 1

Step	Svartidspunkt	Theta	Opgavenummer	Location	Afstand	Score	Ny theta	SEM
1	07MAR18:11:17:54	0,500	0108010420155	0,580	-0,080	1	1,000	-
2	07MAR18:11:19:24	1,000	0108010420129	1,085	-0,085	0	0,500	-
3	07MAR18:11:21:47	0,500	0108010420043	0,464	0,036	1	1,414	1,235
4	07MAR18:11:23:42	1,414	010801000301234804-1	1,505	-0,091	1	2,050	1,174
5	07MAR18:11:26:28	2,050	0108010415180	1,981	0,069	0	1,557	0,948
6	07MAR18:11:29:14	1,557	0108010420122	1,465	0,092	0	1,180	0,845
7	07MAR18:11:31:09	1,180	0108010410399	1,160	0,020	0	0,872	0,786
8	07MAR18:11:33:39	0,872	0108010420028	0,897	-0,025	1	1,140	0,726
9	07MAR18:11:36:52	1,140	0108010420017	1,235	-0,095	1	1,385	0,687
10	07MAR18:11:38:45	1,385	010801000301238997-1	1,445	-0,060	0	1,182	0,647
11	07MAR18:11:41:20	1,182	0108010415109	1,266	-0,084	0	1,000	0,618
12	07MAR18:11:45:13	1,000	0108010415118	1,095	-0,095	1	1,182	0,588
13	07MAR18:11:48:24	1,182	0108010420056	1,259	-0,077	1	1,347	0,566
14	07MAR18:11:49:55	1,347	0108010415153	1,381	-0,034	0	1,202	0,543
15	07MAR18:11:54:22	1,202	0108010420160	1,249	-0,047	1	1,343	0,525
16	07MAR18:11:56:54	1,343	010801000301239000-1	1,367	-0,024	1	1,474	0,511
17	07MAR18:11:59:32	1,474	010801000301234820-1	1,552	-0,078	0	1,357	0,493
18	07MAR18:12:01:19	1,357	0108010415151	1,370	-0,013	0	1,244	0,478

Kilde: Styrelsen for It og Læring

Step1. Run-in perioden. Eleven starter med en dygtighed på 0,5. Algoritmen søger først efter opgaver i opgavebanken i intervallet $0,5 \pm 0,1$, dvs. i intervallet 0,4-0,6. I intervallet findes 28 opgaver (Bilag 1.2), og algoritmen vælger en opgave tilfældig blandt disse. Den valgte opgave har en sværhedsgrad på 0,580. Eleven svarer rigtigt (score=1) og den nye dygtighed forhøjes til 1,0.

Step2. Run-in perioden. Eleven har nu en dygtighed på 1,0 og algoritmen søger først efter opgaver i intervallet 0,9-1,1. I intervallet findes 21 opgaver (Bilag 1.2), og algoritmen vælger en opgave tilfældig blandt disse. Den valgte opgave har en sværhedsgrad på 1,085. Eleven svarer forkert (score=0) og den nye dygtighed sættes til 0,5.

Step3. Run-in perioden. Eleven har nu en dygtighed på 0,5 og algoritmen søger først efter opgaver i intervallet 0,4-0,6. I intervallet findes 28 opgaver (Bilag 1.2). Eleven har tidligere fået én af disse, så der er 27 opgaver tilbage i intervallet, og algoritmen vælger en opgave tilfældig blandt disse. Den valgte opgave har en sværhedsgrad på 0,464. Eleven svarer rigtigt (score=1). Eleven har nu svaret på tre opgaver og elevens dygtighed kan beregnes for første gang. På baggrund af elevens besvarelser på de tre opgaver samt ved anvendelse af opgavernes estimerede tærskelværdier, beregnes elevens dygtighed til 1,414 og den statistiske usikkerhed beregnes til 1,235.

Step4. Eleven har nu en dygtighed på 1,414 og algoritmen søger først efter opgaver i intervallet 1,314-1,514. I intervallet findes 12 opgaver (Bilag 1.2), og algoritmen vælger en opgave tilfældig blandt disse. Den valgte opgave har en sværhedsgrad på 1,505. Eleven svarer rigtigt (score=1), og elevens dygtighed beregnes på baggrund af elevens besvarelser på de fire opgaver samt ved anvendelse af opgavernes estimerede tærskelværdier. Elevens dygtighed beregnes til 2,050 og den statistiske usikkerhed beregnes til 1,174.

Step5. Eleven har nu en dygtighed på 2,050 og algoritmen søger først efter opgaver i intervallet 1,950-2,150. I intervallet findes 1 opgave (Bilag 1.2), og algoritmen vælger denne. Opgaven har en sværhedsgrad på 1,981. Eleven svarer forkert (score=0), og elevens dygtighed beregnes på baggrund af elevens besvarelser på de fem opgaver samt ved anvendelse af opgavernes estimerede tærskelværdier. Elevens dygtighed beregnes til 1,557 og den statistiske usikkerhed beregnes til 0,948.

Step6-Step18. Eleven svarer i alt på 18 opgaver i profilområdet, og efter hver besvarelse vælges en opgave fra opgavebanken med en sværhedsgrad, der ikke ligger mere end $\pm 0,1$ logit fra elevens dygtighed. Dette ses af kolonnen 'Afstand' i Tabel 1.1.

Step1-Step18: Alle 18 opgaver findes i intervallet $\pm 0,1$ logit fra elevens dygtighed.

Eksempel 2

Tabel 1.2 Elevforløb 2

Step	Svartidspunkt	Theta	Opgavenummer	Location	Afstand	Score	Ny heta	SEM
1	24APR18:06:04:43	0,500	010801000301238555-1	0,479	0,021	1	1,000	-
2	24APR18:06:08:44	1,000	0108010410384	0,931	0,069	1	1,500	-
3	24APR18:06:11:57	1,500	010801000301234820-1	1,552	-0,052	1	3,606	2,229
4	24APR18:06:14:17	3,606	0108010410315	3,044	0,562	1	4,791	2,259
5	24APR18:06:18:54	4,791	010801000301239196-1	2,371	2,420	0	3,313	1,193
6	24APR18:06:23:18	3,313	010801000301234837-1	2,263	1,050	0	2,594	0,932
7	24APR18:06:28:25	2,594	0108010420015	2,151	0,443	1	2,887	0,885
8	24APR18:06:29:42	2,887	010801000301234841-1	2,248	0,639	0	2,476	0,774
9	24APR18:06:32:22	2,476	0108010415180	1,981	0,495	0	2,158	0,709
10	24APR18:06:34:32	2,158	010801000301234833-1	1,937	0,221	1	2,359	0,677
11	24APR18:06:35:50	2,359	010801000301234828-1	1,912	0,447	0	2,111	0,634
12	24APR18:06:39:07	2,111	0108010410094	1,832	0,279	0	1,904	0,603
13	24APR18:06:42:43	1,904	0108010410379	1,807	0,097	1	2,063	0,579
14	24APR18:06:43:39	2,063	0108010420095	1,716	0,347	1	2,192	0,560
15	24APR18:06:45:27	2,192	0108010415124	1,713	0,479	1	2,304	0,545

Kilde: Styrelsen for It og Læring

Step1. Run-in perioden. Eleven starter med en dygtighed på 0,5. Algoritmen søger først efter opgaver i opgavebanken i intervallet $0,5 \pm 0,1$, dvs. i intervallet 0,4-0,6. I intervallet findes 28 opgaver (Bilag 1.2), og algoritmen vælger en opgave tilfældig blandt disse. Den valgte opgave har en sværhedsgrad på 0,479. Eleven svarer rigtigt (score=1) og den nye dygtighed forhøjes til 1,0.

Step2. Run-in perioden. Eleven har nu en dygtighed på 1,0 og algoritmen søger først efter opgaver i intervallet 0,9-1,1. I intervallet findes 21 opgaver (Bilag 1.2), og algoritmen vælger en opgave tilfældig blandt disse. Den valgte opgave har en sværhedsgrad på 0,931. Eleven svarer rigtigt (score=1) og den nye dygtighed sættes til 1,5.

Step3. Run-in perioden. Eleven har nu en dygtighed på 1,5 og algoritmen søger først efter opgaver i intervallet 1,4-1,6. I intervallet findes 11 opgaver (Bilag 1.2), og algoritmen vælger en opgave tilfældig blandt disse. Den valgte opgave har en sværhedsgrad på 1,552. Eleven svarer rigtigt (score=1). Eleven har nu svaret på tre opgaver og elevens dygtighed kan beregnes for første gang. På baggrund af elevens besvarelser på de tre opgaver samt ved anvendelse af opgavernes estimerede tærskelværdier, beregnes elevens dygtighed til 3,606 og den statistiske usikkerhed beregnes.

Step4. Eleven har nu en dygtighed på 3,606 og algoritmen søger først efter opgaver i intervallet 3,506-3,706. Der findes ingen opgaver i opgavebanken i dette interval. Dernæst søges i intervallet $\pm 0,2$, dvs. i intervallet 3,406-3,806. Der findes heller ingen opgaver i opgavebanken i dette interval. Dernæst søges i intervallet $\pm 0,5$, dvs. i intervallet 3,106-4,106. Der findes heller ingen

opgaver i opgavebanken i dette interval. Dernæst søges i intervallet $\pm 1,0$, dvs. i intervallet 2,606-4,606. Der findes én opgave i dette interval med en sværhedsgrad på 3,044 (Bilag 1.2) som vælges. Eleven svarer rigtigt (score=1), og elevens dygtighed beregnes på baggrund af elevens besvarelser på de fire opgaver samt ved anvendelse af opgavernes estimerede tærskelværdier. Elevens dygtighed beregnes til 4,791.

Step5. Eleven har nu en dygtighed på 4,791 og algoritmen søger først efter opgaver i intervallet $\pm 0,1$, dernæst i intervallet $\pm 0,2$ og i intervallerne $\pm 0,5$ og $\pm 1,0$. Opgavebanken indeholder ingen opgaver i disse intervaller. Derfor vælges den opgave med en sværhedsgrad, der ligger tættest på elevens dygtighed, og som eleven ikke tidligere har fået. Denne opgave har en sværhedsgrad på 2,371. Eleven svarer forkert (score=0), og elevens dygtighed beregnes på baggrund af elevens besvarelser på de fem opgaver samt ved anvendelse af opgavernes estimerede tærskelværdier. Elevens dygtighed beregnes til 3,313.

Step6. Eleven har nu en dygtighed på 3,313 og algoritmen søger først efter opgaver i intervallet $\pm 0,1$ og i intervallerne $\pm 0,2$, $\pm 0,5$ og $\pm 1,0$. Opgavebanken indeholder ingen opgaver i disse intervaller, som eleven ikke allerede har fået. Derfor vælges den opgave med en sværhedsgrad, der ligger tættest på elevens dygtighed, og som eleven ikke tidligere har fået. Denne opgave har en sværhedsgrad på 2,263. Eleven svarer forkert (score=0), og elevens dygtighed beregnes på baggrund af elevens besvarelser på de seks opgaver samt ved anvendelse af opgavernes estimerede tærskelværdier. Elevens dygtighed beregnes til 2,594.

Step7. Eleven har nu en dygtighed på 2,594 og algoritmen søger først efter opgaver i intervallet $\pm 0,1$ og i intervallet $\pm 0,2$. Opgavebanken indeholder ingen opgaver i disse intervaller, som eleven ikke allerede har fået. Dernæst søges i intervallet $\pm 0,5$, dvs. i intervallet 2,094-3,094. I intervallet findes 5 opgaver (Bilag 1.2). Eleven har tidligere fået tre af disse, så der er 2 opgaver tilbage i intervallet, og algoritmen vælger en opgave tilfældig blandt disse. Den valgte opgave har en sværhedsgrad på 2,151. Eleven svarer rigtigt (score=1), og elevens dygtighed beregnes på baggrund af elevens besvarelser på de syv opgaver samt ved anvendelse af opgavernes estimerede tærskelværdier. Elevens dygtighed beregnes til 2,887.

Step8-Step15. Eleven svarer i alt på 15 opgaver i profilområdet. Forskellen mellem elevens dygtighed og den valgt opgaves sværhedsgrad ses af kolonnen 'Afstand' i Tabel 1.2.

Step1-Step15: Fire opgaver findes i intervallet $\pm 0,1$ logit fra elevens dygtighed, syv opgaver skal findes i intervallet $\pm 0,5$ logit, to opgaver skal findes i intervallet $\pm 1,0$ logit, mens to af opgaverne ligger med en sværhedsgrad mere end $\pm 1,0$ logit fra elevens dygtighed.

Eksempel 3

Tabel 1.3 Elevforløb 3

Step	Svartidspunkt	Theta	Opgavenummer	Location	Afstand	Score	Ny theta	SEM
1	12APR18:07:00:26	0,500	0108010420043	0,464	0,036	1	1,000	-
2	12APR18:07:04:10	1,000	0108010410084	1,081	-0,081	1	1,500	-
3	12APR18:07:09:16	1,500	010801000301234804-1	1,505	-0,005	0	1,741	1,248
4	12APR18:07:10:30	1,741	0108010410094	1,832	-0,091	1	2,384	1,180
5	12APR18:07:12:30	2,384	010801000301239196-1	2,371	0,013	1	2,959	1,152
6	12APR18:07:16:22	2,959	0108010410315	3,044	-0,085	1	3,550	1,148
7	12APR18:07:22:00	3,550	010801000301234837-1	2,263	1,287	0	2,842	0,885
8	12APR18:07:24:07	2,842	010801000301234841-1	2,248	0,594	1	3,089	0,853
9	12APR18:07:25:44	3,089	0108010420015	2,151	0,938	1	3,275	0,831
10	12APR18:07:28:41	3,275	0108010415180	1,981	1,294	0	2,835	0,718
11	12APR18:07:30:48	2,835	010801000301234833-1	1,937	0,898	1	2,973	0,700
12	12APR18:07:35:29	2,973	010801000301234828-1	1,912	1,061	1	3,091	0,686
13	12APR18:07:39:59	3,091	0108010410379	1,807	1,284	0	2,772	0,620
14	12APR18:07:43:02	2,772	0108010420095	1,716	1,056	1	2,866	0,609
15	12APR18:07:47:25	2,866	0108010415124	1,713	1,153	0	2,614	0,564
16	12APR18:07:50:20	2,614	010801000301234824-1	1,685	0,929	1	2,700	0,554
17	12APR18:07:52:16	2,700	010801000301236068-1	1,649	1,051	1	2,776	0,546

Kilde: Styrelsen for It og Læring

Step1. Run-in perioden. Eleven starter med en dygtighed på 0,5. Algoritmen søger først efter opgaver i opgavebanken i intervallet $0,5 \pm 0,1$, dvs. i intervallet 0,4-0,6. I intervallet findes 28 opgaver (Bilag 1.2), og algoritmen vælger en opgave tilfældig blandt disse. Den valgte opgave har en sværhedsgrad på 0,464. Eleven svarer rigtigt (score=1) og den nye dygtighed forhøjes til 1,0.

Step2. Run-in perioden. Eleven har nu en dygtighed på 1,0 og algoritmen søger først efter opgaver i intervallet 0,9-1,1. I intervallet findes 21 opgaver (Bilag 1.2), og algoritmen vælger en opgave tilfældig blandt disse. Den valgte opgave har en sværhedsgrad på 1,081. Eleven svarer rigtigt (score=1) og den nye dygtighed sættes til 1,5.

Step3. Run-in perioden. Eleven har nu en dygtighed på 1,5 og algoritmen søger først efter opgaver i intervallet 1,4-1,6. I intervallet findes 11 opgaver (Bilag 1.2), og algoritmen vælger en opgave tilfældig blandt disse. Den valgte opgave har en sværhedsgrad på 1,505. Eleven svarer forkert (score=0). Eleven har nu svaret på tre opgaver og elevens dygtighed kan beregnes for første gang. På baggrund af elevens besvarelser på de tre opgaver samt ved anvendelse af opgavernes estimerede tærskelværdier, beregnes elevens dygtighed til 1,741 og den statistiske usikkerhed beregnes til 1,248.

Step4. Eleven har nu en dygtighed på 1,741 og algoritmen søger først efter opgaver i intervallet 1,641-1,841. I intervallet findes 6 opgaver (Bilag 1.2), og algoritmen vælger en opgave tilfældig blandt disse. Den valgte opgave har en

sværhedsgrad på 1,832. Eleven svarer rigtigt (score=1), og elevens dygtighed beregnes på baggrund af elevens besvarelser på de fire opgaver samt ved anvendelse af opgavernes estimerede tærskelværdier. Elevens dygtighed beregnes til 2,384.

Step5. Eleven har nu en dygtighed på 2,384 og algoritmen søger først efter opgaver i intervallet 2,284-2,584. I intervallet findes 1 opgave (Bilag 1.2), og algoritmen vælger denne. Opgaven har en sværhedsgrad på 2,371. Eleven svarer rigtigt (score=1), og elevens dygtighed beregnes på baggrund af elevens besvarelser på de fem opgaver samt ved anvendelse af opgavernes estimerede tærskelværdier. Elevens dygtighed beregnes til 2,959.

Step6-Step17. Eleven svarer i alt på 17 opgaver i profilområdet. Forskellen mellem elevens dygtighed og den valgt opgaves sværhedsgrad ses af kolonnen 'Afstand' i Tabel 1.3.

Step1-Step17: Seks opgaver findes i intervallet $\pm 0,1$ logit fra elevens dygtighed, fire opgaver skal findes i intervallet $\pm 1,0$ logit, mens syv af opgaverne ligger med en sværhedsgrad mere end $\pm 1,0$ logit fra elevens dygtighed.

Tjek af beregnet dygtighed og statistisk usikkerhed

For at undersøge om testsystemet beregner elevernes dygtighed og den statistiske usikkerhed (SEM) korrekt, er elevdygtighederne og SEM kontrolberegnet i et kommercielt softwareprogram. Kontrolberegningerne er foretaget i softwareprogrammet RUMM⁷, der er udviklet på University of Western Australia, Perth.

Elevdygtighederne i RUMM er beregnet ved anvendelse af Weighted Maximum Likelihood. I modsætning til testsystemet kan de beregnede elevdygtigheder i RUMM falde udenfor intervallet ± 7 .

Sammenligningen af de beregnede elevdygtigheder og SEM er foretaget for dansk læsning 8. klasse og matematik 6. klasse. Der er anvendt elevbesvarelser fra de obligatoriske test i foråret 2018.

Den absolutte forskel i beregnet elevdygtighed mellem testsystemet og RUMM er:

$$\text{absolut forskel} = \text{abs}(\text{theta}(\text{DNT}) - \text{theta}(\text{RUMM})),$$

hvor $\text{theta}(\text{DNT})$ er elevdygtigheden beregnet i testsystemet og $\text{theta}(\text{RUMM})$ er elevdygtigheden beregnet ved anvendelse af RUMM.

⁷ www.rummlab.com.au

Tabel 1.4 Forskel i estimerede elevdygtigheder mellem testsystemet og RUMM. Gennemsnit (SD)

Fag og klassetrin	Profilområde	Absolut forskel (logit)	
		Alle elever	Ekskl. ekstreme ¹⁾
Dansk læsning 8. klasse	Sprogforståelse	0,02 (0,06)	0,02 (0,03)
	Afkodning	0,02 (0,05)	0,02 (0,02)
	Tekstforståelse	0,02 (0,03)	0,02 (0,02)
Matematik 6. klasse	Tal og algebra	0,03 (0,04)	0,03 (0,03)
	Geometri	0,03 (0,03)	0,03 (0,03)
	Statistik og sandsynlighed	0,02 (0,05)	0,02 (0,03)
I alt		0,02 (0,04)	0,02 (0,03)

1) Ekskl. elever med ekstremt testforløb

Kilde: Styrelsen for It og Læring

Den gennemsnitlige absolutte forskel mellem elevdygtighederne beregnet i testsystemet og beregnet i RUMM er på 0,02-0,03 logit afhængig af profilområde (Tabel 1.4). Elevdygtighederne beregnet i testsystemet er ikke statistisk signifikant forskellig fra elevdygtighederne i RUMM.

Tabel 1.5 viser, at 99,2 % af alle elevdygtighederne beregnes på samme måde (indenfor $\pm 0,1$) i testsystemet som i RUMM. Ses bort fra elever med ekstremt elevforløb, da beregnes 99,3 % af alle elevdygtighederne på samme måde i testsystemet som i RUMM.

Tabel 1.5 Fordeling af absolut forskel i estimerede elevdygtigheder mellem testsystemet og RUMM. Samlet for dansk læsning 8. klasse og matematik 6. klasse 2018

Interval ¹⁾	Alle elever	Ekskl. elever med ekstremt forløb
0,0 – 0,1	99,2 %	99,3 %
0,0 – 0,2	99,7 %	99,8 %
0,0 – 0,3	99,8 %	99,9 %
0,0 – 0,4	99,9 %	99,99 %
0,0 – 0,5	99,9 %	99,999 %
0,5 +	<0,1 %	<0,001 %

1) Forskel på logit-skalaen

Kilde: Styrelsen for It og Læring

De få og små afvigelser, der er mellem de beregnede elevdygtigheder, kan sandsynligvis tilskrives forskel i den anvendte præcision. Desuden kan forskellen ligge i forskellen på estimations metoden, dvs forskellen mellem Weighted Maximum Likelihood og Unweighted Maximum Likelihood.

Figur 1.1 og 1.2 viser den store overensstemmelse, der er, i beregningen af elevdygtighederne i testsystemet og i RUMM. Figurerne viser de beregnede

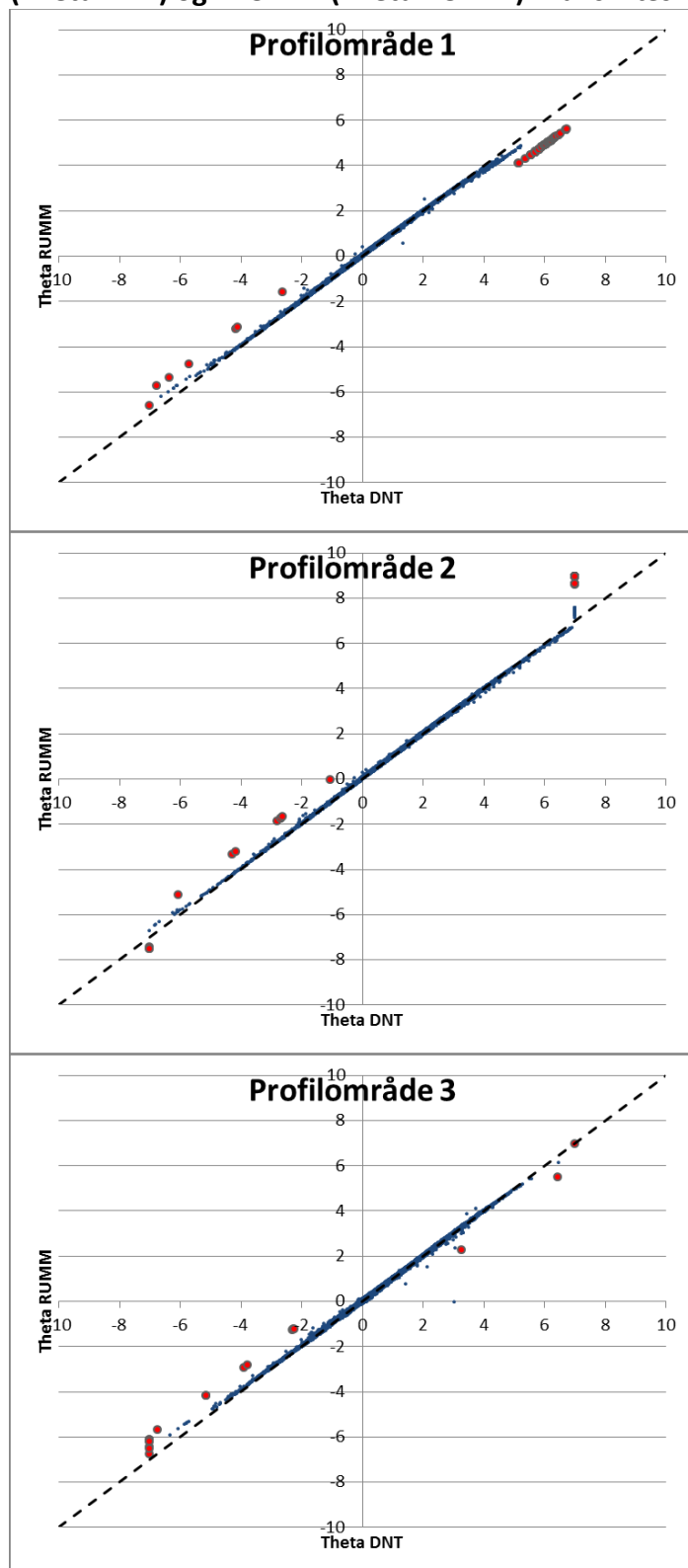
elevdygtigheder i hvert af de tre profilområder i henholdsvis dansk læsning 8. klasse og matematik 6. klasse.

De røde punkter viser beregnede elevdygtigheder for elever med ekstreme forløb.

Figur 1.3 og 1.4 viser tilsvarende fin overensstemmelse mellem beregningerne af SEM i testsystemet og i RUMM. De elevforløb, hvor der er en afvigelse i den beregnede SEM, er primært elevforløb med en SEM over 0,55. Her gælder, at den beregnede statistiske usikkerhed i testsystemet er større end den beregnede usikkerhed i RUMM.

Kontrolberegningerne i RUMM giver således samme elevdygtigheder og SEM som beregnes i testsystemet.

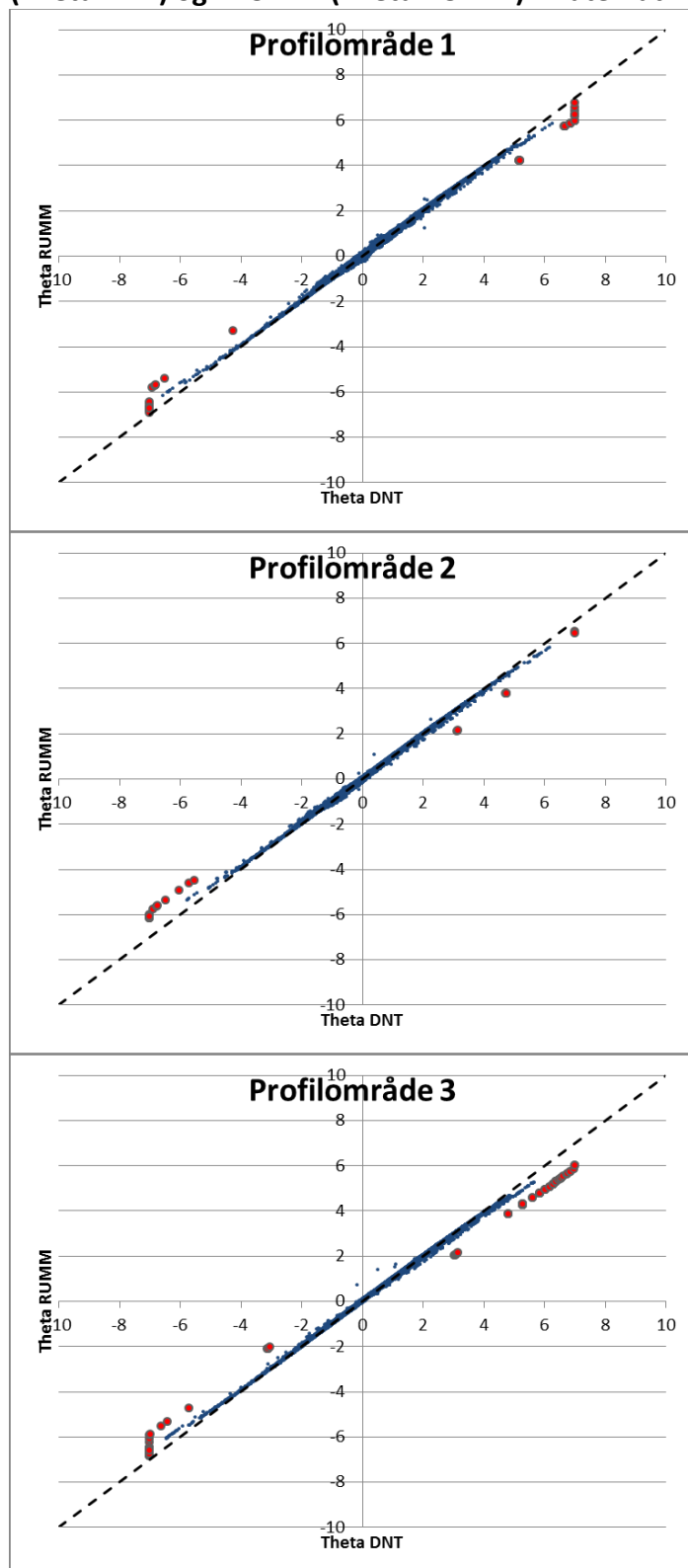
Figur 1.1 Sammenhæng mellem beregnet elevdygtighed i testsystemet (Theta DNT) og i RUMM (Theta RUMM). Dansk læsning 8. klasse



Note: Røde dots er elever med ekstreme besvarelser

Kilde: Styrelsen for It og Læring

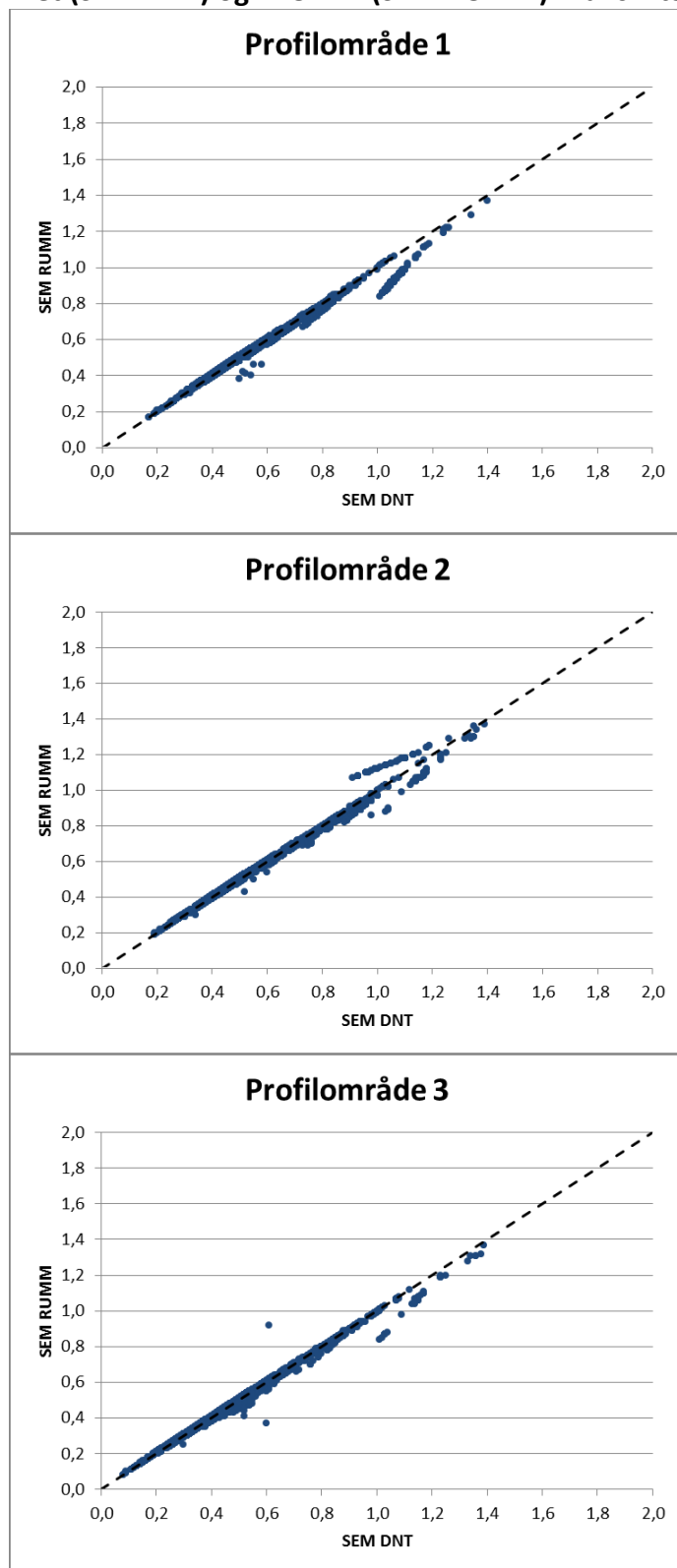
Figur 1.2 Sammenhæng mellem beregnet elevdygtighed i testsystemet (Theta DNT) og i RUMM (Theta RUMM). Matematik 6. klasse



Note: Røde dots er elever med ekstreme besvarelser

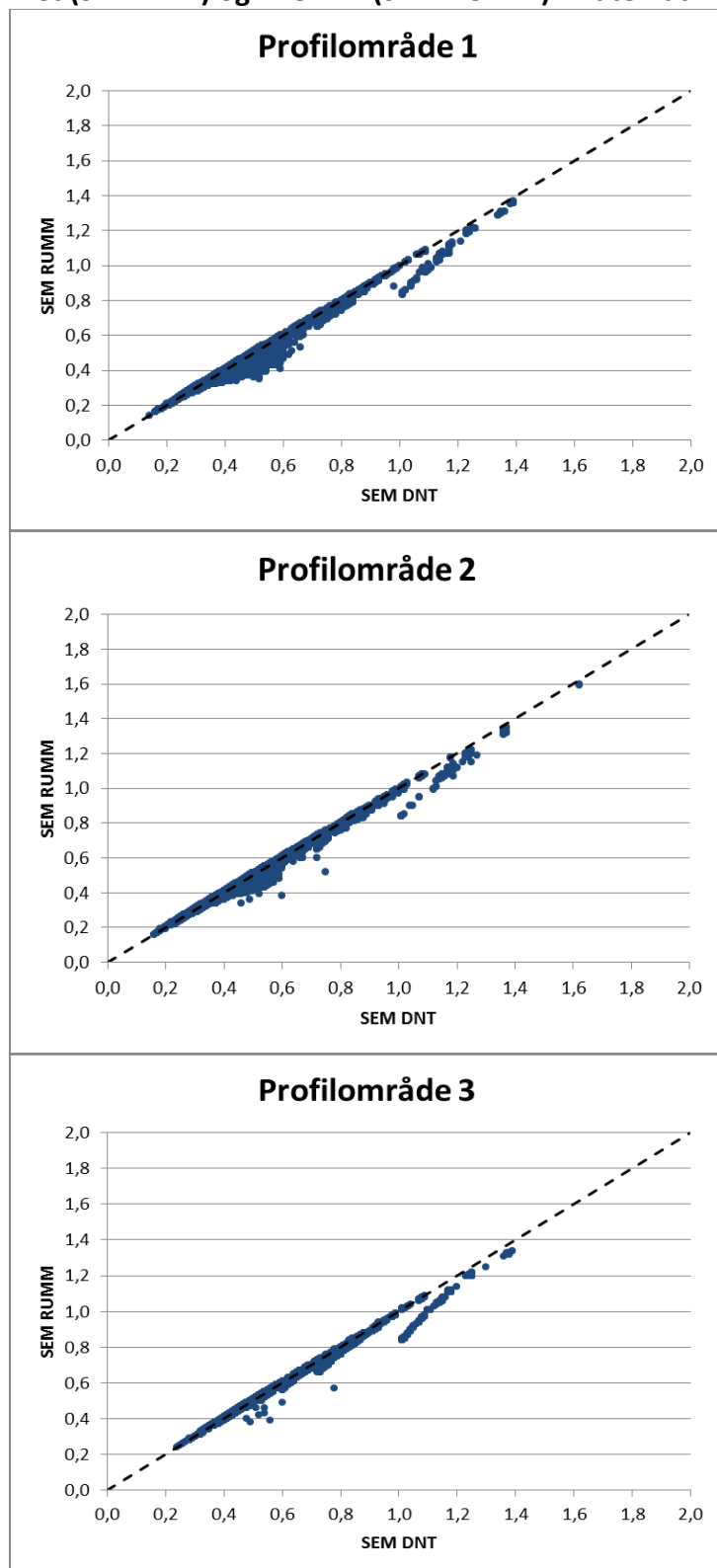
Kilde: Styrelsen for It og Læring

Figur 1.3 Sammenhæng mellem beregnet statistisk usikkerhed i testsystemet (SEM DNT) og i RUMM (SEM RUMM). Dansk læsning 8. klasse



Kilde: Styrelsen for It og Læring

Figur 1.4 Sammenhæng mellem beregnet statistisk usikkerhed i testsystemet (SEM DNT) og i RUMM (SEM RUMM). Matematik 6. klasse



Kilde: Styrelsen for It og Læring

2 Notat 2 De nationale tests måleegenskaber

Sammenfatning

- Der er en tydelig sammenhæng mellem elevernes resultater fra de nationale test i dansk læsning og matematik og elevernes karakterer i såvel standpunktsprøverne i 8. klasse som i folkeskolens prøver i 9. klasse
- De nationale test og folkeskolens prøver når til relativt enslydende vurderinger af elevernes faglige niveau i de områder, hvor der testes
- Der er en sammenhæng mellem elevernes resultater fra de nationale test og resultaterne fra PISA-undersøgelserne

Indledning

I evalueringen af de nationale test skal følgende spørgsmål besvares:

Regner de nationale test rigtigt?

Et element i vurderingen af dette er, at undersøge om elevernes resultater fra de nationale test stemmer overens med elevernes resultater fra andre tilsvarende test og prøver. En sådan egenskab omtales som testenes kriterievaliditet.

Hvis de nationale test beregner elevernes dygtighed forkert, må det forventes, at sammenhængen mellem elevernes beregnede dygtigheder i de nationale test og resultaterne fra andre test og elevvurderinger er begrænset.

For at få en indikation af om testene samlet set ser ud til at måle det samme som andre tilsvarende test og prøver, kan man se på sammenhængen mellem elevernes testresultat i de nationale test og deres efterfølgende præstation i de relevante dele af standpunktsprøverne i 8. klasse samt i folkeskolens prøver i 9. klasse. Endvidere er det muligt at se på sammenhængen mellem elevernes testresultater i de nationale test og elevernes senere PISA resultater. Begge dele er tidligere undersøgt af BUVVM^{8,9,10}.

I notatet beskrives sammenhængen mellem elevernes testresultater og henholdsvis standpunktskarakterer, prøvekarakterer og PISA resultater.

Formålet er således, at vurdere om de nationale test måler den egenskab, der forventes.

Notatet indeholder følgende afsnit:

- Sammenhæng mellem testresultater og prøvekarakterer
- Sammenhæng mellem testresultater og standpunktskarakterer
- Sammenhæng mellem testresultater og PISA test

Sammenhæng mellem testresultater og prøvekarakter

Elevernes karakter i dansk, læsning i folkeskolens prøve i 9. klasse i 2017/2018 sammenholdes med elevernes testresultater i de obligatoriske nationale test i dansk, læsning i 8. klasse i 2016/2017. Tilsvarende sammenholdes elevernes karakter i matematik uden hjælpemidler i folkeskolens

⁸ Børne- og undervisningsministeriet

⁹ <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf16/sep/160912-notat-om-nationale-tests-maaleegenskaber.pdf>

¹⁰ <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf17/jan/170110-kriteriebaserede-test-delrapport-1.pdf>

prøve i 9. klasse i 2017/2018 med elevernes testresultater i de obligatoriske nationale test i matematik i 6. klasse i 2014/2015¹¹ (Figur 2.1 og Figur 2.2).

I Rasch-modellen beregnes elevernes dygtighed på logit-skalaen¹². I testsystemet omregnes resultaterne til den normbaserede skala, der er en simpel percentilskala (1-100)¹³, ved anvendelse af en stykvis sigmoid funktion. I dansk læsning og i matematik omregnes resultaterne endvidere til en kriteriebaseret skala¹⁴ på seks trin¹⁵.

I figurer og tabeller er anvendt den samlede vurdering af elevens testresultat. På den normbaserede skala er den samlede vurdering lig med gennemsnittet af elevens testresultater fra de tre profilområder. På den kriteriebaserede skala er den samlede vurdering¹⁶ ligeledes baseret på elevens kriteriebaserede testresultater fra hvert af de tre profilområder.

¹¹ De nationale test i matematik 8. klasse startede først i 2017/2018

¹² Logits er en transformation med den naturlige logaritme af odds, $p/(1-p)$, hvor p er sandsynligheden for at svare rigtigt på et item.

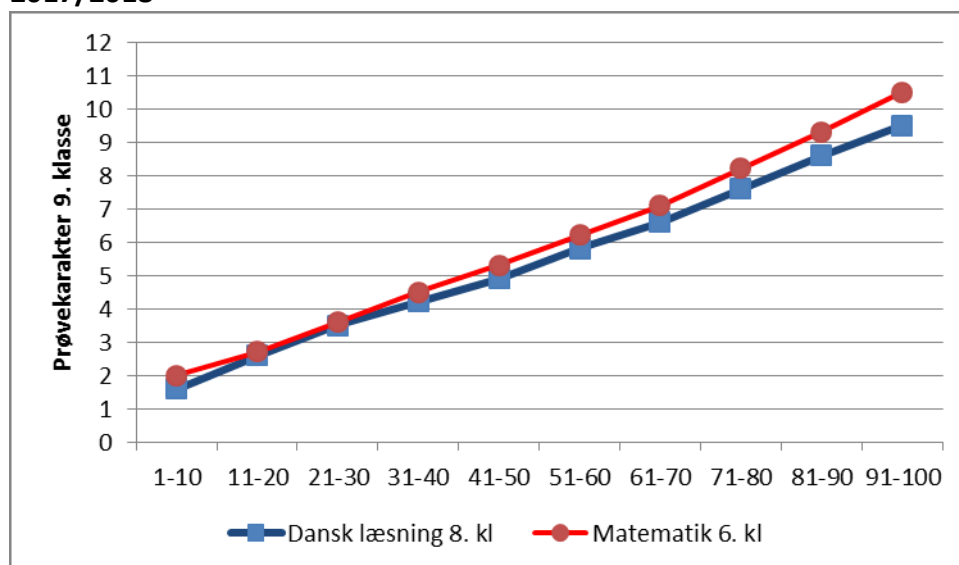
¹³ I formidlingen til elever og forældre bliver resultaterne på 100-skalaen omsat til: En del under gennemsnittet (1-10), under gennemsnittet (11-35), gennemsnittet (35-65), over gennemsnittet (66-90), en del over gennemsnittet (91-100)

¹⁴ <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf14/dec/141218-kriteriebaseret-scoring-den-statistiske-model-bag.pdf>

¹⁵ Ikke tilstrækkelig præstation, mangelfuld præstation, jævn præstation, god præstation, rigtig god præstation og fremragende præstation

¹⁶ Det samlede testresultat opfattes som: 1) Fremragende, hvis præstationen har været fremragende i mindst to profilområder og mindst god i det tredje. 2) Rigtig god, hvis præstationen har været mindst rigtig god i mindst to profilområder og mindst jævn i det tredje. 3) God, hvis præstationen har været mindst god i to profilområder og mindst mangelfuld i det tredje. 4) Jævn, hvis præstationen har været mindst jævn i to profilområder. 5) Mangelfuld, hvis præstationen har været mindst mangelfuld i to profilområder.

Figur 2.1 Sammenhængen mellem resultaterne¹⁾ fra de nationale test (normbaseret skala) og karakterer fra folkeskolens prøver i 9. klasse i 2017/2018



1) Samlet vurdering

Kilde: Styrelsen for It og Læring

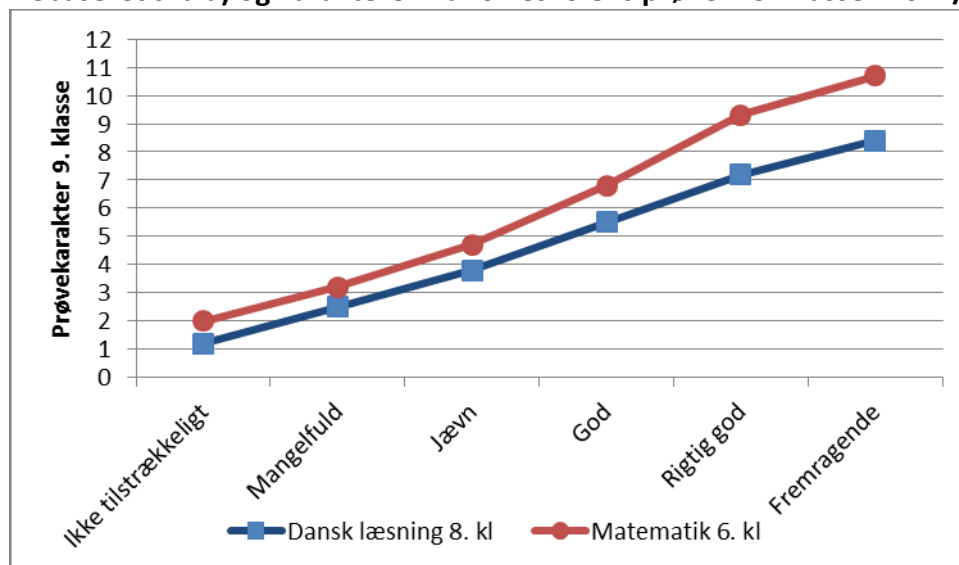
Gruppen af elever, der opnår mellem 31 og 40 point i samlet vurdering i de obligatoriske nationale test i dansk, læsning 8. klasse får i gennemsnit 4,2 i karakter ved folkeskolens prøver i 9. klasse, mens gruppen af elever, der opnår mellem 81 og 90 point i samlet vurdering i de obligatoriske nationale test i dansk, læsning 8. klasse får 8,6 i gennemsnit i karakter ved folkeskolens prøver i 9. klasse (Figur 2.1).

Den samme tydelige sammenhæng ses mellem elevernes testresultater i matematik i 6. klasse og deres karakter tre år senere ved folkeskolens prøve i matematik uden hjælpemidler i 9. klasse.

Karaktergennemsnit med tilhørende spredning fordelt efter elevernes dygtighed i de nationale test findes i Bilag 2.1.

Tilsvarende sammenhæng mellem elevernes testresultater i de nationale test og elevernes karakterer fra folkeskolens prøver i 9. klasse ses, når testresultaterne omregnes til den kriteriebaserede skala (Figur 2.2).

Figur 2.2 Sammenhængen mellem resultaterne¹⁾ fra de nationale test (kriteriebaseret skala) og karakterer fra folkeskolens prøver i 9. klasse i 2017/2018



1) Samlet vurdering

Kilde: Styrelsen for It og Læring

Elevernes fordeling på karakterne ved folkeskolens prøver i 9. klasse i forhold til elevernes testresultater på den kriteriebaserede skala i de nationale test ses i Tabel 2.1.

Blandt de elever, der opnår en mangelfuld præstation i dansk læsning i de nationale test i 8. klasse, får 64 procent højst karakteren 2 i dansk læsning året efter ved folkeskolens prøve i 9. klasse, mens 90 procent højst får karakteren 4.

Blandt de elever, der opnår en rigtig god præstation i dansk læsning i de nationale test i 8. klasse, får 67 procent karakteren 7 eller 10 i dansk læsning året efter ved folkeskolens prøve i 9. klasse.

Blandt de elever, der opnår en fremragende præstation i matematik i de nationale test i 6. klasse, får 87 procent karakteren 10 eller 12 i matematik uden hjælpemidler tre år efter ved folkeskolens prøve i 9. klasse.

Tabel 2.1 Elevernes testresultat¹⁾ i de nationale test sammenholdt med karakteren fra folkeskolens prøve i 9. klasse i 2017/2018. Andel elever (pct.)

Fag og klassetrin	Testresultat	Karakter							I alt
		-3	0	2	4	7	10	12	
Dansk læsning 8. klasse	Ikke tilstrækkelig	1	59	25	9	5	0	0	100
	Mangelfuld	0	27	37	26	9	1	0	100
	Jævn	0	11	29	37	21	2	0	100
	God	0	3	14	33	39	10	1	100
	Rigtig god	0	1	5	18	42	25	8	100
	Fremragende	0	1	4	12	31	31	22	100
Matematik 6. klasse	Ikke tilstrækkelig	0	46	25	19	8	1	0	100
	Mangelfuld	0	16	29	40	14	1	0	100
	Jævn	0	4	16	46	31	3	0	100
	God	0	1	4	23	51	18	4	100
	Rigtig god	0	0	0	4	30	42	24	100
	Fremragende	0	0	0	1	11	31	56	100

1) Samlet vurdering

Kilde: Styrelsen for It og Læring

De nationale test i dansk læsning og i matematik beregner således elevresultater, der ligger relativt tæt op ad elevernes karakterer i folkeskolernes prøve i 9. klasse.

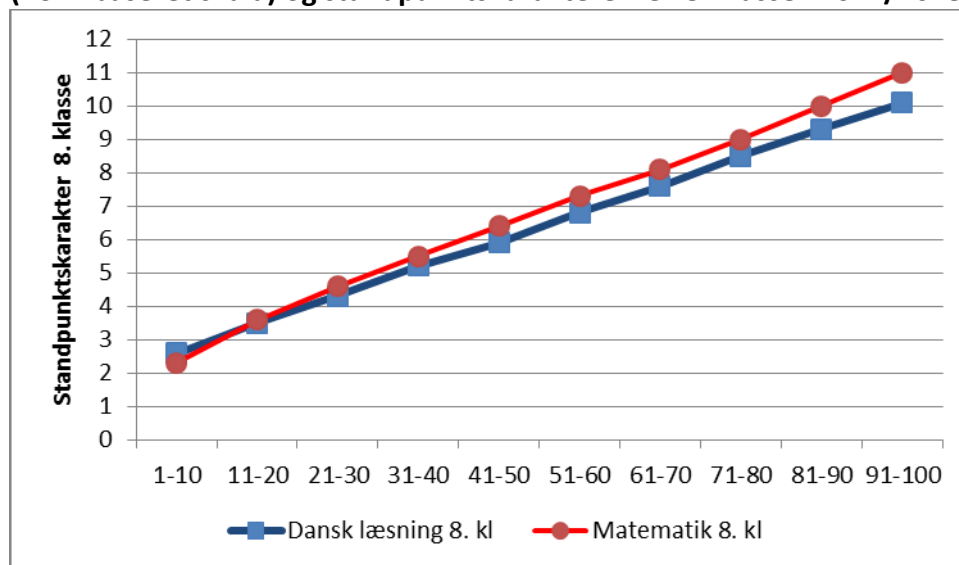
Den samme prædiktive validitet er flere forskere kommet frem til, blandt andet Louise Beuchert & Anne Nandrup fra Aarhus Universitet¹⁷.

Sammenhæng mellem testresultater og standpunktskarakter

Elevernes standpunktskarakter i dansk læsning 8. klasse i 2017/2018 sammenholdes med elevernes testresultater i de obligatoriske nationale test i dansk læsning i 8. klasse samme år. Tilsvarende sammenholdes elevernes standpunktskarakter i matematik uden hjælpemidler i 8. klasse i 2017/2018 med elevernes testresultater i de obligatoriske nationale test i matematik i 8. klasse samme år (Figur 2.3 og Figur 2.4).

¹⁷ Louise V. Beuchert & Anne B. Nandrup. The Danish National Tests at a Glance. Nationaløkonomisk Tidsskrift 2018:2

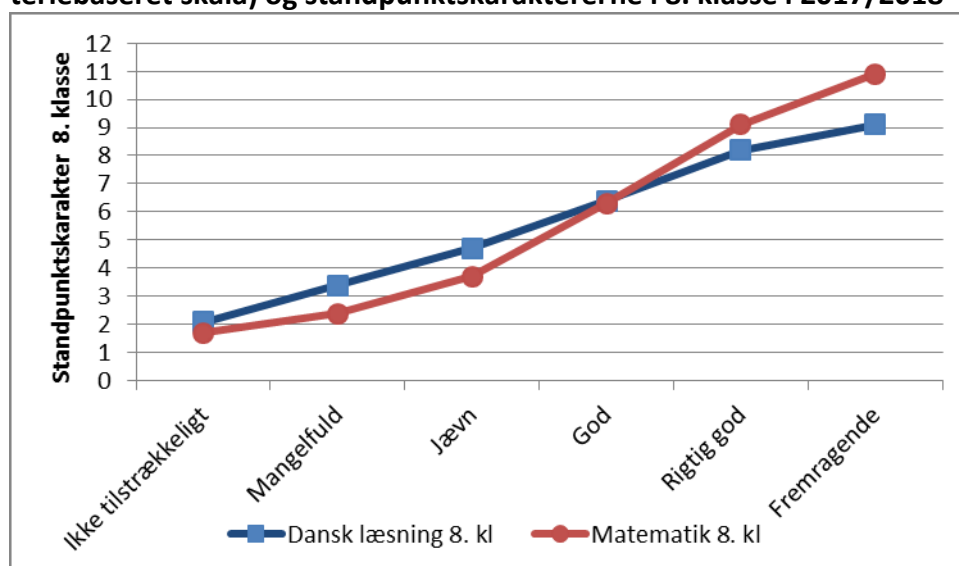
Figur 2.3 Sammenhængen mellem resultaterne¹⁾ fra de nationale test (normbaseret skala) og standpunktskaraktererne i 8. klasse i 2017/2018



1) Samlet vurdering

Kilde: Styrelsen for It og Læring

Figur 2.4 Sammenhængen mellem resultaterne¹⁾ fra de nationale test (kriteriebaseret skala) og standpunktskaraktererne i 8. klasse i 2017/2018



1) Samlet vurdering

Kilde: Styrelsen for It og Læring

Der ses den samme tydelige sammenhæng mellem testresultater og standpunktskarakterer fra 8. klasse som mellem testresultater og prøvekaraktererne i 9. klasse.

Sammenhæng mellem testresultater og PISA test

I en rapport fra konsulentfirmaet DAMVAD udarbejdet i samarbejde med Svend Kreiner i 2014¹⁸ påvises det, at der er en tydelig sammenhæng mellem de resultater, eleverne opnår i de nationale test og i den internationale PISA-undersøgelse, jf. boks 1. Dette gælder både for testene i dansk og matematik.

Boks 1. Uddrag af Damvad-rapport om PISA og de nationale test (s. 5):

”Der er en tydelig sammenhæng mellem resultaterne fra de nationale test og resultaterne fra PISA-undersøgelserne. Sammenhængen kan observeres på tværs af profilområder i både læsning og matematik, men er ikke nødvendigvis jævnt fordelt.”

”Den tydelige sammenhæng mellem resultaterne fra de nationale test og PISA betyder samtidig, at de to test uafhængigt af hinanden når til relativt enslydende vurderinger af elevens faglige niveau. Det er en bekræftelse af, at de nationale test siger noget relevant om elevernes faglige niveau i de områder, hvori de testes.”

I projektet blev anvendt resultaterne fra PISA-undersøgelsen i 2012. Resultaterne fra de elever, der var udtrukket til PISA-undersøgelsen, blev koblet med de samme elevs testresultater fra dansk, læsning i 8. klasse og fra matematik i 6. klasse.

I analysen af sammenhængen i dansk læsning indgik 4.162 elever med resultater i såvel de nationale test som i PISA-undersøgelsen. I analysen fandtes en statistisk signifikant korrelation mellem PISA-resultatet og resultaterne fra hvert af de tre profilområder i de nationale test. Analysen blev foretaget på Rasch-modellens logit-skala og viste den stærkeste korrelation ($r=0,62$) mellem PISAs læsetest og profilområdet ’tekstforståelse’.

I rapporten konkluderedes¹⁹: ”Dette resultat understøtter, at der er en konsistent sammenhæng mellem målingen af profilområder i de nationale test og resultaterne i PISA-undersøgelsen.”

Den faglige færdighed i læsning og matematik, der måles i de nationale test, kan således relateres til den internationale PISA-undersøgelse.

¹⁸ PISA-relateret af de kriteriebaserede nationale test. DAMVAD 2014 (<https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf17/jan/170110-kriteriebaserede-test-delrapport-1.pdf>)

¹⁹ s. 11 i delrapport-1

3 Notat 3 Den statistiske usikkerhed og testenens reliabilitet

Sammenfatning

- Den statistiske usikkerhed på elevernes estimerede dygtighed kan beregnes i de nationale test
- Den statistiske usikkerhed på elevernes estimerede dygtighed vises for den enkelte lærer på resultatvisningerne i testsystemet

- Den gennemsnitlige statistiske usikkerhed på elevernes estimerede dygtighed er 0,46 logit.
- Den gennemsnitlige statistiske usikkerhed er mindst i fysik/kemi i 8. klasse (0,36 i profilområde 3) og størst i matematik i 8. klasse (0,54 i profilområde 3)
- Den statistiske usikkerhed er generelt størst for de dygtigste elever

- 93 procent af alle obligatoriske testforløb i skoleåret 2017/2018 blev afsluttet med en statistisk usikkerhed under 0,55 logit

- Omregnes den statistiske usikkerhed til den normbaserede skala (1-100), da er længden på 68 procent og 95 procent sikkerhedsintervallet på henholdsvis ± 12 og ± 22 point
- På den normbaserede skala er sikkerhedsintervallerne størst på midten af skalaen

- Reliabiliteten, angivet ved Person Separation Index, ligger i intervallet 0,74 – 0,91 for dansk læsning, matematik og engelsk, mens den ligger i intervallet 0,66 – 0,70 for fysik/kemi.
- Reliabiliteten, angivet ved Person Separation Index, ligger over 0,80 i 23 ud af 30 profilområder og under 0,80 i de resterende syv

- Den statistiske sikkerhed kan forbedres ved at forlænge testtiden, således at eleven når at besvare flere opgaver. Øges testtiden, så antallet af opgaver eleverne når at besvare øges fra de nuværende ca. 20 til 40, da kan den bedst mulige statistiske usikkerhed reduceres fra 0,45 til 0,32
- Den statistiske sikkerhed kan endvidere forbedres ved, at øge antallet af polytome opgaver, tilføje flere svære opgaver til opgavebanken samt ved at justere algoritmen i testsystemet så opgaver med størst mulig informationsværdi vælges

Indledning

I evalueringen af de nationale test skal følgende kritik belyses:

1. Den kritik der rejses af den statistiske sikkerhed og reliabiliteten skal adresseres.

1c. Herunder skal det klarlægges om det er muligt at forbedre den adaptive algoritme med henblik på at reducere den statistiske usikkerhed.

I notatet redegøres for den statistiske usikkerhed på de beregnede elevdygtigheder. Ligeledes omsættes den statistiske usikkerhed til sikkerhedsintervaller omkring elevdygtigheden. Sikkerhedsintervallerne omregnes til de forskellige skalaer elevernes dygtighed formidles på. BUVM²⁰ har tidligere undersøgt såvel den statistiske usikkerhed som reliabiliteten^{21,22}. I notatet samles op på tidligere beregninger og suppleres med nye baseret på de seneste obligatoriske test. I notatet præsenteres forslag til forbedringer af de nationale test med henblik på reduktion af den statistiske usikkerhed.

Formålet er således, at beskrive den statistiske usikkerhed og testenes reliabilitet samt komme med forslag til forbedringer af de nationale test.

Notatet indeholder følgende afsnit:

- Den statistiske usikkerhed
- Sikkerhedsintervaller og skalaer til formidling af denne
- Reliabiliteten
- Forslag til forbedringer af den statistiske sikkerhed

Den statistiske usikkerhed

De nationale test er baseret på Rasch-modellen^{23,24}. Rasch-modellen er en sandsynlighedsmodel, hvor sandsynligheden for, at en elev nummer n med dygtighedsparameteren θ_n svarer x rigtige på det polytome item²⁵ nummer i

²⁰ Børne- og undervisningsministeriet

²¹ <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf16/sep/160912-notat-om-nationale-tests-maaleegenskaber.pdf>

²² <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf17/jan/170110-uddybende-bilags-notat-om-de-nationale-tests-maaleegenskaber.pdf>

²³ Rasch, G.: Probabilistic Models for Some Intelligence and Attainment Tests. Danish National Institute for Educational Research, Copenhagen 1960.

²⁴ Karl Bang Christensen, Svend Kreiner, Mounir Mesbah (edt): Rasch Models in Health. Wiley 2013.

²⁵ I Rasch-modellen anvendes begrebet item i stedet for opgave. Et dikotomt item har to svarmuligheder, fx ja/nej eller rigtigt/forkert. Et polytomt item er et item med flere subitems, således at eleven kan score 0, 1, 2, 3, ... rigtige

med item sværhedsparameteren β_i og de såkaldte tærskelværdier τ , samt en størrelse m_i , der angiver den maksimale scoring for det polytome item:

$$P\{X_{ni} = x\} = \frac{e^{-\tau_{1i}-\tau_{2i}\dots-\tau_{xi}+x(\theta_n-\beta_i)}}{\sum_{z=0}^{m_i} e^{-\tau_{1i}-\tau_{2i}\dots-\tau_{zi}+z(\theta_n-\beta_i)}} \quad \{1\}$$

I Rasch-modellen måles itemsværdier og elevdygtigheder på samme skala. Skalaen kaldes logit-skalaen²⁶.

Ved at anvende de estimerede tærskelværdier for opgaverne i opgavebanken kan elevernes dygtighed beregnes. Elevernes dygtighed beregnes som maximum likelihood estimatet af person parameteren, θ .

Anvendelsen af en statistisk model, Rasch-modellen, medfører, at den statistiske usikkerhed på elevdygtighederne kan beregnes. Den statistiske usikkerhed på andre test og prøver i skolen beregnes og formidles ikke.

Den statistiske usikkerhed på elevens estimerede dygtighed betegnes SEM (Standard Error of Measurement).

SEM afhænger både af elevens dygtighed, opgavernes sværhedsgrad samt ikke mindst af antallet af opgaver eleven besvarer. Des flere opgaver eleven besvarer des mindre statistisk usikkerhed på den beregnede dygtighed.

SEM beregnes i testsystemet i lighed med elevens dygtighed efter hver besvarelse på en opgave²⁷.

SEM anvendes i den adaptive algoritme til at angive, hvornår elevens dygtighed er bestemt med den fastsatte præcision så elevens testforløb kan afsluttes. Ligeledes anvendes SEM i formidlingen af elevernes resultater til lærerne.

Den statistiske usikkerhed og testenes stopkriterie

SEM indgår i vurderingen af, hvornår en elev kan afslutte testen. Den enkelte test med 3 profilområder varer som udgangspunkt i 45 minutter, men læreren kan forlænge testtiden for den enkelte elev, hvis den statistiske usikkerhed på elevdygtigheden er for stor.

I den oprindelige løsningsbeskrivelse fra COWI, der udviklede testene, står²⁸:

”Der vælges derfor et stopkriterium der indebærer, at der enten skal være besvaret 20 items på et profil- eller testområde før der kan gives en tilbagemelding, eller der skal være opnået en SE på mindre end 0,6 i det pågældende

²⁶ Logits er en transformation med den naturlige logaritme af odds, $p/(1-p)$, hvor p er sandsynligheden for at svare rigtigt på et item.

²⁷ <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf15/jan/150128-den-adaptive-algoritme-i-de-nationale-test.pdf>

²⁸ Fremme af evalueringskulturen i folkeskolen – Bilag 3A løsningsbeskrivelse (10. juli 2006)

testforløb. Det anbefales at fortsætte testen, hvis testens tidsramme ikke er overskredet selvom stopkriteriet er nået, idet enhver forøgelse af antal besvarede items vil øge sikkerheden for tilbagemeldingen.”

Det endelige stopkriterie i testsystemet blev fastsat til:

- Testtiden er 45 minutter for 3 profilområder
- SEM < 0,55 logit eller eleven skal have besvaret mindst 30 opgaver i hvert profilområde

Når SEM kommer under 0,55 logit i et profilområde, da får eleven ikke flere opgaver i det pågældende profilområde men fortsætter med opgaver i de øvrige profilområder. Når SEM er under 0,55 i alle tre profilområder i en test, fortsætter eleven med opgaver indtil de 45 minutter er gået.

Elevernes teststatus følges af læreren på en monitoreringsskærm under afviklingen af testene. Elevers teststatus ændres fra *gul* til *grøn*, når SEM er under 0,55 logit eller antallet af besvarede opgaver er mindst 30 i alle tre profilområder. Elever med teststatus lig grøn kan umiddelbart afsluttes af læreren efter de 45 minutter. Læreren kan vælge at forlænge elevens test tid, hvis elevens teststatus er lig gul. Læreren kan også vælge at afslutte eleven.

92 procent af eleverne afsluttede den nationale obligatoriske test i skoleåret 2017/2018 med en SEM under 0,55 logit eller havde besvaret mindst 30 opgaver i alle tre profilområder (Tabel 3.1).

Tabel 3.1 Andelen af eleverne, der afslutter testen med en SEM < 0,55¹⁾ i alle tre profilområder (status = 'grøn') i skoleåret 2017/2018.

Test	Afsluttet i 'grøn' ¹⁾
Dansk læsning 2. klasse	94 %
Dansk læsning 4. klasse	95 %
Dansk læsning 6. klasse	96 %
Dansk læsning 8. klasse	91 %
Matematik 3. klasse	87 %
Matematik 6. klasse	89 %
Matematik 8. klasse	80 %
Engelsk 4. klasse	97 %
Engelsk 7. klasse	94 %
Fysik/kemi 8. klasse	98 %
I alt	92 %

1) Elever, hvor SEM er under 0,55 logit eller antallet af besvarede opgaver er mindst 30 i alle tre profilområder.

Kilde: Styrelsen for It og Læring

For den nyeste test i matematik 8. klasse, der startede i skoleåret 2017/2018, er andelen af elever, der afslutter med en SEM under 0,55 logit, kun på 80 procent. Tilsvarende er andelen af elever, der afslutter med en SEM under 0,55 logit i alle tre profilområder i matematik i 3. og 6. klasse, lavere end for

de øvrige test. Læreren kan vælge at forlænge testtiden med henblik på at forbedre den statistiske sikkerhed på elevens beregnede dygtighed.

Den statistiske usikkerhed i de enkelte profilområder

Den gennemsnitlige statistiske usikkerhed på elevdygtigheden i de nationale obligatoriske test i skoleåret 2017/2018 er på 0,46. 93,4 procent af testforløbene afsluttes med en SEM under 0,55.

SEM i hver test i skoleåret 2017/2018 fremgår af Tabel 3.2. Spredningen i SEM er angivet ved henholdsvis 5 % og 95 % percentilerne. Fx er den gennemsnitlige SEM i profilområde 1 i dansk læsning 2. klasse på 0,46, mens henholdsvis 5 % og 95 % percentilerne er på 0,32 og 0,55. Dvs. 5 procent af eleverne opnår et testresultat med en usikkerhed på 0,32 eller derunder, mens 5 procent af eleverne opnår et testresultat med en usikkerhed på 0,55 eller derover. Den gennemsnitlige statistiske usikkerhed er størst i de nye test (nyt profilområde 3²⁹) i matematik i 3. og 6. klasse samt i den nye test i matematik i 8. klasse³⁰. I fysik/kemi i 8. klasse er den gennemsnitlige statistiske usikkerhed på 0,36-0,37 logit.

Tabel 3.2 Den statistiske usikkerhed på den estimerede elevdygtighed i skoleåret 2017/2018. Gennemsnit [5% percentil; 95% percentil] (logit)

Test	Profilområde 1	Profilområde 2	Profilområde 3
Dansk læsning 2. klasse	0,46 [0,32; 0,55]	0,46 [0,32; 0,57]	0,42 [0,26; 0,55]
Dansk læsning 4. klasse	0,47 [0,36; 0,55]	0,48 [0,36; 0,55]	0,44 [0,25; 0,55]
Dansk læsning 6. klasse	0,48 [0,38; 0,55]	0,49 [0,39; 0,55]	0,45 [0,28; 0,54]
Dansk læsning 8. klasse	0,51 [0,40; 0,59]	0,51 [0,41; 0,60]	0,47 [0,31; 0,60]
Matematik 3. klasse	0,48 [0,32; 0,65]	0,49 [0,31; 0,65]	0,53 [0,40; 0,73]
Matematik 6. klasse	0,47 [0,32; 0,58]	0,48 [0,32; 0,62]	0,52 [0,42; 0,66]
Matematik 8. klasse	0,53 [0,43; 0,69]	0,52 [0,40; 0,69]	0,54 [0,43; 0,72]
Engelsk 4. klasse	0,40 [0,26; 0,54]	0,41 [0,25; 0,54]	0,44 [0,29; 0,61]
Engelsk 7. klasse	0,45 [0,27; 0,55]	0,47 [0,31; 0,55]	0,45 [0,29; 0,55]
Fysik/kemi 8. klasse	0,37 [0,23; 0,52]	0,37 [0,22; 0,53]	0,36 [0,21; 0,53]

Kilde: Styrelsen for It og Læring

For testene i alle profilområderne gælder, at den gennemsnitlige statistiske usikkerhed er under de 0,55 logit, der er den fastsatte grænse for, hvornår elevens testresultat vurderes tilstrækkeligt sikkert bestemt indenfor den afsatte tid til testen.

²⁹ I matematik 3. klasse erstattedes profilområde 3 *matematik i anvendelse af statistik og sandsynlighed* i 2015/2016. Tilsvarende ændring skete i matematik i 6. klasse i 2017/2018

³⁰ Matematik i 8. klasse introduceredes i 2017/2018

Den statistiske usikkerhed og antal opgaver

For at vurdere om den statistiske usikkerhed er stor eller lille kan anvendes, at usikkerheden i adaptive test med dikotome³¹ opgaver ikke kan blive mindre end $2/\sqrt{n}$, hvor n er antallet af opgaver (Tabel 3.3).

Tabel 3.3 Sammenhæng mellem antal opgaver og mindst mulige SEM

Antal opgaver	SEM ¹⁾		Antal opgaver	SEM ¹⁾
5	0,89		23	0,42
10	0,63		24	0,41
15	0,52		25	0,40
16	0,50		30	0,37
17	0,49		40	0,32
18	0,47		45	0,30
19	0,46		50	0,28
20	0,45		55	0,27
21	0,44		60	0,26
22	0,43		70	0,24

1) Standard Error of Measurement (SEM)

Kilde: Styrelsen for It og Læring

Der er afsat 45 minutter til hver test. På de 45 minutter får eleverne opgaver i alle tre profilområder, der er knyttet til den enkelte test. Eleverne kan i gennemsnit nå at besvare 19 opgaver i hvert profilområde. Antallet af opgaver eleverne kan nå at besvare varierer fra i gennemsnit 15 til 25 opgaver afhængig af test og klasstrin (Tabel 3.4).

Opgavebanken i de nationale test består både af dikotome og polytome opgaver. En polytom opgave består af flere delopgaver således at eleven kan score 0, 1, 2, 3, ... rigtige. I profilområde 1 i dansk læsning (sprogforståelse) findes udelukkende dikotome opgaver i opgavebanken. Andre profilområder indeholder både dikotome og polytome opgaver. Fx indeholder opgavebanken i fysik/kemi og i engelsk en del polytome opgaver.

Det gennemsnitlige antal delopgaver eleverne når at besvare i hvert profilområde er 23. Antallet af delopgaver eleverne kan nå at besvare varierer fra i gennemsnit 17 til 36 delopgaver afhængig af test og klasstrin (Tabel 3.4).

³¹ Dikotome opgaver har to svarmuligheder, fx ja/nej eller rigtigt/forkert.

Tabel 3.4 Det gennemsnitlige antal opgaver/delopgaver¹⁾ eleverne når at besvare i de obligatoriske test i skoleåret 2017/2018.

Test	Profilområde 1	Profilområde 2	Profilområde 3
Dansk læsning 2. klasse	24/24	24/24	22/29
Dansk læsning 4. klasse	21/21	22/22	20/26
Dansk læsning 6. klasse	20/20	21/21	18/24
Dansk læsning 8. klasse	19/19	19/19	17/21
Matematik 3. klasse	16/19	16/19	19/19
Matematik 6. klasse	15/19	15/19	18/18
Matematik 8. klasse	17/17	16/17	17/17
Engelsk 4. klasse	22/29	22/29	25/29
Engelsk 7. klasse	17/24	18/22	18/25
Fysik/kemi 8. klasse	18/33	18/33	18/36
Samlet pr. profilområde	19/23		

1) Delopgave er et subitem i et polytomt item

Kilde: Styrelsen for It og Læring

Med 19 dikotome opgaver er den mindst mulige statistiske usikkerhed på 0,46 logit, mens der med 23 dikotome opgaver ikke kan opnås en statistisk usikkerhed på mindre end 0,42 (Tabel 3.3).

Den gennemsnitlige statistiske usikkerhed i de nationale test på 0,46 er således ikke langt fra det mest optimale på 0,42 (Tabel 3.3) med det givne antal opgaver.

For yderligere at vurdere om en statistisk usikkerhed på 0,46 ved 23 delopgaver er stor eller lille, kan følgende hentes fra notatet "Om opgavetyper og usikkerhed i de nationale test" (Svend Kreiner, juni 2017³²):

"Tallet 0,45 kan derfor bruges som en benchmark værdi, hvis man både vil vurdere, hvor godt den adaptive algoritme har fungeret for en adaptiv test med 20 dikotome opgaver, og hvor godt en ikke-adaptiv test fungerer for elever med forskellige grader af dygtighed. Det kan for eksempel beregnes, at en ikke-adaptiv test med 20 opgaver, hvor sværhedsgraden er ligeligt fordelt fra -2,5 til +2,5, i bedste fald vil resultere i SEM = 0,54 og i værste fald (for meget dygtige og meget svage elever) med SEM = 0,82. Altså dårligere end en fungerende adaptiv test."

"Eller med andre ord: Hvis den adaptive algoritme fungerer efter hensigten vil usikkerheden på elevdygtigheden i en adaptiv test altid være mindre end usikkerheden i almindelige ikke-adaptive test. Hvor meget mindre afhænger

³² <https://www.uvm.dk/folkeskolen/elevplaner-nationale-test-trivselsmaaling-og-sprogproever/nationale-test/om-de-nationale-test>

af opgavernes sværhedsgrader og af elevernes dygtighed.”³³

Antallet af opgaver og dermed den tid, der afsættes til en test, er helt centralt når den statistiske usikkerhed skal vurderes. Adaptive test giver mulighed for, at mindske denne usikkerhed mest muligt.

Den statistiske usikkerhed og elevdygtigheden

Den observerede statistiske usikkerhed på elevernes dygtighed er generelt mindst for elever med en dygtighed i midten af skalaen og størst for de dygtigste elever.

Tabel 3.5 viser den gennemsnitlige statistiske usikkerhed på elevdygtigheden i dansk læsning 8. klasse og matematik i 6. klasse for elever i hvert af de fem niveauer af den normbaserede skala³⁴.

Bilag 3.1 viser tabeller over SEM for alle de obligatoriske test foråret 2018.

Tabel 3.5 Den statistiske usikkerhed (logit) på den estimerede elevdygtighed skoleåret 2017/2018 i forhold til elevdygtigheden

Fag	Profilmråde	Elevdygtighed på femtrins skalaen ¹⁾				
		1-10	11-35	36-65	66-90	91-100
Dansk læsning 8. klasse	Sprogforståelse	0,49	0,49	0,50	0,50	0,56
	Afkodning	0,49	0,49	0,50	0,51	0,59
	Tekstforståelse	0,44	0,38	0,45	0,51	0,55
Matematik 6. klasse	Tal og algebra	0,45	0,42	0,45	0,48	0,53
	Geometri	0,45	0,42	0,46	0,51	0,56
	Statistik og sandsynlighed	0,50	0,50	0,51	0,53	0,60

1) I formidlingen til elever og forældre bliver resultaterne på 100-skalaen omsat til en fem-trins skala: En del under gennemsnittet (1-10), under gennemsnittet (11-35), gennemsnittet (35-65), over gennemsnittet (66-90), en del over gennemsnittet (91-100)

Kilde: Styrelsen for It og Læring

I fx sprogforståelse i 8. klasse er den gennemsnitlige statistiske usikkerhed på 0,50 for elever med en dygtighed midt på dygtighedsskalaen, mens den er 0,56 for de allerdygtigste elever.

³³ Svend Kreiner (juni 2017). <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf17/sep/170913-om-opgavetyper-og-usikkerhed-i-de-nationale-test.pdf>

³⁴ Målingerne af elevdygtigheden og SEM foregår på logit-skalaen. I formidlingen af resultaterne til lærerne omregnes resultaterne til den normbaserede percentilskala (1-100). I formidlingen til elever og forældre bliver resultaterne på 100-skalaen omsat til en fem-trins skala: En del under gennemsnittet (1-10), under gennemsnittet (11-35), gennemsnittet (35-65), over gennemsnittet (66-90), en del over gennemsnittet (91-100)

Generelt er den statistiske usikkerhed størst for de dygtigste elever, dvs. for elever med et testresultat en del over gennemsnittet. Manglen på svære opgaver i opgavebanken er givetvis en medvirkende årsag til dette. Hvis der mangler svære opgaver i opgavebanken, vil de dygtigste elever kunne opleve at få opgaver, der ligger et stykke væk fra deres dygtighed. Hermed bliver den statistiske usikkerhed på elevens beregnede dygtighed større end, hvis der hele tiden var opgaver med en sværhedsgrad, der passede til elevens dygtighed.

Den statistiske usikkerhed og sikkerhedsintervaller

For hver elevs testresultat kan der beregnes et såkaldt sikkerhedsinterval. I forbindelse med beregning af sikkerhedsintervaller for gennemsnit, fx for en klasses gennemsnit eller en skoles gennemsnit, anvendes ofte et 95 procent sikkerhedsinterval. Et 95 procent sikkerhedsinterval er givet ved *gennemsnit* $\pm 1,96*SE$.

Til angivelse af sikkerhedsinterval for et enkelt individs testresultat anvendes *testresultat* $\pm 1*SEM$, hvilket ca. svarer til et 68 procent sikkerhedsinterval.

Begge usikkerhedsintervaller er beregnet³⁵ på logit-skalaen baseret på den enkelte elevs beregnede statistiske usikkerhed og herefter omsat til og præsenteret på såvel den normbaserede (1-100) skala som på den kriteriebase-rede skala. Usikkerhedsintervallet på den oprindelige dygtighedsskala, logit-skalaen, er blot $\pm 1*SEM$ eller $\pm 1,96*SEM$ for henholdsvis et 68 procent og et 95 procent sikkerhedsinterval.

Den beregnede elevdygtighed er den mest sandsynlige, men det kan ikke afvises med en vis sandsynlighed, at elevens dygtighed afviger fra dette. Des længere væk fra elevens estimerede dygtighed des mindre sandsynligt er denne værdi.

Længden (L) af sikkerhedsintervallet på den normbaserede skala angiver afstanden fra elevens estimerede testresultat til nedre/øvre grænse i intervallet. Sikkerhedsintervallets samlede længde er således $2*L$.

68 procent og 95 procent sikkerhedsintervallet udtrykt på den normbaserede er på henholdsvis ± 12 og ± 22 point (Tabel 3.6)³⁶. Dette svarer til, at en elevs testresultat med stor sandsynlighed ligger i dette interval omkring det estimerede testresultat. Det er vigtigt at understrege, at elevens beregnede score

³⁵ Beregningen af usikkerhedsintervaller er kun foretaget for elever, der har afsluttet testen i teststatus 'grøn', dvs hvor den statistiske usikkerhed på elevens estimerede dygtighed er kommet under 0,55 logit, eller eleven har besvaret mindst 30 opgaver i hvert af de tre profilområder.

³⁶ Der er tale om en lille tilnærmelse, da sikkerhedsintervallerne på percentilskalaen ikke er helt symmetriske

er den mest sandsynlige værdi, men der er en vis sandsynlighed for, at den reelle score afviger fra denne.

Tabel 3.6 Længden af sikkerhedsintervallet i forhold til elevernes dygtighed på den normbaserede (1-100) skala. Samlet for alle de obligatoriske test. Skoleåret 2017/2018

Usikkerhedsinterval	Samlet	Elevdygtighed på femtrins skalaen				
		1: 1-10	2: 11-35	3: 36-65	4: 66-90	5: 91-100
$\pm 1 * SEM^1$	12	3	11	16	12	4
$\pm 1,96 * SEM^2$	22	7	21	28	23	10

1) tilnærmelsesvist et 68 % sikkerhedsinterval

2) tilnærmelsesvist et 95 % sikkerhedsinterval

Note: Der er tale om en tilnærmelse, da sikkerhedsintervallerne på percentilskalaen ikke er helt symmetriske

Kilde: Styrelsen for It og Læring

Målingerne af elevdygtigheden og den statistiske usikkerhed foregår på en såkaldt logit-skala. På denne logit-skala er den statistiske usikkerhed på elevdygtighederne størst for de dygtigste elever og mindst for elever med en dygtighed på midten af skalaen (Tabel 3.5).

Formidlingen af resultaterne til lærerne har siden starten i 2010 foregået på percentilskalaen, 1-100. Omregning fra den grundlæggende logit-skala til percentilskalaen har nogle uheldige egenskaber. Mange elever har en beregnet dygtighed midt på logit-skalaen med en relativt lille forskel imellem sig. Ved omregning til percentilskalaen vil en given forskel i dygtighed strække sig over mange percentiler på midten og over færre i yderområderne af dygtighedsskalaen. En beregnet statistisk usikkerhed på dygtigheden hos elever med en dygtighed på midten af skalaen vil derfor strække sig over flere percentiler end en tilsvarende statistisk usikkerhed hos elever i yderområderne af dygtighedsskalaen. Derfor fremstår den statistiske usikkerhed på elevdygtigheden formidlet på percentilskalaen størst for elever omkring gennemsnittet, hvilket reelt er i modstrid med den faktiske bagvedliggende statistiske usikkerhed.

Fx er 68 procent sikkerhedsintervallet for en elev, der har opnået et testresultat på 50 lig med ± 16 , dvs sikkerhedsintervallet for denne elevs testresultat kunne være [34 - 66], mens 68 procent sikkerhedsintervallet for en elev, der har opnået et testresultat på 95 er lig med ± 4 , dvs sikkerhedsintervallet for denne elevs testresultat kunne være [91 - 99].

Formidlingen af resultaterne til elever og forældre foregår på en femtrins skala, hvor netop det midterste interval (gennemsnittet) er bredest. Dette opvejer til dels denne uheldige konsekvens af en omregning til en percentilskala. Den skala, der skal anvendes til fortsat formidling af elevernes resultater, bør dog genovervejes.

Tilsvarende kan der beregnes sikkerhedsintervaller på den kriteriebaserede skala (Tabel 3.7).

Tabel 3.7 Sikkerhedsintervallet på den kriteriebaserede skala. Samlet for dansk læsning og matematik. Skoleåret 2017/2018

Usikkerhedsinterval	Elevens testresultat ligger med stor sandsynlighed i ...		
	... det beregnede faglige niveau	... det beregnede faglige niveau eller enten niveauet lige under/over (ikke både og)	... det beregnede faglige niveau eller niveauet lige over eller lige under
$\pm 1 * SEM^1$	28 %	65 %	7 %
$\pm 1,96 * SEM^2$	9 %	37 %	54 %

1) tilnærmelsesvist et 68 % sikkerhedsinterval

2) tilnærmelsesvist et 95 % sikkerhedsinterval

Kilde: Styrelsen for It og Læring

Samlet gælder, at 28 procent af elevernes testresultater med 68 procent sandsynlighed sikkerhed ligger indenfor det beregnede faglige niveau, mens 65 procent af elevernes testresultater ligger indenfor det beregnede niveau eller enten niveauet lige under eller niveauet lige over (Tabel 3.7). Således ligger 93 procent af elevernes testresultater med den beregnede statistiske sikkerhed indenfor det beregnede faglige niveau eller i nabo kategorien. De resterende 7 procent af elevernes testresultater har en større usikkerhed, der betyder, at elevens faktiske niveau ikke kan afvises at ligge i både niveauet under og i niveauet over det estimerede niveau.

Bilag 3.1 indeholder sikkerhedsintervallerne udtrykt på den normbaserede skala for alle de obligatoriske test i skoleåret 2017/2018. I bilaget findes også tabeller med sikkerhedsintervallerne på den kriteriebaserede skala for de obligatoriske test i dansk læsning og matematik i 2017/2018.

Reliabiliteten

Reliabiliteten er et udtryk for testens evne til at rangordne eleverne efter elevdygtighed på korrekt måde. Reliabiliteten kan også forklares ved, hvor god en test er til at adskille elever med forskellig dygtighed.

Reliabiliteten er belyst på forskellig vis af BUVVM³⁷. I 2016 blev beregnet en såkaldt test-retest korrelation. Beregningerne var baseret på elevs testresultater fra de frivillige test. I den frivillige testperiode er det muligt at tage den samme test to gange med få dages mellemrum. Gentagelsen af en test skal ske, uden at eleven kan huske det første testforløb, og uden at eleven har lært af den første test eller lært nyt mellem de to testafviklinger. Dette er

³⁷ <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf17/jan/170110-uddybende-bilags-notat-om-de-nationale-tests-maaleegenskaber.pdf>

naturligvis vanskeligt i pædagogiske test herunder i de nationale test. Desuden kan elevers testadfærd, motivation, koncentration mv. spille ind på elevens testresultat. Derfor skal disse test-retest resultater vurderes med stor forsigtighed. BUVM har gennemført analyser af test-retest på en specifik skole, hvor læreren undrede sig over store udsving i nogle af elevernes resultater i to gentagne frivillige test afholdt med syv dages mellemrum i efteråret 2014. Gennemgangen af elevernes testforløb viste, at en stor del af eleverne i andet forsøg besvarede langt flere opgaver uden at anvende længere tid. I elevernes andet forsøg besvarede næsten 50 procent flere opgaver end elever på landsplan i gennemsnit gør. Denne forskel i testadfærd kan betyde, at det er vanskeligt at sammenholde en elevs to test (test-retest).

I 2016 foretog BUVM ligeledes test-retest simuleringer, hvor 5.000 elever med forskellig dygtighed fik simuleret et elevforløb i testsystemet to gange. Disse simuleringer er uafhængig af testadfærd og korrelationen mellem de simulerede testresultater er derfor en beregning af den teoretiske test-retest korrelation, som man ville kunne observere, hvis testen fungerede fuldstændigt som forventet. Simuleringerne måler således om testsystemet og tilhørende opgavebank kan genskabe rangordningen af elevernes testresultater.

I Tabel 3.8 er disse resultater vist for dansk læsning 8. klasse og matematik 6. klasse. Samtlige resultater er tilgængelige på <https://www.uvm.dk/folkeskolen/eleveplaner-nationale-test--trivselsmaaling-og-sprogproever/nationale-test>

Tabel 3.8 Reliabiliteten i de nationale test målt ved test-retest og simuleringer

Test	Profilmråde	Test-retest ¹⁾	Simuleringer ²⁾
Dansk læsning 8. klasse	Sprogforståelse	0,66	0,84
	Afkodning	0,85	0,87
	Tekstforståelse	0,72	0,88
Matematik 6. klasse	Tal og algebra	0,63	0,89
	Geometri	0,65	0,86
	Statistik og sandsynlighed	0,68	0,89

1) Korrelation mellem elevdygtigheden fra to frivillige test efteråret 2015

2) Korrelation mellem elevdygtigheden bestemt ved simuleringer i testsystemet

Kilde: Styrelsen for It og Læring

Baseret på klassisk test teori kan reliabiliteten defineres som forholdet mellem variansen på den sande score og variansen på den observerede score.

Udgangspunktet er:

$$X = T + E$$

hvor X er den observerede elevscore, T er den sande elevscore og E er målefejlen.

Under visse forudsætninger er variansen på den observerede elevscore (X) lig summen af variansen på den sande score (T) og variansen på målefejlen (E)³⁸:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

Reliabiliteten defineres som forholdet mellem variansen på den sande score og variansen på den observerede score:

$$\rho_{xx'} = \frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}$$

I RUMM³⁹ beregnes et indeks for reliabiliteten: *Person Separation Index*^{40,41}.

Person separation indeks (PSI) beregnes som:

$$r_{\theta\theta} = 1 - \frac{\sigma_\varepsilon^2}{\sigma_\theta^2}$$

hvor θ er elevens estimerede dygtighed og σ_θ^2 beregnes som variansen mellem elevernes dygtigheder blandt de elever, der gennemfører en test.

σ_ε^2 er usikkerheden på den enkelte elevs beregnede dygtighed og betegnes SEM², hvor SEM er Standard Error of Measurement.

Reliabiliteten afhænger således ikke kun af SEM men også af spredningen af elevernes dygtighed i den pågældende test. En lille spredning mellem elevernes dygtighed giver en lavere reliabilitet, mens en større spredning i elevernes dygtighed vil give en højere reliabilitet. Reliabiliteten er således gældende for den aktuelle population og er vanskelig at sammenligne med reliabiliteten fra andre test og andre populationer med en anden spredning i elevdygtighederne.

I Tabel 3.9 er reliabiliteten beregnet for dansk læsning 8. klasse og matematik 6. klasse. I beregningerne er anvendt gennemsnittet af SEM på elevdygtighederne i det pågældende profilområde (Tabel 3.2).

³⁸ Davidshofer, Kevin R. Murphy, Charles O. (2005): Psychological testing: principles and application. Pearson/Prentice Hall.

³⁹ RUMM Laboratory Pty Ltd.

⁴⁰ Persons Separation Index udtrykker forholdet mellem usikkerheden på elevdygtigheden på den ene side og spredningen mellem elevernes dygtighed på den anden side

⁴¹ Karl Bang Christensen, Svend Kreiner, Mounir Mesbah (edt): Rasch Models in Health. (2013)

Tabel 3.9 Testenes reliabilitet i skoleåret 2017/2018

Fag	Profilområde	PSI ¹⁾	Simuleringer ²⁾
Dansk læsning 8. klasse	Sprogforståelse	0,83	0,84
	Afkodning	0,85	0,87
	Tekstforståelse	0,84	0,88
Matematik 6. klasse	Tal og algebra	0,83	0,89
	Geometri	0,81	0,86
	Statistik og sandsynlighed	0,84	0,89

1) Person Separation Index (PSI)

2) Korrelation mellem elevdygtigheden bestemt ved simuleringer i testsystemet

Kilde: Styrelsen for It og Læring

Reliabiliteten beregnet ved Person Separation Index er næsten på niveau med de simulerede test-retest korrelationer (Tabel 3.9).

I Bilag 3.2 er reliabiliteten beregnet for alle de obligatoriske test i skoleåret 2017/2018.

Der findes forskellige anbefalinger for niveauet af reliabilitet. En anvendt tommelfingerregel er en reliabilitet på mindst 0,80. I Streiner⁴² anføres, at en optimal reliabilitet ikke bør være under 0,70. I Stewart⁴³ accepteres værdier helt ned til 0,50 dog som værende evidens for tilfredsstillende reliabilitet.

I dansk læsning ligger reliabiliteten i intervallet 0,74 – 0,91. I sprogforståelse (profilområde 1) i 2. og 6. klasse ligger reliabiliteten under 0,80. I alle øvrige test i dansk læsning ligger reliabiliteten på mindst 0,80.

I matematik ligger reliabiliteten i intervallet 0,78 – 0,89. I geometri (profilområde 2) i 3. klasse ligger reliabiliteten under 0,80. I alle øvrige test i matematik ligger reliabiliteten på mindst 0,80.

I engelsk ligger reliabiliteten i intervallet 0,79 – 0,91. I læsning (profilområde 1) i 7. klasse ligger reliabiliteten under 0,80. I alle øvrige test i engelsk ligger reliabiliteten på mindst 0,80.

I fysik/kemi 8. klasse ligger reliabiliteten i intervallet 0,66 – 0,70. Reliabiliteten ligger således relativt lavt i alle tre profilområder til trods for, at spredningen, SEM, på elevdygtighederne også er lille. Den lave reliabilitet skyldes således en meget begrænset spredning i elevernes dygtighed i testen i fysik/kemi.

⁴² Streiner, D. L., G. R. Norman (1995): Health Measurement Scales – A Practical Guide to Their Development and Use. Oxford University Press

⁴³ Stewart, A. L. (1990): Psychometric Considerations in Functional Status Instruments. In Lipkin Jr., M. Functional Status Measurement in Primary Care. Springer-Verlag, NY

I 23 ud af 30 profilområder ligger reliabiliteten, i form af Person Separation Index, over 0,80, mens de resterende syv profilområder har en lavere reliabilitet. Specielt er reliabiliteten lav i fysik/kemi.

Forslag til forbedringer af den statistiske sikkerhed

Antal opgaver

Den statistiske sikkerhed (SEM) kan primært forbedres ved at øge antallet af opgaver den enkelte elev skal besvare (Tabel 3.3). Antallet af point (lig med antal delopgaver) i folkeskolens digitale prøver i 9. klasse i matematik uden hjælpemidler, biologi, geografi m.fl. ligger omkring 50. Med 50 dikotome items kan SEM bringes ned i nærheden af 0,28 (Tabel 3.3). Antallet af opgaver hænger sammen med tiden til den enkelte test.

Polytome opgaver

Øget anvendelse af polytome opgaver, hvor eleverne svarer på flere dikotome delopgaver, vil også have en positiv effekt på SEM. Den statistiske usikkerhed på elevens testresultat er en funktion af den såkaldte test information⁴⁴. Des større test information des mindre statistisk usikkerhed. Test informationen er summen af informationen fra elevens besvarelser af de enkelte opgaver. For dikotome opgaver (rigtig/forkert) er den maksimale informationsværdi på 0,25, og denne opnås, når opgavens sværhedsgrad er lig elevens dygtighed.

Beregninger viser^{45,46}, at polytome opgaver med k del-opgaver kan give op til k^2 gange den information som k dikotome opgaver kan give. Øget anvendelse af polytome opgaver kan derfor mindske den statistiske usikkerhed på den beregnede elevdygtighed.

Sammenhængen mellem opgavetyper og den statistiske usikkerhed i de nationale test er beskrevet af Svend Kreiner²². Af rapporten fremgår også, at der er forskel på de polytome opgavens informationsværdi og at denne varierer med opgavetypen.

Efter hver opgaveafprøvning modtager de faglige opgavekommissioner derfor en opgørelse, der viser den beregnede opgaveinformation for alle de

⁴⁴ www.uvm.dk/-/media/filer/uvm/udd/folke/pdf17/sep/170913-om-opgavetyper-og-usikkerhed-i-de-nationale-test.pdf

⁴⁵ Henrik Albeck et al: Optimizing the precision in estimating student abilities in 'Danish National Tests'. Seventh International Conference on Probabilistic Models for Measurement Developments with Rasch Models, Perth 2018

⁴⁶ Svend Kreiner: On item information and usefulness of partial credit items. Seventh International Conference on Probabilistic Models for Measurement Developments with Rasch Models, Perth 2018

Rasch godkendte opgaver. Dette kan bruges til vurdering af hvilke opgavetyper, der giver mest statistisk information.

Justering af algoritmen

Den adaptive algoritme i testsystemet kan forbedres ved i højere grad end i dag, at vælge opgaver med den størst mulige opgaveinformation. I dag vælger den adaptive algoritme i testsystemet den næste opgave så den bedst muligt passer til elevens dygtighed.

Hvis opgaveparametrene for opgaverne i opgavebanken suppleres med opgavens informationsværdi, da kan algoritmen justeres således, at opgaven med den største informationsværdi, i forhold til hvor elevens estimerede dygtighed er, vælges⁴⁷.

Flere svære opgaver

Den statistiske usikkerhed er størst for de dygtigste elever (Tabel 3.5). Manglen på svære opgaver i opgavebanken har en medvirkende årsag til dette. Når der ikke er tilstrækkeligt med svære opgaver, der matcher elevernes dygtighed, falder den samlede test information og den statistiske usikkerhed stiger.

⁴⁷ www.uvm.dk/-/media/filer/uvm/udd/folke/pdf17/sep/170913-om-opgavetyper-og-usikkerhed-i-de-nationale-test.pdf

4 Notat 4 Opgavebanken og opgavernes sværhedsgrad

Sammenfatning

- I perioden maj 2008 til januar 2019 har der været afholdt 14 opgaveafprøvningsprøver
- I foråret 2019 var der 10.969 opgaver i drift i opgavebanken
- Alle nye opgaver, der tilføjes opgavebanken, passer til Rasch-modellen
- Der er mangel på svære opgaver til de dygtigste elever i flere af profilområderne
- Analyser fra 2018 viser, at under 10 procent af de opgaver, der genafprøves i forbindelse med opgaveafprøvningsprøvene, har ændret deres sværhedsgrad over tid. Analyser fra 2019 viser, at 16 procent af de opgaver, der genafprøves i forbindelse med opgaveafprøvningsprøvene, har ændret deres sværhedsgrad over tid
- Analyser hidtil viser således ingen generel tendens til, at opgavernes sværhedsgrad ændres over tid
- Opgaver med statistisk signifikant ændret sværhedsgrad får denne opdateret i opgavebanken
- Der er forskel på opgavernes estimerede sværhedsgrad når disse beregnes på baggrund af de adaptive testforløb (obligatoriske test) og når de beregnes i lineære afprøvningsforløb (opgaveafprøvningsprøve)
- Andelen af opgaver, hvor den estimerede sværhedsgrad afviger mellem opgaveafprøvningsprøvene og testforløbene ændres generelt ikke i perioden fra 2010 til 2018
- Afvigelsen i den estimerede sværhedsgrad stiger ikke des længere tid, der er mellem testafholdelse og opgaveafprøvningsprøve
- Afvigelsen i den estimerede sværhedsgrad er størst for de svære opgaver

Indledning

I evalueringen af de nationale test skal følgende spørgsmål besvares:

1. Regner de nationale test rigtigt?

1a. Herunder skal det klarlægges om opgavernes sværhedsgrader stadig er korrekt og

1b. om opgaverne passer til Rasch-modellen.

I notatet redegøres for hvor mange opgaver, der er i opgavebanken, hvordan opgaver afprøves og besvarelsene fra opgaveafprøvningerne statistisk analyseres. Opgavebankens sammensætning af opgaver i forhold til opgavernes sværhedsgrad og i forhold til elevernes dygtighed beskrives. I notatet redegøres for, hvordan det undersøges, om opgavernes sværhedsgrad ændres over tid. Endelig belyses forskellen i opgavernes sværhedsgrad når disse fastsættes ud fra data indsamlet under henholdsvis lineær opgaveadministration og under adaptive testafviklinger samt betydningen for elevernes beregnede dygtighed.

Formålet er således at beskrive, hvordan opgaver tilføjes opgavebanken, hvordan opgavernes sværhedsgrader fastlægges samt at vurdere om opgavernes sværhedsgrader er ændret over tid eller stadig er korrekte.

Notatet indeholder følgende afsnit:

- Opgaveafprøvningen
- Opgavebanken
- Stabiliteten af opgavernes sværhedsgrad over tid
- Forskellen i opgavernes sværhedsgrad baseret på lineær og adaptiv testadministration
- Forskel i beregnet elevdygtighed

Opgaveafprøvningen

Opgaverne udarbejdes af faglige opgavekommissioner. Alle opgaver i opgavebanken er afprøvet af elever på det klassetrin, testen er målrettet til. I dag bliver alle nye opgaver afprøvet på ca. 700 elever.

Afprøvning af nye opgaver (udskiftning af opgaver) er omtalt i lov om folkeskolen⁴⁸ §13:

”Stk. 4. Før en test tages i brug og ved udskiftning af opgaver, som indgår i en test, kan undervisningsministeren udpege repræsentative grupper af elever til afprøvning af testenes relevans og pålidelighed m.v.”

⁴⁸ Bekendtgørelse af lov om folkeskolen (LBK nr 1510) af 14. dec 2017

Skoler, der skal deltage i opgaveafprøvningen, udvælges tilfældigt blandt landets folkeskoler. Udvælgelsen sker stratificeret efter skolens beliggenhed (Øst- eller Vestdanmark) og størrelse (under eller over 350 elever på skolen).

Opgaveafprøvningen foregår som en lineær test, hvor eleverne får 2-3 sæt på ca. 30 opgaver i hvert sæt. Et sæt af opgaver kan besvares på 45 minutter.

I forbindelse med afprøvningen af nye opgaver til opgavebanken medtages endvidere et antal af de eksisterende og tidligere godkendte opgaver fra opgavebanken. Dette sikrer, at der sker et overlap mellem blokkene af opgaveafprøvninger, således at nye opgavers sværhedsgrad kan indplaceres på den eksisterende skala. Disse overlappingsopgaver kaldes link-opgaver.

Opgaverne i hvert sæt randomiseres, så opgaverne kommer i forskellig rækkefølge til de enkelte elever.

Afprøvning af opgaver med henblik på anvendelse i de nationale test er foregået siden maj 2008. Der har i alt været afholdt 14 opgaveafprøvninger i perioden maj 2008 til januar 2019.

Bilag 4.1 indeholder tabeller, der viser, hvornår opgavebankens⁴⁹ opgaver er afprøvet. De fleste af opgavebankens opgaver er afprøvet i perioden fra 2008 til 2015. I perioden 2015 til 2018 er der primært afprøvet opgaver til de nye test og profilområder i matematik 3., 6. og 8. klasse samt i engelsk 4. klasse. I januar 2019 er der desuden afprøvet 1.552 opgaver fordelt på næsten alle test og profilområder.

På baggrund af elevernes besvarelser fra opgaveafprøvningen foretages en statistisk analyse, hvor det undersøges om opgaverne passer til Rasch-modellen. Opgaver, der ikke passer til Rasch-modellen fjernes. De resterende opgaver betegnes 'Rasch-godkendte' og tilføjes opgavebanken⁵⁰.

Alle analyser af besvarelser fra opgaveafprøvningerne foretages i analyseprogrammet RUMM⁵¹. Bilag 4.2 viser skærmdumps fra analyserne af opgaver afprøvet i januar 2018. Alle nye opgaver, der tilføjes opgavebanken, passer til Rasch-modellen.

⁴⁹ Der er kun medtaget den del af opgavebanken, der anvendes af de obligatoriske test, dvs dansk læsning, matematik, engelsk samt fysik/kemi

⁵⁰<https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf19/mar/190315-opgaveafprovning-og-beregning-af-opgavernes-svarhedsgrad-i-de-nationale-test.pdf>

⁵¹ www.rummlab.com.au

Opgavebanken

Under afvikling af de nationale test foråret 2019 var der i alt 10.969 opgaver i drift i opgavebanken (Tabel 4.1).

Tabel 4.1 Antal opgaver i opgavebanken foråret 2019

Fag og klassetrin	Antal opgaver i drift
Dansk læsning 2. klasse	729
Dansk læsning 4. klasse	788
Dansk læsning 6. klasse	773
Dansk læsning 8. klasse	821
Matematik 3. klasse	706
Matematik 6. klasse	1.025
Matematik 8. klasse	706
Engelsk 4. klasse	787
Engelsk 7. klasse	640
Fysik/kemi 8. klasse	757
Biologi 8. klasse	764
Geografi 8. klasse	858
Dansk som andetsprog 5. klasse	756
Dansk som andetsprog 7. klasse	859
I alt	10.969

Kilde: Styrelsen for It og Læring

Opgavebankens sammensætning af opgaver i forhold til opgavernes estimerede sværhedsgrad ses i Bilag 4.3.

I dansk læsning er der relativt få svære opgaver i profilområderne *afkodning* og *tekstforståelse*. Dette gør sig ligeledes gældende i det nye profilområde *statistik og sandsynlighed* i matematik 3. og 6. klasse samt i engelsk.

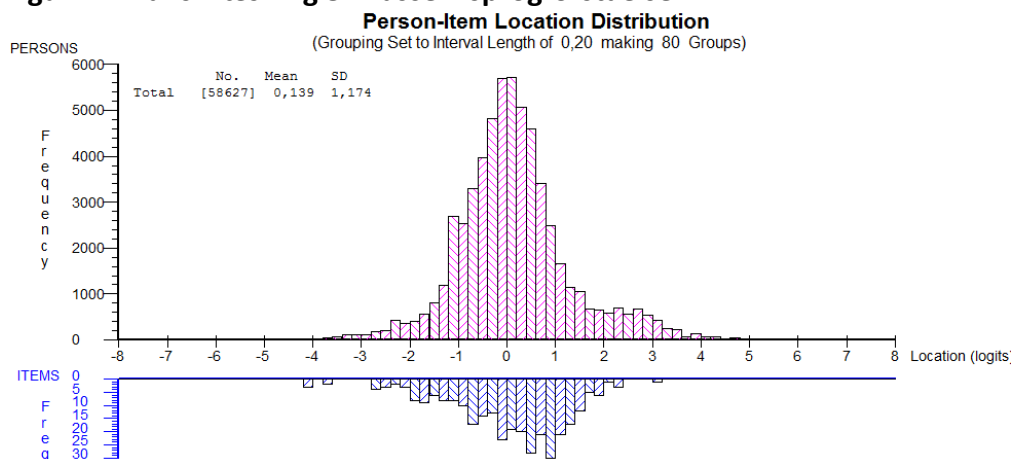
Opgavebankens sammensætning af lette, middelsvære og svære opgaver i forhold til elevgruppens dygtighed kan illustreres med de såkaldte person-item plot, hvor fordelingen af elevernes dygtighed sammenholdes med fordelingen af opgavernes sværhedsgrad (Figur 4.1-4.3).

x-aksen på person-item plottene viser logit-skalaen. Såvel elevernes dygtighed som opgavernes sværhedsgrad måles på samme skala (logit-skala).

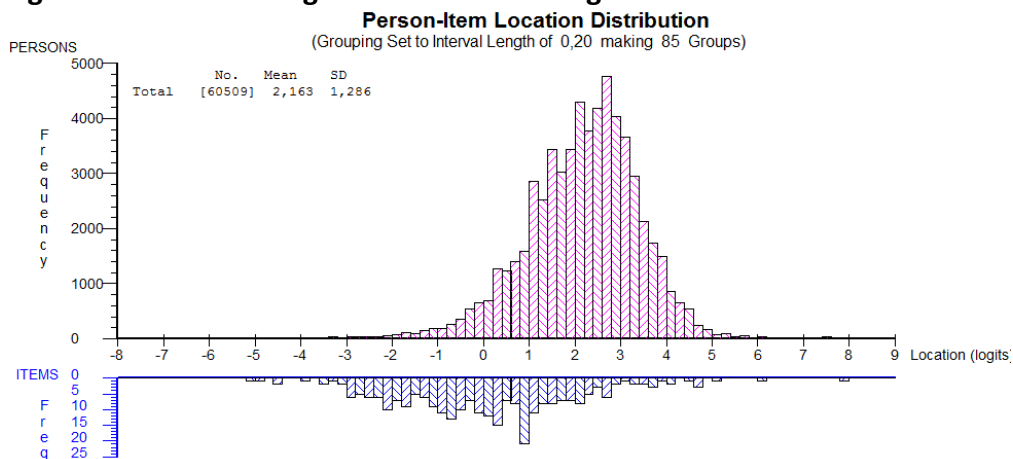
y-aksen på person-item plottene viser antallet af elever (opad) med en given dygtighed og antallet af opgaver (nedad) med en given sværhedsgrad.

Person-item plot for dansk læsning 8. klasse viser, at der mangler svære opgaver til de dygtigste elever i såvel *afkodning* (Figur 4.2) som i *tekstforståelse* (Figur 4.3).

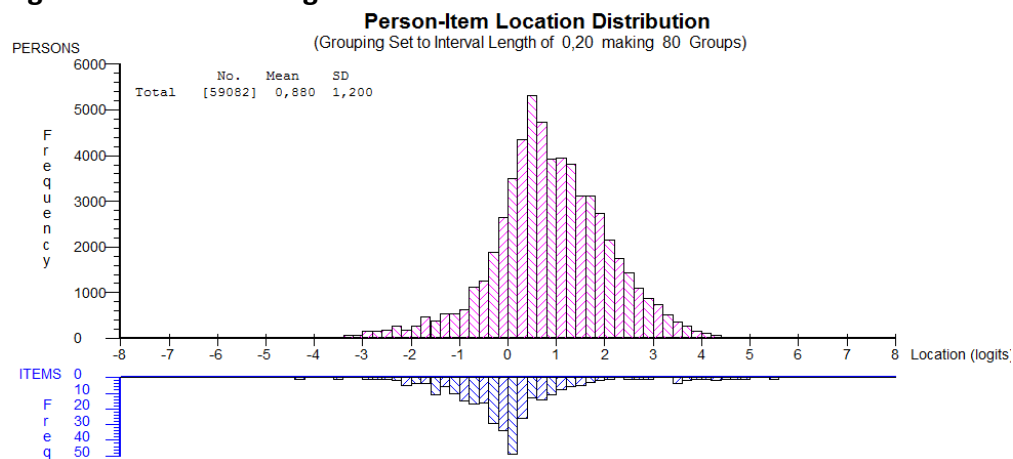
Figur 4.1 Dansk læsning 8. klasse – sprogforståelse



Figur 4.2 Dansk læsning 8. klasse – afkodning



Figur 4.3 Dansk læsning 8. klasse – tekstforståelse



Kilde: Styrelsen for It og Læring

Manglen på svære opgaver gør det primært vanskeligt at skelne de dygtigste og de allerdygtigste elever ved hjælp af testene. Endvidere bliver den statistiske usikkerhed ikke så lille, som den kunne blive, hvis der var tilstrækkeligt med opgaver, der passede til elevernes dygtighed.

I Bilag 4.4 er medtaget person-item plot for alle profilområder for de obligatoriske nationale test foråret 2018.

Person-item figurene viser, at der mangler svære opgaver i flere af profilområderne.

Stabiliteten af opgavernes sværhedsgrad over tid

I forbindelse med afprøvningen af nye opgaver til opgavebanken medtages hver gang et antal af de eksisterende og tidligere godkendte opgaver fra opgavebanken. Dette sikrer, at der sker et overlap mellem blokkene af opgaveafprøvningsopgaver, således at nye opgavers sværhedsgrad kan indplaceres på den eksisterende skala. Disse overlappingsopgaver kaldes link-opgaver. Typisk udvælges 5-10 link-opgaver ved hver opgaveafprøvning.

Anvendelsen af link-opgaver giver endvidere mulighed for, at undersøge om disse opgavers sværhedsgrad er ændret siden tidligere opgaveafprøvningsperioder. For disse opgaver foreligger der nemlig elevbesvarelser fra flere perioder.

I Rasch analysen foretages denne analyse i en Differential Item Functioning (DIF) analyse i RUMM. Der testes for DIF mellem nyeste og tidligere afprøvningsperioder.

I analyserne fra opgaveafprøvning i januar 2018 er der testet for periode DIF mellem afprøvningsperioden i januar 2018 og tidligere afprøvningsperioder (maj 2008 til september 2014) samlet.

I de analyserede profilområder blev der i alt anvendt 208 link-opgaver. 17 (8 procent) af disse udviste statistisk signifikant periode DIF. Disse opgaver er derfor tildelt en ny sværhedsgrad (item split i RUMM) i opgavebanken i skoleåret 2018/2019.

I Bilag 4.5 er opsummeret de statistiske analyser af periode DIF for de analyserede link-opgaver knyttet til profilområderne, hvor der afholdes obligatoriske test.

Analyser af opgaveafprøvningen fra januar 2019 viser, at 48 (16 procent) ud af 296 link-opgaver udviser statistisk signifikant periode DIF. Disse opgaver er derfor tildelt en ny sværhedsgrad (item split i RUMM) i opgavebanken i skoleåret 2019/2020.

Opgavernes sværhedsgrad justeres således efter hver opgaveafprøvning. Det er kun opgaver, hvor der er konstateret statistisk signifikant periode DIF, der får justeret deres opgaveparametre.

Link-opgaverne udvælges blandt opgaver, der ligger på midten af skalaen for opgavernes sværhedsgrad. Analyserne kan således ikke umiddelbart bruges til at udtale sig om, hvorvidt meget lette eller meget svære opgaver har ændret deres sværhedsgrad over tid.

Forskellen i opgavernes sværhedsgrad baseret på lineær og adaptiv testadministration

Sværhedsgraderne på opgaverne i opgavebanken er siden testenes opstart i maj 2008 estimeret på baggrund af opgaveafprøvninger, der foregår som en lineær test.

Elevbesvarelser fra obligatoriske test kan også anvendes til at estimere opgavernes sværhedsgrad. Dette har Jeppe Bundsgaard og Svend Kreiner⁵² (JB & SK) gjort for dansk læsning 8. klasse på baggrund af elevernes besvarelser i de obligatoriske test foråret 2017.

Data fra elevbesvarelser i obligatoriske test er baseret på adaptive forløb, hvor elever, der svarer rigtigt på en opgave, får stillet en sværere opgave næste gang, og elever, der svarer forkert på en opgave, får stillet en lettere opgave næste gang.

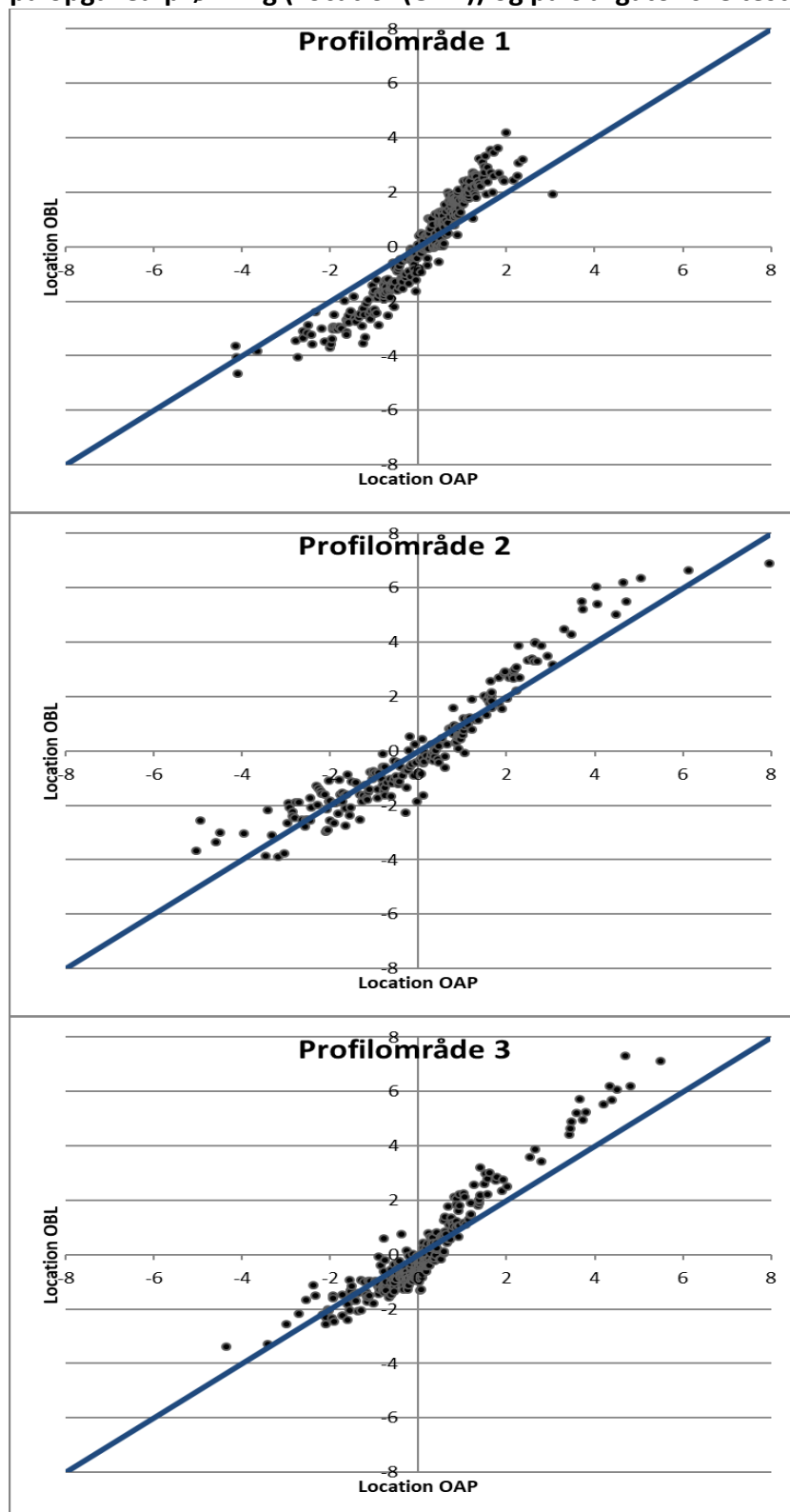
Opgavernes estimerede sværhedsgrad baseret på henholdsvis opgaveafprøvninger (Location (OAP)) og på elevbesvarelser fra de obligatoriske test (Location (OBL)) er vist i Figur 4.4 for dansk læsning i 8. klasse. Data fra elevbesvarelser fra de obligatoriske test stammer fra skoleåret 2017/2018. Sværhedsgraderne baseret på opgaveafprøvningerne er de aktuelle i opgavebanken i testsystemet.

Korrelationen mellem de estimerede opgavesværheder er på 0,95 for alle tre profilområder, mens regressionslinjens hældning er på henholdsvis 1,52, 1,04 og 1,26 i de tre profilområder.

I alle tre profilområder ses, at afvigelsen i opgavernes sværhedsgrad baseret på de to forskellige datasæt er størst for de svære opgaver.

⁵² Jeppe Bundsgaard og Svend Kreiner: Undersøgelse af De Nationale Tests måleegenskaber. Revideret 2. udgave. 25. April 2019

Figur 4.4 Estimerede opgavesværhedsgrader i dansk læsning 8. klasse baseret på opgaveafprøvning (Location(OAP)) og på obligatoriske test (Location(OBL))



Note: Profilområde 1 = Sprogforståelse; Profilområde 2 = Afkodning; Profilområde 3 = Tekstforståelse

Kilde: Styrelsen for It og Læring

Forskellen mellem opgavernes sværhedsgrad baseret på opgaveafprøvnin-
gerne og opgavernes sværhedsgrad, hvis den blev beregnet på baggrund af
elevbesvarelser fra de obligatoriske test, er beregnet for elevbesvarelserne i
de obligatoriske test i henholdsvis skoleårene 2009/2010, 2013/2014 og
2017/2018 i dansk læsning 6. og 8. klasse samt i matematik 3. og 6. klasse. I
matematik er beregningerne kun foretaget for profilområde 1 og 2, da profil-
område 3 blev erstattet med et nyt profilområde i 3. klasse i 2015/2016 og i
6. klasse i 2017/2018.

Forskellen er beregnet som forskellen mellem de estimerede opgavesværhe-
der ved anvendelse af de to forskellige typer af data:

$$\text{forskel} = \text{location(OAP)} - \text{location(OBL)}$$

hvor location(OAP) er opgavens sværhedsgrad baseret på data fra opgaveaf-
prøvnin-gerne og location(OBL) er opgavens sværhedsgrad baseret på data
fra de obligatoriske nationale test.

Desuden er den absolutte forskel beregnet:

$$\text{absolut forskel} = \text{abs}(\text{location(OAP)} - \text{location(OBL)})$$

Ændring over tid

Samlet for opgaverne i dansk læsning og matematik gælder (Tabel 4.2), at
beregningen af opgavernes sværhedsgrad afviger med mindre end 1 logit
mellem besvarelser fra opgaveafprøvnin-gerne og besvarelser fra selve test-
forløbene for 70-79 procent af opgaverne i opgavebanken.

**Tabel 4.2 Fordeling af absolut forskel i estimerede sværhedsgrader mellem
opgaveafprøvnin-gerne og elevbesvarelserne i obligatoriske test. Samlet for
dansk læsning 6. og 8. klasse og matematik 3. og 6. klasse**

Interval ¹⁾	2010	2014	2018
0,0 – 0,5	48 %	41 %	41 %
0,5 – 1,0	31 %	29 %	31 %
1,0 – 1,5	12 %	17 %	17 %
1,5 – 2,0	4 %	7 %	7 %
2,0 – 2,5	2 %	3 %	2 %
2,5 +	3 %	3 %	1 %
Forskel ²⁾	0,01	0,00	0,00
Absolut forskel ³⁾	0,68	0,79	0,75
Antal	2.171	2.333	2.843

1) Absolut forskel på logit-skalaen

2) Gennemsnitlig forskel

3) Gennemsnitlig absolut forskel

Kilde: Styrelsen for It og Læring

Den gennemsnitlige absolutte forskel i opgavernes sværhedsgrad ligger i intervallet 0,68-0,79 for de tre analyserede skoleår.

Beregningerne viser også, at andelen af opgaver, hvor den beregnede sværhedsgrad afviger, ikke ændres markant over tid. Således er fordelingen af afvigelser stort set identisk i perioden 2014 til 2018. I 2010 er der samlet lidt flere opgaver (79 procent), der afviger mindre end 1,0 logit i forhold til i 2018, hvor 72 procent afviger mindre end 1,0 logit. Til gengæld er der en større andel af opgaverne i 2010 (5 procent), der afviger med mere end 2,0 logit, i forhold til 3 procent af opgaverne i 2018.

I Bilag 4.6 er beregningerne for alle profilområderne i dansk læsning 6. og 8. klasse samt i matematik 3. og 6. klasse vist.

I profilområdet sprogforståelse 6. klasse og testforståelse 8. klasse samt i tal og algebra og i geometri i 3. og 6. klasse er andelen, der afviger med mere end 1,0 logit, mindre i 2010 end i såvel 2014 og 2018. I afkodning i 8. klasse er det modsatte tilfældet. Her er andelen, der afviger med mere end 1,0 logit, større end i 2018.

Blandt de analyserede profilområder, ses den største gennemsnitlige forskel mellem opgavernes sværhedsgrad i sprogforståelse i 8. klasse. I dette profilområde er opgavernes sværhedsgrad større, når disse er bestemt ud fra opgaveafprøvningsresultaterne i forhold til, hvis de blev fastsat ud fra besvarelserne i de obligatoriske test i både 2010 og 2014. Opgaverne i dette profilområde vurderes således lettere i de obligatoriske testforløb end ved opgaveafprøvningsresultaterne. Denne forskel er stort set væk i 2018.

De obligatoriske test blev afholdt for første gang i foråret 2010, og besvarelserne fra disse testforløb ligger således relativt tæt på de første opgaveafprøvningsresultater, der blev foretaget i 2008 og 2009. Hvis forskellen mellem opgavernes estimerede sværhedsgrad primært skyldtes, at opgavesværhederne var ændret over tid, ville det forventes, at der var stor overensstemmelse når de to perioder lå tæt på hinanden. Dette er ikke tilfældet (Tabel 4.2). I 2010 afviger sværhedsgraden med mere end 0,5 logit, når denne beregnes baseret på henholdsvis opgaveafprøvningsresultaterne og besvarelserne fra selv testforløbene, for 52 procent af opgaverne i læsning (6. og 8. klasse) og i matematik (3. og 6. klasse) og 21 procent af opgaverne afviger med mere end 1,0 logit i sværhedsgrad.

Opgavernes sværhedsgrad beregnes forskelligt for en del af opgaverne i opgavebanken, når beregningerne baseres på henholdsvis data fra opgaveafprøvningsresultater og på data fra de obligatoriske testforløb. Forskellen ændres ikke fra 2010 til 2018.

Hvis opgavernes (endelige) sværhedsgrad fremadrettet skal baseres på de obligatoriske test, vil det betyde, at elevernes dygtighed skal genberegnes, når alle test er afviklet. Først når alle test er afviklet kan opgavernes aktuelle

sværhedsgrad i populationen fastlægges og elevernes tilhørende dygtighed beregnes.

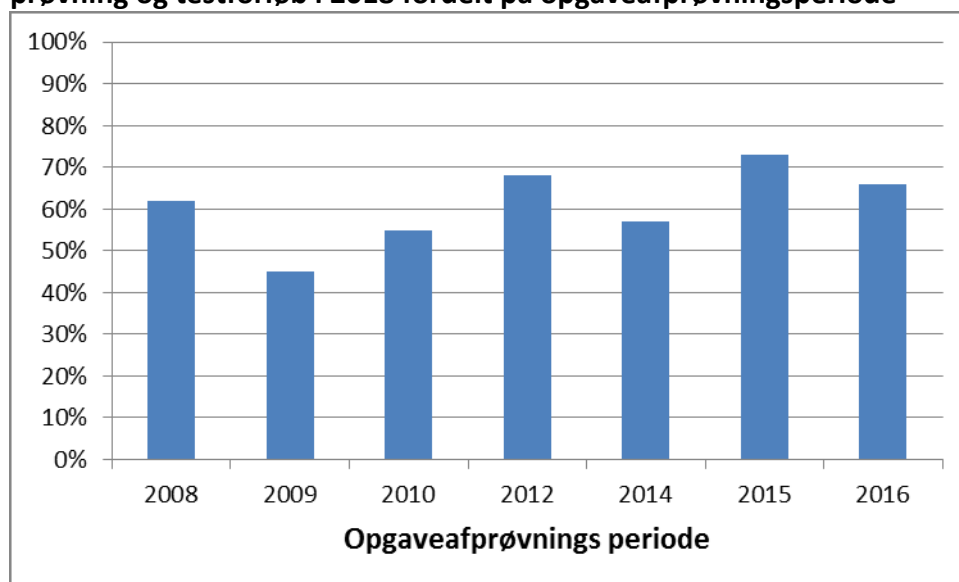
Opgavernes alder

Hvis forskellen i de estimerede sværhedsgrader baseret på henholdsvis opgaveafprøvningsdata og elevforløbene under de obligatoriske test skulle være et udtryk for, at opgavernes sværhedsgrad er ændret over tid, da ville man forvente en større afvigelse des længere tid, der er gået siden opgaverne er afprøvet i forhold til afviklingen af de obligatoriske test. Forskellen ville forventes at være begrænset for de senest afprøvede opgaver.

Dette kan ikke bekræftes (Figur 4.5). For opgaver afprøvet i 2009 og 2010 ligger den beregnede sværhedsgrad i 45 - 55 procent af tilfældene udenfor $\pm 0,5$ logit af den sværhedsgrad, der kan beregnes på baggrund af elevernes besvarelser i de obligatoriske test i foråret 2018. For opgaver afprøvet 5-7 år senere, i 2015 og 2016, ligger den beregnede sværhedsgrad i 66 - 73 procent af tilfældene udenfor $\pm 0,5$ logit af den sværhedsgrad, der kan beregnes på baggrund af elevernes besvarelser i de obligatoriske test i foråret 2018.

Andelen af opgaver med stor afvigelse i estimeret sværhedsgrad baseret på opgaveafprøvningsdata og testforløbsdata stiger således ikke, når den tidsmæssige afstand mellem de to dataindsamlinger øges.

Figur 4.5 Andel opgaver med forskel¹⁾ i sværhedsgrad mellem opgaveafprøvningsperiode og testforløb i 2018 fordelt på opgaveafprøvningsperiode



1) Forskel større end 0,5 logit

Baseret på Bilag 4.6

Kilde: Styrelsen for It og Læring

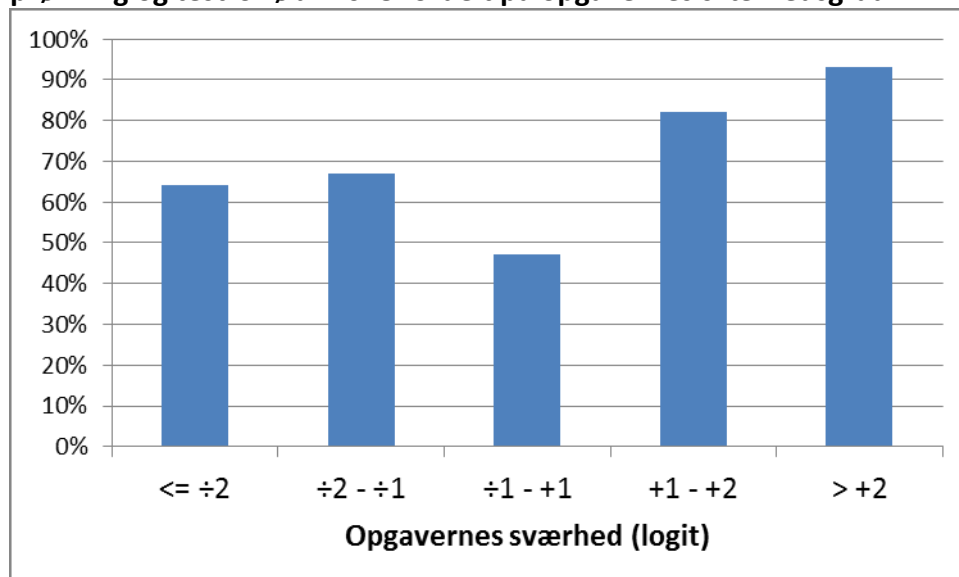
Forskellen i opgavernes sværhedsgrad, når beregningerne baseres på data fra opgaveafprøvninger, og når beregningerne baseres på obligatoriske testforløb, afhænger ikke af, hvor længe siden opgaverne er afprøvet og deres sværhedsgrad er beregnet.

Opgavernes sværhedsgrad

Derimod er der en sammenhæng mellem opgavernes sværhedsgrad og forskellen i de estimerede sværhedsgrader baseret på opgaveafprøvningerne og elevforløbene under de obligatoriske test (Figur 4.6).

47 procent af opgaverne med en estimeret sværhedsgrad i midten af skalaen (logit: $\div 1$; $+1$) afviger med mere end 0,5 logit fra den sværhedsgrad, der kan estimeres på baggrund af elevforløbene i de obligatoriske test i 2018. For de svære opgaver (logit: $> +2$) afviger 93 procent med mere end 0,5 logit.

Figur 4.6 Andel opgaver med forskel¹⁾ i sværhedsgrad mellem opgaveafprøvning og testforløb i 2018 fordelt på opgavernes sværhedsgrad



1) Forskel større end 0,5 logit

Baseret på Bilag 4.6

Kilde: Styrelsen for It og Læring

Den absolutte forskel i opgavernes beregnede sværhedsgrad er mindst for opgaver med en sværhedsgrad midt på skalaen og størst for de lette og de svære opgaver (Tabel 4.3).

Fortegnet på forskellen i opgavernes beregnede sværhedsgrad viser, at de sværeste opgaver, med en estimeret logit-værdi på over $+1,0$, vurderes endnu sværere i de obligatoriske testforløb i 2017/2018. De lettere opgaver, med en estimeret logit-værdi under $\div 1,0$, vurderes omvendt lettere i de obligatoriske testforløb.

**Tabel 4.3 Forskel i estimerede sværhedsgrader mellem opgaveafprøvnin-
gerne og elevbesvarelsene i obligatoriske test 2017/2018 i forhold til opga-
vernes sværhedsgrad. Samlet for dansk læsning 6. og 8. klasse og matema-
tik 3. og 6. klasse. Opgaver i opgavebanken i 2018**

	Opgavernes sværhedsgrad ¹⁾				
	(N=2.843)				
	(÷ ; ÷2] (N=195)	(÷2 ; ÷1] (N=400)	(÷1 ; +1] (N=1.656)	(+1 ; +2] (N=399)	(+2 ; +) (N=193)
Forskel¹⁾	0,30	0,77	0,22	-1,04	-1,57
Abs. Forskel²⁾	0,76	0,85	0,55	1,07	1,59

1) Gennemsnitlig forskel

2) Gennemsnitlig absolut forskel

Kilde: Styrelsen for It og Læring

Der er en sammenhæng mellem opgavernes sværhedsgrad og forskellen i de estimerede sværhedsgrader baseret på opgaveafprøvnings- og elevforløbene under de obligatoriske test.

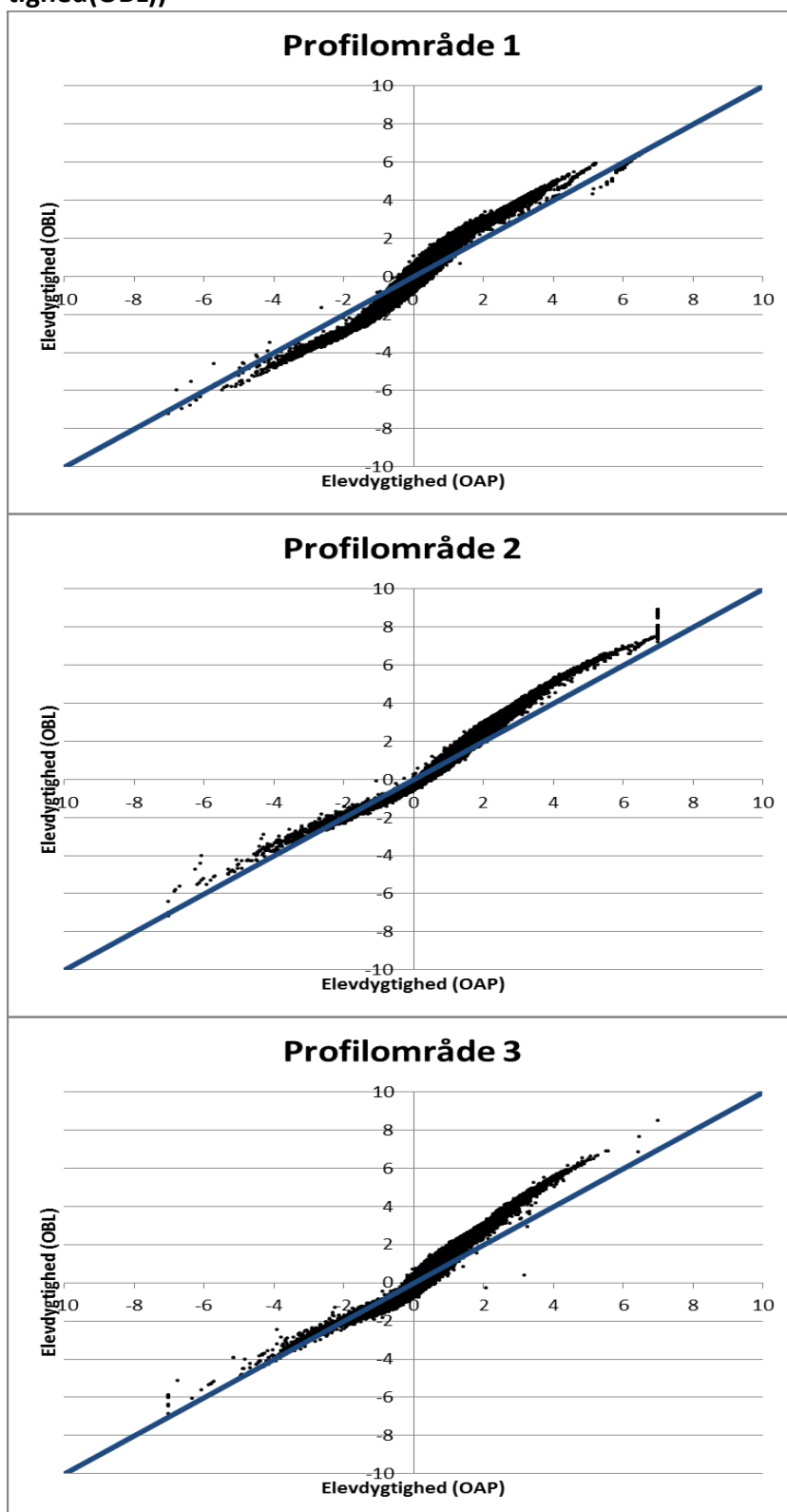
Forskel i beregnet elevdygtighed

Elevdygtighederne i de nationale obligatoriske test i skoleåret 2017/2018 kan beregnes ved at anvende opgavernes estimerede sværhedsgrad baseret på elevbesvarelses fra de obligatoriske test (Figur 4.7). Disse elevdygtigheder er beregnet i RUMM og sammenholdes med de elevdygtigheder, der beregnes i testsystemet. Elevdygtighederne i testsystemet er baseret på opgavesværheder, der er beregnet på baggrund af opgaveafprøvnings-

De dygtigste elever, der får de sværeste opgaver i den adaptive algoritme i dansk læsning 8. klasse, vil få en større beregnet elevdygtighed, hvis der blev anvendt opgavesværheder baseret på de samme elevers besvarelses fra de obligatoriske test. Ændringen i elevdygtighederne er mindre for elever med en elevdygtighed midt i skalaen.

Korrelationen mellem de beregnede elevdygtigheder på logit-skalaen ved de to metoder er på 0,98, 0,99 og 0,98 i de tre profilområder.

Figur 4.7 Beregnede elevdygtigheder i dansk læsning 8. klasse baseret på opgaveafprøvning (Elevdygtighed(OAP)) og på obligatoriske test (Elevdygtighed(OBL))



Note: Profilområde 1 = Sprogforståelse; Profilområde 2 = Afkodning; Profilområde 3 = Tekstforståelse

Kilde: Styrelsen for It og Læring

De høje korrelationer (0,98-0,99) mellem de beregnede elevdygtigheder indikerer, at rangordningen mellem eleverne i stor udstrækning vil være bevaret, hvis elevernes dygtighed blev beregnet baseret på opgavesværheder estimeret i de adaptive obligatoriske test.

Sammenligning af elevernes dygtighed på percentilskalaen kræver, at elevernes dygtigheder baseret på opgavesværheder fra de adaptive obligatoriske test omregnes fra logit-skalaen til en normbaseret percentilskala. Denne omregning skal foretages på samme udgangspopulation som for de nuværende, dvs på resultaterne fra elevernes testafvikling i foråret 2010 (Bilag 1.1).

5 Notat 5 Samling af testresultater fra flere profilområder

Sammenfatning

- Analyser af besvarelsenerne fra de obligatoriske test i 2017/2018 viser, at de tre profilområder i dansk læsning 8. klasse godt kan antages at måle forskellige egenskaber af én og samme færdighed. Analyser viser tilsvarende resultat for matematik 6. klasse
- Som supplement til den beregnede elevdygtighed i hvert af de tre profilområder kan der derfor beregnes én samlet elevdygtighed i henholdsvis dansk læsning 8. klasse og i matematik 6. klasse
- Den statistiske usikkerhed på elevernes estimerede samlede dygtighed er i gennemsnit på ca. 0,30 logit, hvor den i gennemsnit i hvert af de analyserede profilområder ligger på 0,47-0,52 logit
- Det er ikke undersøgt om profilområder i de øvrige nationale test på samme vis måler én og samme færdighed

Indledning

I evalueringen af de nationale test skal det afdækkes:

2. Om sikkerheden i målingerne af elevernes færdigheder kan forbedres ved at kombinere resultater fra forskellige profilområder. Herunder skal det følgende klarlægges:

- a. Kan det påvises, at profilområderne måler forskellige aspekter af den samme bagvedliggende færdighed?
- b. Kan testresultaterne fra profilområderne slås sammen og dermed forbedre sikkerheden i testene.

I notatet beskrives sammenhængen mellem elevernes resultater i de tre profilområder i henholdsvis dansk læsning 8. klasse og matematik 6. klasse. Besvarelserne fra de nationale obligatoriske test i skoleåret 2017/2018 anvendes til analyserne.

Formålet er at vurdere, om elevernes resultater fra tre profilområder kan samles til ét samlet resultat med en større statistisk sikkerhed.

Notatet indeholder følgende afsnit:

- Korrelationen mellem elevresultater
- Antal besvarede opgaver og SEM i profilområderne
- Analyse af mulighed for anvendelse af én Rasch-model pr. test
- Samlet elevdygtighed

Korrelationen mellem elevresultater

De nationale test tester elevernes dygtighed i udvalgte områder og fag. I hvert fag testes eleverne inden for tre hovedområder, der kaldes profilområder. Elevernes dygtighed beregnes i hvert profilområde ud fra de besvarelser eleven har givet på en række opgaver.

Sammenhængen mellem elevernes beregnede dygtighed i de tre profilområder i dansk læsning 8. klasse og i matematik 6. klassen ses i Tabel 5.1. Beregningerne er baseret på elevbesvarelser fra de obligatoriske test i foråret 2018.

Tabel 5.1 Korrelationer mellem elevdygtigheder. Obligatoriske test 2017/2018

Test		Profilområde 2	Profilområde 3
Dansk læsning 8. klasse	Profilområde 1	0,48	0,36
	Profilområde 2	-	0,60
Matematik 6. klasse	Profilområde 1	0,68	0,71
	Profilområde 2	-	0,70

Kilde: Styrelsen for It og Læring

Alle korrelationer er statistisk signifikant forskellige fra nul. I dansk læsning 8. klasse er der specielt en stærk korrelation mellem afkodning (profilområde 2) og tekstforståelse (profilområde 3). I matematik 6. klasse er der en stærk korrelation mellem alle tre profilområder.

Der er således en sammenhæng mellem elevernes dygtighed i de enkelte profilområder i en test. De dygtigste elever i ét profilområde er også blandt de dygtigste i de to andre profilområder.

Antal besvarede opgaver og SEM i profilområderne

Elevernes testresultater afrapporteres i hvert profilområde for sig. I gennemsnit når eleverne at besvare 15-19 opgaver i hvert profilområde i dansk læsning 8. klasse og matematik 6. klasse (Tabel 5.2).

Tabel 5.2 Det gennemsnitlige antal opgaver eleverne når at besvare i de obligatoriske test i 2017/2018.

Test	Profilområde 1	Profilområde 2	Profilområde 3	Samlet
Dansk læsning 8. klasse	19	19	17	55
Matematik 6. klasse	15	15	18	48

Kilde: Styrelsen for It og Læring

Den gennemsnitlige observerede statistiske usikkerhed (SEM) på den beregnede elevdygtighed i dansk læsning 8. klasse og matematik 6. klasse er på 0,47-0,52 logit (Tabel 5.3).

Tabel 5.3 Den gennemsnitlige statistiske usikkerhed (logit) på den estimerede elevdygtighed i skoleåret 2017/2018.

Test	Profilområde 1	Profilområde 2	Profilområde 3
Dansk læsning 8. klasse	0,51	0,51	0,47
Matematik 6. klasse	0,47	0,48	0,52

Kilde: Styrelsen for It og Læring

Hvis elevernes testresultater i to eller tre profilområder kan samles til ét fælles mål for elevens dygtighed, da vil denne beregnede dygtighed være baseret på besvarelser af flere opgaver, og dermed vil den statistiske usikkerhed på elevens resultat være mindre end den statistiske usikkerhed på dygtigheden i hvert af profilområderne.

Analyse af mulighed for anvendelse af én Rasch-model pr. test

Ved at samle den enkelte elevs besvarelser fra alle tre profilområder i henholdsvis dansk læsning 8. klasse og matematik 6. klasse fra de obligatoriske test i 2017/2018 er det undersøgt, om opgaverne fra de tre profilområder kan samles til én opgavebank – én dimension. Hvis elevernes besvarelser af opgaver i de tre profilområder ikke giver anledning til at forkaste hypotesen om én samlet Rasch model, da kan opgavernes parametre estimeres og indplaceres på én samlet logit-skala.

Alle analyser af besvarelsene er foretaget i software programmet RUMM⁵³.

Analyserne i dansk læsning 8. klasse er baseret på besvarelser fra 48.481 elever, mens analyserne i matematik 6. klasse er baseret på besvarelser fra 53.043 elever.

I foråret 2018 indeholdt opgavebanken i dansk læsning 8. klasse 823 opgaver samt 1.019 opgaver i matematik 6. klasse (Tabel 5.4).

Tabel 5.4 Opgavebanken dansk læsning 8. klasse og matematik 6. klasse i foråret 2018

Fag	Profilområde 1	Profilområde 2	Profilområde 3	Samlet
Dansk læsning 8. klasse	308	214	301	823
Matematik 6. klasse	494	263	262	1.019

Kilde: Styrelsen for It og Læring

De statistiske analyser i RUMM viser, at 12 ud af de 823 opgaver udviser misfit til én samlet Rasch model i dansk læsning 8. klasse, og 6 ud af de 1.019 opgaver udviser misfit til én samlet Rasch model i matematik 6. klasse. Disse opgaver skal derfor fjernes inden, der beregnes elevdygtigheder.

Én samlet opgavebank i dansk læsning 8. klasse består således af 811 opgaver og i matematik 6. klasse af 1.013 opgaver.

Baseret på elevbesvarelser fra de obligatoriske test i skoleåret 2017/2018 kan det således ikke afvises, at de tre profilområder i henholdsvis dansk læsning 8. klasse og matematik 6. klasse måler forskellige aspekter af én og samme færdighed.

Hvorvidt dette også gælder for de øvrige test er ikke undersøgt.

Samlet elevdygtighed

Baseret på elevbesvarelsene fra de Rasch godkendte opgaver kan én samlet elevdygtighed i henholdsvis dansk læsning 8. klasse og i matematik 6. klasse beregnes.

Det gennemsnitlige antal opgaver samt den beregnede statistiske usikkerhed ses i Tabel 5.5.

Tabel 5.5 Det gennemsnitlige antal opgaver samt den gennemsnitlige statistiske usikkerhed (SEM) på den estimerede elevdygtighed i skoleåret 2017/2018. Gennemsnit [5% percentil; 95% percentil]

Test	Antal opgaver	SEM
Dansk læsning 8. klasse	55	0,29 [0,22; 0,35]
Matematik 6. klasse	48	0,30 [0,22; 0,37]

Kilde: Styrelsen for It og Læring

⁵³ www.rummlab.com.au

En samlet beregnet elevdygtighed i hver test vil være baseret på ca. 50-60 opgaver. Den statistiske usikkerhed på den samlede elevdygtighed er 0,30 logit, mens usikkerheden på elevdygtigheden i hvert profilområde for sig er på 0,47-0,52 logit.

Korrelationerne mellem elevernes samlede beregnede dygtighed og den beregnede dygtighed i hvert af de tre profilområder i dansk læsning 8. klasse og i matematik 6. klassen er store (Tabel 5.6) og alle statistisk signifikant forskellige fra nul.

Tabel 5.6 Korrelationer mellem samlet beregnet elevdygtighed og elevdygtighederne i hvert af de tre profilområder. Obligatoriske test 2017/2018

Test	Profilområde 1	Profilområde 2	Profilområde 3
Dansk læsning 8. klasse	0,71	0,86	0,82
Matematik 6. klasse	0,89	0,88	0,89

Kilde: Styrelsen for It og Læring

Ud fra rent statistiske betragtninger vil det derfor være muligt at supplere de nuværende resultater i de enkelte profilområder med ét samlet elevresultat med en betydelig større statistisk sikkerhed end, der kan opnås i hvert profilområde for sig.

Der udestår en faglig indholdsmæssig afklaring af muligheden for samling af testresultater fra flere profilområder til ét samlet mål.

Bilag 1.1 Anvendte skalaer til præsentation af elevernes beregnede dygtigheder

Elevdygtigheden og den tilhørende usikkerhed beregnes i den adaptive algoritme i testsystemet på en **logit-skala**, som i princippet går fra minus til plus uendelig, men som i praksis går fra minus 7 til plus 7. I de fleste profilområder er intervallet dog snævrere.

Siden skoleåret 2009/2010 har resultaterne fra de nationale test været formidlet på den **normbaserede skala (1-100)**, der er en **percentilskala**. Percentilskalaen er dannet på baggrund af fordelingen af elevernes testresultater i 2010⁵⁴, hvor en percentilværdi på fx 40 svarer til den elevdygtighed på logit skalaen, hvor 40 procent af testresultaterne i 2010 lå under. En percentilværdi på 90 svarer til den elevdygtighed, hvor 90 procent af eleverne i 2010 scorede under, dvs kun 10 procent af eleverne i 2010 opnåede et bedre testresultat.

I formidlingen til elever og forældre bliver resultaterne på 100-skalaen omsat til en **femtrins skala**:

1. En del under gennemsnittet (1-10)
2. Under gennemsnittet (11-35)
3. Gennemsnittet (35-65)
4. Over gennemsnittet (66-90)
5. En del over gennemsnittet (91-100)

Logit skalaen har den fordel, at det er en interval skala. Forskellen mellem to værdier har samme betydning overalt på skalaen. Således er fx en forbedring eller en forskel fra -2 til -1 den samme som en forbedring fra 0 til +1.

Dette gælder ikke for percentilskalaen. Her vil en forskel eller en forbedring på fx 10 point svarer til en relativ beskeden ændring i elevdygtigheden på logit-skalaen, hvis der er tale om en middeldygtig elev (30-70), mens det svarer til en større ændring i elevdygtigheden på logit-skalaen, hvis der er tale om en meget dygtig elev (80-100) eller en mindre dygtig elev (1-20).

Siden skoleåret 2014/2015 har resultaterne i dansk læsning og matematik desuden været præsenteret på den **kriteriebaserede skala**. Ved den kriteriebaserede scoring omregnes elevens dygtighed på logit-skalaen til en forventet score på et antal udvalgte opgaver.

⁵⁴ Omregning fra logit-skala til percentilskala sker ved anvendelse af en stykvis sigmoid-funktion

Opgavekommissionerne har fastsat kriterier for, hvilket fagligt niveau en score stemmer overens med. I hvert profilområde er defineret seks faglige niveauer:

1. Ikke tilstrækkelig præstation
2. Mangelfuld præstation
3. Jævn præstation
4. God præstation
5. Rigtig god præstation
6. Fremragende præstation

Bilag 1.2 Opgavebanken i dansk læsning 8. klasse – sprogforståelse

Opgaver i opgavebanken i dansk læsning 8. klasse sprogforståelse. Kun opgaver, der var aktive⁵⁵ i foråret 2018.

Tablet 1.2.1 Opgavenummer med tilhørende sværhedsgrad (location)

opgavenummer	location
0108010410311	-4,135
0108010420046	-4,128
0108010410186-1	-4,105
0108010410344	-3,694
0108010420016	-3,657
0108010415103	-2,771
0108010410340	-2,741
0108010420126	-2,616
0108010410369	-2,610
0108010420032	-2,525
0108010410410-1	-2,504
0108010410343	-2,434
0108010410080	-2,399
0108010410316	-2,333
0108010440022	-2,187
0108010415102	-2,133
0108010415171	-2,007
0108010415190	-1,984
0108010415179	-1,958
0108010440027	-1,945
0108010410337	-1,932
0108010420021	-1,922
0108010420101	-1,920

⁵⁵ Tidligere Rasch-godkendte opgaver kan være taget ud af opgavebanken

opgavenummer	location
010801000301236064-1	-1,876
0108010415164	-1,848
0108010410230042	-1,792
0108010410335	-1,790
0108010410376	-1,752
0108010410327	-1,672
0108010415193	-1,641
0108010420034	-1,625
0108010440019	-1,622
0108010420071	-1,619
0108010410385	-1,614
0108010410339	-1,584
0108010420060	-1,573
010801000301239199-1	-1,527
0108010410230045	-1,474
010801000301234830-1	-1,449
0108010410097	-1,415
0108010420064	-1,351
0108010420029	-1,340
0108010415194	-1,322
0108010420044	-1,272
0108010410230025	-1,257
0108010410408-1	-1,247
0108010420041	-1,242
0108010415169	-1,204
0108010410395	-1,192
0108010415157	-1,140
0108010420012	-1,121
0108010420067	-1,086
010801000301238278-1	-1,064

opgavenummer	location
0108010440021	-1,043
0108010410401-1	-1,020
0108010410358	-1,005
0108010420002	-1,000
0108010410402-1	-0,989
010801000301238283-1	-0,944
0108010410098	-0,938
0108010440014	-0,923
0108010410406-1	-0,912
0108010415139	-0,906
0108010410325	-0,871
010801000301239200-1	-0,834
0108010440028	-0,807
0108010410093	-0,793
0108010415178	-0,786
010801000301238279-1	-0,779
0108010415175	-0,778
0108010420042	-0,769
0108010410333	-0,767
0108010420116	-0,739
0108010410230021	-0,733
0108010410230032	-0,713
010801000301238277-1	-0,680
0108010410096	-0,680
010801000301238994-1	-0,652
0108010415117	-0,648
0108010415182	-0,629
0108010410187-1	-0,622
0108010420027	-0,617
0108010410110-1	-0,612

opgavenummer	location
010801000301238285-1	-0,576
0108010440042	-0,574
0108010410230029	-0,570
0108010410230009	-0,567
0108010420084	-0,544
0108010410230020	-0,526
0108010420050	-0,515
0108010420023	-0,472
0108010410411-1	-0,467
0108010420033	-0,466
0108010410230039	-0,437
0108010415140	-0,437
0108010410230014	-0,412
0108010415132	-0,395
0108010420104	-0,384
0108010410230041	-0,376
0108010415129	-0,372
0108010420010	-0,352
0108010420030	-0,325
010801000301239235-1	-0,295
010801000301234835-1	-0,286
0108010410351	-0,283
0108010410400-1	-0,279
0108010410320	-0,256
0108010420161	-0,227
010801000301238995-1	-0,216
0108010410230028	-0,180
010801000301238837-1	-0,178
0108010410414-1	-0,178
0108010410230030	-0,163

opgavenummer	location
0108010440036	-0,143
0108010420150	-0,134
0108010410413-1	-0,108
0108010420003	-0,103
0108010440031	-0,102
010801000301239197-1	-0,073
0108010420141	-0,072
0108010410230018	-0,067
0108010420131	-0,057
0108010440013	-0,057
0108010420045	-0,052
0108010420073	-0,049
0108010440010	-0,044
0108010410230046	-0,021
0108010420058	-0,015
0108010410155-1	-0,001
0108010420136	0,008
0108010420092	0,020
0108010415159	0,050
0108010420048	0,062
0108010420142	0,068
0108010420024	0,074
0108010440037	0,080
0108010415160	0,090
0108010440040	0,102
0108010440025	0,109
0108010410230013	0,116
0108010410230022	0,124
0108010415130	0,132
0108010440001	0,167

opgavenummer	location
0108010420053	0,171
0108010410230038	0,174
0108010415111	0,192
0108010410230008	0,193
0108010410357	0,193
0108010410230017	0,224
010801000301238353-1	0,232
0108010415119	0,237
0108010420094	0,239
0108010420102	0,245
0108010410230023	0,257
0108010410230019	0,271
0108010420061	0,294
0108010410230005	0,305
0108010415122	0,322
010801000301238282-1	0,326
0108010415135	0,347
0108010410397	0,356
0108010420049	0,356
0108010415134	0,380
0108010420066	0,393
0108010420040	0,394
0108010420019	0,398
0108010420151	0,402
0108010410393	0,403
0108010415145	0,409
0108010410083	0,433
010801000301238996-1	0,448
0108010420149	0,458
0108010420059	0,462

opgavenummer	location
010801000301238927-1	0,463
0108010420130	0,463
0108010420043	0,464
0108010420074	0,470
010801000301238555-1	0,479
0108010420112	0,479
0108010420070	0,498
0108010420157	0,499
0108010410398	0,505
0108010410392	0,512
0108010420013	0,516
0108010420113	0,534
0108010420162	0,547
0108010410230031	0,550
0108010420154	0,550
010801000301234832-1	0,554
0108010415133	0,566
0108010410366	0,569
0108010420155	0,580
0108010420018	0,583
010801000301238607-1	0,591
0108010420156	0,614
0108010410373	0,616
010801000301234806-1	0,628
0108010420106	0,635
0108010420014	0,664
0108010420100	0,667
0108010410230034	0,668
0108010415186	0,674
0108010420114	0,696

opgavenummer	location
0108010420120	0,700
0108010415120	0,702
0108010415173	0,714
0108010410230012	0,733
010801000301238839-1	0,735
010801000301238925-1	0,743
010801000301238281-1	0,745
010801000301239195-1	0,760
0108010420140	0,760
010801000301238835-1	0,775
0108010410145-1	0,793
010801000301234815-1	0,800
0108010410088	0,801
0108010420088	0,802
0108010410377	0,805
0108010440045	0,805
010801000301234831-1	0,811
0108010420152	0,817
0108010420031	0,821
010801000301238999-1	0,836
010801000301238873-1	0,843
0108010410328	0,844
0108010420105	0,849
0108010410230024	0,851
0108010420054	0,853
010801000301238836-1	0,862
0108010420147	0,868
0108010420138	0,870
010801000301234836-1	0,875
0108010420132	0,877

opgavenummer	location
010801000301234842-1	0,879
0108010415165	0,893
0108010420028	0,897
0108010410230037	0,901
0108010420068	0,904
0108010410372	0,917
0108010420079	0,920
010801000301234838-1	0,930
0108010440007	0,930
0108010410384	0,931
0108010415158	0,951
010801000301238872-1	1,003
0108010410378	1,009
0108010440006	1,020
0108010415183	1,026
0108010420087	1,026
0108010410230047	1,047
0108010410388	1,049
010801000301238556-1	1,075
010801000301234823-1	1,079
0108010410084	1,081
0108010420129	1,085
0108010415118	1,095
0108010420153	1,097
0108010420109	1,102
0108010440034	1,104
0108010410230035	1,135
0108010410405-1	1,143
0108010420103	1,152
010801000301234840-1	1,155

opgavenummer	location
0108010410399	1,160
0108010415167	1,160
010801000301234789-1	1,202
0108010420017	1,235
010801000301238862-1	1,248
0108010420160	1,249
0108010420056	1,259
0108010415109	1,266
0108010410412-1	1,282
010801000301234825-1	1,288
0108010415113	1,294
010801000301238920-1	1,299
010801000301238923-1	1,306
0108010410368	1,318
010801000301234817-1	1,346
010801000301239000-1	1,367
0108010415151	1,370
0108010415153	1,381
010801000301234826-1	1,396
010801000301238997-1	1,445
010801000301234843-1	1,463
0108010410350	1,464
0108010420122	1,465
010801000301234839-1	1,495
010801000301234804-1	1,505
0108010410407-1	1,524
010801000301236073-1	1,532
010801000301234820-1	1,552
010801000301234802-1	1,563
010801000301234807-1	1,566

opgavenummer	location
010801000301234829-1	1,634
010801000301236068-1	1,649
010801000301234824-1	1,685
0108010415124	1,713
0108010420095	1,716
0108010410379	1,807
0108010410094	1,832
010801000301234828-1	1,912
010801000301234833-1	1,937
0108010415180	1,981
0108010420015	2,151
010801000301234841-1	2,248
010801000301234837-1	2,263
010801000301239196-1	2,371
0108010410315	3,044

Kilde: Styrelsen for It og Læring

Bilag 2.1 Sammenhæng mellem testresultater og karakterer

Elevernes karakter i dansk, læsning i folkeskolens prøve i 9. klasse i 2017/2018 sammenholdes med elevernes testresultater i de obligatoriske nationale test i dansk, læsning i 8. klasse i 2016/2017. Tilsvarende sammenholdes elevernes karakter i matematik uden hjælpemidler i folkeskolens prøve i 9. klasse i 2017/2018 med elevernes testresultater i de obligatoriske nationale test i matematik i 6. klasse i 2014/2015.

Elevernes karakter i dansk, læsning i standpunktsprøven i 8. klasse i 2017/2018 sammenholdes med elevernes testresultater i de obligatoriske nationale test i dansk, læsning i 8. klasse i 2017/2018. Tilsvarende sammenholdes elevernes karakter i matematik uden hjælpemidler i standpunktsprøven i 8. klasse med elevernes testresultater i de obligatoriske nationale test i matematik i 8. klasse i 2017/2018.

Tabellerne indeholder:

- Gns.: Gennemsnitskarakter i gruppen af elever
- SD: Standardafvigelsen

Tabel 2.1.1 Resultaterne fra de nationale test (normbaseret skala) og prøvekaraktererne i 9. klasse i 2017/2018. Nationale test i dansk læsning 8. klasse 2016/2017 samt i matematik 6. klasse 2014/2015

Point	Dansk læsning		Matematik u. hjælpemidler	
	Gns.	SD	Gns.	SD
1-10	1,6	2,0	2,0	2,2
11-20	2,6	2,1	2,7	2,1
21-30	3,5	2,3	3,6	2,0
31-40	4,2	2,4	4,5	2,2
41-50	4,9	2,5	5,3	2,2
51-60	5,8	2,6	6,2	2,2
61-70	6,6	2,6	7,1	2,2
71-80	7,6	2,7	8,2	2,2
81-90	8,6	2,6	9,3	2,2
91-100	9,5	2,5	10,5	1,9

Kilde: Styrelsen for It og Læring

Tabel 2.1.2 Resultaterne fra de nationale test (kriteriebaseret skala) og prøvekaraktererne i 9. klasse i 2017/2018. Nationale test i dansk læsning 8. klasse 2016/2017 samt i matematik 6. klasse 2014/2015

Kriterie	Dansk læsning		Matematik u. hjælpemidler	
	Gns.	SD	Gns.	SD
Ikke tilstrækkelig	1,2	2,0	2,0	2,4
Mangelfuld	2,5	2,2	3,2	2,2
Jævn	3,8	2,4	4,7	2,2
God	5,5	2,6	6,8	2,4
Rigtig god	7,2	2,8	9,3	2,2
Fremragende	8,4	2,9	10,7	1,8

Kilde: Styrelsen for It og Læring

Tabel 2.1.3 Resultaterne fra de nationale test (normbaseret skala) og standpunktsskaraktererne i 8. klasse i 2017/2018

Point	Dansk læsning		Matematik u. hjælpemidler	
	Gns.	SD	Gns.	SD
1-10	2,6	1,8	2,3	1,7
11-20	3,5	1,8	3,6	1,7
21-30	4,3	1,9	4,6	1,8
31-40	5,2	2,1	5,5	2,0
41-50	5,9	2,2	6,4	1,9
51-60	6,8	2,3	7,3	1,9
61-70	7,6	2,4	8,1	1,9
71-80	8,5	2,3	9,0	1,8
81-90	9,3	2,2	10,0	1,5
91-100	10,1	2,0	11,0	1,2

Kilde: Styrelsen for It og Læring

Tabel 2.1.4 Resultaterne fra de nationale test (kriteriebaseret skala) og standpunktsskaraktererne i 8. klasse i 2017/2018

Kriterie	Dansk læsning		Matematik u. hjælpemidler	
	Gns.	SD	Gns.	SD
Ikke tilstrækkelig	2,1	1,7	1,7	1,7
Mangelfuld	3,4	1,8	2,4	1,7
Jævn	4,7	2,0	3,7	1,8
God	6,4	2,4	6,3	2,2
Rigtig god	8,2	2,4	9,1	1,9
Fremragende	9,1	2,4	10,9	1,3

Kilde: Styrelsen for It og Læring

Bilag 3.1 Statistisk usikkerhed på elevdygtighederne

Bilaget indeholder oversigt over den beregnede SEM i forhold til elevernes estimerede dygtighed.

I Rasch-modellen beregnes elevernes dygtighed på logit-skalaen. I testsystemet omregnes resultaterne til den normbaserede skala, der er en simpel percentilskala (1-100), ved anvendelse af en stykvis sigmoid funktion. Elevernes dygtighed er inddelt på femtrins skalaen, der er en gruppering af den normbaserede (1-100) skala.

Tabel 3.1.1 Femtrins skala

Fem trins skala	Normbaseret skala (1-100)
1	1-10
2	11-35
3	35-65
4	66-90
5	91-100

I dansk læsning og matematik omregnes resultaterne endvidere til en kriteriebaseret skala på seks trin.

Tabel 3.1.2 Kriteriebaseret skala

Trin	Kriteriebaseret skala
1	Ikke tilstrækkelig præstation
2	Mangelfuld præstation
3	Jævn præstation
4	God præstation
5	Rigtig god præstation
6	Fremragende præstation

Beregningen af sikkerhedsintervaller er kun foretaget for elever, der har afsluttet testen i teststatus 'grøn', dvs hvor den statistiske usikkerhed på elevens estimerede dygtighed er kommet under 0,55 logit, eller eleven har besvaret mindst 30 opgaver i hvert af de tre profilområder.

Usikkerhedsintervallerne er beregnet på logit-skalaen baseret på den enkelte elevs beregnede statistiske usikkerhed og herefter omsat til og præsenteret på såvel den normbaserede (1-100) skala som på den kriteriebaserede skala.

Tablet 3.1.3 Den statistiske usikkerhed (logit) på den estimerede elevdygtighed i skoleåret 2017/2018 i forhold til elevdygtigheden. Dansk læsning

Klassetrin	Profilområde	Elevdygtighed på femtrins skalaen				
		1-10	11-35	36-65	66-90	91-100
2. klasse	1	0,40	0,43	0,45	0,48	0,54
	2	0,42	0,43	0,46	0,48	0,54
	3	0,43	0,36	0,40	0,45	0,54
4. klasse	1	0,46	0,47	0,47	0,48	0,51
	2	0,46	0,47	0,48	0,49	0,52
	3	0,42	0,34	0,43	0,48	0,52
6. klasse	1	0,48	0,48	0,48	0,48	0,51
	2	0,48	0,48	0,48	0,50	0,56
	3	0,41	0,37	0,46	0,48	0,51
8. klasse	1	0,49	0,49	0,50	0,50	0,56
	2	0,49	0,49	0,50	0,51	0,59
	3	0,44	0,38	0,45	0,51	0,55

Note: Profilområder: 1 (sprogforståelse), 2 (afkodning), 3 (tekstforståelse)

Kilde: Styrelsen for It og Læring

Tablet 3.1.4 Den statistiske usikkerhed (logit) på den estimerede elevdygtighed i skoleåret 2017/2018 i forhold til elevdygtigheden. Matematik

Klassetrin	Profilområde	Elevdygtighed på femtrins skalaen				
		1-10	11-35	36-65	66-90	91-100
3. klasse	1	0,45	0,45	0,45	0,47	0,57
	2	0,43	0,42	0,48	0,52	0,56
	3	0,51	0,50	0,51	0,52	0,63
6. klasse	1	0,45	0,42	0,45	0,48	0,53
	2	0,45	0,42	0,46	0,51	0,56
	3	0,50	0,50	0,51	0,53	0,60
8. klasse	1	0,53	0,51	0,52	0,55	0,60
	2	0,52	0,48	0,52	0,55	0,60
	3	0,54	0,51	0,53	0,56	0,62

Note: Profilområder: 1 (tal og algebra), 2 (geometri), 3 (statistik og sandsynlighed)

Kilde: Styrelsen for It og Læring

Tabel 3.1.5 Den statistiske usikkerhed (logit) på den estimerede elevdygtighed i skoleåret 2017/2018 i forhold til elevdygtigheden. Engelsk

Klassetrin	Profilområde	Elevdygtighed på femtrins skalaen				
		1-10	11-35	36-65	66-90	91-100
4. klasse	1	0,36	0,35	0,38	0,45	0,51
	2	0,32	0,35	0,40	0,47	0,54
	4	0,39	0,38	0,42	0,47	0,66
7. klasse	1	0,36	0,36	0,43	0,50	0,52
	2	0,38	0,40	0,46	0,50	0,51
	3	0,38	0,38	0,43	0,48	0,52

Note: Profilområder: 1 (læsning), 2 (ordforråd), 3 (sprog og sprogbrug), 4 (lytning)

Kilde: Styrelsen for It og Læring

Tabel 3.1.6 Den statistiske usikkerhed (logit) på den estimerede elevdygtighed i skoleåret 2017/2018 i forhold til elevdygtigheden. Fysik/kemi

Klassetrin	Profilområde	Elevdygtighed på femtrins skalaen				
		1-10	11-35	36-65	66-90	91-100
8. klasse	1	0,34	0,31	0,35	0,40	0,46
	2	0,32	0,31	0,33	0,39	0,47
	3	0,32	0,30	0,31	0,37	0,46

Note: Profilområder: 1 (energi og energiomsætning), 2 (fænomener, stoffer og materialer), 3 (anvendelse og perspektiv)

Kilde: Styrelsen for It og Læring

Tabel 3.1.7 Længden af 68 % sikkerhedsintervallet i forhold til elevernes dygtighed på den normbaserede (1-100) skala. Skoleåret 2017/2018. Dansk læsning

Klassetrin	Profilområde	Elevdygtighed på femtrins skalaen				
		1-10	11-35	36-65	66-90	91-100
2. klasse	1	3	12	18	13	6
	2	2	6	9	8	3
	3	2	6	10	9	3
4. klasse	1	3	12	18	14	5
	2	3	9	14	11	4
	3	3	8	14	12	4
6. klasse	1	4	13	19	15	4
	2	3	10	16	13	5
	3	3	11	17	12	4
8. klasse	1	4	15	22	18	3
	2	3	11	15	10	4
	3	2	10	16	13	4

Note: Profilområder: 1 (sprogforståelse), 2 (afkodning), 3 (tekstforståelse)

Note: Der er tale om en lille tilnærmelse, da sikkerhedsintervallerne på percentilskalaen ikke er helt symmetriske

Kilde: Styrelsen for It og Læring

Tabel 3.1.8 Længden af 95 % sikkerhedsintervallet i forhold til elevernes dygtighed på den normbaserede (1-100) skala. Skoleåret 2017/2018. Dansk læsning

Klassetrin	Profilområde	Elevdygtighed på femtrins skalaen				
		1-10	11-35	36-65	66-90	91-100
2. klasse	1	7	23	31	25	13
	2	3	11	18	15	7
	3	4	12	19	17	7
4. klasse	1	7	23	32	26	12
	2	6	18	26	21	9
	3	5	15	26	22	8
6. klasse	1	9	25	33	27	10
	2	6	20	28	24	10
	3	6	20	30	24	10
8. klasse	1	9	29	37	32	8
	2	7	22	27	20	8
	3	5	19	28	24	9

Note: Profilområder: 1 (sprogforståelse), 2 (afkodning), 3 (tekstforståelse)

Note: Der er tale om en lille tilnærmelse, da sikkerhedsintervallerne på percentilskalaen ikke er helt symmetriske

Kilde: Styrelsen for It og Læring

Tabel 3.1.9 Længden af 68 procent sikkerhedsintervallet i forhold til elevernes dygtighed på den normbaserede (1-100) skala. Skoleåret 2017/2018.

Matematik

Klassetrin	Profilområde	Elevdygtighed på femtrins skalaen				
		1-10	11-35	36-65	66-90	91-100
3. klasse	1	3	10	15	11	3
	2	4	14	18	12	4
	3	2	9	14	11	4
6. klasse	1	4	15	20	13	4
	2	4	13	18	14	4
	3	3	12	16	11	4
8. klasse	1	3	10	14	9	3
	2	3	10	15	11	3
	3	3	10	14	10	3

Note: Profilområder: 1 (tal og algebra), 2 (geometri), 3 (statistik og sandsynlighed)

Note: Der er tale om en lille tilnærmelse, da sikkerhedsintervallerne på percentilskalaen ikke er helt symmetriske

Kilde: Styrelsen for It og Læring

Tabel 3.1.10 Længden af 95 procent sikkerhedsintervallet i forhold til elevernes dygtighed på den normbaserede (1-100) skala. Skoleåret 2017/2018.

Matematik

Klassetrin	Profilområde	Elevdygtighed på femtrins skalaen				
		1-10	11-35	36-65	66-90	91-100
3. klasse	1	6	20	26	22	8
	2	8	25	32	25	9
	3	5	18	26	21	9
6. klasse	1	10	26	34	26	9
	2	9	24	32	27	9
	3	8	22	29	22	9
8. klasse	1	7	19	25	18	7
	2	7	20	28	21	7
	3	6	20	26	19	7

Note: Profilområder: 1 (tal og algebra), 2 (geometri), 3 (statistik og sandsynlighed)

Note: Der er tale om en lille tilnærmelse, da sikkerhedsintervallerne på percentilskalaen ikke er helt symmetriske

Kilde: Styrelsen for It og Læring

Tabel 3.1.11 Længden af 68 procent sikkerhedsintervallet i forhold til elevernes dygtighed på den normbaserede (1-100) skala. Skoleåret 2017/2018.

Engelsk

Klassetrin	Profilområde	Elevdygtighed på femtrins skalaen				
		1-10	11-35	36-65	66-90	91-100
4. klasse	1	3	10	11	7	3
	2	3	11	15	10	3
	4	3	10	14	9	4
7. klasse	1	3	11	16	12	5
	2	3	11	16	11	4
	3	3	10	14	10	3

Note: Profilområder: 1 (læsning), 2 (ordforråd), 3 (sprog og sprogbrug), 4 (lytning)

Note: Der er tale om en lille tilnærmelse, da sikkerhedsintervallerne på percentilskalaen ikke er helt symmetriske

Kilde: Styrelsen for It og Læring

Tabel 3.1.12 Længden af 95 procent sikkerhedsintervallet i forhold til elevernes dygtighed på den normbaserede (1-100) skala. Skoleåret 2017/2018.

Engelsk

Klassetrin	Profilområde	Elevdygtighed på femtrins skalaen				
		1-10	11-35	36-65	66-90	91-100
4. klasse	1	6	18	21	14	6
	2	7	21	27	21	7
	4	6	19	25	18	8
7. klasse	1	8	20	28	23	10
	2	7	20	28	21	8
	3	6	18	25	19	7

Note: Profilområder: 1 (læsning), 2 (ordforråd), 3 (sprog og sprogbrug), 4 (lytning)

Note: Der er tale om en lille tilnærmelse, da sikkerhedsintervallerne på percentilskalaen ikke er helt symmetriske

Kilde: Styrelsen for It og Læring

Tabel 3.1.13 Længden af 68 procent sikkerhedsintervallet i forhold til elevernes dygtighed på den normbaserede (1-100) skala. Skoleåret 2017/2018.

Fysik/kemi

Klassetrin	Profilområde	Elevdygtighed på femtrins skalaen				
		1-10	11-35	36-65	66-90	91-100
8. klasse	1	5	15	21	17	6
	2	5	18	24	18	6
	3	6	18	22	17	6

Note: Profilområder: 1 (energi og energiomsætning), 2 (fænomener, stoffer og materialer), 3 (anvendelse og perspektiv)

Note: Der er tale om en lille tilnærmelse, da sikkerhedsintervallerne på percentilskalaen ikke er helt symmetriske

Kilde: Styrelsen for It og Læring

Tabel 3.1.14 Længden af 95 procent sikkerhedsintervallet i forhold til elevernes dygtighed på den normbaserede (1-100) skala. Skoleåret 2017/2018.

Fysik/kemi

Klassetrin	Profilområde	Elevdygtighed på femtrins skalaen				
		1-10	11-35	36-65	66-90	91-100
8. klasse	1	11	27	35	32	15
	2	13	31	38	35	16
	3	14	31	36	33	15

Note: Profilområder: 1 (energi og energiomsætning), 2 (fænomener, stoffer og materialer), 3 (anvendelse og perspektiv)

Note: Der er tale om en lille tilnærmelse, da sikkerhedsintervallerne på percentilskalaen ikke er helt symmetriske

Kilde: Styrelsen for It og Læring

Tabel 3.1.15 68 % sikkerhedsintervallet på den kriteriebaserede skala i forhold til elevernes dygtighed. Skoleåret 2017/2018. Dansk læsning og matematik samlet

Elevens testresultat på den kriteriebaserede skala	Elevens testresultat ligger med stor sandsynlighed i ...		
	... det beregnede faglige niveau	... det beregnede faglige niveau eller enten niveauet lige under/over (ikke både og)	... det beregnede faglige niveau eller niveauet lige over eller lige under
Ikke tilstrækkelig	60 %	40 %	0 %
Mangelfuld	32 %	65 %	3 %
Jævn	6 %	66 %	28 %
God	37 %	61 %	2 %
Meget god	17 %	78 %	4 %
Fremragende	44 %	56 %	0 %

Kilde: Styrelsen for It og Læring

Tabel 3.1.16 95 % sikkerhedsintervallet på den kriteriebaserede skala i forhold til elevernes dygtighed. Skoleåret 2017/2018. Dansk læsning og matematik samlet

Elevens testresultat på den kriteriebaserede skala	Elevens testresultat ligger med stor sandsynlighed i ...		
	... det beregnede faglige niveau	... det beregnede faglige niveau eller enten niveauet lige under/over (ikke både og)	... det beregnede faglige niveau eller niveauet lige over eller lige under
Ikke tilstrækkelig	36 %	62 %	2 %
Mangelfuld	5 %	49 %	46 %
Jævn	0 %	12 %	88 %
God	14 %	37 %	49 %
Meget god	0 %	38 %	62 %
Fremragende	17 %	77 %	6 %

Kilde: Styrelsen for It og Læring

Bilag 3.2 Reliabilitet

Variansen på den observerede elevscore (X) er summen af variansen på den sande score (T) og variansen på målefejlen (E)⁵⁶:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

Reliabiliteten kan defineres som forholdet mellem variansen på den sande score og variansen på den observerede score:

$$\rho_{xx'} = \frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}$$

I RUMM⁵⁷ beregnes reliabiliteten ved anvendelse af Person Separation Index:

$$r_{\theta\theta} = 1 - \frac{\sigma_\varepsilon^2}{\sigma_\theta^2}$$

θ er elevens estimerede dygtighed og σ_θ^2 beregnes som variansen på elevernes dygtigheder blandt de elever, der gennemfører en test.

σ_ε^2 er usikkerheden på den enkelte elevs beregnede dygtighed og betegnes SEM², hvor SEM er Standard Error of Measurement.

SEM beregnes i test- og prøvesystemet i lighed med elevens dygtighed efter hver besvarelse på en opgave.

Der er anvendt resultaterne fra de obligatoriske test i skoleåret 2017/2018 i alle tabellerne.

⁵⁶ Davidshofer, Kevin R. Murphy, Charles O. (2005): Psychological testing: principles and application. Pearson/Prentice Hall.

⁵⁷ RUMM Laboratory Pty Ltd.

Tabel 3.2.1 Den statistiske usikkerhed (SEM) på den estimerede elevdygtighed og beregnet reliabilitet i skoleåret 2017/2018. Dansk læsning

Klassetrin	Profilområde	SEM	Reliabilitet (PSI ¹)
2. klasse	1	0,46	0,74
	2	0,46	0,91
	3	0,42	0,90
4. klasse	1	0,47	0,80
	2	0,48	0,87
	3	0,44	0,88
6. klasse	1	0,48	0,78
	2	0,49	0,85
	3	0,45	0,81
8. klasse	1	0,51	0,83
	2	0,51	0,85
	3	0,47	0,84

1) Person Separation Index (PSI)

Note: Profilområder: 1 (sprogforståelse), 2 (afkodning), 3 (tekstforståelse)

Kilde: Styrelsen for It og Læring

Tabel 3.2.2 Den statistiske usikkerhed (SEM) på den estimerede elevdygtighed og beregnet reliabilitet i skoleåret 2017/2018. Matematik

Klassetrin	Profilområde	SEM	Reliabilitet (PSI ¹)
3. klasse	1	0,48	0,86
	2	0,49	0,78
	3	0,53	0,87
6. klasse	1	0,47	0,83
	2	0,48	0,81
	3	0,52	0,84
8. klasse	1	0,53	0,88
	2	0,52	0,86
	3	0,54	0,89

1) Person Separation Index (PSI)

Note: Profilområder: 1 (tal og algebra), 2 (geometri), 3 (statistik og sandsynlighed)

Kilde: Styrelsen for It og Læring

Tabel 3.2.3 Den statistiske usikkerhed (SEM) på den estimerede elevdygtighed og beregnet reliabilitet i skoleåret 2017/2018. Engelsk

Klassetrin	Profilområde	SEM	Reliabilitet (PSI ¹)
4. klasse	1	0,40	0,91
	2	0,41	0,88
	4	0,44	0,88
7. klasse	1	0,45	0,79
	2	0,47	0,86
	3	0,45	0,88

1) Person Separation Index (PSI)

Note: Profilområder: 1 (læsning), 2 (ordforråd), 3 (sprog og sprogbrug), 4 (lytning)

Kilde: Styrelsen for It og Læring

Tabel 3.2.4 Den statistiske usikkerhed (SEM) på den estimerede elevdygtighed og beregnet reliabilitet i skoleåret 2017/2018. Fysik/kemi

Klassetrin	Profilområde	SEM	Reliabilitet (PSI ¹)
8. klasse	1	0,37	0,68
	2	0,37	0,66
	4	0,36	0,70

1) Person Separation Index (PSI)

Note: Profilområder: 1 (energi og energiomsætning), 2 (fænomener, stoffer og materialer), 3 (anvendelse og perspektiv)

Kilde: Styrelsen for It og Læring

Bilag 4.1 Opgaveafprøvningsperioder

Afprøvning af opgaver med henblik på anvendelse i de nationale test er foregået siden maj 2008.

Bilaget indeholder oversigt over afprøvningsperiode for de opgaver i opgavebanken, der var aktive under afviklingen af de obligatoriske test foråret 2019.

I matematik 3. klasse erstattedes profilområdet *matematik i anvendelse af statistik og sandsynlighed* fra og med skoleåret 2015/2016.

Tilsvarende ændring skete i matematik i 6. klasse i skoleåret 2017/2018.

Matematik i 8. klasse introduceredes fra og med skoleåret 2017/2018.

Engelsk i 4. klasse introduceredes skoleåret 2017/2018.

Tabellerne viser tidspunkt for opgaveafprøvning.

Tabel 4.1.1 Opgavebankens sammensætning i forhold til opgavernes afprøvningsperiode. Dansk læsning

Afprøvningsperiode	Dansk læsning 2. klasse			Dansk læsning 4. klasse			Dansk læsning 6. klasse			Dansk læsning 8. klasse		
	P1 ¹⁾	P2 ²⁾	P3 ³⁾	P1 ¹⁾	P2 ²⁾	P3 ³⁾	P1 ¹⁾	P2 ²⁾	P3 ³⁾	P1 ¹⁾	P2 ²⁾	P3 ³⁾
Maj 2008	0	0	0	0	0	0	0	0	0	0	0	0
Oktober 2008	0	122	89	80	166	82	44	168	75	19	13	3
December 2008	22	23	18	20	0	22	37	38	4	38	52	31
April 2009	66	0	0	6	0	5	24	0	11	88	108	29
September 2009	80	0	21	11	0	23	72	0	61	0	0	111
September 2010	0	0	23	56	53	77	0	0	50	99	0	0
September 2012	34	15	23	11	17	14	10	30	19	30	26	40
September 2014	6	69	16	19	37	32	54	5	24	34	15	38
September 2015	3	0	0	0	0	33	0	0	47	0	0	47
Januar 2016	0	0	0	0	0	0	0	0	0	0	0	0
September 2016	0	0	0	0	0	0	0	0	0	0	0	0
Januar 2017	0	0	0	0	0	0	0	0	0	0	0	0
Januar 2018	34	25	40	24	0	0	0	0	0	0	0	0
I alt	245	254	230	227	273	288	241	241	291	308	214	299

1) Sprogforståelse 2) Afkodning 3) Tekstforståelse

Kilde: Styrelsen for It og Læring

Tabel 4.1.2 Opgavebankens sammensætning i forhold til opgavernes afprøvningsperiode. Matematik

Afprøvningsperiode	Matematik 3. klasse			Matematik 6. klasse			Matematik 8. klasse		
	P1 ¹⁾	P2 ²⁾	P4 ³⁾	P1 ¹⁾	P2 ²⁾	P4 ³⁾	P1 ¹⁾	P2 ²⁾	P4 ³⁾
Maj 2008	68	69	0	252	79	0	0	0	0
Oktober 2008	0	0	0	0	0	0	0	0	0
December 2008	0	30	0	0	0	0	0	0	0
April 2009	0	0	0	0	0	0	0	0	0
September 2009	0	0	0	0	0	0	0	0	0
September 2010	22	35	0	1	25	0	0	0	0
September 2012	59	16	0	49	47	0	0	0	0
September 2014	112	29	0	143	42	0	0	0	0
September 2015	0	0	225	0	0	0	0	0	0
Januar 2016	0	0	0	0	0	0	0	0	0
September 2016	0	0	0	0	96	262	234	249	223
Januar 2017	0	0	0	0	0	0	0	0	0
Januar 2018	0	41	0	29	0	0	0	0	0
I alt	261	220	225	474	289	262	234	249	223

1) Tal og algebra 2) Geometri 3) Statistik og sandsynlighed

Kilde: Styrelsen for It og Læring

Tabel 4.1.3 Opgavebankens sammensætning i forhold til opgavernes afprøvningsperiode. Engelsk

Afprøvningsperiode	Engelsk 4. klasse			Engelsk 7. klasse		
	P1 ¹⁾	P2 ²⁾	P4 ³⁾	P1 ¹⁾	P2 ²⁾	P3 ⁴⁾
Maj 2008	0	0	0	45	43	76
Oktober 2008	0	0	0	0	0	0
December 2008	0	0	0	5	5	4
April 2009	0	0	0	0	0	0
September 2009	0	0	0	8	0	4
September 2010	0	0	0	0	9	0
September 2012	0	0	0	19	37	12
September 2014	0	0	0	26	32	19
September 2015	0	0	0	104	96	98
Januar 2016	0	211	60	0	0	0
September 2016	149	0	131	0	0	0
Januar 2017	70	71	65	0	0	0
Januar 2018	0	31	0	0	0	0
I alt	219	313	256	207	222	213

1) Læsning 2) Ordforråd 3) Lytning 4) Sprog og sprogbrug

Kilde: Styrelsen for It og Læring

Tabel 4.1.4 Opgavebankens sammensætning i forhold til opgavernes afprøvningsperiode. Fysik/kemi

Afprøvningsdato	Fysik/kemi 8. klasse		
	P1 ¹⁾	P2 ²⁾	P4 ³⁾
Maj 2008	67	195	98
Oktober 2008	0	0	0
December 2008	38	0	29
April 2009	0	0	0
September 2009	11	1	11
September 2010	47	7	33
September 2012	22	22	30
September 2014	24	19	26
September 2015	5	16	14
Januar 2016	0	0	0
September 2016	0	0	0
Januar 2017	0	0	0
Januar 2018	4	19	19
I alt	218	279	260

1) Energi og energiomsætning 2) Fænomener, stoffer og materialer 3) Anvendelse og perspektiv

Kilde: Styrelsen for It og Læring

Bilag 4.2 Skærmdumps fra RUMM

På baggrund af elevernes besvarelser fra opgaveafprøvningsene foretages Rasch analyser. Opgaver der ikke passer til modellen fjernes. For de godkendte opgaver estimeres opgavernes tærskelværdier, herunder opgavernes sværhedsgrad.

Bilaget dokumenterer, ved skærmdumps fra RUMM⁵⁸, analyserne foretaget efter opgaveafprøvningsen i januar 2018. Der er kun medtaget skærmdumps fra analyserne af opgaver afprøvet i dansk læsning og matematik:

- Dansk læsning 2. klasse:
 - Sprogforståelse
 - Afkodning
 - Tekstforståelse
- Dansk læsning 4. klasse
 - Sprogforståelse
- Matematik 3. klasse
 - Geometri og måling
- Matematik 6. klasse
 - Tal og algebra

I Raschanalysen forkastes opgaver på baggrund af:

- Itemfit: tester hvor godt opgaverne passer til den overordnede Rasch-model.
- Differentiel item funktion (DIF): tester om elevernes svar på opgaverne afhænger af elevernes køn, geografi (skolens beliggenhed) og skolestørrelse.

I analysen indgår også eksisterende aktive opgaver (linkopgaver) fra tidligere opgaveafprøvnings. Disse opgaver linker de nye opgaver i opgaveafprøvningsen sammen med den eksisterende Rasch-skala. Disse opgaver slettes ikke, men bliver testet for om opgavernes sværhedsgrad har ændret sig fra tidligere opgaveafprøvnings (DIF for periode).

⁵⁸ www.rummlab.com.au

Dansk læsning 2. klasse – sprogforståelse

I opgaveafprøvningen januar 2018 blev 61 opgaver afprøvet, heraf forkastes 27 opgaver og 34 (56 %) opgaver godkendes og tilføjes Opgavebanken.

25 af opgaverne forkastes pga. itemfit og 2 opgaver forkastes pga. DIF på køn.

Analysen indeholder 10 linkopgaver: I0170, I0171, I0173, I0175, I0176, I0182, I0183, I0189, I0192 og I0203.

Initial itemfit-analyse: viser et udpluk af de opgaver der forkastes.

	Seq	Item	Type	Location	SE	FitResid	DF	ChiSq	DF	Prob	F-stat	DF-1	DF-2	Prob
52	290	I0290	Poly	0,541	0,087	4,482	641,70	33,906	9	0,000093
42	280	I0280	Poly	0,669	0,093	2,114	575,96	34,847	9	0,000063
31	269	I0269	Poly	1,603	0,101	2,752	641,70	35,057	9	0,000058
3	173	I0173	Poly	0,647	0,089	3,358	634,83	35,314	9	0,000052
16	254	I0254	Poly	-2,259	0,127	-3,251	641,70	36,828	9	0,000028
63	301	I0301	Poly	-1,934	0,121	-2,907	575,96	39,668	9	0,000009
1	170	I0170	Poly	1,064	0,092	2,187	641,70	40,154	9	0,000007
14	252	I0252	Poly	-2,190	0,123	-3,150	635,81	41,003	9	0,000005
13	251	I0251	Poly	-1,744	0,110	-3,659	641,70	42,408	9	0,000003
45	283	I0283	Poly	-0,849	0,097	-3,769	575,96	42,803	9	0,000002
18	256	I0256	Poly	-0,476	0,093	-5,558	575,96	42,937	9	0,000002
56	294	I0294	Poly	-0,321	0,088	-5,236	636,79	48,176	9	0,000000
41	279	I0279	Poly	-1,438	0,101	-3,841	638,76	48,502	9	0,000000
30	268	I0268	Poly	-0,653	0,095	-6,008	575,96	51,987	9	0,000000
39	277	I0277	Poly	1,395	0,103	3,240	575,96	53,205	9	0,000000
8	189	I0189	Poly	1,475	0,105	3,206	575,96	54,696	9	0,000000
11	249	I0249	Poly	0,673	0,089	5,107	638,76	56,073	9	0,000000
6	182	I0182	Poly	1,013	0,097	4,450	575,96	58,054	9	0,000000
7	183	I0183	Poly	1,098	0,098	4,639	575,96	60,804	9	0,000000
71	309	I0309	Poly	1,027	0,093	4,483	638,76	63,060	9	0,000000
5	176	I0176	Poly	0,895	0,091	5,596	637,77	67,127	9	0,000000

Mean	0,000	0,025	Totals	1,922,058	639
Std Devn	1,020	3,023	Total Prob	0,000000	

Highlight extreme itemfit values. Set +/- 2.5
 below the Bonferroni Select Probability B
0,01
 Adj = 0.000141 for 71

Selector: 5 Item: x010201000301235399-1 1
 File Text Format: Fixed
 < Display Control Sort Items by ... ChiSq Prob Orde Save Copy Resp Types > Chi Squ >

Afsluttende DIF-analyse: skolestørrelse

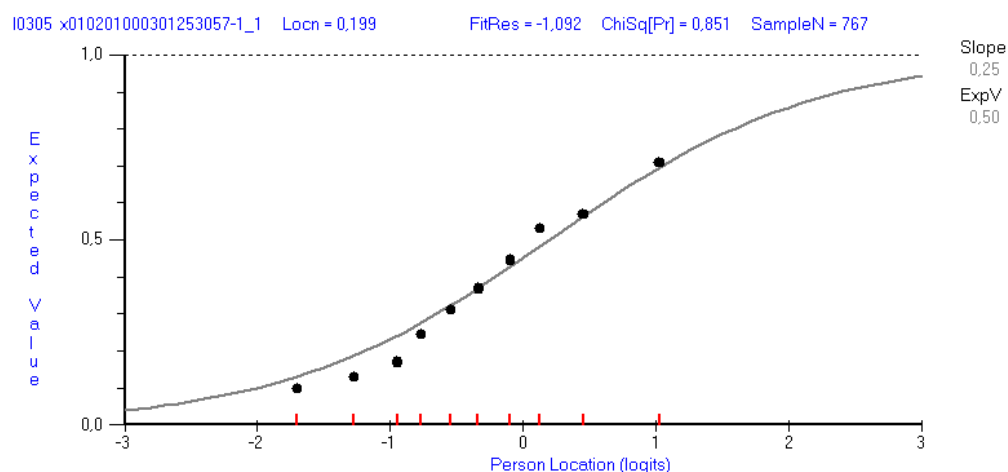
No.	Item	Class Interval				skolestr				Class Interval-by- skolestr			
		MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
11	10250	0.95177	0.97194	9	0.462174	0.99254	1.01357	1	0.314479	1.39473	1.42428	9	0.174071
12	10255	1.28955	1.40676	9	0.181346	1.61660	1.76354	1	0.184667	1.16752	1.27363	9	0.247869
13	10257	1.32323	1.40423	9	0.182421	0.01127	0.01196	1	0.912947	0.84611	0.89790	9	0.526681
14	10258	2.14419	2.46391	9	0.009166	1.71834	1.97457	1	0.160470	1.06678	1.22585	9	0.275880
15	10259	0.69855	0.70426	9	0.705265	0.06033	0.06082	1	0.805272	1.73007	1.74419	9	0.076200
16	10260	1.92092	1.78614	9	0.067687	2.99608	2.78586	1	0.095601	1.17846	1.09577	9	0.363605
17	10261	1.18678	1.24990	9	0.261505	1.44018	1.51678	1	0.218575	0.74600	0.78567	9	0.629768
18	10262	1.31747	1.24737	9	0.263247	0.85134	0.80604	1	0.369680	1.93245	1.82963	9	0.060277
19	10263	2.05697	2.00287	9	0.036686	0.80580	0.78460	1	0.376067	1.52492	1.48481	9	0.149623
20	10264	1.66825	1.92772	9	0.045558	0.69550	0.80368	1	0.370337	0.50814	0.58718	9	0.808166
21	10265	1.34473	1.26358	9	0.253836	0.97180	0.91315	1	0.339683	0.82038	0.77087	9	0.643554
22	10266	2.21209	2.52187	9	0.007634	0.04305	0.04908	1	0.824728	1.15629	1.31822	9	0.223714
23	10267	0.52994	0.56444	9	0.826598	0.29105	0.31000	1	0.577871	1.32879	1.41530	9	0.177617
24	10269	1.56304	1.44547	9	0.164957	0.89726	0.82977	1	0.362688	0.81167	0.75061	9	0.662424
25	10270	2.50579	2.86559	9	0.002547	0.84272	0.96372	1	0.326622	0.81993	0.93766	9	0.491527
26	10271	1.66449	1.88084	9	0.052237	3.36003	3.79677	1	0.051856	0.79845	0.90223	9	0.522893
27	10272	0.97741	0.92517	9	0.502454	1.96967	1.86440	1	0.172591	0.72655	0.68772	9	0.720418
28	10273	2.76346	3.39901	9	0.000445	0.29684	0.36511	1	0.545900	0.90914	1.11823	9	0.347271
29	10274	0.56631	0.57892	9	0.814863	4.25583	4.35065	1	0.037450	1.42790	1.45972	9	0.159582
30	10275	1.09109	1.09969	9	0.360712	0.00431	0.00434	1	0.947501	0.41752	0.42081	9	0.924268
31	10276	0.93754	0.93149	9	0.496913	0.00047	0.00046	1	0.983111	0.89419	0.88842	9	0.535211
32	10278	1.20778	1.33633	9	0.214450	0.84213	0.93176	1	0.334772	0.69178	0.76540	9	0.648654
33	10281	1.11826	1.17231	9	0.309941	2.49763	2.61834	1	0.106144	1.24800	1.30832	9	0.228905
34	10282	1.60322	1.83686	9	0.058837	0.00267	0.00306	1	0.955871	0.86902	0.99566	9	0.442171
35	10284	1.02085	1.06381	9	0.387695	0.06359	0.06626	1	0.796946	1.51651	1.58034	9	0.117325
36	10287	1.69634	1.83894	9	0.058478	0.02364	0.02562	1	0.872883	1.42606	1.54594	9	0.128180
37	10288	1.13410	1.24004	9	0.267328	1.64503	1.79870	1	0.180358	0.75686	0.82756	9	0.590855
38	10289	1.66377	1.90886	9	0.048276	0.43855	0.50316	1	0.478412	0.44887	0.51499	9	0.864128
39	10290	2.04602	1.86872	9	0.053785	0.21112	0.19282	1	0.660732	0.42614	0.38921	9	0.940481
40	10297	2.39006	2.94305	9	0.001972	0.00228	0.00280	1	0.957828	1.69531	2.08756	9	0.028668
41	10298	1.49148	1.64226	9	0.100097	0.27206	0.29956	1	0.584368	0.35505	0.39094	9	0.939584
42	10303	1.64537	1.44309	9	0.165976	0.03658	0.03208	1	0.857866	1.28109	1.12359	9	0.343465
43	10304	0.81340	0.86758	9	0.554145	1.64346	1.75294	1	0.186045	0.38089	0.40626	9	0.931928
44	10305	0.49408	0.53494	9	0.849481	0.65433	0.70845	1	0.400284	1.49009	1.61334	9	0.107666

Main Plot
 ght
 Select Level of Precision
 0.03
 Save
 Copy

Afsluttende DIF-analyse: geografi

No.	Item	Class Interval				oestvest				Class Interval-by- oestvest			
		MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
11	I0250	0.95177	0.96860	9	0.465007	1.76820	1.79946	1	0.180307	1.09621	1.11559	9	0.349381
12	I0255	1.28955	1.38619	9	0.190554	0.27970	0.30066	1	0.583665	0.36991	0.39764	9	0.936357
13	I0257	1.32323	1.40558	9	0.181825	0.00570	0.00605	1	0.938094	0.91026	0.96690	9	0.466317
14	I0258	2.14419	2.44746	9	0.009633	0.14197	0.16205	1	0.687414	0.83429	0.95229	9	0.478843
15	I0259	0.69855	0.70721	9	0.702556	12.29440	12.44671	1	0.000449	0.63100	0.63881	9	0.764107
16	I0260	1.92092	1.76065	9	0.072585	0.00481	0.00441	1	0.947089	0.41771	0.38286	9	0.943480
17	I0261	1.18678	1.24947	9	0.261751	0.30624	0.32241	1	0.570377	0.84959	0.89447	9	0.529766
18	I0262	1.31747	1.23720	9	0.269272	0.23067	0.21662	1	0.641822	1.45592	1.36722	9	0.199682
19	I0263	2.05697	1.97759	9	0.039458	0.52379	0.50358	1	0.478203	0.63282	0.60840	9	0.790421
20	I0264	1.66825	1.93297	9	0.044879	0.02971	0.03443	1	0.852869	0.74530	0.86357	9	0.557751
21	I0265	1.34473	1.26590	9	0.252510	1.48073	1.39393	1	0.238242	0.88670	0.83472	9	0.584273
22	I0266	2.21209	2.52367	9	0.007593	0.00516	0.00588	1	0.938919	1.20429	1.37391	9	0.196208
23	I0267	0.52994	0.55856	9	0.831258	0.07746	0.08164	1	0.775194	0.66385	0.69971	9	0.709472
24	I0269	1.56304	1.45817	9	0.159869	0.00963	0.00898	1	0.924534	1.57287	1.46735	9	0.156267
25	I0270	2.50579	2.85956	9	0.002606	0.00044	0.00050	1	0.982422	0.78513	0.89598	9	0.528414
26	I0271	1.66449	1.86444	9	0.054698	0.55129	0.61752	1	0.432304	0.62087	0.69545	9	0.713323
27	I0272	0.97741	0.92273	9	0.504603	0.03044	0.02874	1	0.865409	0.74571	0.70400	9	0.705539
28	I0273	2.76346	3.38767	9	0.000448	1.45655	1.78556	1	0.181949	0.59022	0.72353	9	0.687545
29	I0274	0.56631	0.57413	9	0.818753	0.06674	0.06766	1	0.794884	1.37962	1.39867	9	0.185209
30	I0275	1.09109	1.10140	9	0.359466	0.12402	0.12519	1	0.723605	0.51238	0.51722	9	0.862601
31	I0276	0.93754	0.94973	9	0.481050	13.15376	13.32475	1	0.000290	0.77921	0.78934	9	0.626349
32	I0278	1.20778	1.33874	9	0.213244	0.39259	0.43515	1	0.509726	0.85617	0.94894	9	0.481721
33	I0281	1.11826	1.15472	9	0.321757	0.48652	0.50239	1	0.478709	0.44974	0.46441	9	0.898543
34	I0282	1.60322	1.86607	9	0.054217	0.37457	0.43598	1	0.509309	1.77962	2.07140	9	0.030065
35	I0284	1.02085	1.05303	9	0.396037	0.56353	0.58130	1	0.446079	0.76988	0.79415	9	0.621869
36	I0287	1.69634	1.83926	9	0.058423	7.43629	8.06280	1	0.004675	0.61392	0.66564	9	0.740367
37	I0288	1.13410	1.23575	9	0.269892	0.68229	0.74345	1	0.388896	0.64310	0.70074	9	0.708526
38	I0289	1.66377	1.94189	9	0.043955	1.67138	1.95077	1	0.163055	1.24428	1.45228	9	0.162542
39	I0290	2.04602	1.89771	9	0.049564	0.31722	0.29423	1	0.587710	1.59100	1.47567	9	0.153071
40	I0297	2.39006	2.90093	9	0.002279	0.88842	1.07831	1	0.299472	0.77674	0.94277	9	0.487080
41	I0298	1.49148	1.68361	9	0.089695	2.80398	3.16518	1	0.075758	1.47655	1.66675	9	0.093810
42	I0303	1.64537	1.43591	9	0.168911	0.82335	0.71854	1	0.396941	0.79873	0.69705	9	0.711900
43	I0304	0.81340	0.87629	9	0.546231	0.48079	0.51796	1	0.472008	1.09615	1.18091	9	0.304533
44	I0305	0.49408	0.52857	9	0.854255	1.00969	1.08017	1	0.299066	0.66715	0.71373	9	0.696596

Eksempel på ICC-kurve for en godkendt opgave.



Kilde: Styrelsen for It og Læring

Dansk læsning 2. klasse – afkodning

I opgaveafprøvningen januar 2018 blev 32 opgaver afprøvet, heraf forkastes 7 opgaver og 25 (78 %) opgaver godkendes og tilføjes Opgavebanken.

6 af opgaverne forkastes pga. itemfit og 1 opgave forkastes pga. DIF på geografi.

Analysen indeholder 10 linkopgaver: I0036, I0050, I0091, I0094, I0097, I0151, I0153, I0238, I0246 og I0248.

Initiel itemfit-analyse: viser et udpluk af de opgaver der forkastes.

Seq	Item	Type	Location	SE	FitResid	DF	ChiSq	DF	Prob	F-stat	DF-1	DF-2	Prob
29	311	I0311	Poly	-1,408	0,124	0,079	702,90	15,703	9	0,073342
27	309	I0309	Poly	-0,106	0,096	-1,061	702,90	15,807	9	0,071026
37	319	I0319	Poly	-0,588	0,104	-1,875	702,90	15,886	9	0,069304
39	321	I0321	Poly	-0,066	0,096	-1,755	702,90	16,137	9	0,064076
4	94	I0094	Poly	2,127	0,095	-1,553	651,44	16,869	9	0,050812
7	153	I0153	Poly	2,507	0,096	2,303	702,90	18,016	9	0,034985
30	312	I0312	Poly	-1,340	0,127	0,775	649,50	18,262	9	0,032250
6	151	I0151	Poly	0,975	0,088	3,588	702,90	18,407	9	0,030736
9	246	I0246	Poly	1,374	0,091	2,911	652,41	19,576	9	0,020715
13	295	I0295	Poly	-1,146	0,116	0,670	702,90	19,671	9	0,020055
14	296	I0296	Poly	-0,068	0,100	-3,229	652,41	21,020	9	0,012563
3	91	I0091	Poly	1,569	0,088	-0,991	702,90	23,642	9	0,004907
26	308	I0308	Poly	-0,556	0,108	-2,473	652,41	23,919	9	0,004431
28	310	I0310	Poly	-0,117	0,101	-3,864	651,44	24,149	9	0,004073
12	294	I0294	Poly	0,383	0,095	-3,320	651,44	24,405	9	0,003706
24	306	I0306	Poly	-1,346	0,127	2,146	652,41	27,069	9	0,001362
2	50	I0050	Poly	1,882	0,093	1,975	652,41	28,841	9	0,000689
17	299	I0299	Poly	1,932	0,090	-0,617	702,90	29,829	9	0,000469
15	297	I0297	Poly	0,519	0,090	-4,217	701,93	29,934	9	0,000450
25	307	I0307	Poly	-0,822	0,109	1,442	702,90	32,249	9	0,000180
8	238	I0238	Poly	1,443	0,088	2,715	702,90	35,945	9	0,000041
Mean			0,000		-0,336	Totals		662,971	378	<input type="checkbox"/> Highlight extreme itemfit values. Set +/- 2.5			
Std Devn			1,235		1,907	Total Prob		0,000000	<input checked="" type="checkbox"/> below the Bonferroni Select Probability B				
Selector 8 Item x010202000301239054-1 1										File Text Format		Adi = 0.000238 for 42	
										<input checked="" type="radio"/> Fixed		0,01	
< Display Control										Sort Items by ...		Save	
										ChiSq Prob Orde		Tab Delim	
										Copy		Resp Types >	
										Chi Squ >			

Afsluttende itemfit-analyse.

Seq	Item	Type	Location	SE	FitResid	DF	ChiSq	DF	Prob	F-stat	DF-1	DF-2	Prob
12	295	I0295	Poly	-1,165	0,116	0,549	696,04	12,541	9	0,184494
22	309	I0309	Poly	-0,133	0,095	-1,316	696,04	12,877	9	0,168272
11	293	I0293	Poly	0,106	0,092	-1,831	696,04	13,433	9	0,143980
24	312	I0312	Poly	-1,358	0,127	0,607	643,91	13,473	9	0,142337
18	303	I0303	Poly	-1,055	0,113	0,967	696,04	16,233	9	0,062176
1	36	I0036	Poly	1,415	0,091	-1,250	646,81	16,654	9	0,054416
6	151	I0151	Poly	0,930	0,087	3,654	696,04	17,181	9	0,045957
31	320	I0320	Poly	-1,666	0,137	-0,447	646,81	17,243	9	0,045044
32	321	I0321	Poly	-0,105	0,095	-2,071	696,04	17,622	9	0,039823
33	322	I0322	Poly	-0,269	0,102	-2,499	644,88	17,752	9	0,038159
25	313	I0313	Poly	0,642	0,088	-1,086	695,08	18,177	9	0,033173
8	238	I0238	Poly	1,405	0,087	2,432	696,04	18,387	9	0,030941
26	314	I0314	Poly	-0,717	0,110	-1,980	646,81	18,823	9	0,026743
7	153	I0153	Poly	2,446	0,096	2,044	696,04	21,206	9	0,011767
14	299	I0299	Poly	1,864	0,089	-0,971	696,04	21,430	9	0,010874
9	246	I0246	Poly	1,327	0,091	2,401	646,81	21,487	9	0,010655
4	94	I0094	Poly	2,028	0,094	-1,469	645,84	22,306	9	0,007959
21	306	I0306	Poly	-1,372	0,127	1,886	646,81	22,393	9	0,007715
23	311	I0311	Poly	-1,417	0,123	-0,307	696,04	22,410	9	0,007668
2	50	I0050	Poly	1,837	0,093	1,845	646,81	23,402	9	0,005355
3	91	I0091	Poly	1,494	0,087	-1,529	696,04	28,434	9	0,000807
Mean			0,000		-0,149	Totals		499,963	315	<input type="checkbox"/> Highlight extreme itemfit values. Set +/- 2.5			
Std Devn			1,315		1,611	Total Prob		0,000000	<input checked="" type="checkbox"/> below the Bonferroni Select Probability B				
Selector 3 Item x01020204621-5 1										File Text Format		Adi = 0.000286 for 35	
										<input checked="" type="radio"/> Fixed		0,01	
< Display Control										Sort Items by ...		Save	
										ChiSq Prob Orde		Tab Delim	
										Copy		Resp Types >	
										Chi Squ >			

Afsluttende DIF-analyse: køn

No.	Item	Class Interval				koen				Class Interval-by- koen			
		MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
11	I0293	1.48776	1.81385	9	0.062467	1.37687	1.67865	1	0.195545	0.69722	0.85003	9	0.570100
12	I0295	1.38600	1.31543	9	0.224914	0.31241	0.29651	1	0.586264	1.07280	1.01818	9	0.423579
13	I0298	0.69891	0.98049	9	0.454794	0.42996	0.60320	1	0.437636	0.35180	0.49353	9	0.879345
14	I0299	2.26220	2.60968	9	0.005730	0.23936	0.27613	1	0.599421	0.52943	0.61076	9	0.788477
15	I0300	1.11715	1.14981	9	0.325060	0.12138	0.12492	1	0.723863	1.80157	1.85424	9	0.055972
16	I0301	1.27909	1.52015	9	0.136621	1.02671	1.22020	1	0.269712	0.50349	0.59838	9	0.798922
17	I0302	0.28961	0.28386	9	0.979129	0.84080	0.82411	1	0.364320	1.03263	1.01213	9	0.428595
18	I0303	1.77565	1.61819	9	0.106041	1.60272	1.46059	1	0.227252	2.34533	2.13735	9	0.024593
19	I0304	0.49426	0.52243	9	0.858806	0.35421	0.37440	1	0.540829	1.16780	1.23437	9	0.270629
20	I0305	0.64258	0.78675	9	0.628754	4.72013	5.77920	1	0.016463	1.07089	1.31117	9	0.227151
21	I0306	2.17551	1.58387	9	0.116185	0.48686	0.35445	1	0.551812	1.70867	1.24399	9	0.264891
22	I0309	1.38625	1.63999	9	0.100132	0.01966	0.02326	1	0.878790	1.00871	1.19334	9	0.295990
23	I0311	2.37949	2.66143	9	0.004844	0.56295	0.62966	1	0.427749	0.91494	1.02335	9	0.419408
24	I0312	1.49893	1.39478	9	0.186569	1.17158	1.09017	1	0.296832	1.43866	1.33870	9	0.213208
25	I0313	1.99435	2.25766	9	0.017122	0.28153	0.31870	1	0.572572	0.68287	0.77303	9	0.641555
26	I0314	2.04050	2.83750	9	0.002764	0.23616	0.32840	1	0.566805	0.93947	1.30642	9	0.229840
27	I0315	0.85303	0.75665	9	0.656815	0.50744	0.45011	1	0.502504	0.60504	0.53668	9	0.848231
28	I0316	1.20155	1.53761	9	0.130876	0.23320	0.29842	1	0.585062	1.70288	2.17916	9	0.021803
29	I0317	1.00128	1.18289	9	0.302760	0.66859	0.78986	1	0.374445	1.08622	1.28324	9	0.242235
30	I0318	0.51302	0.58341	9	0.811287	3.18231	3.61899	1	0.057563	0.76284	0.86752	9	0.554128
31	I0320	1.90334	2.26802	9	0.016686	0.05618	0.06694	1	0.795906	2.50529	2.98531	9	0.001717
32	I0321	1.83177	2.34750	9	0.013024	0.05116	0.06556	1	0.797982	1.02902	1.31873	9	0.223191
33	I0322	1.78783	2.48354	9	0.008588	5.00440	6.95179	1	0.008566	0.53459	0.74262	9	0.669861
34	I0323	1.07702	1.22643	9	0.275272	0.12764	0.14534	1	0.703132	0.60004	0.68328	9	0.724485
35	I0324	0.38900	0.41421	9	0.927843	0.44770	0.47671	1	0.490163	0.39969	0.42559	9	0.921646

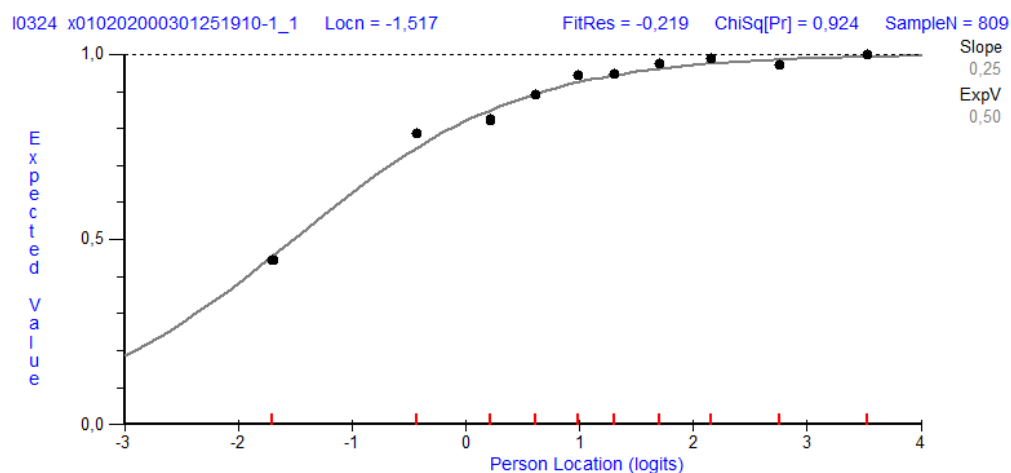
Afsluttende DIF-analyse: skolestørrelse

No.	Item	Class Interval				skolestr				Class Interval-by- skolestr			
		MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
11	I0293	1.48776	1.82200	9	0.061067	5.81441	7.12065	1	0.007807	0.49015	0.60026	9	0.797348
12	I0295	1.38600	1.30948	9	0.228042	0.82688	0.78123	1	0.377061	0.64281	0.60733	9	0.791390
13	I0298	0.69891	0.99515	9	0.442552	0.08602	0.12248	1	0.726462	1.14827	1.63498	9	0.101647
14	I0299	2.26220	2.66065	9	0.004857	5.53060	6.50471	1	0.010959	1.23492	1.45243	9	0.161886
15	I0300	1.11715	1.13174	9	0.337617	1.33018	1.34755	1	0.246131	0.54678	0.55392	9	0.834923
16	I0301	1.27909	1.55227	9	0.125860	3.73028	4.52699	1	0.033716	1.55754	1.89019	9	0.050430
17	I0302	0.28961	0.28868	9	0.977871	0.15545	0.15494	1	0.693994	2.34007	2.33254	9	0.013701
18	I0303	1.77565	1.60786	9	0.108960	0.14020	0.12695	1	0.721715	1.95875	1.77366	9	0.069800
19	I0304	0.49426	0.51810	9	0.861973	0.72245	0.75730	1	0.384502	0.55584	0.58266	9	0.811907
20	I0305	0.64258	0.77695	9	0.637899	1.23141	1.48892	1	0.222791	0.65666	0.79397	9	0.622025
21	I0306	2.17551	1.57875	9	0.117742	2.30374	1.67180	1	0.196486	1.18514	0.86005	9	0.560954
22	I0309	1.38625	1.63672	9	0.101004	0.29543	0.34880	1	0.554979	0.84676	0.99975	9	0.438656
23	I0311	2.37949	2.64866	9	0.005065	0.55831	0.62146	1	0.430771	0.57951	0.64506	9	0.758730
24	I0312	1.49893	1.38120	9	0.192758	1.56145	1.43880	1	0.230767	0.63559	0.58566	9	0.809436
25	I0313	1.99435	2.25226	9	0.017408	2.29319	2.58974	1	0.108004	0.29451	0.33259	9	0.964221
26	I0314	2.04050	2.83491	9	0.002801	0.18084	0.25125	1	0.616351	0.89813	1.24779	9	0.262649
27	I0315	0.85303	0.77262	9	0.641928	0.02358	0.02136	1	0.883866	2.47421	2.24100	9	0.018015
28	I0316	1.20155	1.51973	9	0.136957	0.54529	0.68968	1	0.406584	1.00533	1.27155	9	0.248951
29	I0317	1.00128	1.17431	9	0.308399	0.14755	0.17305	1	0.677529	0.66222	0.77666	9	0.638170
30	I0318	0.51302	0.58155	9	0.812815	3.00811	3.40994	1	0.065256	0.57858	0.65587	9	0.749111
31	I0320	1.90334	2.25445	9	0.017393	1.97617	2.34072	1	0.126526	1.92727	2.28281	9	0.015940
32	I0321	1.83177	2.33575	9	0.013487	0.00008	0.00010	1	0.992709	0.72908	0.92968	9	0.498412
33	I0322	1.78783	2.50573	9	0.008006	0.60768	0.85170	1	0.356426	1.48217	2.07733	9	0.029495
34	I0323	1.07702	1.23174	9	0.272048	0.89819	1.02722	1	0.311161	0.80912	0.92536	9	0.502212
35	I0324	0.38900	0.41872	9	0.925419	0.11017	0.11859	1	0.730683	1.16719	1.25637	9	0.257638

Afsluttende DIF-analyse: skolestørrelse

No.	Item	Class Interval				oestvest				Class Interval-by- oestvest			
		MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
11	I0293	1.48776	1.80187	9	0.064579	0.13645	0.16526	1	0.684498	0.41056	0.49724	9	0.876838
12	I0295	1.38600	1.32952	9	0.217641	0.00047	0.00045	1	0.982902	1.97752	1.89694	9	0.049480
13	I0298	0.69891	0.98024	9	0.455004	0.08129	0.11402	1	0.735744	0.37727	0.52913	9	0.853854
14	I0299	2.26220	2.62884	9	0.005382	0.01843	0.02142	1	0.883640	1.04605	1.21559	9	0.281942
15	I0300	1.11715	1.13101	9	0.338126	1.65641	1.67698	1	0.195791	0.46506	0.47084	9	0.894454
16	I0301	1.27909	1.53650	9	0.131053	2.41526	2.90133	1	0.088939	1.04593	1.25643	9	0.257431
17	I0302	0.28961	0.28281	9	0.979398	0.47480	0.46365	1	0.496176	0.79848	0.77972	9	0.635306
18	I0303	1.77565	1.58737	9	0.114945	0.17789	0.15903	1	0.690181	0.84422	0.75470	9	0.658631
19	I0304	0.49426	0.52859	9	0.854254	6.99216	7.47782	1	0.006420	1.22620	1.31137	9	0.227229
20	I0305	0.64258	0.78278	9	0.632455	0.08831	0.10758	1	0.743017	1.26319	1.53881	9	0.130277
21	I0306	2.17551	1.59525	9	0.112796	8.99023	6.59233	1	0.010456	1.47207	1.07943	9	0.375754
22	I0309	1.38625	1.62860	9	0.103191	0.60675	0.71282	1	0.398796	0.48306	0.56750	9	0.824225
23	I0311	2.37949	2.66674	9	0.004780	2.18507	2.44885	1	0.118049	0.87328	0.97870	9	0.456215
24	I0312	1.49893	1.39507	9	0.186448	1.67445	1.55843	1	0.212339	1.39877	1.30185	9	0.232289
25	I0313	1.99435	2.27232	9	0.016382	3.06650	3.49391	1	0.062011	0.81680	0.93065	9	0.497560
26	I0314	2.04050	2.84698	9	0.002699	0.45224	0.63098	1	0.427284	1.08836	1.51852	9	0.137372
27	I0315	0.85303	0.76258	9	0.651288	0.03659	0.03271	1	0.856512	1.34075	1.19860	9	0.292630
28	I0316	1.20155	1.51943	9	0.137070	1.67570	2.11901	1	0.145967	0.86832	1.09804	9	0.361883
29	I0317	1.00128	1.18048	9	0.304340	0.86337	1.01788	1	0.313363	0.92971	1.09610	9	0.363190
30	I0318	0.51302	0.57487	9	0.818246	0.00707	0.00792	1	0.929067	0.17369	0.19464	9	0.994733
31	I0320	1.90334	2.20623	9	0.020105	0.06057	0.07021	1	0.791116	0.80743	0.93592	9	0.493004
32	I0321	1.83177	2.33982	9	0.013322	0.74869	0.95634	1	0.328451	0.75212	0.96073	9	0.471486
33	I0322	1.78783	2.49698	9	0.008233	3.73235	5.21279	1	0.022731	0.95490	1.33366	9	0.215731
34	I0323	1.07702	1.22568	9	0.275739	0.42756	0.48657	1	0.485689	0.52469	0.59710	9	0.799994
35	I0324	0.38900	0.41392	9	0.927993	0.01501	0.01597	1	0.899454	0.40158	0.42731	9	0.920690

Eksempel på ICC-kurve over godkendt opgave.



Kilde: Styrelsen for It og Læring

Dansk læsning 2. klasse – tekstforståelse

I opgaveafprøvnningen januar 2018 blev 60 opgaver afprøvet, heraf forkastes 7 opgaver og 53 (88 %) opgaver godkendes og tilføjes Opgavebanken.

Alle 7 opgaver forkastes pga. itemfit.

Analysen indeholder 10 linkopgaver: I0150, I0151, I0152, I0153, I0155, I0157, I0158, I0159, I0160 og I0161.

Initial itemfit-analyse: viser et udpluk af de opgaver der forkastes.

Seq	Item	Type	Location	SE	FitResid	DF	ChiSq	DF	Prob	F-stat	DF-1	DF-2	Prob
65	328	I0328	Poly	-0,155	0,025	4,577	840,97	22,163	9	0,008377
45	308	I0308	Poly	-0,062	0,032	3,977	778,31	22,727	9	0,006841
15	278	I0278	Poly	-0,147	0,026	5,856	774,39	23,238	9	0,005684
30	293	I0293	Poly	0,041	0,027	3,472	772,44	24,092	9	0,004158
68	331	I0331	Poly	0,256	0,074	-2,604	836,07	24,586	9	0,003467
32	295	I0295	Poly	-0,263	0,078	-3,541	839,99	25,534	9	0,002435
69	332	I0332	Poly	1,513	0,082	-1,587	772,44	26,555	9	0,001656
10	161	I0161	Poly	-0,653	0,084	-2,661	829,22	26,656	9	0,001594
34	297	I0297	Poly	-0,057	0,080	-3,487	757,75	32,905	9	0,000139
7	158	I0158	Poly	-0,756	0,085	-3,583	839,99	39,531	9	0,000009
6	157	I0157	Poly	0,469	0,077	-4,171	761,67	41,667	9	0,000004
4	153	I0153	Poly	1,034	0,075	-2,191	841,95	44,153	9	0,000001
27	290	I0290	Poly	-0,319	0,035	4,886	775,37	46,918	9	0,000000
23	286	I0286	Poly	-0,231	0,078	-5,011	840,97	48,173	9	0,000000
1	150	I0150	Poly	-0,355	0,079	-2,497	839,99	48,843	9	0,000000
8	159	I0159	Poly	-0,035	0,079	-5,450	777,33	52,520	9	0,000000
70	333	I0333	Poly	-0,891	0,093	-3,461	757,75	54,517	9	0,000000
2	151	I0151	Poly	1,081	0,078	-3,568	774,39	55,648	9	0,000000
20	283	I0283	Poly	-0,041	0,033	6,512	835,09	59,583	9	0,000000
25	288	I0288	Poly	-0,110	0,081	-6,203	757,75	71,266	9	0,000000
33	296	I0296	Poly	-0,588	0,086	-5,323	781,25	77,176	9	0,000000

Mean 0,000 1,063 Totals 1.497,717 630
Std Devn 0,468 3,068 Total Prob 0,000000

Selector Item x010203000301252960-1 1
33
Sort Items by...
ChiSq Prob Ord

File Text Format
Fixed
Tab Delim
Copy
Resp Types >
Chi Squ >

Highlight extreme fit-resid values. Set +/- 2.5
below the Bonferroni Select Probability B
0,01
Adi = 0.000143 for 70

Afsluttende itemfit-analyse: de seks statistisk signifikante opgaver er linkopgaver, der ikke slettes.

Seq	Item	Type	Location	SE	FitResid	DF	ChiSq	DF	Prob	F-stat	DF-1	DF-2	Prob
40	309	I0309	Poly	0,141	0,029	1,192	758,12	17,652	9	0,039428
27	294	I0294	Poly	0,274	0,030	3,793	761,05	18,213	9	0,032787
30	299	I0299	Poly	-0,032	0,029	0,418	773,75	18,219	9	0,032715
25	292	I0292	Poly	-0,050	0,025	5,849	838,23	18,665	9	0,028197
56	325	I0325	Poly	-0,158	0,032	2,823	839,21	18,937	9	0,025737
26	293	I0293	Poly	0,007	0,027	3,288	770,82	19,591	9	0,020614
12	275	I0275	Poly	0,252	0,030	3,460	777,66	23,761	9	0,004697
59	328	I0328	Poly	-0,186	0,025	4,406	839,21	25,120	9	0,002842
44	313	I0313	Poly	-0,087	0,025	3,955	837,25	25,400	9	0,002560
62	331	I0331	Poly	0,208	0,074	-2,529	834,32	26,542	9	0,001665
28	295	I0295	Poly	-0,304	0,078	-3,553	838,23	26,642	9	0,001603
39	308	I0308	Poly	-0,095	0,031	3,658	776,68	26,794	9	0,001513
23	289	I0289	Poly	0,061	0,028	-0,201	838,23	27,051	9	0,001372
10	161	I0161	Poly	-0,693	0,084	-2,472	827,48	27,439	9	0,001183
63	332	I0332	Poly	1,459	0,082	-1,639	770,82	28,867	9	0,000682
4	153	I0153	Poly	0,985	0,075	-2,265	840,18	41,333	9	0,000004
6	157	I0157	Poly	0,421	0,077	-4,131	759,10	44,155	9	0,000001
7	158	I0158	Poly	-0,794	0,085	-3,643	838,23	44,591	9	0,000001
1	150	I0150	Poly	-0,395	0,079	-2,519	838,23	49,132	9	0,000000
2	151	I0151	Poly	1,030	0,078	-3,602	772,77	50,105	9	0,000000
8	159	I0159	Poly	-0,080	0,079	-5,392	775,71	50,831	9	0,000000

Mean 0,000 1,180 Totals 1.098,870 567
Std Devn 0,466 2,531 Total Prob 0,000000

Selector Item x0102030540008-1 1
8
Sort Items by...
ChiSq Prob Ord

File Text Format
Fixed
Tab Delim
Copy
Resp Types >
Chi Squ >

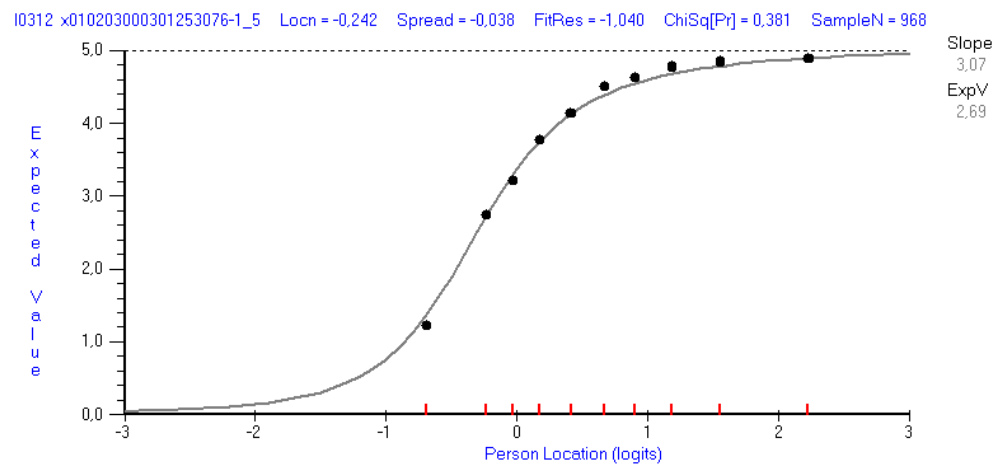
Highlight extreme fit-resid values. Set +/- 2.5
below the Bonferroni Select Probability B
0,01
Adi = 0.000159 for 63

Afsluttende DIF-analyse: køn

		Class Interval				køen				Class Interval-by- køen			
No.	Item	MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
11	I0274	1.69288	1.61940	9	0.105333	4.51428	4.31834	1	0.038021	0.37046	0.35438	9	0.955983
12	I0275	2.61133	2.22169	9	0.018990	0.00539	0.00459	1	0.946066	0.79152	0.67341	9	0.733460
13	I0276	1.33415	1.19644	9	0.293861	4.41819	3.96213	1	0.046901	1.03468	0.92788	9	0.499934
14	I0277	0.76127	0.51119	9	0.867109	0.22288	0.14966	1	0.698933	0.53750	0.36094	9	0.953289
15	I0278	2.03292	1.15501	9	0.321189	1.49358	0.84858	1	0.357237	1.91652	1.08888	9	0.368404
16	I0279	0.50624	0.40033	9	0.935107	4.06681	3.21596	1	0.073329	1.29668	1.02539	9	0.417655
17	I0280	0.93119	1.03003	9	0.413840	0.15147	0.16754	1	0.682412	0.83247	0.92083	9	0.506083
18	I0281	0.97764	0.81946	9	0.598311	0.50725	0.42517	1	0.514570	1.29121	1.08229	9	0.373325
19	I0282	1.39948	1.21645	9	0.281250	0.05746	0.04995	1	0.823244	1.58999	1.38204	9	0.191990
20	I0284	1.18012	1.23808	9	0.268026	2.39109	2.50853	1	0.113626	0.44625	0.46817	9	0.896271
21	I0285	1.52867	1.36850	9	0.198311	0.01134	0.01016	1	0.919841	2.93604	2.62842	9	0.005362
22	I0287	0.68860	0.71313	9	0.697196	0.02326	0.02408	1	0.876755	0.56833	0.58858	9	0.807135
23	I0289	2.86246	3.03497	9	0.001407	1.31895	1.39845	1	0.237314	1.08104	1.14619	9	0.327102
24	I0291	1.78633	1.61624	9	0.106425	0.39172	0.35442	1	0.551810	1.32992	1.20328	9	0.289490
25	I0292	1.82221	1.20866	9	0.285925	1.41587	0.93914	1	0.332780	2.13096	1.41345	9	0.177716
26	I0293	2.16130	1.80311	9	0.064185	0.23355	0.19485	1	0.659028	2.39296	1.99638	9	0.037068
27	I0294	1.97172	1.64449	9	0.098763	16.92971	14.12008	1	0.000173	0.26714	0.22280	9	0.991280
28	I0295	2.69466	3.39136	9	0.000436	1.33303	1.67769	1	0.195590	0.58954	0.74197	9	0.670501
29	I0298	0.79615	0.74597	9	0.666782	2.07915	1.94811	1	0.163169	0.85677	0.80277	9	0.613814
30	I0299	1.72003	1.71498	9	0.081729	2.42594	2.41881	1	0.120298	0.53061	0.52905	9	0.854016
31	I0300	1.70938	1.44984	9	0.162746	0.94449	0.80109	1	0.371052	0.79850	0.67726	9	0.729974
32	I0301	1.29445	1.11217	9	0.351161	0.06434	0.05528	1	0.814198	1.76237	1.51421	9	0.138287
33	I0302	1.54327	1.13289	9	0.336505	2.69524	1.97854	1	0.159942	1.41522	1.03889	9	0.406664
34	I0303	1.33417	0.86583	9	0.555604	0.55237	0.35847	1	0.549548	1.52570	0.99013	9	0.446560
35	I0304	0.66354	0.59938	9	0.798184	0.99395	0.89784	1	0.343640	1.13869	1.02859	9	0.414983
36	I0305	1.03010	0.78456	9	0.630791	1.28730	0.98046	1	0.322405	1.20317	0.91638	9	0.510072
37	I0306	1.46200	1.17466	9	0.308017	0.13735	0.11036	1	0.739807	2.59066	2.08151	9	0.028896
38	I0307	1.84314	2.00836	9	0.035666	1.09842	1.19689	1	0.274263	1.03431	1.12703	9	0.340509
39	I0308	2.95608	2.38241	9	0.011594	0.33420	0.26934	1	0.603906	1.35868	1.09501	9	0.363839
40	I0309	2.05334	2.00340	9	0.036338	7.35357	7.17470	1	0.007551	2.44734	2.38782	9	0.011421
41	I0310	1.63620	1.18071	9	0.303880	0.93963	0.67805	1	0.410490	0.38078	0.27478	9	0.981425
42	I0311	1.64379	1.57210	9	0.119361	0.04970	0.04753	1	0.827463	1.50664	1.44093	9	0.166315
43	I0312	0.95457	1.07066	9	0.382174	0.51980	0.58301	1	0.445377	0.88313	0.99054	9	0.446222
44	I0313	2.90946	2.21170	9	0.019496	0.00007	0.00005	1	0.994366	1.65301	1.25657	9	0.256984
45	I0314	1.03669	0.80148	9	0.615024	0.05098	0.03941	1	0.842698	0.44483	0.34391	9	0.960073
46	I0315	1.90924	1.32511	9	0.219718	1.14314	0.79339	1	0.373362	1.75143	1.21557	9	0.281779
47	I0316	1.05587	0.75128	9	0.661835	2.09469	1.49043	1	0.222500	1.18078	0.84016	9	0.579136
48	I0317	0.74046	0.71922	9	0.691578	2.63994	2.56423	1	0.109728	0.68192	0.66236	9	0.743390
49	I0318	0.94872	0.75587	9	0.657554	0.25565	0.20368	1	0.651893	2.32825	1.85498	9	0.055546
50	I0319	0.38029	0.32947	9	0.965383	4.13611	3.58336	1	0.058722	2.12584	1.84174	9	0.057478
51	I0320	1.32387	1.32030	9	0.222159	10.49023	10.46190	1	0.001269	0.71189	0.70997	9	0.700111
52	I0321	1.45455	1.23856	9	0.267802	2.32259	1.97769	1	0.160058	0.54032	0.46009	9	0.901365
53	I0322	0.70100	0.49517	9	0.878365	0.00412	0.00291	1	0.957139	1.86376	1.31653	9	0.223949
54	I0323	1.39799	1.24820	9	0.262029	0.01270	0.01134	1	0.915304	1.75518	1.56711	9	0.120905
55	I0324	1.29136	1.11547	9	0.348918	0.00312	0.00269	1	0.958782	1.50238	1.29775	9	0.234117
56	I0325	2.25166	1.81259	9	0.062347	1.37300	1.10526	1	0.293412	0.58192	0.46845	9	0.896133
57	I0326	1.17710	1.24096	9	0.266308	0.25928	0.27334	1	0.601228	0.64835	0.68353	9	0.724303
58	I0327	1.02365	1.08237	9	0.373294	0.00913	0.00966	1	0.921724	0.41364	0.43736	9	0.915049
59	I0328	2.66795	1.84651	9	0.056696	0.03920	0.02713	1	0.869269	1.79590	1.24296	9	0.264965
60	I0329	1.05421	1.04778	9	0.399850	3.78966	3.76654	1	0.052638	0.66277	0.65873	9	0.746645
61	I0330	1.83458	1.47592	9	0.152539	0.40064	0.32231	1	0.570378	1.03578	0.83329	9	0.585512
62	I0331	2.75448	3.15632	9	0.000920	0.00362	0.00415	1	0.948799	0.63825	0.73137	9	0.680353
63	I0332	2.99234	3.43137	9	0.000371	0.04867	0.05581	1	0.813272	1.05810	1.21334	9	0.283154

Main Plot splavnings: 63 ght ...
 Select Level of Precision Dimension: 1x-105 Input 0.03 Save Copy

Eksempel på ICC-kurve over godkendt opgave.



Kilde: Styrelsen for It og L ring

Dansk læsning 4. klasse – sprogforståelse

I opgaveafprøvningen januar 2018 afprøves 51 opgaver, heraf forkastes 27 opgaver og 24 (56 %) opgaver godkendes og tilføjes Opgavebanken.

23 af opgaverne forkastes pga. itemfit og 4 opgaver forkastes pga. DIF på geografi.

Analysen indeholder 10 linkopgaver: I0012, I0014, I0032, I0061, I0067, I0078, I0197, I0200, I0203 og I0204.

Initial itemfit-analyse: viser et udpluk af de opgaver der forkastes.

	Seq	Item	Type	Location	SE	FitResid	DF	ChiSq	DF	Prob	F-stat	DF-1	DF-2	Prob
15	217	I0217	Poly	-0,746	0,085	-2,609	742,51	28,021	9	0,000946
38	240	I0240	Poly	-0,047	0,076	-3,456	793,45	28,051	9	0,000935
32	234	I0234	Poly	-2,391	0,129	-2,372	793,45	28,651	9	0,000742
46	248	I0248	Poly	0,746	0,084	2,893	661,21	29,535	9	0,000526
17	219	I0219	Poly	-0,523	0,080	-4,144	792,47	34,728	9	0,000067
58	260	I0260	Poly	-2,420	0,143	-1,971	661,21	36,851	9	0,000028
48	250	I0250	Poly	1,653	0,091	1,697	742,51	37,932	9	0,000018
54	256	I0256	Poly	-0,850	0,087	-4,289	743,49	39,385	9	0,000010
59	261	I0261	Poly	-0,065	0,076	-4,063	793,45	40,178	9	0,000007
16	218	I0218	Poly	0,878	0,085	2,128	661,21	41,177	9	0,000005
23	225	I0225	Poly	0,501	0,076	5,060	793,45	41,950	9	0,000003
34	236	I0236	Poly	-1,226	0,099	-2,631	660,23	42,806	9	0,000002
22	224	I0224	Poly	1,415	0,092	3,036	661,21	45,242	9	0,000001
26	228	I0228	Poly	1,978	0,096	3,313	792,47	46,657	9	0,000000
44	246	I0246	Poly	-0,215	0,077	-5,450	792,47	49,691	9	0,000000
21	223	I0223	Poly	-1,718	0,107	-3,882	742,51	57,544	9	0,000000
20	222	I0222	Poly	1,740	0,090	3,388	793,45	59,376	9	0,000000
42	244	I0244	Poly	0,621	0,079	5,839	742,51	62,168	9	0,000000
13	215	I0215	Poly	1,017	0,087	5,869	661,21	91,009	9	0,000000
57	259	I0259	Poly	-0,195	0,080	6,780	742,51	99,466	9	0,000000
45	247	I0247	Poly	1,463	0,088	6,800	742,51	132,654	9	0,000000
				Mean	0,000	0,263	Totals	1.697,662	549					
				Std Devn	1,067	2,978	Total Prob	0,000000						

Highlight extreme fit/resid values. Set +/- 2.5
 below the Bonferroni Select Probability B
0,01
 Adi = 0.000164 for 61

Selector: 45 Item: x010401000301253072-1 1
 File Text Format: Fixed Tab Delim

Afsluttende itemfit-analyse

Seq	Item	Type	Location	SE	FitResid	DF	ChiSq	DF	Prob	F-stat	DF-1	DF-2	Prob
18	231	I0231	Poly	1,285	0,084	1,678	778,43	13,006	9	0,162340
23	241	I0241	Poly	-0,929	0,088	-0,659	730,20	13,323	9	0,148509
30	253	I0253	Poly	-0,291	0,081	-1,987	731,16	13,578	9	0,138135
20	235	I0235	Poly	0,798	0,081	3,090	730,20	14,163	9	0,116643
2	14	I0014	Poly	0,394	0,077	0,314	778,43	14,399	9	0,108815
12	214	I0214	Poly	-1,498	0,099	-1,044	729,24	14,685	9	0,099958
26	245	I0245	Poly	-0,157	0,085	-0,103	651,10	14,894	9	0,093900
7	197	I0197	Poly	0,550	0,080	-0,130	730,20	15,611	9	0,075467
34	263	I0263	Poly	-2,204	0,129	-1,418	651,10	16,371	9	0,059525
3	32	I0032	Poly	0,299	0,077	1,465	777,47	16,982	9	0,049007
14	217	I0217	Poly	-0,826	0,086	-2,051	730,20	18,211	9	0,032803
15	220	I0220	Poly	1,548	0,091	2,261	730,20	18,438	9	0,030419
10	204	I0204	Poly	0,329	0,084	-2,204	651,10	18,726	9	0,027623
11	213	I0213	Poly	-0,213	0,078	-1,546	778,43	18,880	9	0,026238
9	203	I0203	Poly	0,970	0,088	-0,381	651,10	19,540	9	0,020971
5	67	I0067	Poly	0,342	0,077	4,569	778,43	19,630	9	0,020339
27	249	I0249	Poly	-0,778	0,083	-2,022	777,47	19,725	9	0,019689
28	251	I0251	Poly	-1,696	0,111	-1,006	651,10	20,059	9	0,017552
4	61	I0061	Poly	0,261	0,079	4,612	729,24	21,522	9	0,010525
22	240	I0240	Poly	-0,111	0,077	-3,019	778,43	22,518	9	0,007377
16	221	I0221	Poly	0,270	0,084	4,360	651,10	23,740	9	0,004734

Mean	0,000	0,422	Totals	490,272	306
Std Devn	0,904	1,980	Total Prob	0,000000	

Selector: 16 Item: x010401000301252976-1
 Sort Items by: ChiSq Prob Order
 File Text Format: Fixed
 Tab Delim:
 Copy: Resp Types: Chi Squ:

Afsluttende DIF-analyse: køn

No.	Item	Class Interval				køen				Class Interval-by- køen			
		MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
11	I0213	2,05059	2,30981	9	0,014512	4,59934	5,18075	1	0,023094	0,90449	1,01883	9	0,422919
12	I0214	1,58078	1,82761	9	0,060019	0,20067	0,23201	1	0,630185	0,83637	0,96696	9	0,466115
13	I0216	0,57058	0,57361	9	0,819379	12,24525	12,31039	1	0,000475	0,83852	0,84299	9	0,576546
14	I0217	1,92509	2,32938	9	0,013724	0,00074	0,00090	1	0,976362	1,65147	1,99830	9	0,036914
15	I0220	2,01471	1,76815	9	0,070775	0,11621	0,10198	1	0,749563	1,48598	1,30412	9	0,230765
16	I0221	2,65700	2,33655	9	0,013529	1,40449	1,23511	1	0,266826	0,82686	0,72714	9	0,684216
17	I0226	0,61757	0,62002	9	0,780574	0,30282	0,30403	1	0,581527	0,91261	0,91623	9	0,510233
18	I0231	1,53213	1,45480	9	0,160666	6,03685	5,73214	1	0,016883	1,82837	1,73609	9	0,077128
19	I0233	0,88826	0,85231	9	0,568034	0,01749	0,01679	1	0,896947	1,02543	0,98392	9	0,451905
20	I0235	1,59760	1,42499	9	0,173067	0,30478	0,27185	1	0,602257	0,70657	0,63023	9	0,771747
21	I0239	1,24723	1,18490	9	0,301592	0,14340	0,13623	1	0,712164	1,18587	1,12660	9	0,341227
22	I0240	2,51764	2,98731	9	0,001665	1,62900	1,93289	1	0,164850	0,33542	0,39800	9	0,936288
23	I0241	1,44354	1,57992	9	0,117076	0,04855	0,05314	1	0,817722	1,27370	1,39404	9	0,186581
24	I0242	1,07100	1,01364	9	0,427353	5,66515	5,36175	1	0,020892	1,28236	1,21368	9	0,283271
25	I0243	0,96959	1,05770	9	0,392093	0,43115	0,47033	1	0,493035	1,52175	1,66003	9	0,094695
26	I0245	1,67392	1,75466	9	0,073703	0,92143	0,96588	1	0,326073	0,79873	0,83725	9	0,581884
27	I0249	2,08679	2,47321	9	0,008734	1,94266	2,30238	1	0,129585	0,58752	0,69631	9	0,712658
28	I0251	2,15424	2,54248	9	0,007136	0,33162	0,39139	1	0,531796	0,47818	0,56435	9	0,826695
29	I0252	1,24383	1,26434	9	0,252632	1,01364	1,03035	1	0,310398	1,18447	1,20400	9	0,288977
30	I0253	1,48451	1,70898	9	0,083159	4,36781	5,02824	1	0,025243	1,07462	1,23711	9	0,268709
31	I0254	0,99110	1,13631	9	0,334391	0,06189	0,07096	1	0,790047	1,51392	1,73572	9	0,077596
32	I0255	1,28362	1,37618	9	0,194618	2,80330	3,00544	1	0,083368	0,65914	0,70667	9	0,703155
33	I0258	1,11382	1,15340	9	0,322246	2,44990	2,53696	1	0,111621	0,58169	0,60236	9	0,795645
34	I0263	1,75530	2,29119	9	0,015546	0,48059	0,62731	1	0,428627	0,48311	0,63061	9	0,771357

Select Level of Precision: 0.03
 Save: Copy:

Afsluttende DIF-analyse: skolestørrelse

No.	Item	Class Interval				skolestr				Class Interval-by- skolestr			
		MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
11	I0213	2,05059	2,30354	9	0,014781	3,80429	4,27357	1	0,039037	0,78150	0,87791	9	0,544596
12	I0214	1,58078	1,81895	9	0,061500	0,00252	0,00290	1	0,957102	0,52153	0,60010	9	0,797509
13	I0216	0,57058	0,56509	9	0,826220	0,69734	0,69063	1	0,406196	0,81132	0,80352	9	0,613126
14	I0217	1,92509	2,30105	9	0,014946	0,46846	0,55995	1	0,454527	0,76605	0,91566	9	0,510746
15	I0220	2,01471	1,76086	9	0,072202	0,03437	0,03004	1	0,862428	1,10897	0,96924	9	0,464173
16	I0221	2,65700	2,34270	9	0,013263	0,25499	0,22483	1	0,635541	1,17172	1,03312	9	0,411675
17	I0226	0,61757	0,62221	9	0,778686	0,64574	0,65059	1	0,420163	1,16196	1,17069	9	0,310702
18	I0231	1,53213	1,43558	9	0,168482	0,15889	0,14888	1	0,699720	1,24897	1,17026	9	0,310877
19	I0233	0,88826	0,85283	9	0,567558	3,25285	3,12307	1	0,077664	0,71169	0,68329	9	0,724453
20	I0235	1,59760	1,43519	9	0,168788	3,86167	3,46908	1	0,062938	0,96360	0,86564	9	0,555782
21	I0239	1,24723	1,17725	9	0,306594	0,49321	0,46553	1	0,495289	0,64945	0,61301	9	0,786517
22	I0240	2,51764	3,00410	9	0,001557	0,37155	0,44334	1	0,505700	0,88708	1,05848	9	0,391495
23	I0241	1,44354	1,58241	9	0,116317	3,38312	3,70857	1	0,054529	1,02051	1,11868	9	0,346670
24	I0242	1,07100	1,00931	9	0,430882	1,44089	1,35790	1	0,244338	1,42180	1,33990	9	0,212577
25	I0243	0,96959	1,04482	9	0,402157	0,06639	0,07154	1	0,789168	0,57580	0,62048	9	0,780212
26	I0245	1,67392	1,80999	9	0,063293	0,40462	0,43752	1	0,508561	2,97864	3,22077	9	0,000790
27	I0249	2,08679	2,46897	9	0,008835	0,36936	0,43701	1	0,508760	0,63597	0,75244	9	0,660748
28	I0251	2,15424	2,54491	9	0,007083	0,11312	0,13364	1	0,714823	0,56135	0,66315	9	0,742615
29	I0252	1,24383	1,26237	9	0,253757	0,32691	0,33178	1	0,564774	1,12689	1,14369	9	0,328925
30	I0253	1,48451	1,69431	9	0,086506	0,11923	0,13608	1	0,712313	0,93000	1,06143	9	0,389298
31	I0254	0,99110	1,12518	9	0,342236	0,22794	0,25877	1	0,611128	0,86775	0,98514	9	0,450882
32	I0255	1,28362	1,36200	9	0,201326	0,03759	0,03988	1	0,841783	0,11841	0,12564	9	0,999058
33	I0258	1,11382	1,15820	9	0,318982	0,82293	0,85572	1	0,355226	1,11193	1,15623	9	0,320320
34	I0263	1,75530	2,29304	9	0,015444	0,01123	0,01467	1	0,903602	0,58021	0,75796	9	0,655583

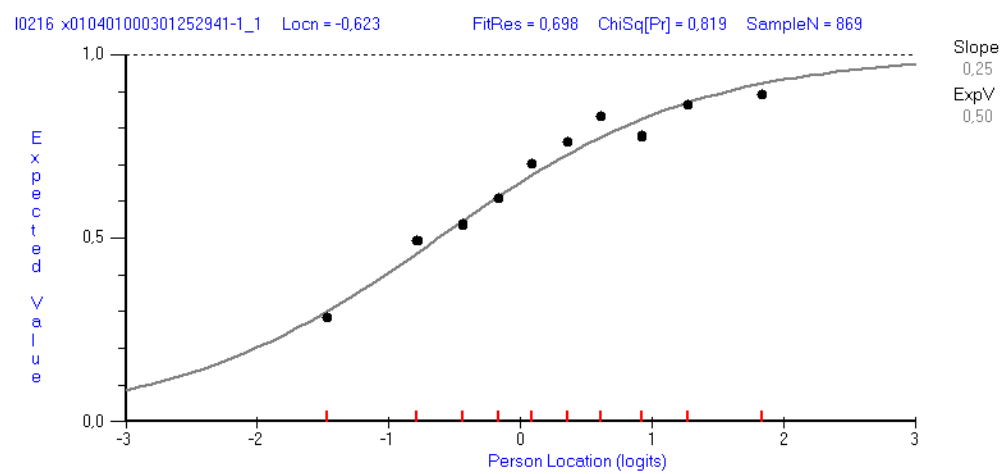
< Main Plot solovina: 34 ght Select Level of Precision Denominator: 14-102 Prob=0,000294 Adjust 0,03 Save Copy

Afsluttende DIF-analyse: geografi

No.	Item	Class Interval				oestvest				Class Interval-by- oestvest			
		MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
11	I0213	2,05059	2,28939	9	0,015445	1,69353	1,89075	1	0,169518	0,53487	0,59716	9	0,800009
12	I0214	1,58078	1,85367	9	0,055793	0,00034	0,00040	1	0,984152	1,85292	2,17279	9	0,022064
13	I0216	0,57058	0,56853	9	0,823470	0,08564	0,08534	1	0,770285	1,41286	1,40779	9	0,180319
14	I0217	1,92509	2,30021	9	0,014996	0,26766	0,31982	1	0,571897	0,76341	0,91217	9	0,513847
15	I0220	2,01471	1,75626	9	0,073126	0,05676	0,04948	1	0,824055	0,86107	0,75061	9	0,662443
16	I0221	2,65699	2,39383	9	0,011344	0,09742	0,08777	1	0,767119	2,95234	2,65993	9	0,004922
17	I0226	0,61757	0,61877	9	0,781649	1,24246	1,24487	1	0,264898	0,64286	0,64411	9	0,759596
18	I0231	1,53213	1,43414	9	0,169081	1,33695	1,25144	1	0,263621	1,02443	0,95891	9	0,472939
19	I0233	0,88826	0,85388	9	0,566594	0,03611	0,03471	1	0,852262	1,16310	1,11808	9	0,347300
20	I0235	1,59760	1,43022	9	0,170860	0,40914	0,36627	1	0,545225	1,03052	0,92255	9	0,504648
21	I0239	1,24723	1,18395	9	0,302201	0,89063	0,84545	1	0,358195	1,04181	0,98895	9	0,447698
22	I0240	2,51764	2,99594	9	0,001596	0,25094	0,29862	1	0,584916	0,70073	0,83385	9	0,584977
23	I0241	1,44354	1,56062	9	0,123060	0,52889	0,57179	1	0,449795	0,29459	0,31848	9	0,969069
24	I0242	1,07100	1,00447	9	0,434857	2,99903	2,81272	1	0,093985	0,87604	0,82162	9	0,596339
25	I0243	0,96959	1,05482	9	0,394335	0,97353	1,05910	1	0,303743	1,24285	1,35209	9	0,206117
26	I0245	1,67392	1,74304	9	0,076074	1,64827	1,71633	1	0,190626	0,25508	0,26561	9	0,983481
27	I0249	2,08679	2,45826	9	0,009161	0,53108	0,62562	1	0,429207	0,29621	0,34894	9	0,958130
28	I0251	2,15424	2,54071	9	0,007162	1,12548	1,32739	1	0,249688	0,34708	0,40935	9	0,930411
29	I0252	1,24383	1,25083	9	0,260458	1,36516	1,37285	1	0,241671	0,21770	0,21892	9	0,991833
30	I0253	1,48451	1,68932	9	0,087675	1,07367	1,22180	1	0,269364	0,61194	0,69637	9	0,712584
31	I0254	0,99110	1,12801	9	0,340231	0,71322	0,81174	1	0,367938	0,97439	1,10899	9	0,353860
32	I0255	1,28362	1,37886	9	0,193368	0,35111	0,37716	1	0,539302	1,09015	1,17104	9	0,310365
33	I0258	1,11382	1,16202	9	0,316404	6,15960	6,42614	1	0,011447	0,79566	0,83009	9	0,588455
34	I0263	1,75530	2,29920	9	0,015163	1,32229	1,73202	1	0,188614	0,58388	0,76480	9	0,649213

< Main Plot solovina: 34 ght Select Level of Precision Denominator: 14-102 Prob=0,000294 Adjust 0,03 Save Copy

Eksempel på ICC-kurve for godkendt opgave.



Kilde: Styrelsen for It og L ring

Matematik 3. klasse – geometri og måling

I opgaveafprøvningen januar 2018 afprøves 57 opgaver, heraf forkastes 16 opgaver og 41 (72 %) opgaver godkendes og tilføjes Opgavebanken.

Alle 16 opgaver forkastes pga. itemfit.

Analysen indeholder 10 linkopgaver: I0034, I0058, I0102, I0120, I0156, I0163, I0168, I0169, I0181 og I0185.

Initial itemfit-analyse: viser et udpluk af de opgaver der forkastes.

	Seq	Item	Type	Location	SE	FitResid	DF	ChiSq	DF	Prob	F-stat	DF-1	DF-2	Prob
6	163	I0163	Poly	1,329	0,078	3,195	815,48	30,046	9	0,000431
28	221	I0221	Poly	-0,250	0,040	2,864	815,48	30,832	9	0,000316
53	246	I0246	Poly	1,158	0,082	-4,615	726,18	31,160	9	0,000278
24	217	I0217	Poly	0,671	0,081	3,460	728,14	31,920	9	0,000206
36	229	I0229	Poly	1,408	0,083	-2,861	728,14	32,262	9	0,000179
11	204	I0204	Poly	-1,355	0,115	-2,105	726,18	32,427	9	0,000168
55	248	I0248	Poly	-0,505	0,088	-1,605	814,50	34,994	9	0,000060
8	169	I0169	Poly	1,148	0,082	5,107	726,18	39,919	9	0,000008
47	240	I0240	Poly	1,382	0,083	3,535	726,18	42,302	9	0,000003
27	220	I0220	Poly	-0,185	0,041	2,988	728,14	43,367	9	0,000002
37	230	I0230	Poly	1,206	0,077	-5,774	815,48	44,044	9	0,000001
30	223	I0223	Poly	0,981	0,081	-5,825	728,14	45,961	9	0,000001
9	181	I0181	Poly	0,985	0,081	6,481	726,18	48,670	9	0,000000
49	242	I0242	Poly	-0,027	0,081	-3,621	815,48	51,725	9	0,000000
34	227	I0227	Poly	0,351	0,078	-5,741	815,48	52,943	9	0,000000
31	224	I0224	Poly	0,670	0,077	-5,325	815,48	53,629	9	0,000000
35	228	I0228	Poly	0,797	0,081	-6,166	726,18	54,473	9	0,000000
46	239	I0239	Poly	0,270	0,079	-4,657	815,48	58,161	9	0,000000
16	209	I0209	Poly	0,181	0,054	5,492	815,48	66,920	9	0,000000
5	156	I0156	Poly	1,308	0,082	5,499	726,18	67,101	9	0,000000
4	120	I0120	Poly	1,171	0,082	6,022	728,14	69,647	9	0,000000

Mean	0,000	-0,177	Totals	1,674,285	603
Std Devn	1,355	2,969	Total Prob	0,000000	

Highlight extreme fit/resid values. Set +/- 2.5
 below the Bonferroni
 Select Probability B: 0.01
 Adi = 0.000149 for 67

Selector: 4 Item: x0203020410030002-3 1
 Sort Items by: ChiSq Prob Ord
 File Text Format: Fixed
 Tab Delim:

Afsluttende itemfit-analyse: de to statistisk signifikante opgaver er link opgaver, der ikke slettes.

Seq	Item	Type	Location	SE	FitResid	DF	ChiSq	DF	Prob	F-stat	DF-1	DF-2	Prob
25	221	I0221	Poly	-0,089	0,039	0,731	807,97	15,403	9	0,080437
24	220	I0220	Poly	-0,022	0,040	1,050	724,05	15,857	9	0,069921
29	231	I0231	Poly	-0,398	0,093	-2,487	720,15	16,515	9	0,056876
21	217	I0217	Poly	0,812	0,080	2,753	724,05	16,520	9	0,056795
31	234	I0234	Poly	-1,704	0,137	-1,197	720,15	16,647	9	0,054536
45	254	I0254	Poly	0,374	0,078	0,130	807,97	17,225	9	0,045304
33	236	I0236	Poly	-1,335	0,113	-1,567	807,97	17,295	9	0,044292
49	258	I0258	Poly	-2,255	0,170	0,457	720,15	17,356	9	0,043431
26	222	I0222	Poly	-1,007	0,109	-1,949	721,13	17,694	9	0,038900
28	226	I0226	Poly	-0,267	0,090	-1,650	724,05	17,816	9	0,037370
27	225	I0225	Poly	1,512	0,082	-1,450	721,13	19,602	9	0,020536
11	205	I0205	Poly	0,865	0,080	-0,596	724,05	19,651	9	0,020197
42	251	I0251	Poly	0,560	0,077	-2,703	807,97	19,962	9	0,018150
39	245	I0245	Poly	0,457	0,077	-1,956	807,97	21,222	9	0,011701
37	243	I0243	Poly	0,692	0,081	-3,157	721,13	21,799	9	0,009538
8	169	I0169	Poly	1,287	0,081	3,854	721,13	23,158	9	0,005852
6	163	I0163	Poly	1,467	0,077	1,805	807,97	24,362	9	0,003765
38	244	I0244	Poly	1,386	0,081	-2,663	724,05	27,099	9	0,001347
9	181	I0181	Poly	1,124	0,080	5,335	721,13	31,970	9	0,000202
5	156	I0156	Poly	1,441	0,081	4,384	721,13	38,661	9	0,000013
4	120	I0120	Poly	1,310	0,081	4,954	724,05	48,520	9	0,000000

Mean	0,000	-0,025	Totals	758,095	459
Std Devn	1,454	1,899	Total Prob	0,000000	

Highlight extreme fit values. Set +/- 2.5
 below the Bonferroni Select Probability B
0,01
 Adi = 0.000196 for 51

Selector: 4 Item: x0203020410030002-3 1
 Sort Items by: ChiSq Prob Orde Save File Text Format: Fixed Copy Resp Types > Chi Squ >

Afsluttende DIF-analyse: køn

No.	Item	Class Interval				køen				Class Interval-by- køen			
		MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
11	I0205	2.20105	2.35236	9	0.012792	7.57786	8.09879	1	0.004565	0.45967	0.49127	9	0.880957
12	I0206	0.50758	0.50557	9	0.871094	10.79185	10.74901	1	0.001073	1.57666	1.57040	9	0.119791
13	I0207	1.68076	1.55308	9	0.125541	5.25885	4.85933	1	0.027816	1.09058	1.00773	9	0.432065
14	I0208	0.51120	0.50964	9	0.868136	10.96793	10.93443	1	0.001001	0.94690	0.94401	9	0.485870
15	I0210	0.74109	0.74885	9	0.664076	10.55274	10.66335	1	0.001151	0.65063	0.65745	9	0.747751
16	I0211	0.81527	1.15889	9	0.318667	0.00720	0.01023	1	0.919416	1.40973	2.00390	9	0.036349
17	I0212	0.76794	1.15952	9	0.318051	0.00817	0.01233	1	0.911614	1.03931	1.56927	9	0.120140
18	I0213	1.13169	1.44262	9	0.165797	0.11636	0.14833	1	0.700260	0.58571	0.74664	9	0.666136
19	I0214	1.38155	1.30359	9	0.231093	0.29505	0.27840	1	0.597904	0.57098	0.53876	9	0.846677
20	I0216	0.39620	0.39150	9	0.939440	1.46447	1.44708	1	0.229387	1.27173	1.25663	9	0.257256
21	I0217	2.07476	1.93376	9	0.044499	3.89961	3.63460	1	0.056988	1.28360	1.19637	9	0.293987
22	I0218	0.94297	0.99190	9	0.445021	0.17863	0.18789	1	0.664770	0.93763	0.98628	9	0.449713
23	I0219	1.10657	1.04845	9	0.399421	1.67479	1.58682	1	0.208183	0.59032	0.55931	9	0.830752
24	I0220	1.67093	1.58928	9	0.114302	10.31928	9.81502	1	0.001815	0.67808	0.64494	9	0.758847
25	I0221	1.51224	1.48113	9	0.150399	0.25785	0.25255	1	0.615430	1.34381	1.31616	9	0.224216
26	I0222	1.90843	2.50634	9	0.007919	0.08744	0.11483	1	0.734809	0.83698	1.09921	9	0.360849
27	I0225	2.15437	2.42944	9	0.010071	5.85025	6.59721	1	0.010408	1.36628	1.54073	9	0.129567
28	I0226	1.97682	2.31835	9	0.014217	1.56347	1.83358	1	0.176141	0.56985	0.66829	9	0.738047
29	I0231	1.77959	2.28092	9	0.015935	0.22722	0.29123	1	0.589613	1.16263	1.49017	9	0.147301
30	I0233	1.11186	1.20320	9	0.289427	0.38860	0.42052	1	0.516861	0.77930	0.84332	9	0.576227
31	I0234	1.73683	2.23387	9	0.018371	1.60646	2.06620	1	0.151028	0.90373	1.16235	9	0.316340
32	I0235	1.44704	1.59776	9	0.111806	4.14419	4.57583	1	0.032756	2.72713	3.01118	9	0.001553
33	I0236	1.78586	2.28194	9	0.015778	0.29824	0.38108	1	0.537204	0.52251	0.66766	9	0.738662
34	I0237	1.07366	1.25140	9	0.260300	0.11093	0.12929	1	0.719292	0.80893	0.94284	9	0.486889
35	I0238	1.31828	1.34662	9	0.208999	0.40985	0.41866	1	0.517812	0.58948	0.60215	9	0.795772
36	I0241	0.89139	1.08662	9	0.370191	0.58049	0.70763	1	0.400505	0.82388	1.00433	9	0.434848
37	I0243	2.35326	2.84835	9	0.002656	8.57897	10.38387	1	0.001337	0.85916	1.03991	9	0.406141
38	I0244	2.86059	3.41643	9	0.000400	5.96723	7.12672	1	0.007776	0.99615	1.18971	9	0.298268
39	I0245	2.34535	2.66236	9	0.004782	0.03072	0.03487	1	0.851916	0.81025	0.91977	9	0.507043
40	I0247	1.29942	1.43004	9	0.170984	0.38885	0.42794	1	0.513214	0.82123	0.90378	9	0.521331
41	I0250	1.15115	1.25968	9	0.255486	1.86513	2.04097	1	0.153551	0.61689	0.67505	9	0.731950
42	I0251	2.23642	2.59551	9	0.005924	1.13065	1.31219	1	0.252340	0.36841	0.42757	9	0.920658
43	I0252	1.24620	1.32005	9	0.222442	0.05701	0.06039	1	0.805950	1.55109	1.64301	9	0.099264
44	I0253	1.28041	1.42096	9	0.174819	1.40548	1.55975	1	0.212109	0.67188	0.74563	9	0.667071
45	I0254	1.89904	1.94743	9	0.042616	0.16646	0.17071	1	0.679589	0.86119	0.88314	9	0.539842
46	I0255	1.18342	1.29016	9	0.238377	0.00429	0.00467	1	0.945528	0.97585	1.06386	9	0.387457
47	I0256	1.55920	1.67777	9	0.090494	1.56879	1.68809	1	0.194273	1.05863	1.13914	9	0.332241
48	I0257	0.98875	1.05731	9	0.392366	8.89759	9.51451	1	0.002113	2.05025	2.19241	9	0.020694
49	I0258	1.95558	1.83228	9	0.059289	6.44345	6.03717	1	0.014247	1.25048	1.17163	9	0.310130
50	I0259	1.36748	1.34114	9	0.211719	1.66936	1.63721	1	0.201121	1.00577	0.98641	9	0.449714
51	I0260	0.41385	0.40890	9	0.930743	0.03904	0.03857	1	0.844333	0.63373	0.62614	9	0.775338

Select Level of Precision
 ght
 Prob=0.000100
 Adjust

Afsluttende DIF-analyse: skolestørrelse

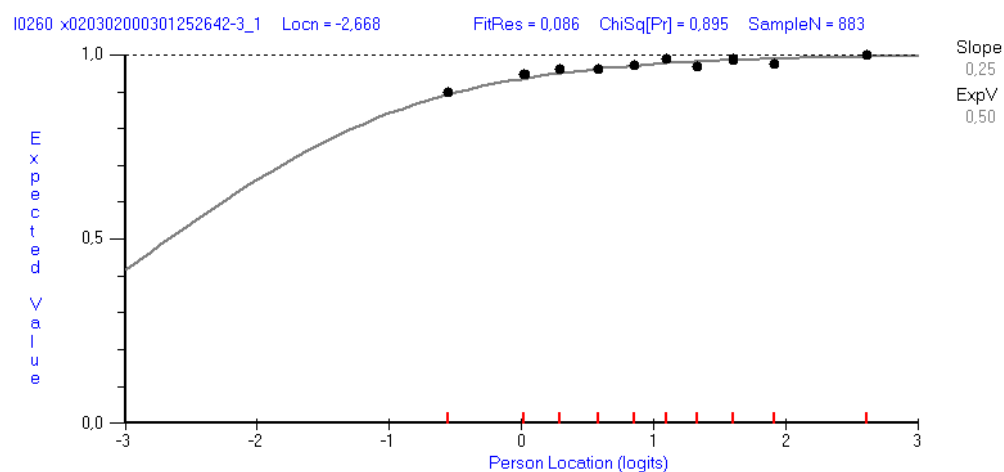
No.	Item	Class Interval				skolestr				Class Interval -by- skolestr			
		MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
11	I0205	2.20105	2.38367	9	0.011613	0.80768	0.87470	1	0.349965	2.19788	2.38023	9	0.011741
12	I0206	0.50758	0.50242	9	0.873308	6.58653	6.51958	1	0.010844	1.47970	1.46465	9	0.156730
13	I0207	1.68076	1.53637	9	0.131032	0.18899	0.17275	1	0.677808	0.71399	0.65266	9	0.752017
14	I0208	0.51120	0.49907	9	0.875588	0.53906	0.52627	1	0.468423	0.40071	0.39120	9	0.939585
15	I0210	0.74109	0.74121	9	0.671185	0.00002	0.00002	1	1.000000	1.00905	1.00922	9	0.430848
16	I0211	0.81527	1.13683	9	0.333841	0.06420	0.08952	1	0.764871	0.30829	0.42989	9	0.919296
17	I0212	0.76794	1.15927	9	0.318218	0.32908	0.49678	1	0.481111	0.99125	1.49638	9	0.144749
18	I0213	1.13169	1.45327	9	0.161497	1.03377	1.32753	1	0.249628	0.94205	1.20974	9	0.285530
19	I0214	1.38155	1.30136	9	0.232286	0.17314	0.16309	1	0.686442	0.43890	0.41343	9	0.928311
20	I0216	0.39620	0.38858	9	0.940844	0.09336	0.09156	1	0.762301	0.81631	0.80060	9	0.615850
21	I0217	2.07476	1.92477	9	0.045666	0.35194	0.32649	1	0.567920	1.27543	1.18322	9	0.302486
22	I0218	0.94297	0.99361	9	0.443588	0.71223	0.75048	1	0.386592	1.02572	1.08081	9	0.374385
23	I0219	1.10657	1.05925	9	0.391017	0.08203	0.07852	1	0.779362	1.62840	1.55876	9	0.123706
24	I0220	1.67093	1.56692	9	0.121118	1.11393	1.04459	1	0.307092	0.49765	0.46667	9	0.897185
25	I0221	1.51224	1.48150	9	0.150259	2.96352	2.90327	1	0.088798	1.06645	1.04478	9	0.402161
26	I0222	1.90843	2.49757	9	0.008137	0.83127	1.08789	1	0.297279	0.54074	0.70767	9	0.702212
27	I0225	2.15437	2.39784	9	0.011130	0.88794	0.98828	1	0.320504	0.98399	1.09519	9	0.363813
28	I0226	1.97682	2.31557	9	0.014325	0.01607	0.01882	1	0.890959	0.65966	0.77270	9	0.641857
29	I0231	1.77959	2.27551	9	0.016208	2.21701	2.83482	1	0.092682	0.79339	1.01449	9	0.426550
30	I0233	1.11186	1.22376	9	0.276618	0.76427	0.84118	1	0.359335	2.13137	2.34586	9	0.012960
31	I0234	1.73683	2.21198	9	0.019649	0.96911	1.23423	1	0.266955	0.36082	0.45953	9	0.901685
32	I0235	1.44704	1.54056	9	0.129623	0.59978	0.63854	1	0.424500	0.42346	0.45082	9	0.907040
33	I0236	1.78586	2.28816	9	0.015465	0.03985	0.05105	1	0.821298	0.74195	0.95064	9	0.480043
34	I0237	1.07366	1.25629	9	0.257452	0.44778	0.52395	1	0.469396	1.03878	1.21549	9	0.281951
35	I0238	1.31828	1.38367	9	0.191342	12.27563	12.88451	1	0.000342	1.37421	1.44237	9	0.165888
36	I0241	0.89139	1.10237	9	0.358529	2.34909	2.90508	1	0.088725	1.56750	1.93850	9	0.043891
37	I0243	2.35326	2.81681	9	0.002920	4.29094	5.13617	1	0.023729	0.59646	0.71396	9	0.696419
38	I0244	2.86059	3.37326	9	0.000458	0.06100	0.07194	1	0.788639	0.79274	0.93481	9	0.493895
39	I0245	2.34534	2.65358	9	0.004903	0.37322	0.42227	1	0.515997	0.51050	0.57759	9	0.816171
40	I0247	1.29942	1.42674	9	0.172371	0.12035	0.13214	1	0.716332	0.68252	0.74939	9	0.663576
41	I0250	1.15115	1.25955	9	0.255563	0.49619	0.54292	1	0.461468	0.76132	0.83301	9	0.585778
42	I0251	2.23642	2.60713	9	0.005705	0.01011	0.01179	1	0.913525	0.83792	0.97681	9	0.457680
43	I0252	1.24620	1.30931	9	0.228076	3.30650	3.47395	1	0.062746	0.57112	0.60004	9	0.797549
44	I0253	1.28041	1.41515	9	0.177321	0.00888	0.00982	1	0.921122	0.53054	0.58637	9	0.808918
45	I0254	1.89904	1.94932	9	0.042371	3.69531	3.79317	1	0.051792	0.55391	0.56858	9	0.823448
46	I0255	1.18342	1.28838	9	0.239344	0.11257	0.12255	1	0.726391	0.86270	0.93921	9	0.490055
47	I0256	1.55920	1.67057	9	0.092243	0.01196	0.01281	1	0.909868	0.91040	0.97543	9	0.458954
48	I0257	0.98875	1.03063	9	0.413383	0.00651	0.00679	1	0.934344	0.86543	0.90209	9	0.522779
49	I0258	1.95559	1.82116	9	0.061171	1.54437	1.43821	1	0.230826	1.27499	1.18735	9	0.299814
50	I0259	1.36748	1.33447	9	0.215061	0.00202	0.00197	1	0.964524	0.78212	0.76325	9	0.650675
51	I0260	0.41385	0.41215	9	0.929040	0.86913	0.86555	1	0.352470	1.25825	1.25308	9	0.259088

Main Plot
 ght
 Select Level of Precision
 Print
 Save
 Copy

Afsluttende DIF-analyse: geografi

No.	Item	Class Interval				oestvest				Class Interval-by-oestvest			
		MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
11	I0205	2.20105	2.36885	9	0,012148	1,07506	1,15702	1	0,282442	1,70461	1,83456	9	0,058907
12	I0206	0.50758	0.49674	9	0,877268	0,02746	0,02687	1	0,869824	1,17007	1,14506	9	0,327935
13	I0207	1.68076	1.54672	9	0,127606	1,73692	1,59839	1	0,206551	1,12642	1,03659	9	0,408784
14	I0208	0.51120	0.50774	9	0,869488	0,32416	0,32196	1	0,570607	1,82864	1,81627	9	0,061988
15	I0210	0.74109	0.74080	9	0,671567	0,06510	0,06508	1	0,798682	0,95714	0,95676	9	0,474858
16	I0211	0.81527	1.14528	9	0,327973	0,04188	0,05884	1	0,808417	0,73499	1,03249	9	0,412034
17	I0212	0.76794	1.14954	9	0,324852	0,47559	0,71192	1	0,399061	0,47122	0,70537	9	0,704352
18	I0213	1.13169	1.43587	9	0,168573	0,03718	0,04718	1	0,828135	0,30018	0,38086	9	0,944466
19	I0214	1.38155	1.32144	9	0,221715	10,19512	9,75148	1	0,001875	0,61896	0,59203	9	0,804238
20	I0216	0.39620	0.38950	9	0,940401	0,00002	0,00002	1	0,994775	1,01996	1,00272	9	0,436176
21	I0217	2.07476	1.92595	9	0,045506	0,01147	0,01065	1	0,917911	1,36642	1,26841	9	0,250487
22	I0218	0.94297	0.99204	9	0,444897	0,00710	0,00747	1	0,931169	0,96910	1,01953	9	0,422316
23	I0219	1.10657	1.06607	9	0,385754	7,00663	6,75021	1	0,009581	1,39378	1,34277	9	0,210916
24	I0220	1.67093	1.57443	9	0,118803	3,22097	3,03495	1	0,081905	0,67143	0,63266	9	0,769623
25	I0221	1.51224	1.47872	9	0,151322	0,04688	0,04584	1	0,830500	1,21784	1,19084	9	0,297334
26	I0222	1.90843	2.50152	9	0,008032	2,08486	2,73278	1	0,098753	0,49765	0,65231	9	0,752325
27	I0225	2.15437	2.38603	9	0,011529	0,51511	0,57050	1	0,450314	0,67000	0,74205	9	0,670405
28	I0226	1.97686	2.36301	9	0,012372	3,33300	3,98413	1	0,046307	1,66630	1,99183	9	0,037642
29	I0231	1.77959	2.26289	9	0,016834	0,15678	0,19935	1	0,655376	0,67445	0,85761	9	0,563135
30	I0233	1.11186	1.21305	9	0,283242	0,17823	0,19445	1	0,659339	1,47658	1,61096	9	0,107760
31	I0234	1.73683	2.21811	9	0,019273	0,01179	0,01505	1	0,902349	0,64031	0,81774	9	0,599918
32	I0235	1.44704	1.56291	9	0,122381	1,56971	1,69539	1	0,193309	1,39293	1,50446	9	0,142078
33	I0236	1.78586	2.30271	9	0,014789	2,24798	2,89858	1	0,089043	0,93957	1,21150	9	0,284203
34	I0237	1.07366	1.25259	9	0,259609	2,35034	2,74204	1	0,098164	0,62557	0,72983	9	0,681744
35	I0238	1.31828	1.36046	9	0,202262	0,75928	0,78358	1	0,376347	1,34980	1,39299	9	0,187098
36	I0241	0.89139	1,07898	9	0,375934	0,00121	0,00146	1	0,969498	0,42220	0,51105	9	0,867129
37	I0243	2.35326	2.79803	9	0,003126	0,10824	0,12870	1	0,719868	0,61335	0,72928	9	0,682263
38	I0244	2.86059	3.40378	9	0,000410	2,76927	3,29511	1	0,069913	1,10176	1,31096	9	0,227182
39	I0245	2.34535	2.66123	9	0,004799	2,97707	3,37803	1	0,066423	0,44924	0,50975	9	0,868127
40	I0247	1.29942	1,43462	9	0,169066	0,07825	0,08639	1	0,768912	1,08886	1,20216	9	0,290302
41	I0250	1.15115	1,25260	9	0,259589	0,00091	0,00099	1	0,974910	0,40950	0,44559	9	0,910185
42	I0251	2.23642	2,59485	9	0,005941	0,00677	0,00785	1	0,929419	0,47377	0,54970	9	0,838338
43	I0252	1.24620	1,30474	9	0,230494	5,27087	5,51848	1	0,019100	0,08674	0,09081	9	0,999751
44	I0253	1.28041	1,43518	9	0,168835	0,49960	0,55999	1	0,454517	1,48892	1,66889	9	0,092658
45	I0254	1.89904	1,94304	9	0,043141	0,65279	0,66792	1	0,414025	0,60915	0,62327	9	0,777822
46	I0255	1.18342	1,29192	9	0,237403	2,27255	2,48090	1	0,115665	0,82391	0,89944	9	0,525221
47	I0256	1.55920	1,67131	9	0,092065	0,69629	0,74636	1	0,387931	0,86754	0,92992	9	0,498177
48	I0257	0.98875	1,03207	9	0,412226	0,44983	0,46954	1	0,493388	0,93637	0,97740	9	0,457174
49	I0258	1.95559	1,80619	9	0,063767	0,00011	0,00010	1	0,992621	0,73646	0,68020	9	0,727295
50	I0259	1.36748	1,33396	9	0,215312	0,16355	0,15954	1	0,689719	0,73269	0,71474	9	0,695701
51	I0260	0.41385	0,41021	9	0,930061	0,32979	0,32688	1	0,567646	0,89074	0,88289	9	0,540066

Eksempel på ICC-kurve for en godkendt opgave.



Kilde: Styrelsen for It og L ring

Matematik 6. klasse – tal og algebra

I opgaveafprøvningen januar 2018 afprøves 61 opgaver, heraf forkastes 32 opgaver og 29 (48 %) opgaver godkendes og tilføjes Opgavebanken.

31 opgaver forkastes pga. itemfit og 1 opgave forkastes pga. DIF på geografi.

Analysen indeholder 10 linkopgaver: I0030, I0143, I0200, I0263, I0289, I0307, I0366, I0389, I0404 og I0458.

Initial itemfit-analyse: viser et udpluk af de opgaver der forkastes.

Seq	Item	Type	Location	SE	FitResid	DF	ChiSq	DF	Prob	F-stat	DF-1	DF-2	Prob
35	531	I0531	Poly	0,415	0,078	-3,381	884,11	36,655	9	0,000030
7	366	I0366	Poly	-0,013	0,081	3,543	862,47	36,764	9	0,000029
17	513	I0513	Poly	-1,085	0,094	2,608	883,12	36,774	9	0,000029
52	548	I0548	Poly	1,048	0,080	-0,888	844,77	38,671	9	0,000013
41	537	I0537	Poly	0,814	0,078	-5,129	884,11	39,553	9	0,000009
58	554	I0554	Poly	-1,929	0,120	1,879	842,80	43,753	9	0,000002
45	541	I0541	Poly	0,640	0,079	-5,839	862,47	45,390	9	0,000001
12	508	I0508	Poly	-1,183	0,097	2,981	863,46	45,991	9	0,000001
51	547	I0547	Poly	3,100	0,107	-3,285	862,47	46,940	9	0,000000
38	534	I0534	Poly	0,109	0,080	-5,554	884,11	53,234	9	0,000000
19	515	I0515	Poly	0,582	0,080	3,870	843,79	60,843	9	0,000000
36	532	I0532	Poly	-0,093	0,082	-5,253	861,49	62,126	9	0,000000
40	536	I0536	Poly	-0,154	0,084	-4,705	845,75	64,765	9	0,000000
44	540	I0540	Poly	-0,084	0,081	-5,249	884,11	65,684	9	0,000000
39	535	I0535	Poly	0,616	0,079	-7,091	861,49	67,801	9	0,000000
15	511	I0511	Poly	-1,312	0,100	4,329	861,49	70,204	9	0,000000
23	519	I0519	Poly	-2,055	0,120	2,095	884,11	70,670	9	0,000000
10	458	I0458	Poly	0,457	0,078	6,104	883,12	71,030	9	0,000000
2	143	I0143	Poly	1,194	0,081	5,791	842,80	73,569	9	0,000000
24	520	I0520	Poly	-2,239	0,129	3,447	862,47	76,879	9	0,000000
61	557	I0557	Poly	-1,091	0,097	5,193	844,77	127,823	9	0,000000
Mean			0,000		-0,159		Totals	2,246,341	639				
Std Devn			1,276		3,068		Total Prob	0,000000					

Selector: 61 Item: x020601000301248057-4 1
 Sort Items by: ChiSq Prob Orde
 File Text Format: Fixed
 below the Bonferroni: Select Probability B: 0,01
 Adi = 0.000141 for 71

Afsluttende itemfit-analyse

Seq	Item	Type	Location	SE	FitResid	DF	ChiSq	DF	Prob	F-stat	DF-1	DF-2	Prob
7	366	I0366	Poly	0,286	0,079	2,046	845,53	12,069	9	0,209421
9	404	I0404	Poly	0,104	0,080	-1,040	865,92	12,167	9	0,204056
20	521	I0521	Poly	-1,974	0,135	0,233	829,03	12,538	9	0,184625
18	517	I0517	Poly	-1,454	0,113	-0,049	845,53	12,927	9	0,165918
19	518	I0518	Poly	-2,275	0,149	1,299	828,06	13,996	9	0,122484
31	549	I0549	Poly	1,682	0,078	-1,345	863,98	14,293	9	0,112266
13	510	I0510	Poly	-0,731	0,093	-1,005	865,92	14,872	9	0,094526
28	529	I0529	Poly	-0,112	0,083	-2,066	844,56	15,017	9	0,090468
12	509	I0509	Poly	-0,022	0,084	1,156	828,06	15,022	9	0,090337
4	263	I0263	Poly	1,224	0,077	-2,205	844,56	15,104	9	0,088108
35	555	I0555	Poly	-1,364	0,109	1,218	866,89	15,286	9	0,083364
33	551	I0551	Poly	1,944	0,082	1,278	828,06	16,523	9	0,056736
26	527	I0527	Poly	-0,821	0,097	-1,106	829,03	16,673	9	0,054089
24	525	I0525	Poly	-1,269	0,106	0,915	866,89	16,793	9	0,052052
10	458	I0458	Poly	0,750	0,076	3,860	865,92	17,533	9	0,041000
6	307	I0307	Poly	1,031	0,075	-2,000	865,92	17,537	9	0,040947
5	289	I0289	Poly	0,573	0,078	0,732	828,06	19,326	9	0,022560
37	561	I0561	Poly	0,553	0,077	-1,560	865,92	20,596	9	0,014570
38	564	I0564	Poly	-0,077	0,082	-2,255	865,92	20,721	9	0,013948
39	566	I0566	Poly	0,531	0,079	-1,408	829,03	23,013	9	0,006169
2	143	I0143	Poly	1,455	0,078	3,907	826,12	28,651	9	0,000742
Mean			0,000		-0,096		Totals	485,349	351				
Std Devn			1,128		1,502		Total Prob	0,000003					

Selector: 2 Item: x02060108033-3 1
 Sort Items by: ChiSq Prob Orde
 File Text Format: Fixed
 below the Bonferroni: Select Probability B: 0,01
 Adi = 0.000256 for 39

Afsluttende DIF-analyse: køn

No.	Item	Class Interval				køen				Class Interval-by- køen			
		MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
11	I0507	0.33155	0.32342	9	0.967473	0.13595	0.13262	1	0.715828	1.94332	1.89565	9	0.049289
12	I0509	1.63452	1.55883	9	0.123364	6.39761	6.10137	1	0.013698	0.52440	0.50012	9	0.874938
13	I0510	1.56439	1.79186	9	0.065981	1.25337	1.43562	1	0.231165	0.35643	0.40826	9	0.931107
14	I0512	0.92877	0.98846	9	0.447853	2.70014	2.87368	1	0.090421	0.71031	0.75596	9	0.657471
15	I0513	0.68758	0.64374	9	0.759987	0.61793	0.57853	1	0.447108	1.20373	1.12698	9	0.340493
16	I0514	0.55899	0.64095	9	0.762439	0.89850	1.03022	1	0.310401	0.54730	0.62754	9	0.774148
17	I0516	0.63370	0.68124	9	0.726418	7.10546	7.63853	1	0.005839	1.12160	1.20575	9	0.287681
18	I0517	1.34351	1.39804	9	0.184469	0.76603	0.79712	1	0.372220	0.84992	0.88442	9	0.538651
19	I0518	1.60553	1.20828	9	0.286174	0.75104	0.56521	1	0.452379	1.93927	1.45944	9	0.158702
20	I0521	1.36018	1.33971	9	0.212114	0.16053	0.15811	1	0.691008	1.24002	1.22136	9	0.278038
21	I0522	1.26830	1.14188	9	0.330019	0.04312	0.03882	1	0.843835	2.20254	1.98301	9	0.038343
22	I0523	0.54916	0.64896	9	0.755367	1.90317	2.24901	1	0.134079	1.05436	1.24597	9	0.263156
23	I0524	0.89010	0.98392	9	0.451657	4.92337	5.44236	1	0.019887	0.95703	1.05791	9	0.391856
24	I0525	1.75547	1.60471	9	0.109396	0.89534	0.81845	1	0.365894	1.48004	1.35294	9	0.205504
25	I0526	0.50781	0.56177	9	0.828896	0.21888	0.24214	1	0.622787	0.55055	0.60906	9	0.790019
26	I0527	1.85365	2.19040	9	0.020783	1.05206	1.24318	1	0.265186	0.94219	1.11336	9	0.350298
27	I0528	0.53202	0.61197	9	0.787552	0.43858	0.50448	1	0.477723	0.74565	0.85770	9	0.562992
28	I0529	1.65978	2.00378	9	0.036122	0.01937	0.02339	1	0.878439	0.58929	0.71142	9	0.698796
29	I0530	0.96529	0.96845	9	0.464740	3.49083	3.50224	1	0.061647	0.62720	0.62925	9	0.772659
30	I0539	1.09177	1.20957	9	0.285366	0.61666	0.68320	1	0.408723	0.90660	1.00443	9	0.434618
31	I0549	1.56089	1.77783	9	0.068601	0.74477	0.84829	1	0.357297	1.59277	1.81414	9	0.062020
32	I0550	1.11310	1.25183	9	0.259721	3.14925	3.54173	1	0.060190	0.87669	0.98595	9	0.449934
33	I0551	2.13485	2.04533	9	0.032018	5.02940	4.81850	1	0.028433	2.01168	1.92732	9	0.045099
34	I0553	1.07681	1.03626	9	0.408837	0.82961	0.79837	1	0.371829	0.77682	0.74757	9	0.665296
35	I0555	1.65942	1.43834	9	0.167120	2.66951	2.31386	1	0.128580	1.36158	1.18018	9	0.304140
36	I0556	0.82711	0.80518	9	0.611566	0.11583	0.11276	1	0.737088	0.81456	0.79296	9	0.622952
37	I0561	2.25643	2.57005	9	0.006373	0.00189	0.00215	1	0.963191	1.20587	1.37347	9	0.195670
38	I0564	2.31678	2.88663	9	0.002289	6.78066	8.44845	1	0.003742	1.03100	1.28458	9	0.241038
39	I0566	2.51919	2.86817	9	0.002437	1.93801	2.20648	1	0.137796	1.07622	1.22531	9	0.275608

Select Level of Precision
 ght
 adjust
 0.03

< Main Plot solavinda: 39

Afsluttende DIF-analyse: skolestørrelse

No.	Item	Class Interval			skolestr				Class Interval-by- skolestr				
		MS	F	DF	MS	F	DF	Prob	MS	F	DF	Prob	
11	I0507	0.33155	0.31990	9	0.968647	0.24673	0.23806	1	0.625735	0.83802	0.80857	9	0.608409
12	I0509	1.63452	1.55714	9	0.123893	0.59705	0.56879	1	0.450952	1.06327	1.01293	9	0.427653
13	I0510	1.56439	1.79296	9	0.065778	1.04922	1.20253	1	0.273128	0.43098	0.49395	9	0.879228
14	I0512	0.92877	0.98988	9	0.446672	0.51976	0.55396	1	0.456916	1.07746	1.14636	9	0.325619
15	I0513	0.68758	0.64133	9	0.762115	1.05326	0.98240	1	0.321874	0.76589	0.71437	9	0.696086
16	I0514	0.55899	0.64922	9	0.755137	5.10390	5.92770	1	0.015097	1.13081	1.31333	9	0.225591
17	I0516	0.63370	0.67140	9	0.735320	0.18487	0.19587	1	0.658199	0.56769	0.60146	9	0.796448
18	I0517	1.34351	1.43394	9	0.168998	1.12091	1.19635	1	0.274359	3.08537	3.29305	9	0.000574
19	I0518	1.60553	1.20150	9	0.290451	2.20506	1.65017	1	0.199289	1.08400	0.81122	9	0.605958
20	I0521	1.36018	1.34217	9	0.210884	5.08064	5.01337	1	0.025408	0.86566	0.85419	9	0.566223
21	I0522	1.26830	1.13515	9	0.334723	3.28674	2.94170	1	0.086680	1.20406	1.07766	9	0.376670
22	I0523	0.54916	0.65233	9	0.752381	0.11254	0.13368	1	0.714737	1.66743	1.98068	9	0.038635
23	I0524	0.89010	0.97409	9	0.459947	0.02219	0.02429	1	0.876137	0.65492	0.71672	9	0.693911
24	I0525	1.75547	1.61673	9	0.105989	6.32904	5.82882	1	0.015971	1.66473	1.53316	9	0.131671
25	I0526	0.50781	0.56902	9	0.823117	2.93960	3.29394	1	0.069906	1.33544	1.49641	9	0.144628
26	I0527	1.85365	2.19462	9	0.020518	1.19983	1.42053	1	0.233659	1.07651	1.27452	9	0.246741
27	I0528	0.53202	0.62002	9	0.780659	0.48826	0.56902	1	0.450849	1.83368	2.13696	9	0.024350
28	I0529	1.65978	2.00381	9	0.036120	0.05347	0.06455	1	0.799505	0.58680	0.70842	9	0.701561
29	I0530	0.96529	0.98593	9	0.449971	12.52464	12.79241	1	0.000360	1.26113	1.28809	9	0.239193
30	I0539	1.09177	1.23674	9	0.268695	10.40474	11.78638	1	0.000608	1.65677	1.87677	9	0.052073
31	I0549	1.56089	1.75476	9	0.073098	0.37138	0.41751	1	0.518372	0.51858	0.58300	9	0.811789
32	I0550	1.11310	1.24698	9	0.262567	4.11751	4.61273	1	0.032023	0.44221	0.49539	9	0.878221
33	I0551	2.13485	2.03042	9	0.033459	0.00010	0.00010	1	0.992052	1.86079	1.76976	9	0.070221
34	I0553	1.07681	1.03441	9	0.410309	1.37346	1.31939	1	0.251016	0.54099	0.51969	9	0.860974
35	I0555	1.65942	1.43196	9	0.169757	0.35678	0.30787	1	0.579110	1.11995	0.96644	9	0.466413
36	I0556	0.82711	0.80165	9	0.614858	1.65937	1.60829	1	0.205092	0.21496	0.20834	9	0.993217
37	I0561	2.25643	2.56178	9	0.006543	0.56302	0.63921	1	0.424221	0.86891	0.98649	9	0.449463
38	I0564	2.31678	2.85496	9	0.002520	0.00215	0.00265	1	0.958950	0.92167	1.13577	9	0.334285
39	I0566	2.51919	2.85168	9	0.002572	0.24861	0.28143	1	0.595929	0.79313	0.89780	9	0.526606

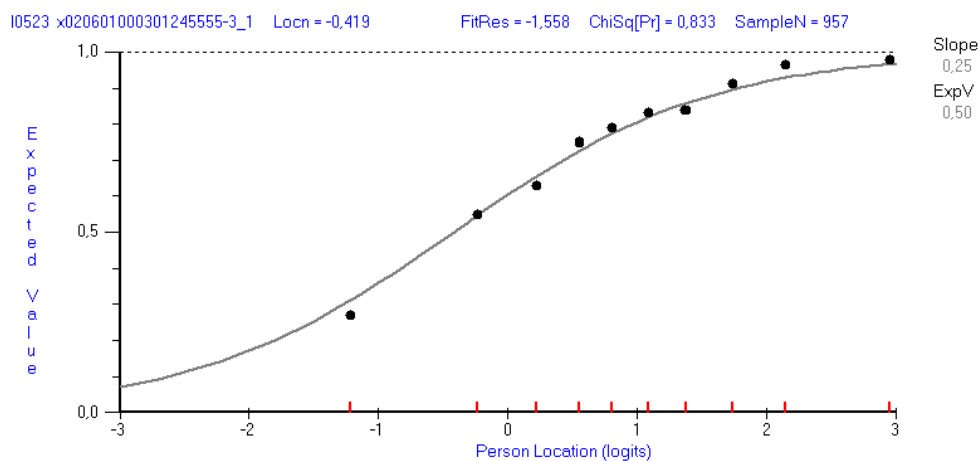
Select Level of Precision
 0.000000, 0.00010, 0.00050, 0.00100, 0.00250, 0.00500, 0.01000, 0.02500, 0.05000, 0.10000, 0.25000, 0.50000, 1.00000, 2.50000, 5.00000, 10.00000, 25.00000, 50.00000, 100.00000, 250.00000, 500.00000, 1000.00000

< Main Plot solvning: 39 ght ind edukt 0.03 Save Copy

Afsluttende DIF-analyse: geografi

No.	Item	Class Interval				oestvest				Class Interval-by- oestvest			
		MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
11	I0507	0.33155	0.31979	9	0.968685	0.68593	0.66159	1	0.416233	0.75305	0.72632	9	0.685039
12	I0509	1.63452	1.56190	9	0.122386	5.61784	5.36823	1	0.020754	0.80137	0.76576	9	0.648331
13	I0510	1.56439	1.80326	9	0.063917	0.00570	0.00657	1	0.935520	1.02978	1.18702	9	0.299676
14	I0512	0.92877	0.98667	9	0.449351	0.24394	0.25914	1	0.610863	0.82546	0.87693	9	0.545453
15	I0513	0.68758	0.64848	9	0.755804	0.30911	0.29153	1	0.589390	1.99379	1.88039	9	0.051471
16	I0514	0.55899	0.64333	9	0.760340	0.01495	0.01720	1	0.895700	0.95135	1.09489	9	0.363795
17	I0516	0.63370	0.67328	9	0.733625	1.05651	1.12250	1	0.289674	0.72632	0.77168	9	0.642806
18	I0517	1.34351	1.39083	9	0.187723	0.00000	0.00000	1	1.000000	0.46433	0.48069	9	0.888165
19	I0518	1.60553	1.20248	9	0.289839	0.73639	0.55152	1	0.457903	1.34729	1.00907	9	0.430809
20	I0521	1.36018	1.33087	9	0.216574	0.00665	0.00651	1	0.935701	0.63160	0.61798	9	0.782387
21	I0522	1.26830	1.12243	9	0.343732	0.21529	0.19053	1	0.662582	0.31809	0.28150	9	0.979788
22	I0523	0.54916	0.65449	9	0.750456	0.08965	0.10685	1	0.743828	1.93334	2.30415	9	0.014690
23	I0524	0.89010	0.97456	9	0.459545	2.01214	2.20308	1	0.138122	0.47500	0.52008	9	0.860682
24	I0525	1.75547	1.60989	9	0.107909	0.28833	0.26442	1	0.607229	1.88872	1.73209	9	0.077782
25	I0526	0.50781	0.56494	9	0.826380	0.15375	0.17105	1	0.679305	1.03630	1.15289	9	0.322480
26	I0527	1.85365	2.19528	9	0.020496	0.25077	0.29699	1	0.585931	1.20546	1.42762	9	0.171684
27	I0528	0.53202	0.60974	9	0.789449	2.38645	2.73505	1	0.098536	0.22138	0.25372	9	0.986016
28	I0529	1.65978	2.02095	9	0.034353	0.05895	0.07178	1	0.788800	1.24954	1.52144	9	0.135741
29	I0530	0.96529	0.96741	9	0.465625	0.60447	0.60580	1	0.436597	0.84900	0.85086	9	0.569278
30	I0539	1.09176	1.20836	9	0.286120	0.48708	0.53910	1	0.463001	0.83725	0.92666	9	0.500935
31	I0549	1.56089	1.78014	9	0.068169	0.25503	0.29085	1	0.589818	1.75716	2.00398	9	0.036076
32	I0550	1.11310	1.25456	9	0.258122	1.38396	1.55985	1	0.212032	1.25635	1.41602	9	0.176581
33	I0551	2.13485	2.02308	9	0.034168	0.09640	0.09135	1	0.762522	1.49700	1.41862	9	0.175506
34	I0553	1.07681	1.04245	9	0.403934	0.07007	0.06784	1	0.794594	1.44462	1.39852	9	0.184255
35	I0555	1.65942	1.43842	9	0.167090	0.51337	0.44500	1	0.504889	1.60726	1.39321	9	0.186594
36	I0556	0.82711	0.80651	9	0.610328	0.00043	0.00042	1	0.983933	0.98710	0.96251	9	0.469790
37	I0561	2.25643	2.56893	9	0.006398	0.30594	0.34831	1	0.555215	1.13503	1.29222	9	0.236838
38	I0564	2.31678	2.87513	9	0.002371	1.07177	1.33007	1	0.249112	1.35443	1.68085	9	0.089360
39	I0566	2.51919	2.87201	9	0.002410	4.86100	5.54178	1	0.018796	0.86013	0.98059	9	0.454452

Eksempel på ICC-kurve for en godkendt opgave.



Kilde: Styrelsen for It og Læring

Bilag 4.3 Opgavebankens sammensætning i forhold til opgavernes sværhedsgrad

Tabel 4.3.1 Opgavebankens sammensætning i forhold til opgavernes sværhedsgrad

		Femtrinsskala					I alt
		* ¹⁾	** ²⁾	*** ³⁾	**** ⁴⁾	***** ⁵⁾	
Fag		Antal (Pct.)	Antal (Pct.)	Antal (Pct.)	Antal (Pct.)	Antal (Pct.)	
Dansk læsning 2. klasse	P1 ⁶⁾	83 (34 %)	43 (18 %)	33 (13 %)	50 (20 %)	36 (15 %)	245
	P2 ⁷⁾	58 (23 %)	111 (44 %)	47 (19 %)	27 (11 %)	11 (4 %)	254
	P3 ⁸⁾	15 (7 %)	111 (48 %)	72 (31 %)	32 (14 %)	0 (0 %)	230
Dansk læsning 4. klasse	P1 ⁶⁾	40 (18 %)	51 (22 %)	45 (20 %)	71 (31 %)	20 (9 %)	227
	P2 ⁷⁾	159 (58 %)	69 (25 %)	20 (7 %)	15 (5 %)	10 (4 %)	273
	P3 ⁸⁾	43 (15 %)	131 (45 %)	80 (28 %)	26 (9 %)	8 (3 %)	288
Dansk læsning 6. klasse	P1 ⁶⁾	35 (15 %)	45 (19 %)	47 (20 %)	72 (30 %)	42 (17 %)	241
	P2 ⁷⁾	169 (70 %)	48 (20 %)	13 (5 %)	4 (2 %)	7 (3 %)	241
	P3 ⁸⁾	60 (21 %)	149 (51 %)	53 (18 %)	19 (7 %)	10 (3 %)	291
Dansk læsning 8. klasse	P1 ⁶⁾	50 (16 %)	56 (18 %)	54 (18 %)	91 (30 %)	57 (19 %)	308
	P2 ⁷⁾	134 (63 %)	44 (21 %)	20 (9 %)	8 (4 %)	8 (4 %)	214
	P3 ⁸⁾	53 (18 %)	140 (47 %)	64 (21 %)	24 (8 %)	18 (6 %)	299
Matematik 3. klasse	P1 ⁹⁾	47 (18 %)	114 (44 %)	51 (20 %)	28 (11 %)	21 (8 %)	261
	P2 ¹⁰⁾	77 (35 %)	50 (23 %)	42 (19 %)	31 (14 %)	20 (9 %)	220
	P4 ¹¹⁾	41 (18 %)	112 (50 %)	49 (22 %)	23 (10 %)	0 (0 %)	225
Matematik 6. klasse	P1 ⁹⁾	192 (41 %)	89 (19 %)	65 (14 %)	75 (16 %)	53 (11 %)	474
	P2 ¹⁰⁾	47 (16 %)	76 (26 %)	50 (17 %)	60 (21 %)	56 (19 %)	289
	P4 ¹¹⁾	94 (36 %)	60 (23 %)	59 (23 %)	43 (16 %)	6 (2 %)	262
Matematik 8. klasse	P1 ⁹⁾	22 (9 %)	70 (30 %)	70 (30 %)	56 (24 %)	16 (7 %)	234
	P2 ¹⁰⁾	41 (16 %)	65 (26 %)	60 (24 %)	45 (18 %)	38 (15 %)	249
	P4 ¹¹⁾	66 (30 %)	63 (28 %)	33 (15 %)	29 (13 %)	32 (14 %)	223
Engelsk 4. klasse	P1 ¹²⁾	41 (19 %)	57 (26 %)	81 (37 %)	30 (14 %)	10 (5 %)	219
	P2 ¹³⁾	156 (50 %)	66 (21 %)	46 (15 %)	42 (13 %)	3 (1 %)	313
	P4 ¹⁴⁾	145 (57 %)	65 (25 %)	38 (15 %)	4 (2 %)	4 (2 %)	256
Engelsk 7. klasse	P1 ¹²⁾	24 (12 %)	99 (48 %)	42 (20 %)	29 (14 %)	13 (6 %)	207
	P2 ¹³⁾	43 (19 %)	61 (27 %)	64 (29 %)	39 (18 %)	15 (7 %)	222
	P4 ¹⁵⁾	27 (13 %)	96 (45 %)	50 (23 %)	31 (15 %)	9 (4 %)	213
Fysik/kemi 8. klasse	P1 ¹⁶⁾	9 (4 %)	42 (19 %)	49 (22 %)	58 (27 %)	60 (28 %)	218
	P2 ¹⁷⁾	28 (10 %)	51 (18 %)	58 (21 %)	61 (22 %)	81 (29 %)	279
	P3 ¹⁸⁾	31 (12 %)	40 (15 %)	54 (21 %)	70 (27 %)	65 (25 %)	260

¹⁾ Percentilværdi: 1-10 ²⁾ Percentilværdi: 11-35 ³⁾ Percentilværdi: 36-65 ⁴⁾ Percentilværdi: 66-90 ⁵⁾ Percentilværdi: 91-100

⁶⁾ Sprogforståelse ⁷⁾ Afkodning ⁸⁾ Tekstforståelse ⁹⁾ Tal og algebra ¹⁰⁾ Geometri ¹¹⁾ Statistik og sandsynlighed

¹²⁾ Læsning ¹³⁾ Ordforråd ¹⁴⁾ Lytning ¹⁵⁾ Sprog og sprogbrug

¹⁶⁾ Energi og energiomsætning ¹⁷⁾ Fænomener, stoffer og materialer ¹⁸⁾ Anvendelser og perspektiver

Kilde: Styrelsen for It og Læring

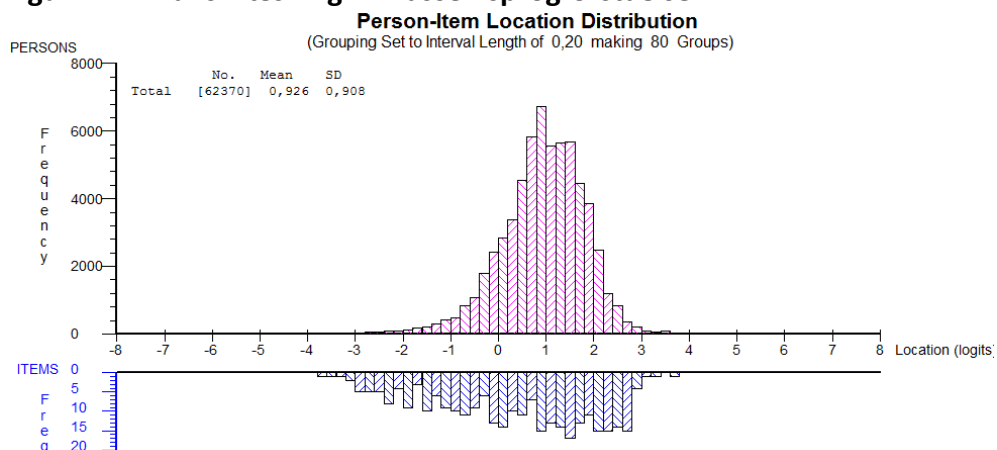
Bilag 4.4 Sammenhæng mellem elevernes dygtighed og opgavernes sværhedsgrad

Figureerne viser sammenhængen mellem opgavernes sværhedsgrad og elevernes dygtighed i de obligatoriske test i foråret 2018 (person-item distribution). Begge dele er udtrykt på Rasch (logit) skalaen.

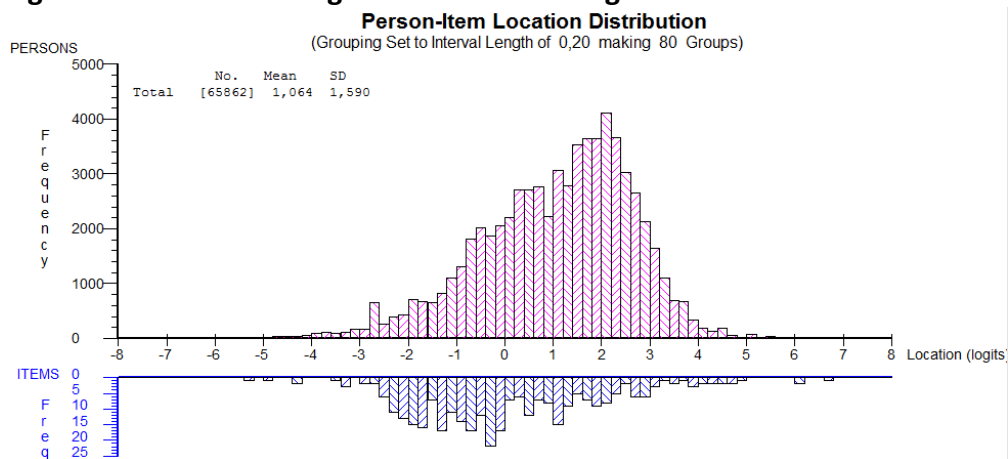
Figureerne er hentet fra analyserne foretaget med RUMM⁵⁹.

⁵⁹ www.rummlab.com.au

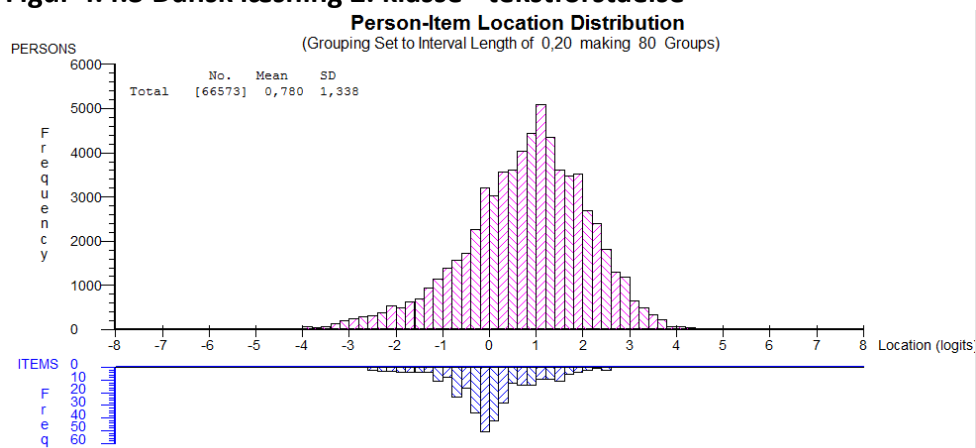
Figur 4.4.1 Dansk læsning 2. klasse - sprogforståelse



Figur 4.4.2 Dansk læsning 2. klasse - afkodning

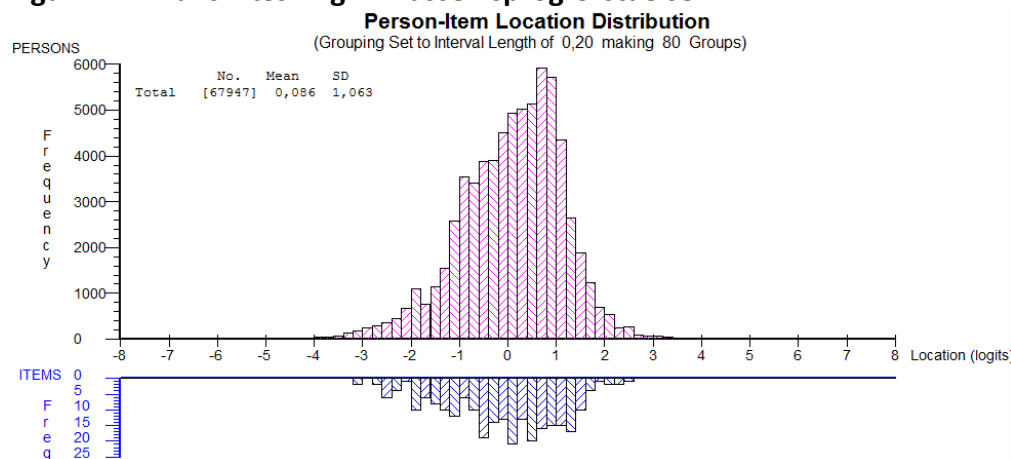


Figur 4.4.3 Dansk læsning 2. klasse - tekstforståelse

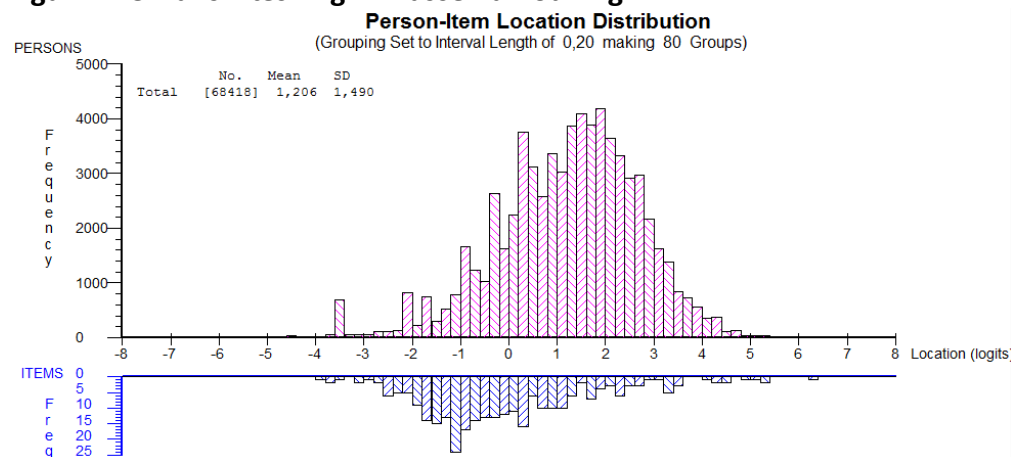


Kilde: Styrelsen for It og Læring

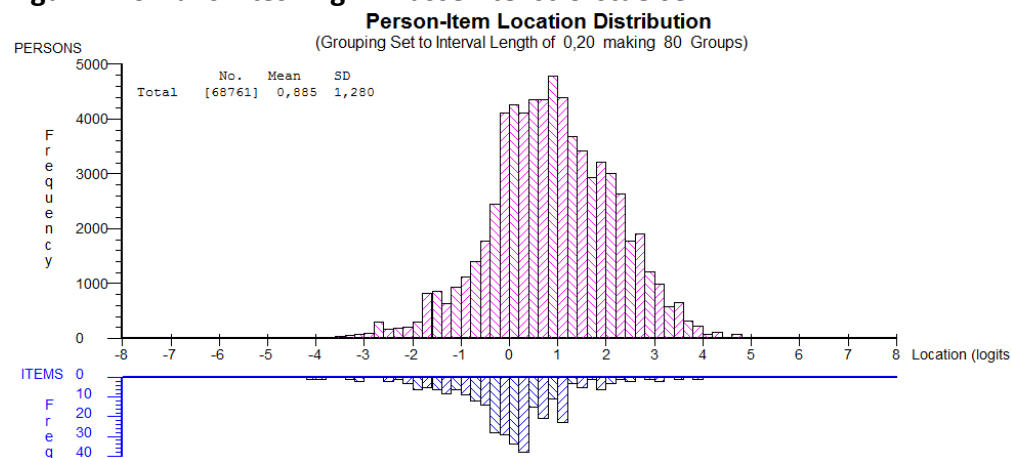
Figur 4.4.4 Dansk læsning 4. klasse - sprogforståelse



Figur 4.4.5 Dansk læsning 4. klasse - afkodning

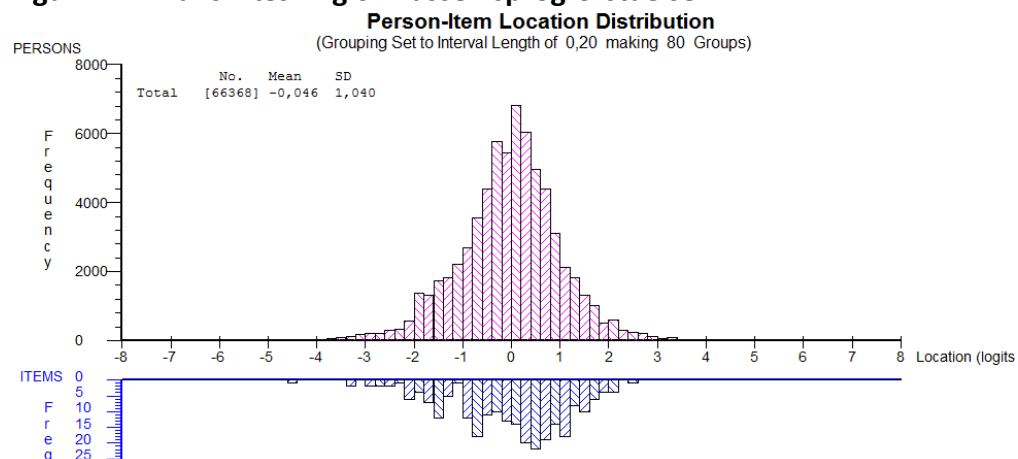


Figur 4.4.6 Dansk læsning 4. klasse - tekstforståelse

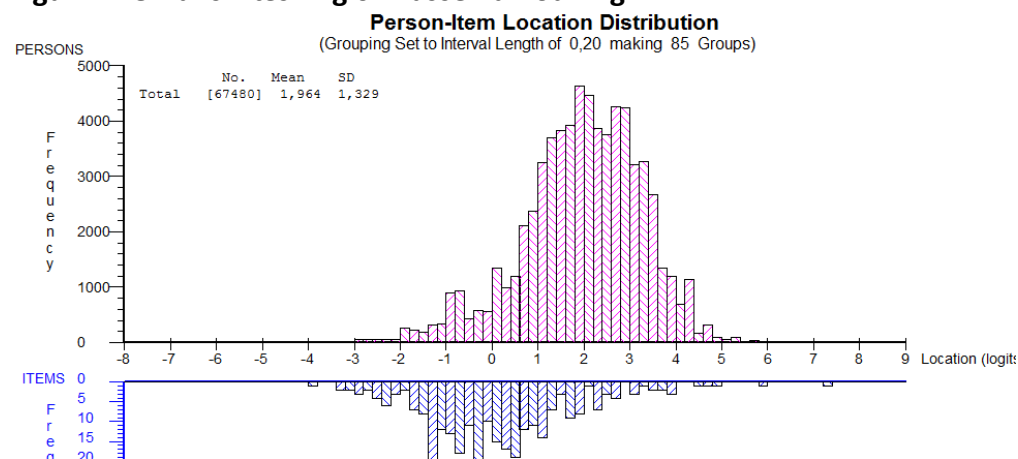


Kilde: Styrelsen for It og Læring

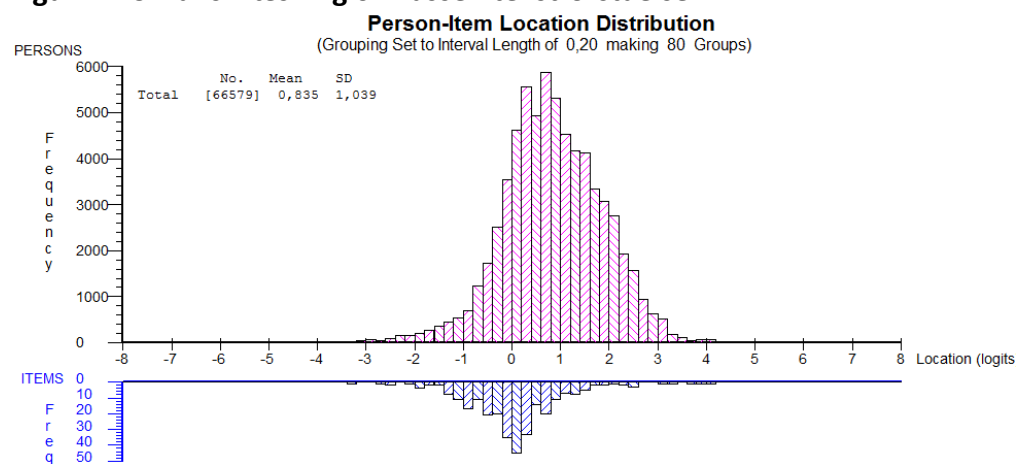
Figur 4.4.7 Dansk læsning 6. klasse - sprogforståelse



Figur 4.4.8 Dansk læsning 6. klasse - afkodning

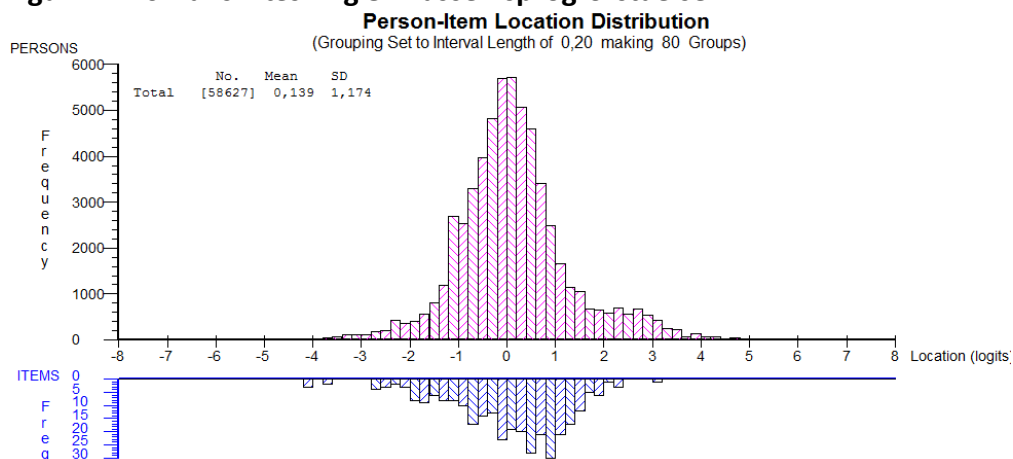


Figur 4.4.9 Dansk læsning 6. klasse - tekstforståelse

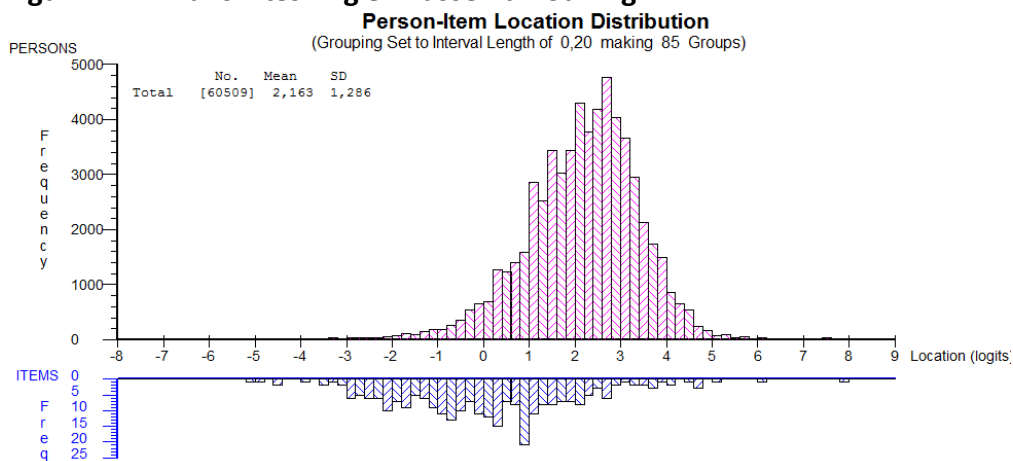


Kilde: Styrelsen for It og Læring

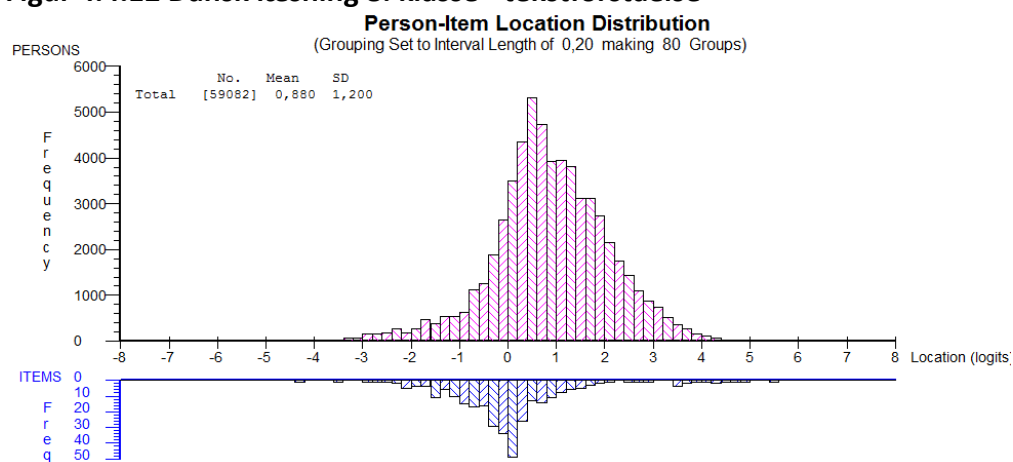
Figur 4.4.10 Dansk læsning 8. klasse - sprogforståelse



Figur 4.4.11 Dansk læsning 8. klasse - afkodning

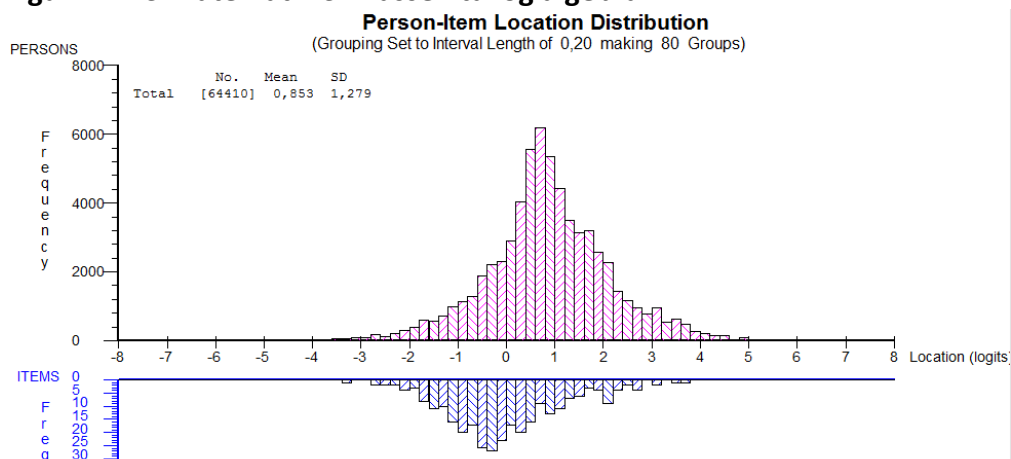


Figur 4.4.12 Dansk læsning 8. klasse - tekstforståelse

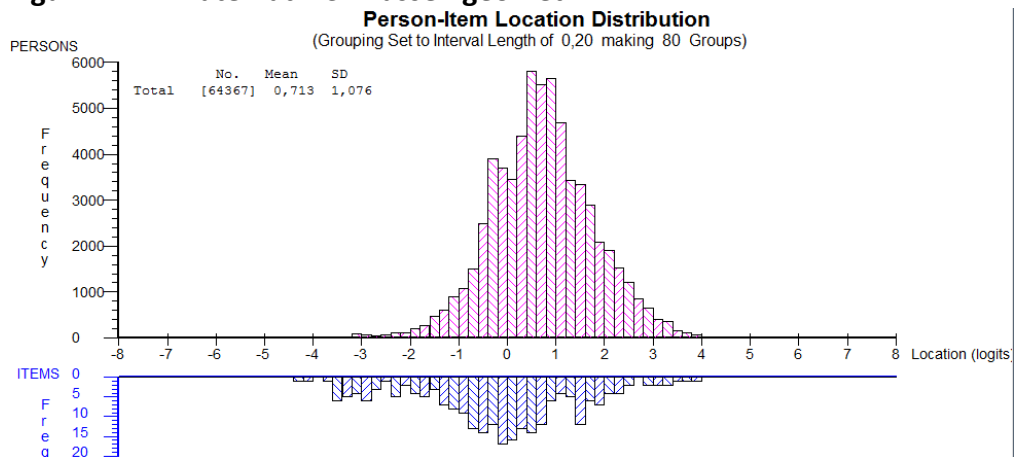


Kilde: Styrelsen for It og Læring

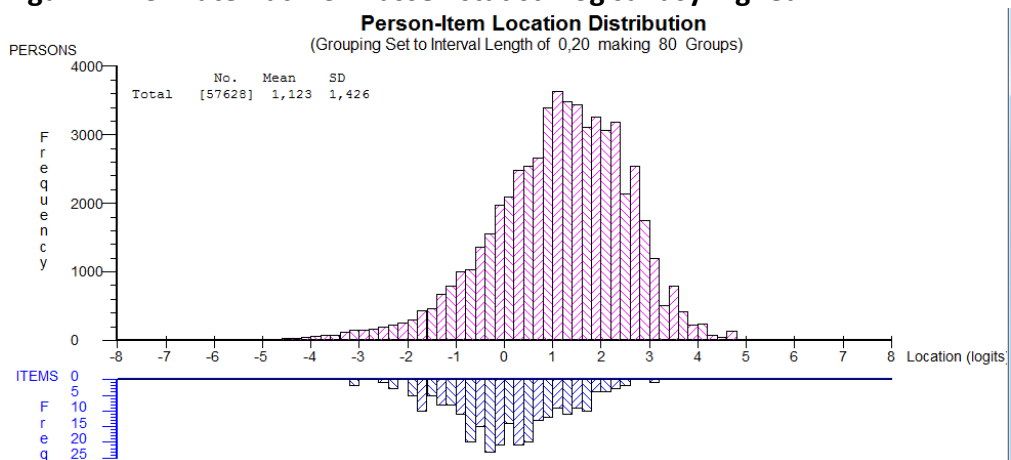
Figur 4.4.13 Matematik 3. klasse - tal og algebra



Figur 4.4.14 Matematik 3. klasse - geometri

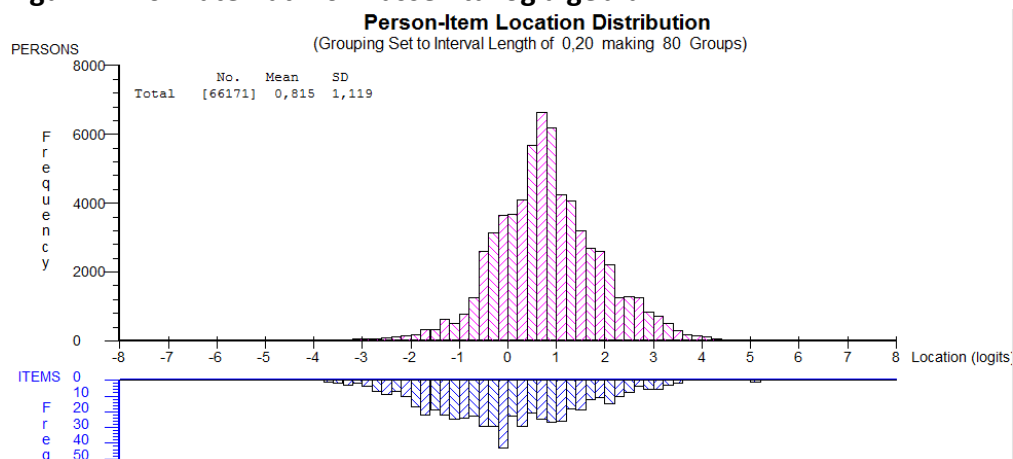


Figur 4.4.15 Matematik 3. klasse - statistik og sandsynlighed

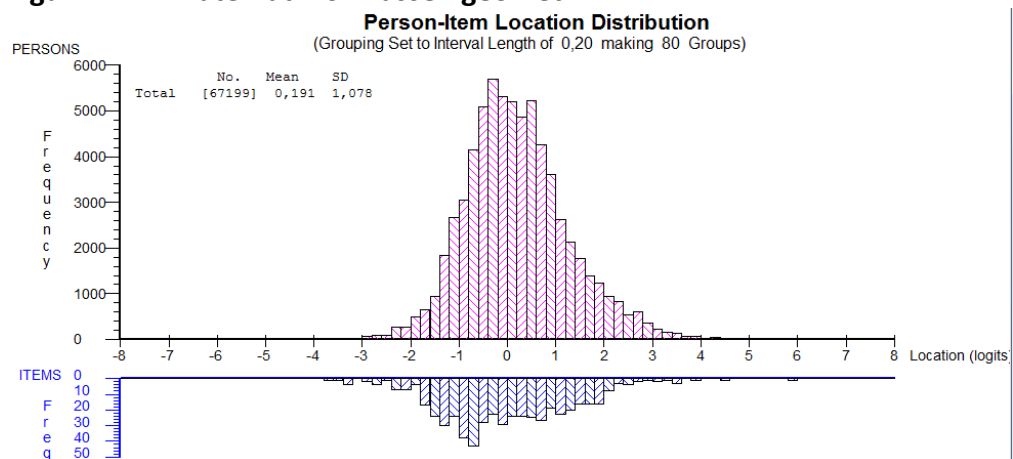


Kilde: Styrelsen for It og L ring

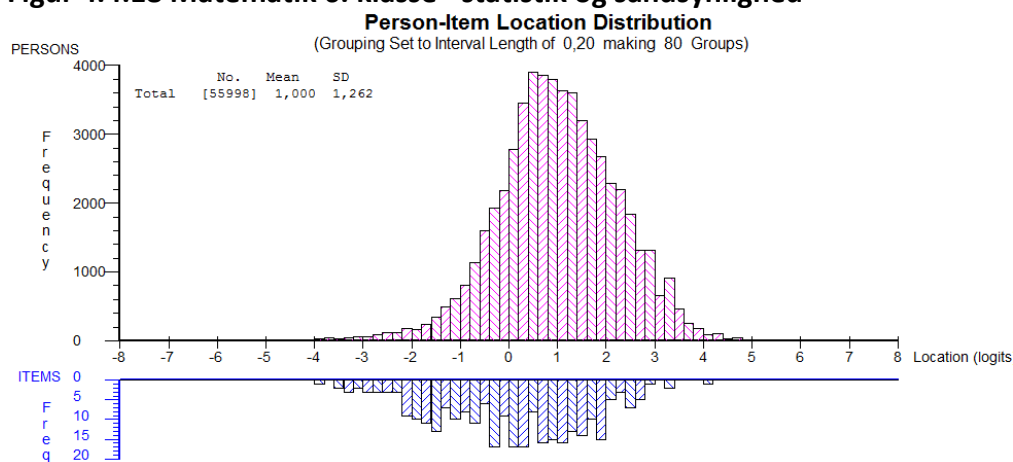
Figur 4.4.16 Matematik 6. klasse - tal og algebra



Figur 4.4.17 Matematik 6. klasse - geometri

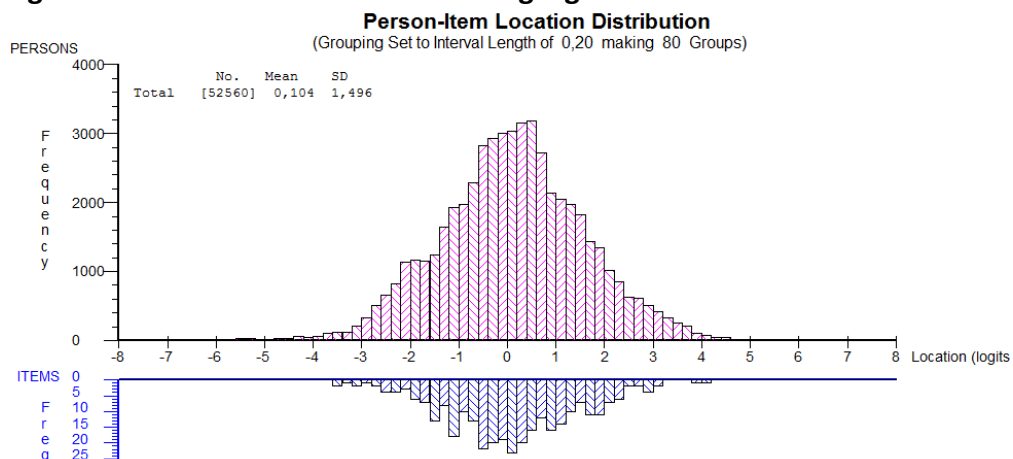


Figur 4.4.18 Matematik 6. klasse - statistik og sandsynlighed

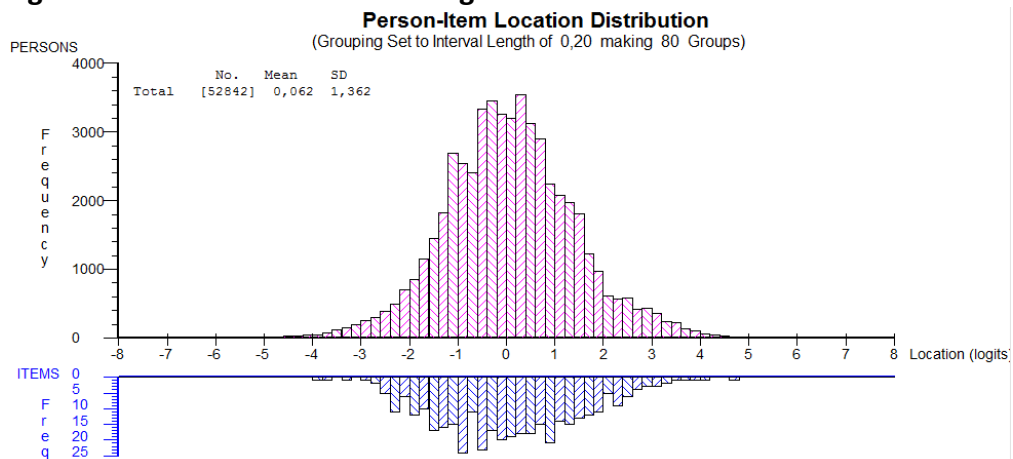


Kilde: Styrelsen for It og Læring

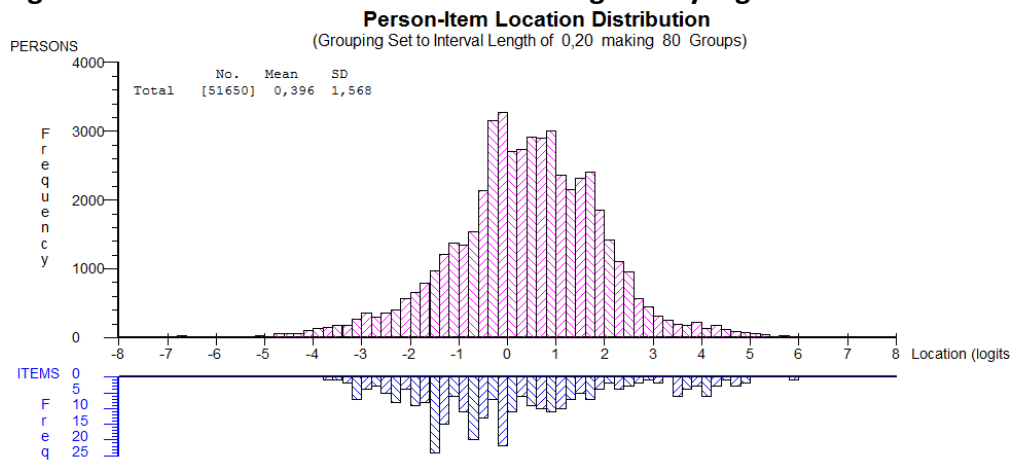
Figur 4.4.19 Matematik 8. klasse - tal og algebra



Figur 4.4.20 Matematik 8. klasse - geometri

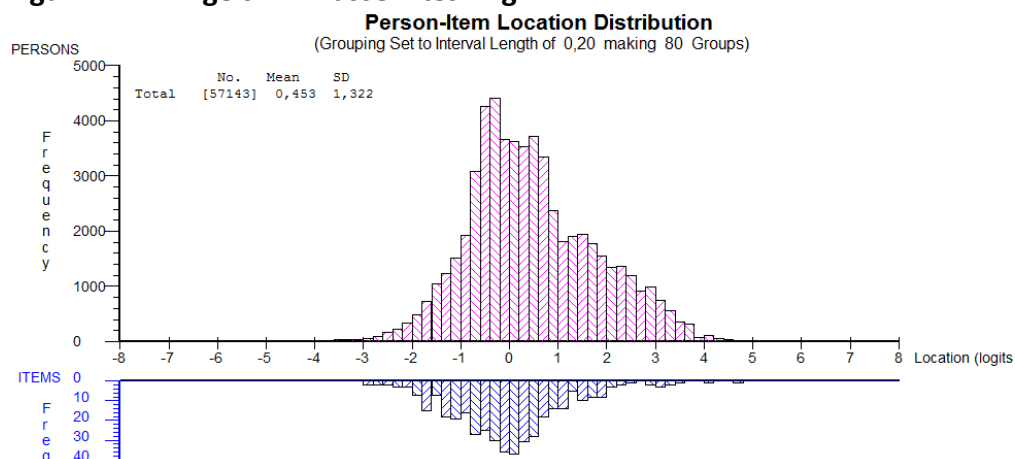


Figur 4.4.21 Matematik 8. klasse - statistik og sandsynlighed

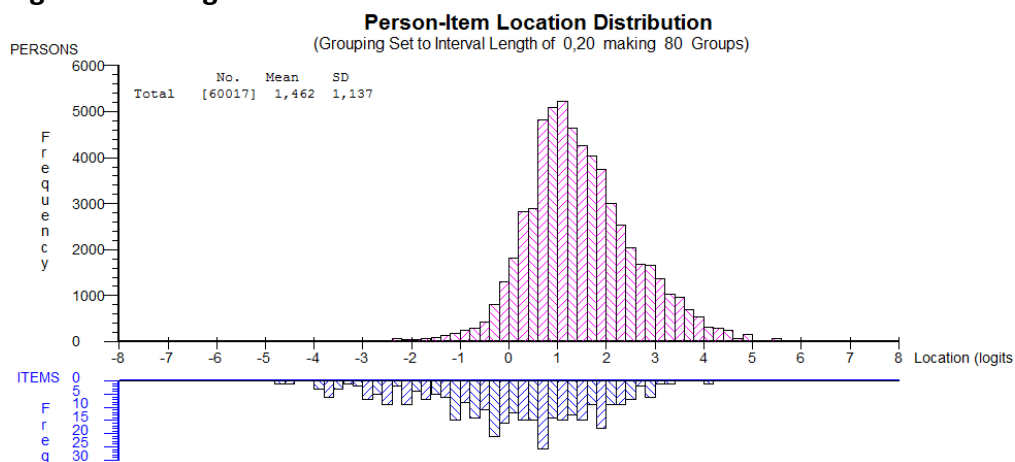


Kilde: Styrelsen for It og Læring

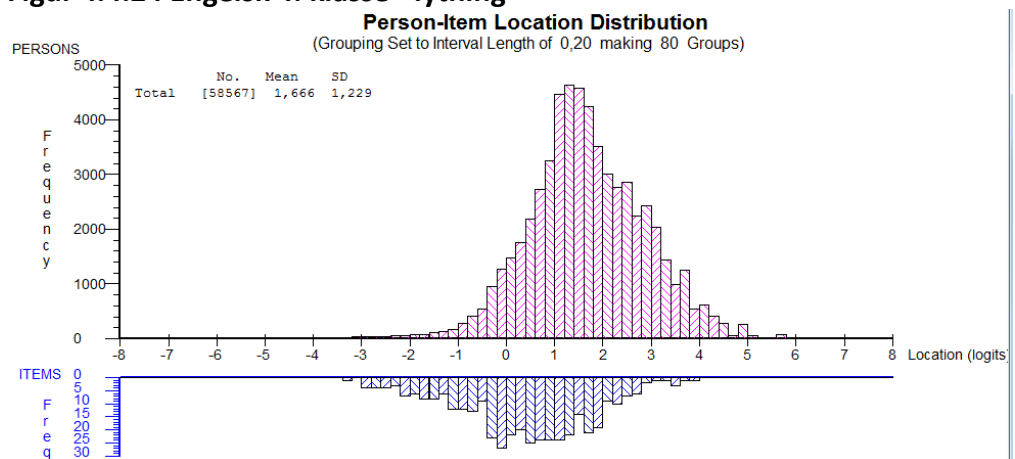
Figur 4.4.22 Engelsk 4. klasse - læsning



Figur 4.4.23 Engelsk 4. klasse - ordforråd

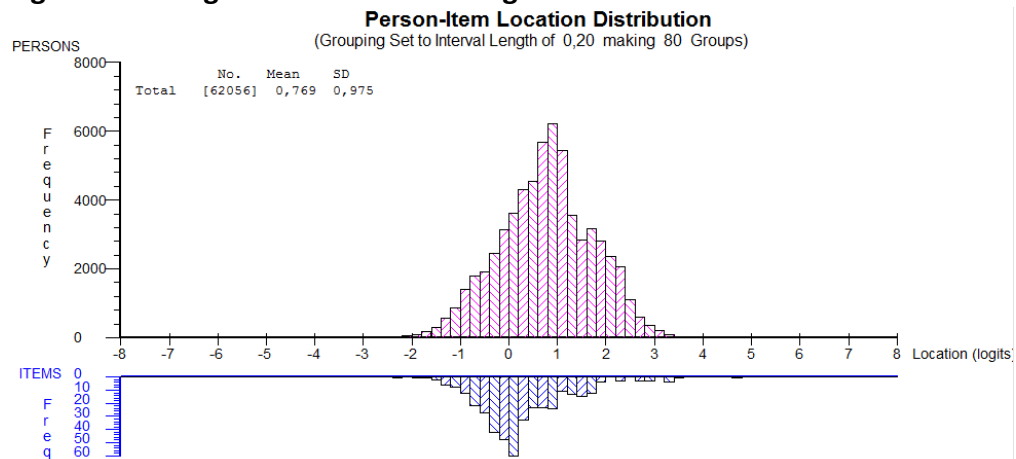


Figur 4.4.24 Engelsk 4. klasse - lytning

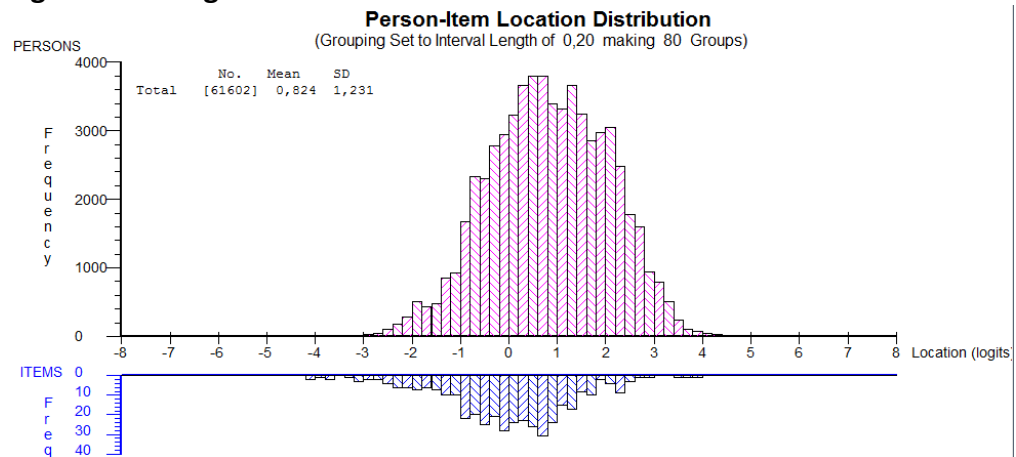


Kilde: Styrelsen for It og Læring

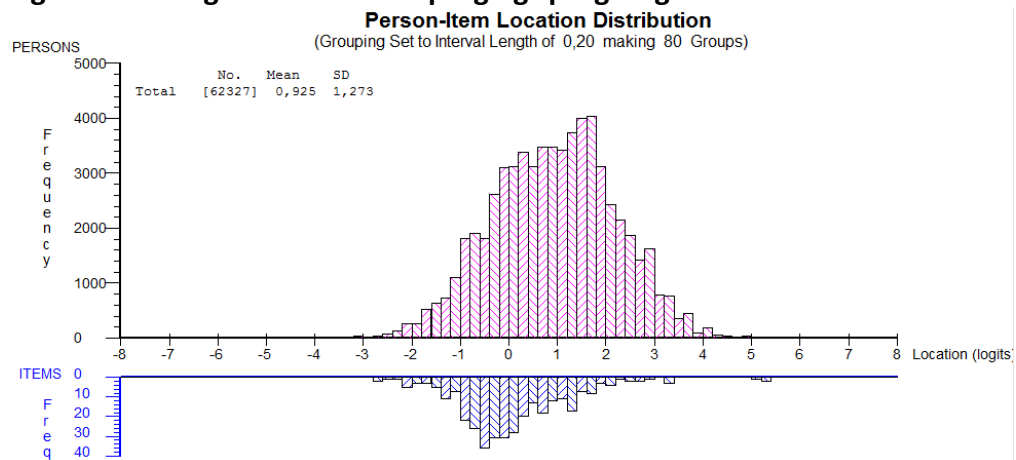
Figur 4.4.25 Engelsk 7. klasse - læsning



Figur 4.4.26 Engelsk 7. klasse - ordforråd

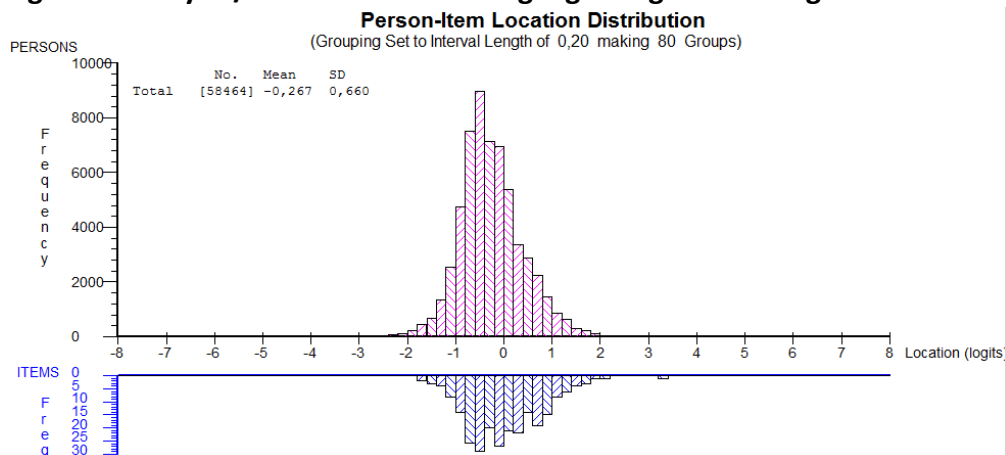


Figur 4.4.27 Engelsk 7. klasse - sprog og sprogbrug

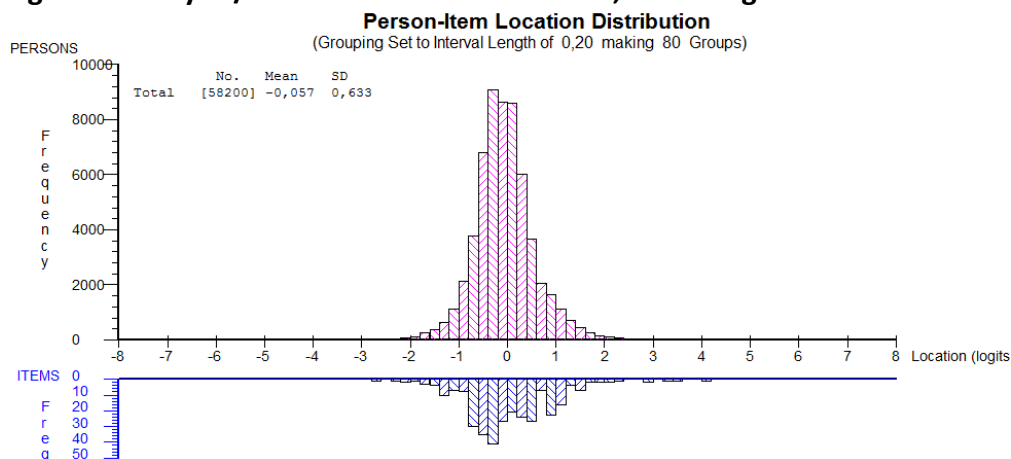


Kilde: Styrelsen for It og Læring

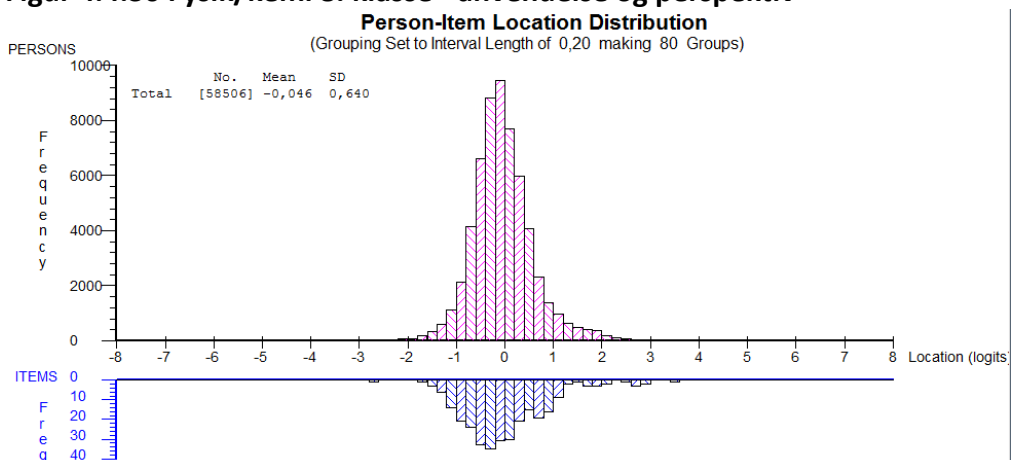
Figur 4.4.28 Fysik/kemi 8. klasse - energi og energiomsætning



Figur 4.4.29 Fysik/kemi 8. klasse - fænomener, stoffer og materialer



Figur 4.4.30 Fysik/kemi 8. klasse - anvendelse og perspektiv



Kilde: Styrelsen for It og Læring

Bilag 4.5 Undersøgelse af link-opgavernes ændrede sværhedsgrad

I forbindelse med afprøvningen af nye opgaver til opgavebanken medtages hver gang et antal af de eksisterende og tidligere godkendte opgaver fra opgavebanken. Dette sikrer, at der sker et overlap mellem blokkene af opgaveafprøvningsperioder, således at nye opgavers sværhedsgrad kan indplaceres på den eksisterende skala. Disse overlappingsopgaver kaldes link-opgaver. Typisk udvælges 5-10 link-opgaver ved hver opgaveafprøvning.

Anvendelsen af link-opgaver giver endvidere mulighed for, at undersøge om disse opgavers sværhedsgrad er ændret siden tidligere opgaveafprøvningsperioder.

I Rasch analysen foretages denne analyse i en Differential Item Functioning (DIF) analyse. Der testes for DIF mellem ny og tidligere afprøvningsperioder.

I den seneste analyserede opgaveafprøvning fra januar 2018 indgik opgaver fra følgende profilområder:

- Dansk læsning 2. klasse: sprogforståelse, afkodning og tekstforståelse
- Dansk læsning 4. klasse: sprogforståelse
- Matematik 3. klasse: geometri
- Matematik 6. klasse: tal og algebra
- Engelsk 4. klasse: ordforråd
- Fysik/kemi 8. klasse: energi og energiomsætning, fænomener, stoffer og materialer samt anvendelse og perspektiv

Der er testet for periode DIF mellem afprøvningsperioden i januar 2018 og tidligere afprøvningsperioder samlet.

Tabellerne indeholder:

- Opgavenummer
- Maj2008 – Sep2014: opgavens estimerede sværhedsgrad (location) baseret på tidligere opgaveafprøvningsperioder samt antal besvarelser
- Jan2018: opgavens nye sværhedsgrad (location) hvis denne afviger statistisk signifikant samt antal besvarelser fra opgaveafprøvningsperioden i januar 2018
- ANOVA for periode DIF: F-ratio samt tilhørende p-værdi fra ANOVA test for periode DIF. Der er anvendt Bonferroni korrektion for multiple test.

Tabel 4.5.1 Dansk læsning 2. klasse - sprogforståelse

Opgavenummer	Maj2008 - Sep2014		Jan2018		ANOVA for periode DIF	
	location	Antal	Location	Antal	F-ratio	P-værdi
010201000301235192-1	1,73	780	-	653	0,01	n.s.
010201000301235195-1	1,22	698	-	646	0,08	n.s.
010201000301235202-1	1,72	782	-	644	4,12	n.s.
010201000301235397-1	1,11	667	-	653	0,65	n.s.
010201000301235399-1	1,87	783	-	647	2,07	n.s.
010201000301235416-1	2,07	781	-	586	3,50	n.s.
010201000301235418-1	1,79	699	-	586	0,09	n.s.
010201000301235196-1	1,75	751	-	586	7,09	n.s.
010201000301235205-1	1,78	753	-	653	0,01	n.s.
010201000301235423-1	1,05	750	-	586	1,79	n.s.

Kilde: Styrelsen for It og Læring

Tabel 4.5.2 Dansk læsning 2. klasse - afkodning

Opgavenummer	Maj2008 - Sep2014		Jan2018		ANOVA for periode DIF	
	location	Antal	Location	Antal	F-ratio	P-værdi
01020204628-4	1,03	428	-	670	0,83	n.s.
01020204122-2	1,21	440	-	670	7,27	n.s.
01020204621-5	2,66	2107	-	721	4,75	n.s.
01020204626-5	2,19	603	-	669	1,71	n.s.
0102020410001-4	0,57	2446	-	720	0,68	n.s.
010202000301235179-1	1,05	652	-	721	1,88	n.s.
010202000301235189-1	1,92	653	-	721	4,12	n.s.
010202000301239054-1	0,92	717	-	721	0,20	n.s.
010202000301239443-1	1,54	734	-	670	10,79	n.s.
010202000301239453-1	0,76	779	0,40	670	14,14	0,002

Kilde: Styrelsen for It og Læring

Tabel 4.5.3 Dansk læsning 2. klasse - tekstforståelse

Opgavenummer	Maj2008 - Sep2014		Jan2018		ANOVA for periode DIF	
	location	Antal	Location	Antal	F-ratio	P-værdi
0102030510270006-3	0,06	587	-	858	10,38	n.s.
0102030510270009-1	0,86	586	-	791	0,06	n.s.
0102030510270010-1	-0,14	589	-	778	1,69	n.s.
0102030510270011-1	0,81	589	-	860	0,02	n.s.
0102030510270013-1	2,11	1377	-	793	2,21	n.s.
0102030540001-2	0,14	586	-	777	0,95	n.s.
0102030540002-2	-0,71	1413	-	858	0,46	n.s.
0102030540008-1	-0,04	588	-	794	1,22	n.s.
0102030540012-1	0,40	594	-	780	0,98	n.s.
0102030540012-3	-0,52	584	-	847	2,23	n.s.

Kilde: Styrelsen for It og Læring

Tabel 4.5.4 Dansk læsning 4. klasse - sprogforståelse

Opgavenummer	Maj2008 - Sep2014		Jan2018		ANOVA for periode DIF	
	location	Antal	Location	Antal	F-ratio	P-værdi
0104010110595-2	0,64	677	-	757	4,03	n.s.
0104010110599-3	0,49	680	-	807	5,12	n.s.
0104010110586-2	0,03	2017	-	806	0,19	n.s.
0104010110572-2	0,13	653	-	756	0,59	n.s.
0104010110599-2	0,13	663	-	807	0,06	n.s.
0104010110597-4	0,18	1438	-	674	1,21	n.s.
010401000301238258-1	-0,03	830	-	757	1,47	n.s.
010401000301238266-1	0,17	819	-	806	0,19	n.s.
010401000301238272-1	0,00	833	0,91	675	13,24	0,003
010401000301238275-1	0,09	820	-	675	0,66	n.s.

Kilde: Styrelsen for It og Læring

Tabel 4.5.5 Matematik 3. klasse - geometri

Opgavenummer	Maj2008 - Sep2014		Jan2018		ANOVA for periode DIF	
	location	Antal	Location	Antal	F-ratio	P-værdi
02030204403-1	0,24	539	-	742	0,87	n.s.
02030204403-2	0,01	1258	-	738	10,24	n.s.
02030205101-4	0,47	1369	0,12	827	49,19	0,00001
0203020410030002-3	0,51	1212	-	742	1,22	n.s.
020302000301235338-5	0,53	551	-	739	3,34	n.s.
020302000301235338-2	0,69	740	-	828	0,51	n.s.
020302000301235795-1	0,67	742	-	828	0,11	n.s.
020302000301235797-2	0,84	705	-	739	2,83	n.s.
020302000301239628-4	0,64	854	-	739	3,59	n.s.
020302000301235860-1	0,66	778	-	742	0,00	n.s.

Kilde: Styrelsen for It og Læring

Tabel 4.5.6 Matematik 6. klasse – tal og algebra

Opgavenummer	Maj2008 - Sep2014		Jan2018		ANOVA for periode DIF	
	location	Antal	Location	Antal	F-ratio	P-værdi
02060108011-11	0,66	1745	-	852	5,98	n.s.
02060108033-3	0,62	1759	1,30	851	22,91	0,00001
02060107003-22	0,59	2914	-	871	0,28	n.s.
02060106003-15	0,45	1751	-	870	5,87	n.s.
02060106007-4	0,39	853	-	544	0,49	n.s.
020601000301234766-1	0,49	680	-	892	2,84	n.s.
020601000301238714-1	0,36	665	-	871	6,50	n.s.
020601000301239117-1	0,33	650	-	870	0,42	n.s.
020601000301239365-1	0,34	652	-0,05	892	18,12	0,00003
020601000301238641-1	0,72	545	-	892	2,27	n.s.

Kilde: Styrelsen for It og Læring

Tabel 4.5.7 Engelsk 4. klasse – ordforråd

Opgavenummer	Maj2008 - Sep2014		Jan2018		ANOVA for periode DIF	
	location	Antal	Location	Antal	F-ratio	P-værdi
060402000301244229-1	1,42	850	-	695	0,17	n.s.
060402000301244205-2	1,21	657	-	698	2,90	n.s.
060402000301244420-1	1,32	807	-	698	1,34	n.s.
060402000301244255-2	1,83	615	-	695	3,89	n.s.
060402000301244239-1	1,29	763	-	697	2,61	n.s.
060402000301244253-2	1,46	764	-	717	0,00	n.s.
060402000301249843-2	1,43	1065	-	695	0,12	n.s.
060402000301249780-1	1,50	604	-	718	1,16	n.s.
060402000301249783-1	1,46	603	-	719	0,93	n.s.
060402000301249789-1	1,56	604	-	718	0,01	n.s.

Kilde: Styrelsen for It og Læring

Tabel 4.5.8 Fysik/kemi 8. klasse – energi og energiomsætning

Opgavenummer	Maj2008 - Sep2014		Jan2018		ANOVA for periode DIF	
	location	Antal	Location	Antal	F-ratio	P-værdi
03080115016-21	-0,35	476	-	660	1,19	n.s.
0308010713087-1	-0,33	464	-	660	0,53	n.s.
0308010713090-1	-0,47	2349	-	660	2,06	n.s.
0308011411120-2	-0,63	498	-	661	0,14	n.s.
0308011413178-1	-0,65	498	-	660	5,22	n.s.

Kilde: Styrelsen for It og Læring

Tabel 4.5.9 Fysik/kemi 8. klasse – fænomener, stoffer og materialer

Opgavenummer	Maj2008 - Sep2014		Jan2018		ANOVA for periode DIF	
	location	Antal	Location	Antal	F-ratio	P-værdi
03080204008-1	0,17	418	-	482	9,56	n.s.
03080204008-2	-0,31	414	-	482	0,46	n.s.
03080204008-5	0,58	481	-	482	1,65	n.s.
03080204008-6	0,40	485	-	482	6,37	n.s.
03080204034-3	-0,36	2291	-	482	0,18	n.s.
0308020411001-1	-0,79	482	-	482	1,66	n.s.
0308020412004-1	-0,02	437	-	481	0,15	n.s.
0308020412004-3	-0,78	1575	-	454	1,17	n.s.
0308020412008-1	-0,84	484	-	453	2,70	n.s.
0308020412010-2	-1,71	417	-	453	1,58	n.s.

Kilde: Styrelsen for It og Læring

Tabel 4.5.10 Fysik/kemi 8. klasse – anvendelse og perspektiv

Opgavenummer	Maj2008 - Sep2014		Jan2018		ANOVA for periode DIF	
	location	Antal	Location	Antal	F-ratio	P-værdi
0308030113007-3	-0,07	2321	-	723	3,12	n.s.
03080320010-3	0,08	2458	-	662	2,04	n.s.
03080311012-2	-0,28	2345	-	661	0,44	n.s.
03080310002-2	-0,35	2215	-	661	0,58	n.s.
03080318028-21	-0,11	2347	-	725	5,11	n.s.
0308031912168-1	-0,42	562	-	723	0,48	n.s.
0308032013174-1	-0,12	2442	-	726	1,15	n.s.
0308032211138-1	-0,09	2451	-	723	0,60	n.s.
0308032011136-1	-0,33	500	-0,73	726	14,12	0,002
030803000301239811-1	0,19	537	-	724	0,28	n.s.

Kilde: Styrelsen for It og Læring

Bilag 4.6 Forskel i opgavernes sværhedsgrad

Sværhedsgraderne på opgaverne i opgavebanken er estimeret på baggrund af opgaveafprøvninger. I en opgaveafprøvning afprøver ca. 700 elever opgaver i et lineært afprøvningsforløb.

Elevbesvarelser fra obligatoriske test kan også anvendes til at estimere opgavernes sværhedsgrad.

Data fra elevbesvarelser i obligatoriske test er baseret på adaptive forløb, hvor elever, der svarer rigtigt på en opgave, får stillet en sværere opgave næste gang, og elever, der svarer forkert på en opgave, får stillet en lettere opgave næste gang. Alle elever starter med middelsvære opgaver, dvs opgaver der ligger midt på sværhedsskalaen for det pågældende fag og profilområde.

Dette bilag indeholder tabeller og figurer, der viser forskellen mellem opgavernes sværhedsgrad baseret på opgaveafprøvningerne og opgavernes sværhedsgrad, hvis den blev beregnet på baggrund af elevbesvarelser fra de obligatoriske test. Sammenligningen er foretaget for elevbesvarelserne i de obligatoriske test i foråret 2010, 2014 og 2018 i dansk læsning 6. og 8. klasse samt i matematik 3. og 6. klasse. I matematik er beregningerne kun foretaget for profilområde 1 og 2, da profilområde 3 blev erstattet med et nyt profilområde i 3. klasse i 2015/2016 og i 6. klasse i 2017/2018.

I tabellerne anvendes forskellen i en opgaves sværhedsgrad, når den beregnes ved anvendelse af data fra henholdsvis opgaveafprøvninger, location(OAP), og fra de obligatoriske test, location(OBL), dvs

$$\text{forskel} = \text{location(OAP)} - \text{location(OBL)}$$

Desuden er den absolutte forskel beregnet:

$$\text{absolut forskel} = \text{abs}(\text{location(OAP)} - \text{location(OBL)})$$

Tabellerne og figurerne viser:

- Fordelingen af forskellen i estimeret sværhedsgrad for de enkelte profilområder og samlet for testene i dansk læsning og matematik
- Fordelingen af forskellen i estimeret sværhedsgrad i forhold til hvilket år opgaverne er afprøvet
- Fordelingen af forskellen i estimeret sværhedsgrad i forhold til opgavernes sværhedsgrad

Tabel 4.6.1 Fordeling af absolut forskel i estimerede sværhedsgrader mellem opgaveafprøvningsgraderne og elevbesvarelserne i obligatoriske test. Dansk læsning 6. klasse

Interval ¹⁾	Sprogforståelse			Afkodning			Tekstforståelse		
	2010	2014	2018	2010	2014	2018	2010	2014	2018
0,0 – 0,5	42 %	28 %	29 %	46 %	45 %	54 %	70 %	68 %	65 %
0,5 – 1,0	41 %	33 %	37 %	35 %	33 %	29 %	26 %	23 %	28 %
1,0 – 1,5	13 %	26 %	25 %	16 %	16 %	12 %	4 %	9 %	4 %
1,5 – 2,0	5 %	9 %	7 %	3 %	5 %	4 %			2 %
2,0 – 2,5		3 %	2 %			1 %			1 %
2,5 +		1 %							
Forskel²⁾	0,04	0,00	-0,03	0,04	0,05	-0,02	-0,13	-0,12	0,05
Abs. forskel³⁾	0,65	0,88	0,82	0,60	0,63	0,58	0,38	0,41	0,47
Antal	224	264	241	231	229	241	185	222	292

1) Absolut forskel på logit-skalaen

2) Gennemsnitlig forskel

3) Gennemsnitlig absolut forskel

Kilde: Styrelsen for It og Læring

Tabel 4.6.2 Fordeling af absolut forskel i estimerede sværhedsgrader mellem opgaveafprøvningsgraderne og elevbesvarelserne i obligatoriske test. Dansk læsning 8. klasse

Interval ¹⁾	Sprogforståelse			Afkodning			Tekstforståelse		
	2010	2014	2018	2010	2014	2018	2010	2014	2018
0,0 – 0,5	29 %	25 %	34 %	26 %	30 %	52 %	83 %	77 %	54 %
0,5 – 1,0	39 %	31 %	38 %	22 %	18 %	35 %	16 %	19 %	34 %
1,0 – 1,5	29 %	25 %	21 %	11 %	14 %	8 %	1 %	4 %	10 %
1,5 – 2,0	3 %	17 %	5 %	9 %	9 %	4 %	0 %	1 %	2 %
2,0 – 2,5		1 %	1 %	10 %	10 %	0 %			0 %
2,5 +				23 %	18 %				0 %
Forskel²⁾	0,48	0,42	0,04	-0,09	-0,10	-0,03	-0,05	-0,05	0,04
Abs. forskel³⁾	0,74	0,92	0,76	1,52	1,39	0,55	0,31	0,36	0,54
Antal	178	162	308	217	212	214	207	193	301

1) Absolut forskel på logit-skalaen

2) Gennemsnitlig forskel

3) Gennemsnitlig absolut forskel

Kilde: Styrelsen for It og Læring

Tabel 4.6.3 Fordeling af absolut forskel i estimerede sværhedsgrader mellem opgaveafprøvningsgraderne og elevbesvarelsesgraderne i obligatoriske test. Matematik 3. klasse

Interval ¹⁾	Tal og algebra			Geometri og måling		
	2010	2014	2018	2010	2014	2018
0,0 – 0,5	62 %	50 %	46 %	51 %	43 %	35 %
0,5 – 1,0	30 %	29 %	30 %	33 %	30 %	29 %
1,0 – 1,5	3 %	15 %	18 %	12 %	16 %	19 %
1,5 – 2,0	2 %	4 %	4 %	4 %	6 %	11 %
2,0 – 2,5	1 %	1 %	1 %	1 %	4 %	6 %
2,5 +	2 %	1 %	2 %	1 %	2 %	1 %
Forskel ²⁾	-0,01	0,03	-0,04	-0,06	-0,07	-0,01
Abs. forskel ³⁾	0,49	0,63	0,66	0,60	0,75	0,87
Antal	202	230	290	180	223	199

1) Absolut forskel på logit-skalaen

2) Gennemsnitlig forskel

3) Gennemsnitlig absolut forskel

Kilde: Styrelsen for It og Læring

Tabel 4.6.4 Fordeling af absolut forskel i estimerede sværhedsgrader mellem opgaveafprøvningsgraderne og elevbesvarelsesgraderne i obligatoriske test. Matematik 6. klasse

Interval ¹⁾	Tal og algebra			Geometri og måling		
	2010	2014	2018	2010	2014	2018
0,0 – 0,5	43 %	27 %	28 %	40 %	30 %	25 %
0,5 – 1,0	33 %	36 %	27 %	37 %	34 %	24 %
1,0 – 1,5	15 %	21 %	26 %	15 %	19 %	24 %
1,5 – 2,0	6 %	8 %	11 %	5 %	8 %	21 %
2,0 – 2,5	2 %	5 %	5 %	3 %	5 %	3 %
2,5 +	1 %	3 %	3 %	1 %	3 %	3 %
Forskel ²⁾	-0,03	-0,04	-0,03	-0,02	-0,01	0,05
Abs. forskel ³⁾	0,72	0,92	1,00	0,73	0,92	1,06
Antal	347	333	494	200	265	263

1) Absolut forskel på logit-skalaen

2) Gennemsnitlig forskel

3) Gennemsnitlig absolut forskel

Kilde: Styrelsen for It og Læring

Tabel 4.6.5 Fordeling af absolut forskel i estimerede sværhedsgrader mellem opgaveafprøvningsgraderne og elevbesvarelserne i obligatoriske test. Samlet for dansk læsning 6. og 8. klasse og matematik 3. og 6. klasse. Kun opgaver der har været aktive i hele perioden

Interval ¹⁾	2010	2014	2018
0,0 – 0,5	49 %	41 %	43 %
0,5 – 1,0	31 %	30 %	31 %
1,0 – 1,5	12 %	16 %	15 %
1,5 – 2,0	4 %	7 %	7 %
2,0 – 2,5	2 %	3 %	2 %
2,5 +	2 %	3 %	2 %
Forskel ²⁾	0,03	0,03	0,06
Absolut forskel ³⁾	0,69	0,79	0,74
Antal	1.601	1.601	1.601

1) Absolut forskel på logit-skalaen

2) Gennemsnitlig forskel

3) Gennemsnitlig absolut forskel

Kilde: Styrelsen for It og Læring

Tabel 4.6.6 Fordeling af absolut forskel i estimerede sværhedsgrader mellem opgaveafprøvningsgraderne og elevbesvarelserne i obligatoriske test foråret 2018 i forhold til perioden for opgavernes afprøvningsperiode. Samlet for dansk læsning 6. og 8. klasse og matematik 3. og 6. klasse. Opgaver i opgavebanken i 2018

Interval ¹⁾	Periode for opgaveafprøvningsperiode (N=2.843)						
	2008 (N=1.123)	2009 (N=507)	2010 (N=246)	2012 (N=326)	2014 (N=503)	2015 (N=94)	2016 (N=44)
0,0 – 0,5	38 %	55 %	45 %	32 %	43 %	27 %	34 %
0,5 – 1,0	31 %	30 %	30 %	25 %	31 %	65 %	20 %
1,0 +	32 %	15 %	25 %	43 %	27 %	9 %	45 %
Forskel ²⁾	0,03	0,14	-0,05	-0,32	0,06	0,41	0,33
Abs. forskel ³⁾	0,82	0,55	0,70	0,88	0,72	0,65	0,88

1) Absolut forskel på logit-skalaen

2) Gennemsnitlig forskel

3) Gennemsnitlig absolut forskel

Kilde: Styrelsen for It og Læring

Tabel 4.6.7 Fordeling af absolut forskel i estimerede sværhedsgrader mellem opgaveafprøvningsgraderne og elevbesvarelserne i obligatoriske test foråret 2018 i forhold til opgavernes sværhedsgrad. Samlet for dansk læsning 6. og 8. klasse og matematik 3. og 6. klasse. Opgaver i opgavebanken i 2018

Interval ¹⁾	Opgavernes sværhedsgrad ¹⁾ (N=2.843)				
	(÷ ; ÷2] (N=195)	(÷2 ; ÷1] (N=400)	(÷1 ; +1] (N=1.656)	(+1 ; +2] (N=399)	(+2 ; +) (N=193)
0,0 – 0,5	36 %	33 %	53 %	18 %	7 %
0,5 – 1,0	36 %	26 %	34 %	30 %	16 %
1,0 +	28 %	42 %	13 %	52 %	78 %
Forskel²⁾	0,30	0,77	0,22	-1,04	-1,57
Abs. forskel³⁾	0,76	0,85	0,55	1,07	1,59

1) Absolut forskel på logit-skalaen

2) Gennemsnitlig forskel

3) Gennemsnitlig absolut forskel

Kilde: Styrelsen for It og Læring