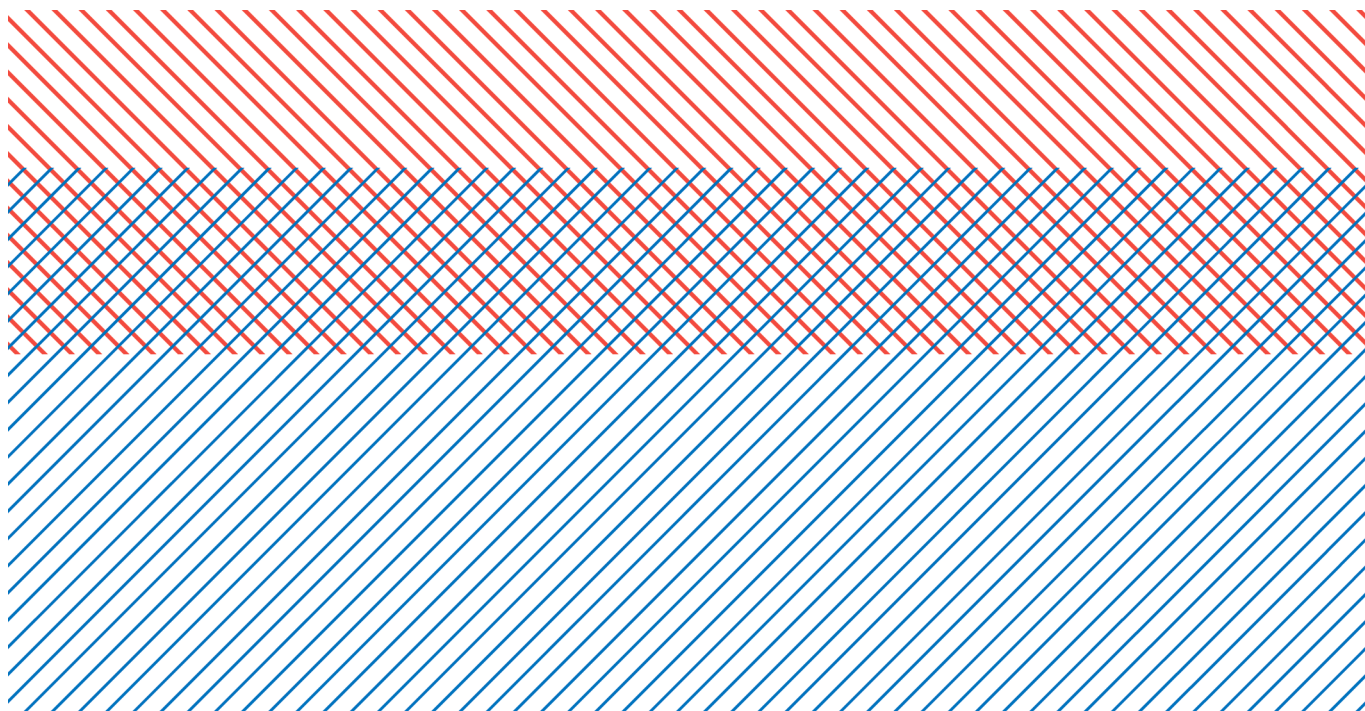


Notat

Metodenotat for skalavalideringer

Teknisk bilag til rapporten "Lærere og pædagogers oplevelse af den længere og mere varierede skoledag, 2018"



Peter Rohde Skov

Metodenotat for skalavalideringer – Teknisk bilag til rapporten "Lærere og pædagogers oplevelse af den længere og mere varierede skoledag, 2018"

© VIVE og forfatterne, 2019

e-ISBN: 978-87-7119-624-5

Projekt: 11013

VIVE – Viden til Velfærd

Det Nationale Forsknings- og Analysecenter for Velfærd

Herluf Trolles Gade 11, 1052 København K

www.vive.dk

VIVEs publikationer kan frit citeres med tydelig kildeangivelse.

Forord

Dette notat er et teknisk bilag til rapporten "Lærere og pædagogers oplevelse af den længere og mere varierede skoledag i folkeskolereformens fjerde år" (Jensen, Skov & Thranholm, 2018).

Notatet dokumenterer de beregninger og tekniske analyser, der ligger bag de indeks, som anvendes i rapporten.

Hanne Søndergård Pedersen

Konst. forsknings- og analysechef, VIVE Børn og Uddannelse

2019

Indhold

1	Indledning og sammenfatning	6
1.1	Læsevejledning	7
2	Validering af indeks ved hjælp af Rasch-modellen	8
2.1	De grafiske modeller	9
2.2	Statistiske test-of-fit	11
3	Brug af Fælles Mål	13
3.1	Graf	13
3.2	Globale test	14
3.3	Item fits	15
3.4	Oversigt over LD og DIF	16
4	Åben skole	18
4.1	Graf	19
4.2	Globale test	19
4.3	Item fits	20
4.4	Oversigt over LD og DIF	20
5	Brug af it	21
5.1	Graf	22
5.2	Globale test	22
5.3	Item fits	23
5.4	Oversigt over LD og DIF	23
6	Fokus på faglighed	24
6.1	Graf	25
6.2	Globale test	25
6.3	Item fits	26
6.4	Oversigt over LD og DIF	26
7	Brug af test og elevplaner	27
7.1	Graf	27
7.2	Globale test	28
7.3	Item fits	28
7.4	Oversigt over LD og DIF	28
8	Brug af feedbacksamtaler og elevvurdering	30
8.1	Graf	31
8.2	Globale test	31
8.3	Item fits	32
8.4	Oversigt over LD og DIF	32
9	Anvendelse af Ekstra støtte i undervisningen	33
9.1	Graf	34

9.2	Globale test.....	34
9.3	Item fits.....	35
9.4	Oversigt over LD og DIF	35
10	Samarbejde mellem pædagoger og øvrigt personale	36
10.1	Graf	36
10.2	Globale test.....	37
10.3	Item fits.....	37
10.4	Oversigt over LD og DIF	37
	Litteratur	38

1 Indledning og sammenfatning

Dette notat indeholder dokumentationen af de skalavalideringer, der er foretaget i forbindelse med rapporteringen af lærere og pædagogers besvarelser i rapporten "Lærernes og pædagogernes oplevelse af den længere og mere varierede skoledag i folkeskolereformens fjerde år" (Jensen, Skov & Thranholm, 2018). En skala er et andet ord for et indeks, hvorfor skalavalidering handler om, hvorvidt de anvendte indeks i Jensen, Skov & Thranholm (2018) giver mening i statistisk forstand.

En skala i statistiske analyser er med andre ord en kondensering af en række spørgsmål inden for et enkelt emne. Skalaer anvendes ofte, når man har flere spørgsmål, der handler om samme emne – et emne, som kan være vanskeligt at måle ved et enkelt spørgsmål, og hvor svarene på de enkelte spørgsmål peger i samme retning. I dette notat beskriver vi de test, som vi har anvendt til at afgøre, hvorvidt en gruppe af spørgsmål (items) kan kondenseres til en skala. De anvendte skalaer blev konstrueret første gang i 2015 i forbindelse med offentliggørelsen af den første af fire årlige statusrapporter om lærernes og pædagogernes besvarelser, se Jacobsen & Kreiner (2015) for den første validering af skalaer. I de efterfølgende år har der alene været tale om en genberegning af skalaer for fortsat at sikre, om de er valide i teknisk forstand.

Formålet med at validere de enkelte skalaer (eller indeks) er som nævnt at tage stilling til følgende spørgsmål:

- Er svarene på de enkelte spørgsmål indbyrdes afhængige, og kan denne afhængighed forklares 100 % af den bagvedliggende fællesnævner (faktor)?
- Svarer respondenter med forskelligt køn, alder mv. ens på de stillede spørgsmål, eller betyder baggrundsforholdene noget for disse svar?
- Bidrager alle de udvalgte spørgsmål væsentligt til skalaen?

1.1 Sammenfatning

Vi anvender en grafisk log-lineær Rasch-model til at konstruere de enkelte skalaer. Det er en psykometrisk model, der måler adfærd (Kreiner, 2007; Kreiner & Christensen, 2002, 2004, 2007, 2011), og alle analyser er foretaget ved hjælp af programmet DIGRAM (Kreiner, 2003; Kreiner & Nielsen, 2016). Ud over at af- eller bekræfte, om en gruppe spørgsmål kan samles til en skala, tager denne model også højde for, at der kan være systematiske forskelle i svarene for forskellige grupper af respondenter. Det kan fx være, at ældre lærere i højere grad svarer i én retning, mens yngre lærere svarer i en anden. Dette benævnes *differentiel item funktion (DIF)* i de statistiske analyser.

Rasch-modellen analyserer også *lokal afhængighed (LD)*. LD er, når den bagvedliggende fællesnævner (skalaen) ikke kan forklare hele den indbyrdes afhængighed mellem de spørgsmål, som indgår i skalaen. Eksempelvis består skalaen for brug af fælles mål af svarene på tre spørgsmål. Bland disse tre spørgsmål skal svaret på spørgsmålet "Jeg nedbryder målbeskrivelsen i Fælles Mål til konkrete mål for eleverne" være uafhængigt af svarene på spørgsmålet "Jeg gennemfører min undervisning med udgangspunkt i Fælles Mål" og svaret på spørgsmålet "Jeg evaluerer elevernes udbytte af undervisningen i forhold til Fælles Mål (LPQ99)". Hvis fx svarene på det første af spørgsmålene kan forudsige svarene på de næste to spørgsmål, vil der være tale om LD, hvilket kan medføre skævheder i skalaen.

Rasch-modellen kan dog ikke tage højde for systematiske forskelle, som vi umiddelbart ikke har en variabel for. Det vil sige, at den tekniske model ikke kan være bedre end de tilgængelige data. Vi vil derfor aldrig være i stand til at forklare hele sammenhængen mellem de udvalgte spørgsmål.

Analyserne viser, at de foreslåede skalaer både har DIF og LD. Analyserne viser dog også, at svarene passer til de log-lineære Rasch-modeller. Derfor kan den samlede score over de enkelte spørgsmål i skalaerne betragtes som et statistisk udtryk for det, som skalaerne skal måle. I de tilfælde, hvor der var tale om DIF, var det imidlertid nødvendigt at foretage justeringer, således at værdierne blev sammenlignelige på tværs af forskellige baggrundsvariable (defineret ved køn, alder eller andre variable, der var årsag til DIF). Vi har fundet følgende skalaer valide:

- Brug af Fælles Mål
- Åben skole
- Brug af IT
- Fokus på faglighed
- Brug af test og elevplaner
- Brug af feedbacksamtaler og elevvurdering
- Anvendelse af ekstra støtte i undervisningen
- Samarbejde mellem pædagoger og øvrigt personale.

I forhold til tidligere rapporteringer har vi ikke fundet, at indekset 'Lærersamarbejde og pædagogisk faglig sparring' er validt.

1.2 Læsevejledning

I kapitel 2 giver vi en kort introduktion til validering af skalaer ved hjælp af Rasch-modellen, og i de efterfølgende kapitler 3-10 gennemgår vi valideringen af de enkelte indeks ét efter ét.

For hvert af disse kapitler præsenteres først de spørgsmål fra spørgeskemaet (items), der er anvendt til at måle de pågældende skalaer, samt antallet af respondenter. Derefter illustrerer en figur den anvendte psykometriske model. Endelig præsenteres de testresultater, som dokumenterer figurens indhold.

Gennemgangen af den første skala, 'Brug af Fælles Mål', indeholder en kort beskrivelse af, hvordan resultaterne skal læses, mens gennemgangen af de øvrige indeks kun viser testresultaterne uden yderligere beskrivelse.

Vi anvender løbende en række forkortelser i notatet. De er anført i Tabel 1.1.

Tabel 1.1 Anvendte forkortelser i notatet

Forkortelse	Betydning
GLLRM	Graphical Log-Linear Rasch Model
LD	Lokal afhængighed
DIF	Differential Item Functioning
IRT	Item Response Theory
CLR	Conditional Likelihood Ratio
p	P-værdi
DF	Frihedsgrader

2 Validering af skalaer ved hjælp af Rasch-modellen

I brugen af Rasch-modellen anvendes ofte benævnelsen skalaer frem for indeks. I de efterfølgende tekniske beskrivelser anvender vi derfor også benævnelsen skala og konstruktion af skalaer, som anvendes i Jensen, Skov & Thranholm (2018).

I det følgende skelner vi imellem items, der er spørgsmål, som definerer en skala, og baggrundsvARIABLE, der er karakteristika, som kan påvirke svarene på disse items.

Der skal mindst tre items til at danne en skala. Endvidere skal der være tilstrækkelig stor sammenhæng mellem disse items til, at de tilsammen kan anvendes til at måle den samme skala. Men når spørgsmål sammensættes til en skala, kan det være, at deres indbyrdes afhængighed ikke udelukkende kan forklares af én bagvedliggende faktor, der ikke er direkte observerbar i data, og dermed er latent. Denne form for afhængighed mellem variablene benævnes *lokal afhængighed* (LD). En antagelse bag en Rasch-model er, at items, der definerer skalaen, er betinget uafhængige, givet den bagvedliggende faktor. De forskellige items i en skala skal dermed kun være afhængige gennem den bagvedliggende faktor og ellers ikke være afhængige af hinanden. Det skyldes, at det ellers er "for let" at forudsige svar og scorer på skalaen (Kreiner & Christensen, 2007). Rasch-modellen er i stand til at analysere, hvorvidt der er tale om LD, og alle modeller i vores analyser er karakteriseret af dette.

LD er i sig selv ikke direkte afgørende for, hvorvidt de enkelte items kan samles til en skala eller ej, men det påvirker skalaens pålidelighed. Da vi har mange respondenter i vores analysegrundlag, opvejer det dog de potentielle skævheder, som LD kan forårsage. Vi foretager dog fortsat analyser af LD for at sikre, at enkelte spørgsmål ikke udelukkes eller inkluderes på et forkert grundlag. Analyserne af LD kan forklare, hvorfor enkelte items fungerer bedre end andre i skalaen. I vores analyser varierer antallet af respondenter mellem 8.771 til 10.711 – på nær en analyse, som kun bygger på 1.910 besvarelser. Alle skalaer bygger derfor på svar fra mange respondenter. Vurderingen af styrken af den lokale sammenhæng rapporteres sammen med resultaterne af de statistiske test, der dokumenter sammenhængens eksistens.

Det kan imidlertid være problematisk for en skala, hvis forskellige grupper af respondenter svarer systematisk forskelligt på de samme items. Man kan eksempelvis tænke sig, at mænd og kvinder har forskellige opfattelser af specifikke handlinger. Det vil sige, at der kan være en DIF, der skævrider den målte sammenhæng mellem den generelle faktor (som man ønsker at finde) og de bagvedliggende variable. Rasch-modellen kan korrigere for DIF, således at man kan undlade at fjerne spørgsmål med DIF eller undgå, at målingerne bliver behæftet med systematiske fejl.

LD og DIF kan forklare de fleste tegn på, at i øvrigt gode spørgsmål ser ud til at fungere dårligt i skalaer. Det kan naturligvis ikke afvises, at der stadig er dårlige spørgsmål tilbage efter korrektion for de to andre forhold. Dette undersøges også i de tekniske analyser.

Vi har valideret følgende skalaer, hvor tallene i parentes definerer de teoretisk set mulige minimums- og maksimumsværdier:

- Brug af Fælles Mål (0-12)
- Åben skole (0-9)
- Brug af it (0-20)

- Fokus på faglighed (0-12)
- Brug af test og elevplaner (0-15)
- Brug af feedbacksamtaler og elevvurdering¹
- Anvendelse af ekstra støtte i undervisningen
- Samarbejde mellem pædagoger og øvrigt personale (0-15).

Med undtagelse af 'Samarbejde mellem pædagoger og øvrigt personale' er alle skalaer beregnet med udgangspunkt i besvarelsene fra dansk og matematiklærere, mens 'Samarbejde mellem pædagoger og øvrigt personale' er beregnet på baggrund af besvarelses fra pædagogerne.²

Bortset fra skalaerne for 'Åben skole', 'Brug af feedbacksamtaler og elevvurdering' samt 'Anvendelse af ekstra støtte i undervisningen' er skalaerne konstrueret, således at lav score er udtryk for høj grad af anvendelse af det pågældende emne.³

Hvis man ønsker at afrapportere skalaerne med værdier fra 0 til 1, som det er gjort i rapporten Jensen, Skov & Thranholm (2018), skal de omregnes, således at værdien 1 er udtryk for det mest positive og værdien 0 for det mindst positive.

Bortset fra de tre ovennævnte skalaer er de beregnet på baggrund af følgende formel:

$$Skala_{omskalaeret\ score} = \frac{(Skala_{maksimumscore} - Skala_{score})}{Skala_{maksimumscore}}$$

Eksempelvis vil reskaleringen for 'Brug af Fælles Mål' kunne ske ved:

$$Brug\ af\ Fælles\ Mål_{omskaleret\ score} = \frac{(12 - Brug\ af\ Fælles\ Mål_{score})}{12}$$

2.1 De grafiske modeller

Nedenstående Figur 2.1 viser alle forbindelser imellem baggrundsvariable og items for skalaen for "Brug af Fælles Mål". Denne graf laves for alle indeks, dog med udskiftning af de – for indekset – relevante items. Der er to betingelser, der definerer den grafiske model. Den første betingelse er, at variablene skal kunne beskrive årsagsforhold mellem hinanden. Dette gøres ved at dele variablene op i blokke som markeret med kasser i Figur 2.1. Disse opdelinger svarer til vores forventning om, hvorvidt en variabel er en baggrundsvariabel af betydning eller ej (såsom alder, køn). Den anden betingelse er, at variablene er betinget uafhængige, når vi har taget højde for variable på samme eller foregående niveau.

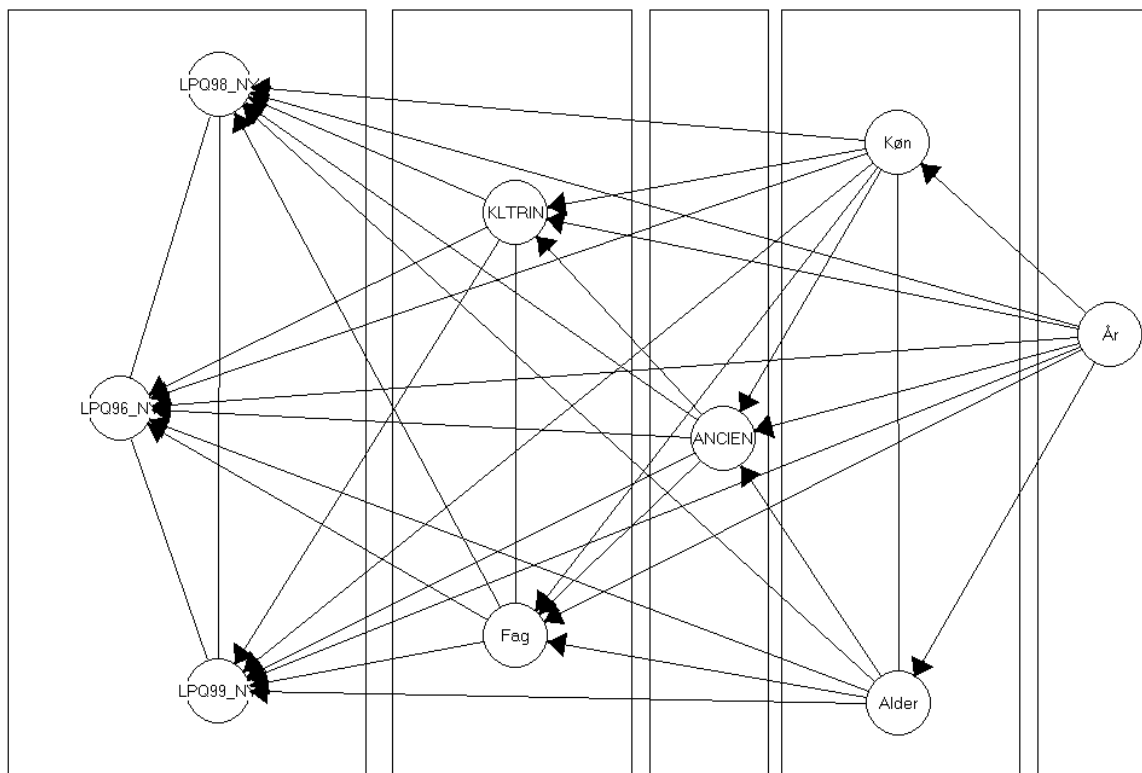
¹ I forhold til tidligere rapporter analyserer Jensen, Skov & Thranholm (2018) ikke skalaerne for 'Brug af feedbacksamtaler og elevvurdering' samt 'Anvendelse af ekstra støtte i undervisningen'. For begge skalaer gælder, at selvom der i teknisk forstand er tale om en sammenhæng mellem de enkelte spørgsmål, kan man sætte spørgsmålstejn ved, hvorvidt skalaerne giver indholdsmæssig mening. Eksempelvis indgår der i skalaen for 'Brug af feedbacksamtaler og elevvurdering' en kombination af holdningsspørgsmål og spørgsmål om, hvor hyppigt lærerne foretager bestemte handlinger. Der er derfor tale om spørgsmål med forskellige dimensioner, som samtidig også benytter vidt forskellige svarskalaer. I bilaget til Jensen, Skov & Thranholm (2018) viser vi dog disse to skalaer, og den tilhørende dokumentation for disse skalaer er derfor præsenteret i dette notat.

² Det har ikke været muligt at beregne en skala for 'Lærersamarbejde og pædagogisk faglig sparring'. Det skyldes, at data ikke længere passer på en Rasch-model og derfor ikke vil lade sig estimere. Dokumentation for denne skala er derfor ikke medtaget i dette tekniske notat.

³ Ud fra de beregnede og korrigerede skalaer trækker vi person-estimer, som er de estimer, der fremkommer, når man tager højde for DIF. Vi anvender person-estimerne i afrapporteringen af skalaerne, som er gennemsnit af disse person-estimer.

Figuren bruger kanter eller pile til at angive forhold imellem variable. Variable på samme niveau har ingen pil for enden, da der ikke antages at være nogen retningsbetingelse i årsagsniveauet.

Figur 2.1 Graf før modelsøgning for skalaen "Brug af Fælles Mål"



Anm.: Variable er repræsenteret ved punkter, og variablene antages at være betinget uafhængige givet alle de andre variable, hvis der ikke er nogen linje eller pil, der forbinder dem. "-" præfiks og "_n" suffiks angiver, at variablene er vendt om henholdsvis sammensat ud fra andre variable i spørgeskemaet. Pile antages at referere til sammenhænge, mens kanter uden retning repræsenterer symmetriske relationer uden simpel forklaring. De hvide punkter er items; det grønne punkt repræsenterer den generelle faktor, mens de grå punkter er baggrundsvARIABLE. Anciën = anciennitet; Kltrin = klassetrin.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

For hver skala giver Rasch-modellen en grafisk illustration af den indbyrdes sammenhæng mellem de udvalgte variable, hvor sammenhængen mellem fx svar på item A, B og C kan forklares af andre mellemliggende, bagvedliggende og sideordnede variable, som eksempelvis angivet i Figur 2.1, de tre items LPQ96, LPQ98 og LPQ99. I vores tilfælde ligger survey-år forud for baggrundsvARIABLENE alder og køn, mens alder og køn er på et sideordnet niveau. Survey-år ligger forud for de øvrige variable, og dermed antager vi, at der kan være forskelle over tid, i hvordan lærerne besvarer de spørgsmål. Når baggrundsvARIABLENE er på sideordnet niveau, så antages de ikke at påvirke hinanden som årsager. Alder og køn ligger forud for anciennitet i vores modeller, mens anciennitet ligger forud for det fag, som lærerne underviser i, samt klassetrin. En anden betegnelse for denne indbyrdes sammenhæng er baggrundsvARIABLENES rekursive niveau, hvor fag og klassetrin ligger på samme rekursive niveau, mens køn og alder ligger på samme rekursive niveau.

Et item (som er et svar på de udvalgte spørgsmål til skalaen) kan være direkte påvirket af køn, alder eller år. Det kan også være påvirket af den mellemliggende variabel – lærerens anciennitet. Anciennitet kan derimod ikke påvirke lærernes køn eller alder. År kan have en direkte påvirkning på

lærernes svar på et item, da forståelsen af et spørgsmål kan ændre sig over tid, eksempelvis som følge af en ændring i praksis eller lovgivning.

Sammenhængene mellem baggrundsvARIABLE og items beregnes ved hjælp af gamma-koefficienter (γ). Gamma-koefficienter er rangkorrelationer beregnet for ordinale kategoriale data, som det er tilfældet for items på følgeforskningens lærerdata⁴.

For hver skala opstiller vi først en model, der antager, at alle baggrundsvARIABLE påvirker lærernes svar på de udvalgte items. Vi foretager herefter en modelsøgning, som analyserer de indbyrdes sammenhænge mellem de udvalgte items og baggrundsvARIABLENE. Det vil sige, at vi ser på resultaterne af de statistiske test. Herefter justerer vi modellen, så vi kun bibeholder sammenhænge mellem variablene, som er nogenlunde stærke. Det vil sige, at vi fjerner pile eller kanter, der er statistisk usikre eller svage. Modelsøgningsprocessen foretages ad flere omgange for at få den model, som forklarer mest muligt af den indbyrdes sammenhæng mellem de udvalgte items og baggrundsvARIABLE. Vi fremstiller resultaterne fra den endelige model i dette notat. Da de statistiske beregninger bygger på en modelsøgningsproces, hvor vi anvender alle tilgængelige data i 2018, så kan resultaterne ikke direkte sammenlignes med resultater præsenteret i tidligere tekniske notater eller rapporter, der omhandler de samme skalaer.

2.2 Statistiske test-of-fit

Alle vores modeller er karakteriseret af DIF, og Tabel 2.2 viser, hvilke baggrundsvARIABLE vi anvender til at analysere DIF. I den følgende gennemgang af de enkelte skalaer præsenterer vi for det første resultaterne fra de statistiske test, der dokumenterer, at der er DIF. For det andet inkluderer vi også de partielle korrelationer eller gamma-værdier, hvor der er kontrolleret for den generelle (fælles) faktor. Vi inkluderer disse korrelationer for at dokumentere, om der er tale om en stærk eller svag DIF-effekt.

Tabel 2.1 BaggrundsvARIABLE anvendt i analysen af DIF

VARIABLE	Kategorier
År	2014; 2015; 2016; 2017; 2018
Køn	Kvinde; Mand
Alder	17-39 år; 40-49 år; 50-84 år
Anciennitet	0-9 år; 10-19 år; 20+ år
Klassetrin	Indskoling; Mellemtrin; Udskoling
Fag	Dansklærer; Dansk- og Matematiklærer; Matematiklærer

Anm.: Kategorierne angiver kodningen af baggrundsvARIABLENE i skalavalideringen. Spørgsmål, der ikke indgår i skalaen: "Samarbejde mellem pædagoger og øvrigt personale", da denne skala beregnes blandt pædagoger.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

⁴ Gamma-koefficienten gør brug af konkordans og diskordans og forklarer forholdet imellem variablene. Konkordans betyder, at der er den samme tendens mellem to variable. Hvis den ene variabels værdi stiger, så gør den anden det også. Ved et diskordant forhold falder værdien eller scoren på en variabel, når den anden variabel stiger. En positiv sammenhæng mellem variablene er givet af konkordansen, da denne optræder hyppigere end diskordante forhold. Omvendt forholder det sig for negative sammenhænge, hvor der er flere diskordante forhold end konkordante. Ved at standardisere forholdet mellem to størrelser er det lettere at fortolke data, og man får derved en værdi for gamma, der falder i mellem -1 og +1. Endvidere angiver tegnet, om der er tale om et positivt eller negativt forhold mellem variablene. Størrelsen på værdien for gamma angiver endvidere, hvor stærk sammenhængen er mellem variablene. En absolut værdi under 0,15 anses for at være svag; en værdi mellem 0,15 og 0,35 anses for at være moderat, mens en værdi over 0,35 anses for at være stærk.

Vi anvender en lang række statistiske test for at undersøge, om skalaerne passer på data. Testene omfatter CLR-test for homogenitet og DIF (Andersen, 1973), beregning af 'item fit statistics' (Kreiner, 2011) samt CLR-test for LD og DIF for specifikke kombinationer af items og variable (Kelderman, 1984).

Når vi foretager mange statistiske test, er der mulighed for at få type-I fejl, dvs. falsk-positive resultater. Vi korrigerer derfor for denne multiple testning ved hjælp af Benjamini-Hochberg-metoder (Benjamini & Hochberg, 1995), som kontrollerer, at raten af falsk-positive resultater ikke overstiger 5 %.

Vi afviser en model, hvis vi har mere end 1 % falsk-positive resultater i de globale test og test, der understøtter både LD og DIF eller fravær af homogenitet mellem variablene. I notatet præsenteres alene de endelige modeller, hvor vi ikke har yderligere specifikke test, der taler imod modellerne.

Det er samtidig vigtigt at skelne mellem signifikans og relevans. Af den årsag ledsages signifikant evidens for LD og DIF altid af gamma-koefficienter, der fortæller, hvor stærk afhængigheden eller DIF-effekten er. Gamma-koefficienter under 0,15 betragtes som regel som svage sammenhænge, mens koefficienter over 0,35 er stærke sammenhænge. Da det ikke giver anledning til beregningsmæssige problemer at inkludere de svage effekter, er disse altid inkluderet i forbindelse med justering af skalaværdierne.

Analyserne er udført ved hjælp af programmet DIGRAM v. 4.01.9 (Kreiner, 2003; Kreiner & Nielsen, 2016).

3 Brug af Fælles Mål

Boksen nedenfor viser de items, der indgår i skalaen for brug af Fælles Mål.

Items i skala for brug af Fælles Mål 2014-2018 (spørgsmål til lærere)

- ✓ Hvor enig eller uenig er du i nedenstående udsagn?
- Jeg nedbryder målbeskrivelsen i Fælles Mål til konkrete mål for eleverne (LPQ96)
 - Jeg gennemfører min undervisning med udgangspunkt i Fælles Mål (LPQ98)
 - Jeg evaluerer elevernes udbytte af undervisningen i forhold til Fælles Mål (LPQ99).

Benyttede svarkategorier: Helt enig, Delvis enig, Hverken enig eller uenig, Delvis uenig, Helt uenig.

Forud for de tre spørgsmål i skalaen er lærerne blevet spurgt om, hvorvidt de bruger eller ikke bruger Fælles Mål. De lærere, som har svaret, at de generelt ikke bruger Fælles Mål, er ikke blevet spurgt til, hvordan de bruger Fælles Mål i henhold til de tre spørgsmål. Værdien på skalaen kan derfor indikere en højere brug af Fælles Mål, end det reelt er tilfældet, idet de, der slet ikke bruger Fælles Mål, ikke indgår i analysen.

Besvarelsenerne fra 2014-2015 er ikke direkte sammenlignelige med besvarelsenerne fra 2016-2018, da lærerne i 2015 blev spurgt om Fælles Mål både fra 2009 og 2014, mens de i 2016-2018 alene er spurgt om deres brug af Fælles Mål 2014.

Note: $n = 10.711$ lærere.

Kilde: Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale.

3.1 Graf

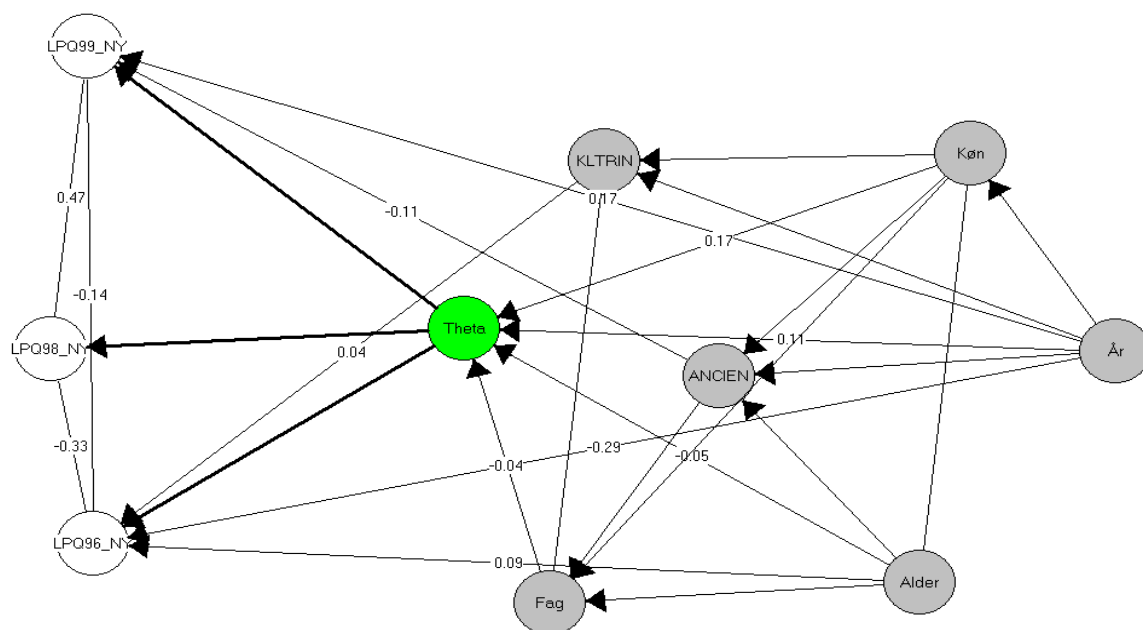
Den grafiske log-lineære Rasch-model viser sammenhængen mellem svarene på de tre spørgsmål, der indgår i Brug af Fælles Mål, samt de udvalgte baggrundsvariable. Figuren bruger kanter eller pile til at angive forhold imellem variable. Variable på samme niveau har ingen pil for enden, da der her ikke antages at være nogen retningsbetingelse i årsagsniveauet. Der er således tale om korrelationskoefficienter. Pile uden retning angiver endvidere LD imellem de items, der definerer skalaen. En LD angives ved pile imellem items, der ikke har nogle retningspile. Pile, der er markeret med fed, indikerer sammenhængen mellem den latente/bagvedliggende faktor for skalaen (theta i figuren) og de tre items, der definerer skalaen.

Figur 3.1 viser, at skalaen Brug af Fælles Mål afhænger af køn, alder, anciennitet, klassetrin og alder. Sammenhængene fremgår af pilene, der går fra baggrundsvariablene direkte til items. Eksempelvis er der en direkte sammenhæng imellem år, LPQ96 og LPQ99. De angivne tal ved pilene i Figur 3.1 indikerer, om der er en positiv sammenhæng imellem baggrundsvariable og items. Et negativt tal mellem to items eller baggrundsvariable eller items og baggrundsvariable indikerer en negativ sammenhæng. Dette er eksempelvis tilfældet for sammenhængen mellem år og item'et LPQ96, hvor der er en gamma-koefficient på -0,29. Figur 3.1 viser også, at der er LD mellem LPQ98 og LPQ99, samt LD mellem LPQ96 og LPQ98. Det fremgår af linjerne mellem de tre items.

Figur 3.1 viser, at item'et LPQ98 ('Jeg gennemfører min undervisning med udgangspunkt i Fælles Mål') ikke har sammenhæng med baggrundsvariablene. Når der er en sammenhæng, som går igennem en mellemliggende baggrundsvariabel, så indikerer det, at item'et fungerer forskelligt i grupper, der er defineret af disse baggrundsvariable. Eksempelvis viser figuren, at LPQ99 fungerer forskelligt

i grupper defineret af anciennitet og alder, hvilket betyder, at nogle grupper af lærere svarer forskelligt fra andre grupper af lærere, som også er defineret ved alder og anciennitet. Tilsvarende gør sig gældende for LPQ96, hvor der er svar, som varierer over baggrundsvariablene år og klasstrin. Figur 3.1 viser, at lærernes svar på de tre items afhænger af alder, anciennitet, køn og klasstrin. Med disse forskelle mellem forbindelserne imellem items og baggrundsvariable er der altså grund til at anvende korrigerede skalaer, som det sker i Jensen, Skov & Thranholm (2018).

Figur 3.1 IRT-graf for GLLRM for brug af Fælles Mål



Anm.: Variable er repræsenteret ved punkter, og variablene antages at være betinget uafhængige givet alle de andre variable, hvis der ikke er nogen linje eller pil, der forbinder dem. "-" præfiks og "_n" suffiks angiver, at variablene er vendt om henholdsvis sammensat ud fra andre variable i spørgeskemaet. Pile antages at referere til sammenhænge, mens kanter uden retning repræsenterer symmetriske relationer uden simpel forklaring. De hvide punkter er items; det grønne punkt repræsenterer den generelle faktor, mens de grå punkter er baggrundsvariable. Ancien = anciennitet; Kltrin = klasstrin.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

3.2 Globale test

En global test er en overordnet test mellem den statistiske model og data og er en test af afvigelser mellem data og den statistiske model. Den viser hermed, hvor godt den foreslåede statistiske model passer på data. Tabel 3.1 viser denne test for homogenitet (CLR-test) samt test for DIF for de items, der indgår i skalaen Brug af Fælles Mål.

Den første række i Tabel 3.1 (scoregrupper) beskriver den globale test for homogenitet, mens de resterende rækker beskriver den anden globale test for DIF. Parameter-estimerne for baggrundsvariablene (klasstrin, fag, anciennitet, køn, alder og år) er beregnet i én samlet model.

Tabel 3.1 Globale test for homogenitet og DIF

	CLR	DF	p-værdi	Delta-værdi
Scoregrupper	150671,3	115	0,000	1506,32
Klassetrin	2943,0	206	0,000	623,46
Fag	246,7	230	0,215	3,41
Anciennitet	185,3	206	0,846	22,60
Køn	146,2	115	0,026	
Alder	260,4	214	0,017	7,50
År	1660,1	332	0,000	522,08

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

Første kolonne viser selve værdien af den globale test (CLR), mens DF viser modellens frihedsgrader, p-værdien viser signifikansniveauet, mens delta-værdien viser, hvor meget usikkerhed der er forbundet med det enkelte CLR-estimat. Generelt gælder det, at jo højere delta-værdi, desto større usikkerhed er der omkring parameter-estimatet. Jo mere kompleks en model er, jo mere usikkerhed er der ofte forbundet med modellen. Delta-værdien er også et udtryk for tilpasning imellem modellen og data. Delta-værdien kan anses for at være den maksimale afstand i mellem modellens fittede og observerede værdier. Da delta-værdien skal være tæt på 0, er de delta-værdier, der er vist for de globale test, udtryk for store usikkerheder for teststørrelsen. Eksempelvis er der en delta-værdi på 1506,32 for den globale test for homogenitet, hvilket må anses for at være langt fra de observerede værdier. Relativt til denne værdi er værdien for fag (3,41) lavere, men stadig ikke tæt på 0.

Tabellen viser, at de globale test indeholder signifikante værdier, som ikke går væk, hvis vi ændrer i modellen. Det ser vi ud fra p-værdien. P-værdierne angiver sandsynligheden for, at resultatet *ikke* er en tilfældighed, og vi anvender her en grænse for p-værdien på 0,05. Det vil sige, at man er mindst 95 % sikker på, at homogenitetstesten ikke er tilfældig. Vi finder lave p-værdier i både den globale test for homogenitet (scoregrupper) og i den globale test for DIF, når det gælder klassetrin, køn, alder og år. Generelt gælder, at jo lavere p-værdi, jo mere sikkert er et resultat. Det er her vigtigt at understrege, at statistisk signifikans ikke er udtryk for en effekts størrelse eller betydning, men blot er et mål for præcision.

Tabellens delta-værdier viser, at testene for homogenitet og yderligere DIF er mindre pålidelige, da modellen var for kompleks til, at parameter-estimerne kunne færdig-estimeres. Modellen accepteres, da test-værdierne understøtter lav grad af LD, DIF eller lav grad af homogenitet.

I analysen af en skala står den ovenstående globale test ikke alene – den efterfølges typisk af test for item fits, DIF og LD, fordi de globale test kan være behæftet med usikkerheder (Kreiner, 2003). Vi beskriver disse test i næste afsnit.

3.3 Item fits

Item fits-test er en test, der vurderer, hvorvidt de observerede værdier for de enkelte items passer overens med Rasch-modellens forventede værdier (for yderligere om denne test henvises til Kreiner, 2011). Tabel 3.2 viser resultaterne af denne item-fit-test. Tabellen sammenligner de observerede og forventede korrelationer mellem de enkelte items og restscoren defineret som summen af de øvrige items, når der er kontrolleret for den fælles faktor, som i dette tilfælde definerer Brug af Fælles Mål.

Sammenligningen sker på baggrund af p- og delta-værdien (sammenligningens usikkerhed). Vi var i stand til at estimere modellen, og der er derfor ikke nogen delta-værdier i tabellen.

Tabel 3.2 Item-fits i form af sammenligninger mellem observerede og forventede korrelationer mellem items og restscore

Item	Observeret	Forventet	p-værdi	Delta
LPQ96	0,654	0,654	0,934	
LPQ98	0,750	0,750	0,985	
LPQ99	0,791	0,790	0,857	

Anm.: Observeret og forventet hentyder til observerede og forventede korrelationer mellem items og restscores. Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

I samtlige tilfælde er der god tilpasning mellem det, der er observeret, og det, der forventes af de enkelte items. Det understøttes af de høje p-værdier, som er udtryk for overensstemmelser imellem de observerede og forventede værdier – og dermed om der er statistisk forskel på størrelserne. I tilfældet med forskelle mellem de observerede og forventede værdier ønsker vi ikke en statistisk signifikant teststørrelse (p-værdi), der er mindre end 0,05, da dette indikerer at forskellene mellem observerede og forventede værdier er store og usikre. Tabellen viser meget høje p-værdier, og vi accepterer derfor modellens antagelse om homogenitet.

3.4 Oversigt over LD og DIF

Den anden test, der ofte foretages i forlængelse af de globale test, er test for LD og DIF for den udvalgte skala. Denne test fremgår af de to følgende tabeller, hvor Tabel 3.3 analyserer LD mellem de udvalgte items, der indgår i skalaen, mens Tabel 3.4 analyserer DIF mellem de udvalgte items og baggrundsvariable.

For begge tabeller gælder, at analysen består af to dele. Den første del viser, hvorvidt CLR-testen er signifikant. Den anden del viser den statistiske sammenhæng mellem de enkelte variabler og angives ved hjælp af de partielle gamma-værdier. I beregningen af de partielle gammaværdier er der både taget højde for de øvrige variabler og den forventede fælles (latente) faktor. Eksempelvis viser første række i Tabel 3.3 en statistisk sammenhæng på 0,269 mellem de to spørgsmål LPQ96 og LPQ98, når vi samtidig tager højde for lærernes anciennitet, klassetrin alder, køn, undersøgelsesår samt en fælles bagvedliggende faktor for deres brug af Fælles Mål. Det vil sige at gamma-værdien udtrykker den statistiske sammenhæng man kan forvente blandt lærere med samme anciennitet, alder og køn og som i lige så høj grad anvender Fælles Mål.

Tabel 3.3 Oversigt over resultater, der dokumenterer LD

Item 1	Item 2	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ96	LPQ98	178,42	16	0,001	0,269	
LPQ96	LPQ99	326,66	16	0,001	0,451	
LPQ98	LPQ99	657,98	16	0,001	0,736	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

Tabellen viser, at der er LD mellem alle tre items, idet p-værdierne alle er under 0,05. Beregningen af de partielle gamma-korrelationer afslører, at LD mellem LPQ96 og LPQ98 er moderat (partiell gamma = 0,269), og at LD mellem LPQ96 og LPQ99 (partiell gamma = 0,451) samt mellem LPQ98 og LPQ99 (partiell gamma = 0,736) er stærk.

Skalaernes pålidelighed svækkes af LD. Selvom det sker, styrkes skalaernes pålidelighed af det forholdsvis store antal respondenter ($n = 10.711$), der ligger bag skalaberegningen.

Tabel 3.4 viser evidensen for DIF. Her er der tale om, at DIF er højsignifikant ($p < 0,01$) i forhold til alle items og baggrundsvARIABLE. Som vist af de partielle gamma-korrelationer, er DIF svage (dvs. mindre end $\pm 0,15$) til moderate (dvs. større end $\pm 0,15$, men mindre end $\pm 0,35$). Eksempelvis er der en statistisk signifikant DIF mellem LPQ96 og Klassesetrin ($p < 0,001$), hvor der er en gamma-værdi på 0,021. Der er også en statistisk signifikant DIF mellem LPQ96 og År ($p < 0,001$), hvor der er en gamma-værdi på -0,21 og en tilsvarende usikkerhed om estimatet med en delta-værdi på 437,9.

Tabel 3.4 Oversigt over resultater, der dokumenterer DIF

Item	Baggrund	CLR	DF	p-værdi	Partiell gamma	Delta-værdi
LPQ96	Klassesetrin	33,44	8	0,001	0,021	
LPQ96	Alder	56,26	8	0,001	0,028	
LPQ96	År	751,87	16	0,001	-0,210	437,9
LPQ99	Anciennitet	21,49	8	0,006	-0,015	
LPQ99	År	103,64	16	0,000	0,049	205,6

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

Overordnet set peger de partielle gamma-værdier på, at DIF har betydning for skalaen, men at betydningen af DIF er svag. Vi tager dog højde for alle DIF-relationer i forbindelse med beregningen af de justerede skalaværdier, hvorfor vi korrigerer for disse skævheder, når vi beregner person-estimerne, som vi anvender til de videre analyser, der præsenteres i Jensen, Skov & Thranholm (2018).

4 Åben skole

Boksen nedenfor viser de items, der indgår i skalaen for Åben skole.

Items i skala for Åben skole 2014-2018 (spørgsmål til lærere)

- ✓ Hvor ofte har du besøg af en voksen udefra i din [dansk/matematik] undervisning i [Klassebetegnelse1] (fx medarbejder fra lokal virksomhed, gymnasium, forening eller lignende)? (LPQ90)
- ✓ Hvor ofte tager du eleverne med på besøg uden for skolen (fx på museum, bondegård eller virksomhed)? (LPQ91)
- ✓ Hvor ofte finder din [dansk/matematik] undervisning i [klassebetegnelse1] sted uden for skolen (i sportsklubben, naturen, byen eller andet)? (LPQ93)

Benyttede svarkategorier: Hver dag, 2-4 gange om ugen, 1 gang om ugen, 2-3 gange om måneden, 1 gang om måneden, Hver anden måned, 1-5 gange på 1 år, Aldrig.

Svarene på disse items er kategoriseret på følgende måde:

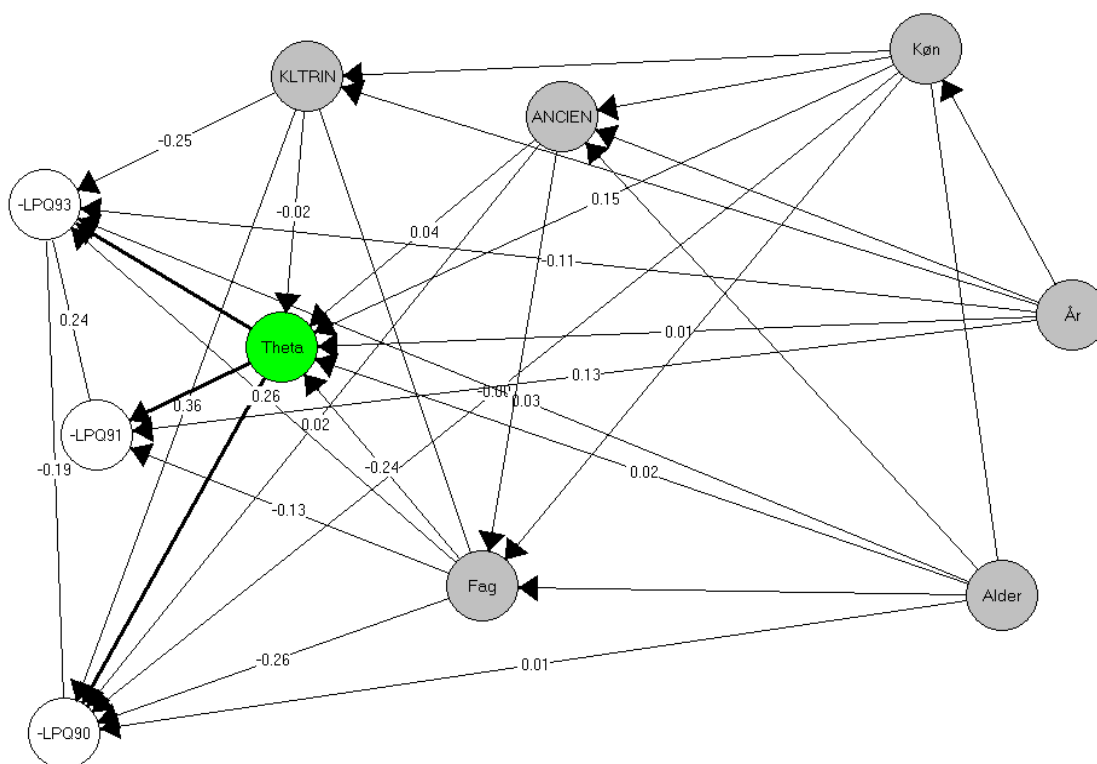
1. Hver dag, 2-4 gange om ugen, 1 gang om ugen
2. 2-3 gange om måneden, 1 gang om måneden
3. Hver anden måned, 1-5 gange på 1 år
4. Aldrig.

Note: n = 10.041 lærere.

Kilde: Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale.

4.1 Graf

Figur 4.1 IRT-graf for GLLRM for Åben skole



Anm.: Variable er repræsenteret ved punkter og antages at være betinget uafhængige, givet alle andre variable, hvis der ikke er en linje eller pil, der forbinder dem. "-" præfiks og "_n" suffiks angiver, at variablene er vendt om henholdsvis sammensat ud fra andre variable i spørgeskemaet. Pile antages at referere til sammenhænge, mens kanter uden retning repræsenterer symmetriske relationer uden simpel forklaring. De hvide punkter er items; det grønne punkt repræsenterer den generelle faktor, og de grå punkter er baggrundsvariable. Ancienni = anciennitet; Klassetrin = klassetrin.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

4.2 Globale test

Tabel 4.1 Globale test for homogenitet og DIF

	CLR	DF	p-værdi	Delta-værdier
Scoregrupper	250,4	268	0,773	291,415
Klassetrin	554,5	456	0,001	
Fag	602,1	416	0,000	4,021
Anciennitet	607,7	494	0,000	3,653
Køn	353,7	261	0,000	6,906
Alder	567,2	480	0,004	
År	907,2	864	0,150	7,736

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

4.3 Item fits

Tabel 4.2 Item fits i form af sammenligninger mellem observerede og forventede korrelationer mellem items og restscorer

Item	Observeret	Forventet	p-værdi	Delta-værdi
LPQ90	0,362	0,349	0,297	
LPQ91	0,551	0,546	0,647	
LPQ93	0,472	0,488	0,162	

Anm.: Observeret og forventet hentyder til observerede og forventede korrelationer mellem items og restscorer. Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

4.4 Oversigt over LD og DIF

Tabel 4.3 Oversigt over resultater, der dokumenterer LD

Item 1	Item 2	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ90	LPQ93	264,45	49	0,000	0,235	
LPQ91	LPQ93	1046,57	49	0,000	0,582	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

Tabel 4.4 Oversigt over resultater, der dokumenterer DIF

Item	Baggrund	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ90	Klassetrin	326,41	14	0,000	-0,226	
LPQ90	Fag	304,63	14	0,000	0,271	
LPQ90	Anciennitet	33,95	14	0,002	-0,042	
LPQ90	Køn	19,85	7	0,006	-0,095	
LPQ90	Alder	33,08	14	0,003	-0,026	
LPQ91	Fag	156,09	14	0,000	0,239	
LPQ91	År	86,46	28	0,000	-0,090	
LPQ93	Klassetrin	90,13	14	0,000	0,208	
LPQ93	Fag	57,32	14	0,000	-0,042	
LPQ93	Alder	37,53	14	0,001	-0,037	
LPQ93	År	197,89	28	0,000	0,111	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

5 Brug af it

Boksen nedenfor viser de items, der indgår i skalaen for Brug af it.

Items i skala for Brug af it 2014-2018 (spørgsmål til lærere)

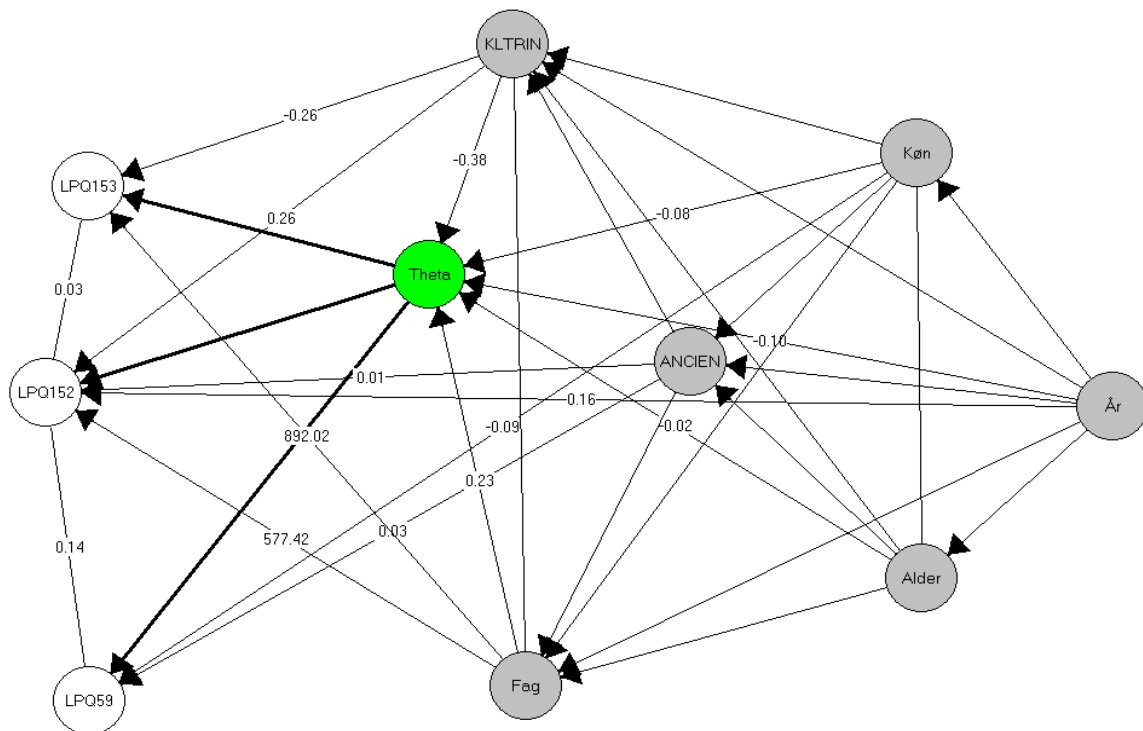
- ✓ Hvor stor vægt lægger du på nedenstående opgaver i dit daglige arbejde? (benyttede svarkategorier: Meget stor vægt, Stor vægt, Nogen vægt, Lille vægt, Ingen vægt)
 - Bruge it til at understøtte undervisningen (LPQ59).
- ✓ Hvor ofte bruger du apps, digitale bøger eller andre digitale læremidler i din [dansk/matematik] undervisning i [klassebetegnelse1]? (benyttede svarkategorier: Hver lektion, Hver dag, 2-4 gange om ugen, 1 gang om ugen, 2-3 gange om måneden, 1 gang om måneden, Hver anden måned, 1-5 gange på et år, Aldrig) (LPQ152).
- ✓ Hvor ofte tilrettelægger du [dansk/matematik] undervisningen i [klassebetegnelse1], så eleverne skal bruge internettet til at søge informationer? (benyttede svarkategorier: Hver lektion, Hver dag, 2-4 gange om ugen, 1 gang om ugen, 2-3 gange om måneden, 1 gang om måneden, Hver anden måned, 1-5 gange på et år, Aldrig) (LPQ153).

Note: n = 8.771 lærere.

Kilde: Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale.

5.1 Graf

Figur 5.1 IRT graf for GLLRM for brug af IT



Anm.: Variable er repræsenteret ved punkter og antages at være betinget uafhængige givet alle de andre variable, hvis der ikke er nogen linje eller pil, der forbinder dem. "-" præfiks og "_n" suffiks angiver, at variablene er vendt om henholdsvis sammensat ud fra andre variable i spørgeskemaet. Pile antages at referere til sammenhænge, mens kanter uden retning repræsenterer symmetriske relationer uden simpel forklaring. De hvide punkter er items; det grønne punkt repræsenterer den generelle faktor, mens de grå punkter er baggrundsvariable. Ancienni = anciennitet; Klasseset = klasseset.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

5.2 Globale test

Tabel 5.1 Globale test for homogenitet og DIF

	CLR	DF	p-værdi	Delta-værdi
Scoregrupper	167,0	202	0,966	
Klasseset	508,8	332	0,000	
Fag	465,2	320	0,000	2,273
Anciennitet	381,9	332	0,031	3,082
Køn	213,1	198	0,219	
Alder	442,6	404	0,090	
År	673,1	696	0,726	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

5.3 Item fits

Tabel 5.2 Item fits i form af sammenligninger mellem observerede og forventede korrelationer mellem items og restscorer

Item	Observeret	Forventet	p-værdi	Delta-værdi
LPQ59	0,482	0,481	0,892	
LPQ152	0,563	0,561	0,766	
LPQ153	0,506	0,504	0,897	

Anm.: Observeret og forventet hentyder til observerede og forventede korrelationer mellem items og restscorer. Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

5.4 Oversigt over LD og DIF

Tabel 5.3 Oversigt over resultater, der dokumenterer LD

Item 1	Item 2	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ59	LPQ152	208,67	28	0,000	0,399	
LPQ152	LPQ153	942,26	49	0,000	0,484	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

Tabel 5.4 Oversigt over resultater, der dokumenterer DIF

Item	Baggrund	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ59	Anciennitet	18,96	8	0,015	0,077	
LPQ59	Køn	12,76	4	0,013	-0,001	
LPQ152	Klassetrin	189,44	14	0,000	0,087	
LPQ152	Fag	108,61	14	0,000	0,000	
LPQ152	Anciennitet	35,30	14	0,001	-0,033	
LPQ152	År	96,90	28	0,000	-0,131	
LPQ153	Klassetrin	94,54	14	0,000	-0,468	
LPQ153	Fag	550,24	14	0,000	0,000	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

6 Fokus på faglighed

Boksen nedenfor viser de items, der indgår i skalaen for Fokus på faglighed.

Items i skala for Fokus på faglighed 2014-2018 (spørgsmål til lærere)

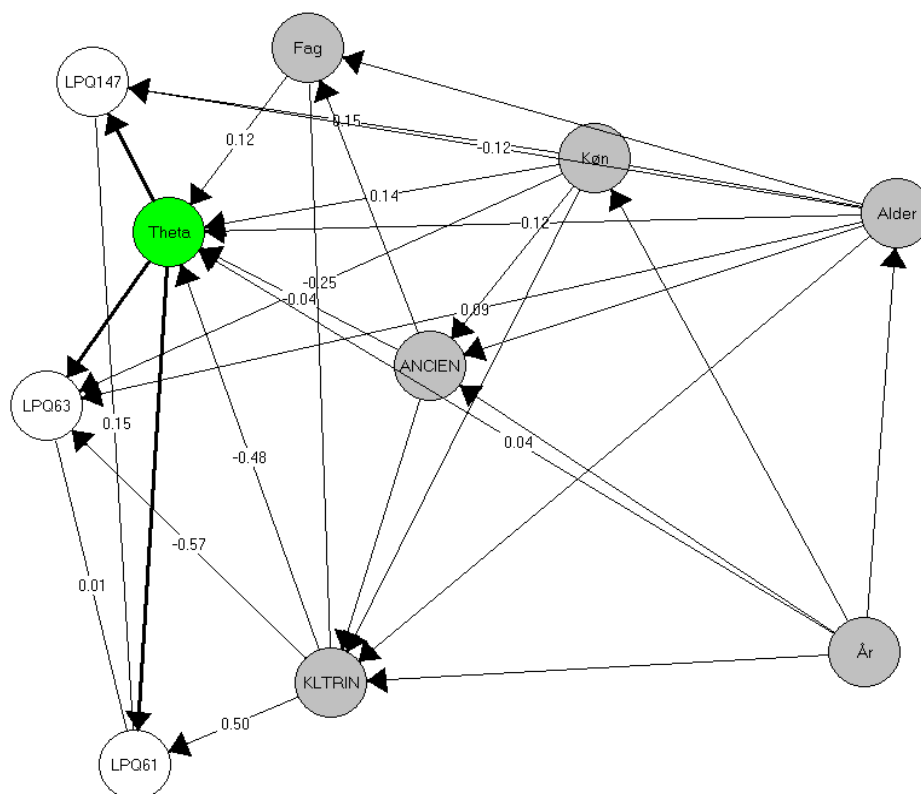
- ✓ Hvor stor vægt lægger du på nedenstående opgaver i dit daglige arbejde? (benyttede svarkategorier: Meget stor vægt, Stor vægt, Nogen vægt, Lille vægt, Ingen vægt)
 - At udvikle elevernes faglige viden og færdigheder (LPQ61)
 - At forberede og motivere eleverne til en ungdomsuddannelse (LPQ63).
- ✓ Hvordan vil du beskrive din undervisning i [dansk/matematik] i [klassebetegnelse 1] på en skala fra 1 (helt enig med A) til 5 (helt enig med B), og hvor 3 er neutral? (benyttede svarkategorier A: Jeg giver udtryk for høje forventninger til mine elevers faglige præstationer/B: Jeg giver ikke udtryk for høje forventninger til mine elevers faglige præstationer) (LPQ147).

Note: n = 9.762 lærere.

Kilde: Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale.

6.1 Graf

Figur 6.1 IRT graf for GLLRM for Fokus på faglighed



Anm.: Variable er repræsenteret ved punkter og antages at være betinget uafhængige givet alle de andre variable, hvis der ikke er nogen linje eller pil, der forbinder dem. "-" præfiks og "_n" suffiks angiver, at variablene er vendt om henholdsvis sammensat ud fra andre variable i spørgeskemaet. Pile antages at referere til sammenhænge, mens kanter uden retning repræsenterer symmetriske relationer uden simpel forklaring. De hvide punkter er items; det grønne punkt repræsenterer den generelle faktor, mens de grå punkter er baggrundsvariable. Ancienni = anciennitet; Klassestr = klassestrin.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

6.2 Globale test

Tabel 6.1 Globale test for homogenitet og DIF

	CLR	DF	p-værdi	Delta-værdi
Scoregrupper	33,4	67	1,000	150,202
Klassestrin	104,5	96	0,260	
Fag	169,6	134	0,020	
Anciennitet	98,3	134	0,991	44,805
Køn	82,1	59	0,025	
Alder	86,3	102	0,867	
År	299,9	268	0,088	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

6.3 Item fits

Tabel 6.2 Item fits i form af sammenligninger mellem observerede og forventede korrelationer mellem items og restscorer

Item	Observeret	Forventet	p-værdi	Delta-værdi
LPQ61	0,424	0,426	0,869	
LPQ63	0,334	0,330	0,727	
LPQ147	0,326	0,320	0,603	

Anm.: Observeret og forventet hentyder til observerede og forventede korrelationer mellem items og restscorer. Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

6.4 Oversigt over LD og DIF

Tabel 6.3 Oversigt over resultater, der dokumenterer LD

Item 1	Item 2	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ61	LPQ63	222,40	12	0,000	0,411	
LPQ61	LPQ147	98,30	12	0,000	0,373	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

Tabel 6.4 Oversigt over resultater, der dokumenterer DIF

Item	Baggrund	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ61	Klassetrin	67,33	6	0,000	0.201	
LPQ63	Klassetrin	977,40	8	0,000	-0.625	
LPQ63	Køn	20,15	4	0,001	0,000	
LPQ63	Alder	32,15	8	0,001	-0.109	
LPQ147	Køn	23,96	4	0,001	0,000	
LPQ147	Alder	28,55	8	0,001	-0.106	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

7 Brug af test og elevplaner

Boksen nedenfor viser de items, der indgår i skalaen for Brug af test og elevplaner.

Items i skala for Brug af test og elevplaner 2014-2018 (spørgsmål til lærere)

✓ I hvor høj grad er du enig i følgende udsagn?

- Jeg anvender resultater fra test (fx nationale test) til vurdering af den enkelte elev (LPQ141)
- Jeg oplever, at elevplanen er et redskab, der understøtter elevens læring (LPQ142)
- Jeg oplever, at elevplanen er et godt udgangspunkt for dialog med forældre om elevens faglige udvikling (LPQ143).

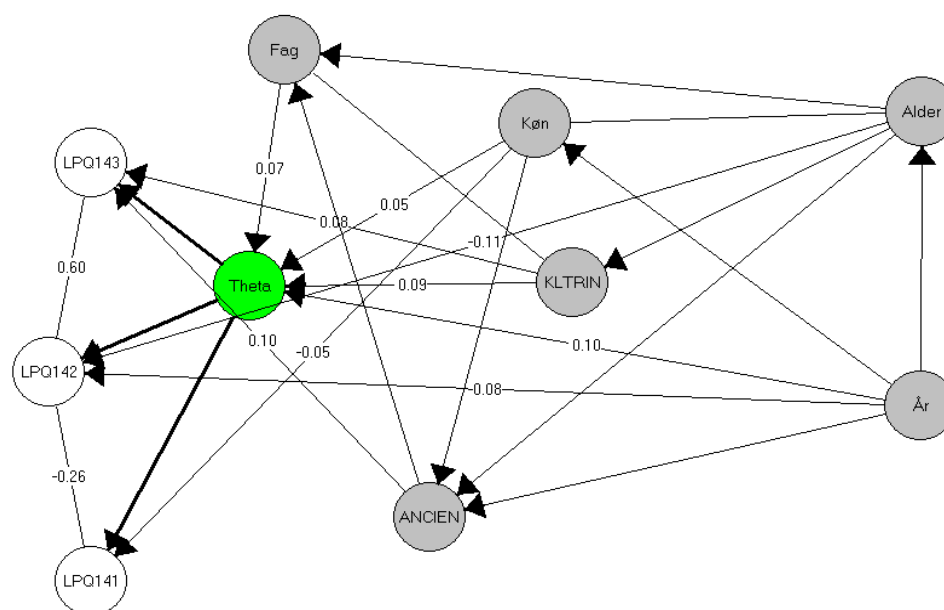
Benyttede svarkategorier: I meget høj grad, i høj grad, i nogen grad, i lav grad, i meget lav grad, slet ikke.

Note: n = 8.773 lærere.

Kilde: Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale.

7.1 Graf

Figur 7.1 IRT graf for GLLRM for test og elevplaner



Anm.: Variable er repræsenteret ved punkter og antages at være betinget uafhængige givet alle de andre variable, hvis der ikke er nogen linje eller pil, der forbinder dem. "-" præfiks og "_n" suffiks angiver, at variablene er vendt om henholdsvis sammensat ud fra andre variable i spørgeskemaet. Pile antages at referere til sammenhænge, mens kanter uden retning repræsenterer symmetriske relationer uden simpel forklaring. De hvide punkter er items, det grønne punkt repræsenterer den generelle faktor, mens de grå punkter er baggrundsvARIABLE. Ancienni = anciennitet; Klassestr = klassestrin.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

7.2 Globale test

Tabel 7.1 Globale test for homogenitet og DIF

	CLR	DF	p-værdi	Delta-værdi
Scoregrupper	112,4	117	0,603	
Klassetrin	214,6	204	0,292	
Fag	334,6	234	0,000	
Anciennitet	227,1	204	0,128	
Køn	131,6	112	0,100	
Alder	284,2	214	0,001	
År	389,1	388	0,475	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

7.3 Item fits

Tabel 7.2 Item fits i form af sammenligninger mellem observerede og forventede korrelationer mellem items og restscorer

Item	Observeret	Forventet	p-værdi	Delt-værdi
LPQ141	0,340	0,339	0,891	
LPQ142	0,709	0,703	0,434	
LPQ143	0,668	0,666	0,790	

Anm.: Observeret og forventet hentyder til observerede og forventede korrelationer mellem items og restscorer. Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

7.4 Oversigt over LD og DIF

Tabel 7.3 Oversigt over resultater, der dokumenterer LD

Item 1	Item 2	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ141	LPQ142	478,51	25	0,000	0,000	
LPQ142	LPQ143	3727,64	25	0,000	0,797	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

Tabel 7.4 Oversigt over resultater, der dokumenterer DIF

Item	Baggrund	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ141	Køn	16,50	5	0,006	0,000	
LPQ142	Alder	40,63	10	0,000	0,000	
LPQ142	År	47,04	20	0,001	-0,018	
LPQ143	Klassetrin	43,41	10	0,049	0,068	
LPQ143	Anciennitet	47,82	10	0,000	0,000	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat. Modellen forværres og kompliceres af at fjerne DIF mellem LPQ143 og anciennitet, selvom relationen ikke er signifikant, når man tager højde for multipel testning. Derfor beholdes relationen i modellen.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

8 Brug af feedbacksamtaler og elevvurdering

Boksen nedenfor viser de items, der indgår i skalaen for Brug af feedbacksamtaler og elevvurdering.

Items i skala for Brug af feedbacksamtaler og elevvurdering 2014-2018 (spørgsmål til lærere)

- ✓ Hvor enig eller uenig er du i nedenstående udsagn? (benyttede svarkategorier: I meget høj grad, I høj grad, I nogen grad, I lav grad, I meget lav grad, Slet ikke)
 - Jeg har jævnligt samtaler med hver enkelt elev, om de når deres læringsmål (LPQ140).
- ✓ Hvor ofte vurderer du, hvor den enkelte elev befinder sig i forhold til læringsmålene? (benyttede svarkategorier: Hver dag, 2-4 gange om ugen, 1 gang om ugen, 2-3 gange om måneden, 1 gang om måneden, Hver anden måned, 1-5 gange på et år, Aldrig) (LPQ159).
- ✓ Hvor ofte har du samtaler med hver enkelt elev om elevens udbytte af [dansk/matematik] undervisningen? (benyttede svarkategorier: Hver dag, 2-4 gange om ugen, 1 gang om ugen, 2-3 gange om måneden, 1 gang om måneden, Hver anden måned, 1-5 gange på et år, Aldrig) (LPQ161).

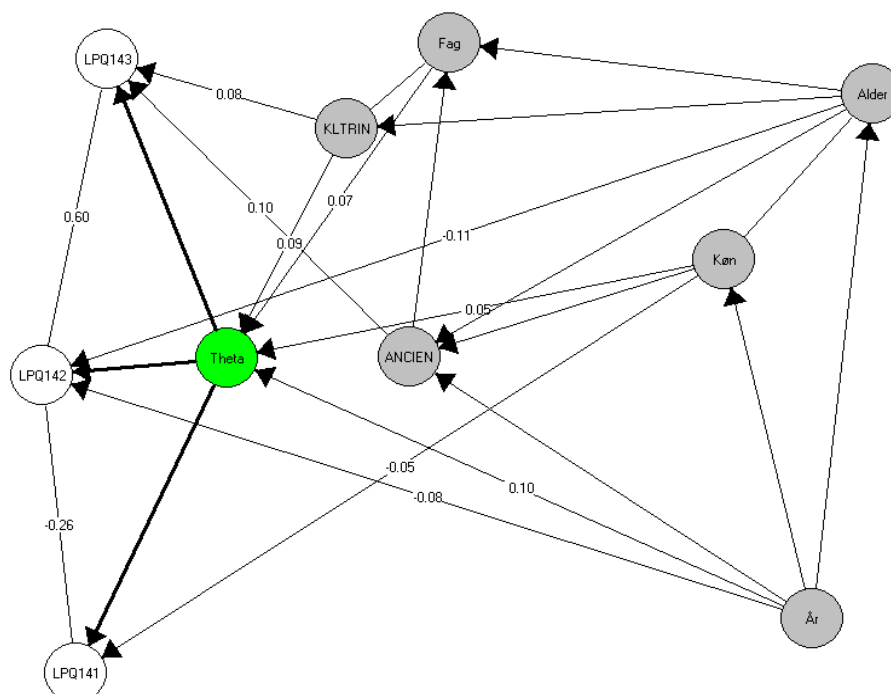
Note: n = 10.292 lærere.

OBS: Skalaen er ikke medtaget i rapporten om den længere og mere varierede skoledag i årene 2014-2018 (Jensen, Skov & Thranholm (2018)) på grund af manglende konceptuel overensstemmelse mellem items for skalaen. Der er i stedet rapporteret på enkeltitems.

Kilde: Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale.

8.1 Graf

Figur 8.1 IRT graf for GLLRM for feedbacksamtaler og elevvurdering



Anm.: Variable er repræsenteret ved punkter og antages at være betinget uafhængige givet alle de andre variable, hvis der ikke er nogen linje eller pil, der forbinder dem. "LPQ" præfiks og "_n" suffiks angiver, at variablene er vendt om henholdsvis sammensat ud fra andre variable i spørgeskemaet. Pile antages at referere til sammenhænge, mens kanter uden retning repræsenterer symmetriske relationer uden simpel forklaring. De hvide punkter er items, det grønne punkt repræsenterer den generelle faktor, mens de grå punkter er baggrundvariable. Ancienni = anciennitet; Klassesetrin = klassesetrin.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

8.2 Globale test

Tabel 8.1 Globale test for homogenitet og DIF

	CLR	DF	p-værdi	Delta-værdi
Scoregrupper	112,4	117	0,603	
Klassesetrin	214,6	204	0,292	
Fag	334,6	234	0,000	
Anciennitet	227,1	204	0,128	
Køn	131,6	112	0,100	
Alder	284,2	214	0,001	
År	389,1	388	0,475	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

8.3 Item fits

Tabel 8.2 Item fits i form af sammenligninger mellem observerede og forventede korrelationer mellem items og restscorer

Item	Observeret	Forventet	p-værdi	Delta-værdi
LPQ140	0,340	0,339	0,891	
LPQ159	0,709	0,703	0,434	
LPQ161	0,668	0,666	0,790	

Anm.: Observeret og forventet hentyder til observerede og forventede korrelationer mellem items og restscorer. Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

8.4 Oversigt over LD og DIF

Tabel 8.3 Oversigt over resultater, der dokumenterer LD

Item 1	Item 2	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ140	LQP159	478,51	25	0,000	0,453	
LPQ159	LPQ161	3727,64	25	0,000	0,485	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

Tabel 8.4 Oversigt over resultater, der dokumenterer DIF

Item	Baggrund	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ140	Køn	16,50	5	0,006	0,094	
LPQ159	Alder	40,63	10	0,000	-0,042	
LPQ159	År	47,04	20	0,001	-0,013	
LPQ161	Klassetrin	43,41	10	0,000	0,070	
LPQ161	Anciennitet	47,82	10	0,000	0,068	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

9 Anvendelse af ekstra støtte i undervisningen

Boksen nedenfor viser de items, der indgår i skalaen for Anvendelse af ekstra støtte i undervisningen.

Items i skala for Anvendelse af ekstra støtte i undervisningen, 2014-2018 (spørgsmål til lærere)

✓ Har der i [klassens] seneste 4 ugers sammenhængende undervisningsperiode været tilknyttet flere lærere/pædagoger/anden ressourceperson i nogen af eller i alle timerne i [dansk/matematik]?

- Støtte til hele klassen fra en læreruddannet (tolærerordning) (LPQ166)
- Støtte til hele klassen fra en ikke-læreruddannet (pædagog eller anden ressourceperson) (LPQ167)
- Støtte til en enkelt elev eller få elever med særlige behov (støtteperson) (LPQ168).

Benyttede svarkategorier: Angiv det samlede antal lektioner inden for de seneste 4 undervisningsuger. Dette tal har vi efterfølgende delt med fire.

Svarene på disse items er kategoriseret på følgende måde: 0-4 timer, 5-9 timer, 10-19 timer, 20-29 timer og 30-160 timer.

De anvendte spørgsmål er stillet til de lærere, hvis klasse inden for den seneste 4 ugers sammenhængende undervisningsperiode har været tilknyttet flere lærere, pædagoger eller andre ressourcepersoner i nogle eller alle af timerne i dansk eller matematik (LPQ163). Såfremt lærerne svarer 'Nej' til at have været tilknyttet støtte, er de placeret i kategorien med 0-4 timers støtte for hver af de tre støttemuligheder. Derfor er spørgsmålene angivet med "_n" i dette afsnit.

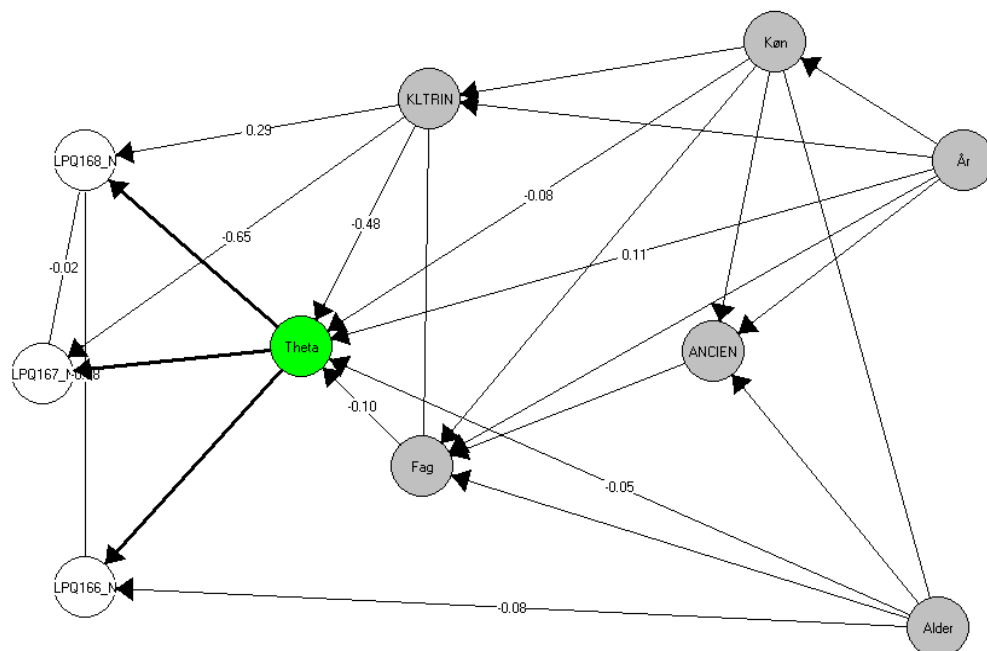
Note: n = 9.190 lærere.

OBS: Skalaen er ikke medtaget i rapporten om den længere og mere varierede skoledag i årene 2014-2018 (Jensen, Skov & Thranholm (2018)) på grund af manglende konceptuel overensstemmelse mellem items for skalaen. Der er i stedet rapporteret på enkeltitems.

Kilde: Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale.

9.1 Graf

Figur 9.1 IRT graf for GLLRM for anvendelse af ekstra støtte i undervisningen



Anm.: Variable er repræsenteret ved punkter og antages at være betinget uafhængige givet alle de andre variable, hvis der ikke er nogen linje eller pil, der forbinder dem. "-" præfiks og "_n" suffiks angiver, at variablene er vendt om henholdsvis sammensat ud fra andre variable i spørgeskemaet. Pile antages at referere til sammenhænge, mens kanter uden retning repræsenterer symmetriske relationer uden simpel forklaring. De hvide punkter er items; det grønne punkt repræsenterer den generelle faktor, mens de grå punkter er baggrundsvariable. Ancienni = anciennitet; Klassetr = klassetrin.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

9.2 Globale test

Tabel 9.1 Globale test for homogenitet og DIF

	CLR	DF	p-værdi	Delta-værdi
Scoregrupper	10,2	67	1,000	27,369
Klassetrin	99,9	86	0,145	
Fag	75,5	67	0,223	
Anciennitet	136,2	134	0,431	
Køn	99,6	67	0,006	
Alder	105,1	118	0,797	
År	222,8	268	0,980	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

9.3 Item fits

Tabel 9.2 Item fits i form af sammenligninger mellem observerede og forventede korrelationer mellem items og restscorer

Item	Observeret	Forventet	p-værdi	Delta-værdi
LPQ166_n	0,696	0,695	0,966	
LPQ167_n	0,696	0,697	0,997	
LPQ168_n	0,665	0,665	0,986	

Anm.: Observeret og forventet hentyder til observerede og forventede korrelationer mellem items og restscorer. Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

9.4 Oversigt over LD og DIF

Tabel 9.3 Oversigt over resultater, der dokumenterer LD

Item 1	Item 2	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ166_n	LPQ167_n	55,33	16	0,000	0,553	
LPQ167_n	LPQ168_n	66,18	16	0,000	0,362	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

Tabel 9.4 Oversigt over resultater, der dokumenterer DIF

Item	Baggrund	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ166_n	Alder	23,25	8	0,003	0,000	
LPQ167_n	Klassetrin	263,85	8	0,000	-0,776	
LPQ168_n	Klassetrin	38,66	8	0,000	-0,317	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

10 Samarbejde mellem pædagoger og øvrigt personale

Boksen nedenfor viser de items, der indgår i skalaen for Samarbejde mellem pædagoger og øvrigt personale.

Items i skala for Samarbejde mellem pædagoger og øvrigt personale 2014-2018 (spørgsmål til pædagoger)

✓ Hvor enig er du i nedenstående udsagn?

- Jeg drøfter jævnligt elevernes udbytte af undervisningen med mine kolleger (LPQ127)
- Jeg observerer af og til undervisningen (LPQ129)
- Vi (klassens team) gennemgår sammen klassens resultater i test og prøver (LPQ131).

Benyttede svarkategorier: I meget høj grad, i høj grad, i nogen grad, i lav grad, i meget lav grad, slet ikke.

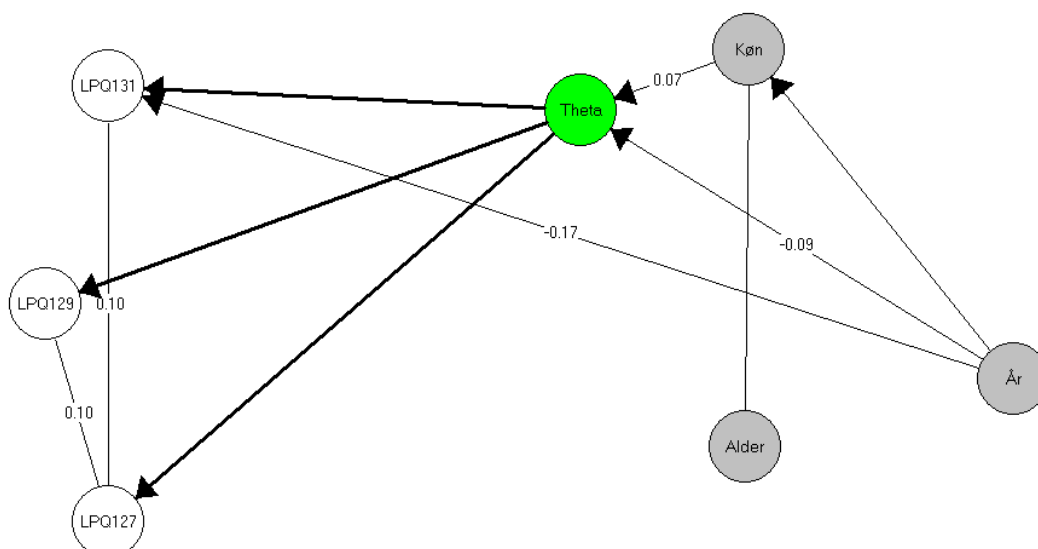
Besvarelserne fra 2014-2015 er ikke direkte sammenlignelige med besvarelserne fra 2016-2018, da spørgsmålsformuleringen er ændret på LPQ131. I 2016 er formuleringen ændret fra 'Vi gennemgår sammen klassens resultater i test og prøver'.

Note: n = 1.910 lærere.

Kilde: Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale.

10.1 Graf

Figur 10.1 IRT graf for GLLRM for samarbejde mellem pædagoger og øvrigt personale



Anm.: Variable er repræsenteret ved punkter og antages at være betinget uafhængige givet alle de andre variable, hvis der ikke er nogen linje eller pil, der forbinder dem. "-" præfiks og "_n" suffiks angiver, at variablene er vendt om henholdsvis sammensat ud fra andre variable i spørgeskemaet. Pile antages at referere til sammenhænge, mens kanter uden retning repræsenterer symmetriske relationer uden simpel forklaring. De hvide punkter er items; det grønne punkt repræsenterer den generelle faktor, mens de grå punkter er baggrundsfaktorer.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

10.2 Globale test

Tabel 10.1 Globale test for homogenitet og DIF

	CLR	DF	p-værdi	Delta-værdi
Scoregrupper	71,2	81	0,775	0,324
Køn	78,1	81	0,570	12,510
Alder	155,5	162	0,629	0,040
År	228,6	224	0,403	0,082

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

10.3 Item fits

Tabel 10.2 Item fits i form af sammenligninger mellem observerede og forventede korrelationer mellem items og restscorer

Item	Observeret	Forventet	p-værdi	Delta-værdi
LPQ127	0,557	0,543	0,456	
LPQ129	0,448	0,449	0,965	
LPQ131	0,497	0,491	0,716	

Anm.: Observeret og forventet hentyder til observerede og forventede korrelationer mellem items og restscorer. Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

10.4 Oversigt over LD og DIF

Tabel 10.3 Oversigt over resultater, der dokumenterer LD

Item 1	Item 2	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ127	LPQ129	175,68	25	0,000	0,362	
LPQ127	LPQ131	91,76	25	0,000	0,257	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

Tabel 10.4 Oversigt over resultater, der dokumenterer DIF

Item	Baggrund	CLR	DF	p-værdi	Partiel gamma	Delta-værdi
LPQ131	År	59,80	20	0,000	0,087	

Anm.: Delta angiver, at parameterestimatet ikke kunne færdigestimeres. Højere delta værdi betyder mere usikkert resultat.

Kilde: Undervisningsministeriets dataindsamling til evaluering af folkeskolereformen 2014-2018. Spørgeskema til lærere, børnehaveklasseledere og andet pædagogisk personale. Beregninger foretaget af VIVE.

Litteratur

- Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika* 38(1): 123-140.
- Benjamini, Y. & Hochberg, Y. (1995) Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1): 289-300.
- Jacobsen, R. H. & Kreiner, S. (2015). *En længere og mere varieret skoledag – Metodenotat for indeksberegninger*. København: KORA – Det Nationale Institut for Kommuners og Regioners Analyse og Forskning.
- Jensen, V.M., Skov, P. R. & Thranholm, E. (2018). *Lærere og pædagogers oplevelse af den længere og mere varierede skoledag i folkeskolereformens fjerde år – Kortlægning 2018*. København: VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.
- Kelderman, H. (1984). Loglinear Rasch model tests. *Psychometrika* 49(2): 223-245.
- Kreiner, S. (2003). *Introduction to DIGRAM*. København: Københavns Universitet, Biostatistisk afdeling.
- Kreiner, S. (2007). Validity and objectivity: Reflections on the role and nature of Rasch models. *Nordic Psychology* 59(3): 268-298.
- Kreiner, S. (2011). A note on item-restscore association in Rasch models. *Applied Psychological Measurement* 35(7): 557-561.
- Kreiner, S. & Christensen, K. B. (2002). Graphical Rasch Models. In Mesbah, M. et al. (Ed.), *Statistical Methods for Quality of Life Studies. Design, Measurement and Analysis* (pp.169-184). Dordrecht: Kluwer Academic Publishers.
- Kreiner, S. & Christensen, K. B. (2004). Analysis of local dependency and multidimensionality in graphical loglinear Rasch models, *Communications in Statistics* 33(6): 1239-1276.
- Kreiner, S. & Christensen, K. B. (2007). Validity and Objectivity in health-related Scales: Analysis by Graphical Loglinear Rasch models. In von Davier, M & Carstensen, C. H. (Ed.), *Multivariate and Mixture Distribution Rasch Models* (pp. 329-346). New York: Springer.
- Kreiner, S. & Christensen, K. B. (2011). Item Screening in Graphical Loglinear Rasch models, *Psychometrika* 76(2): 228-256.
- Kreiner, S. & Nielsen, T. (2016). *Item Analysis in DIGRAM 3.42. Part 1: Guided Tours*. København: Københavns Universitet, Biostatistisk afdeling.



DET NATIONALE FORSKNINGS-
OG ANALYSECENTER FOR VELFÆRD