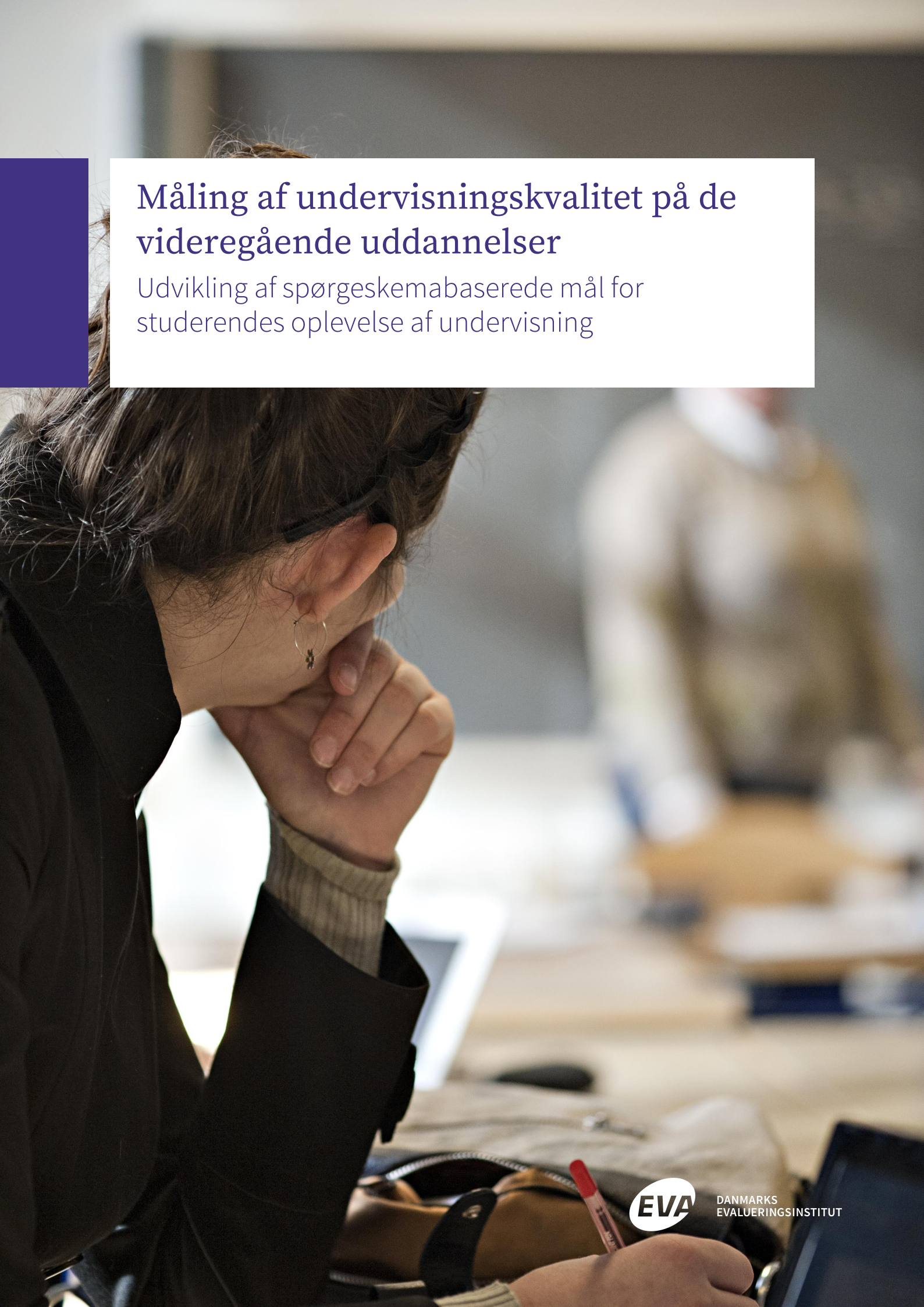
Måling af undervisningskvalitet på de videregående uddannelser

Udvikling af spørgeskemabaserede mål for studerendes oplevelse af undervisning



DANMARKS EVALUERINGSINSTITUT

INDHOLD

Måling af undervisningskvalitet på de videregående uddannelser

1	Resumé	4
2	Indledning	7
2.1	Formål og fokus	7
2.2	Baggrund	7
2.3	Syv korte skalaer om oplevelser af undervisning på uddannelsesniveau	8
2.4	Fra begreber til måling af undervisningskvalitet	8
2.5	Rapportens opbygning	8
3	Centrale aspekter af undervisning på de videregående uddannelser	9
3.1	EVA's mål for undervisningskvalitet på de videregående uddannelser	10
3.2	Udviklingen af skalaer om kvalitet af undervisning	12
4	Valideringsanalyse	17
4.1	Intro til valideringsanalyser	17
4.2	Analysestrategi	18
4.3	Resultater af valideringsanalyser	18
4.4	Skalakonstruktion	18
4.5	Dokumentation for de enkelte skalaer	19
	Appendiks A – Litteraturliste	54

1 Resumé

Danmarks Evalueringsinstitut (EVA) præsenterer i denne rapport konkrete bud på, hvordan man kan måle de studerendes oplevelse af syv centrale aspekter af undervisning på de videregående uddannelser.

Relevans

Undervisning er en hjørnesten i kvaliteten på de videregående uddannelser, og studerendes oplevelse af undervisning er et centralt perspektiv i vurderingen af undervisningskvalitet. Derfor er der også på mange videregående uddannelser stort fokus på at fremme god undervisning gennem løbende systematisk kvalitetsudviklingsarbejde, der inddrager de studerende. Herunder er der en stigende interesse for at bruge data til at arbejde med kvalitetsudviklingen. Mange af de tilgængelige indikatorer, der fx bruges internationalt til at rangordne forskellige institutioner eller til ekstern kvalitetssikring, bygger ikke på data, der direkte handler om uddannelseskvalitet, men fx data om forholdet mellem antallet af undervisere og antallet af studerende. Og når de studerendes perspektiv indgår, er det ofte i form af data fra studentertilfredshedsmålinger (Loukkola et al, 2020). Det er derfor relevant at arbejde systematisk med måling af undervisningskvalitet – herunder om de måleredskaber, vi anvender, er valide.

Faglig kontekst

Der er kun i begrænset omfang arbejdet med at udvikle validerede mål for studerendes oplevelse af undervisningskvalitet i en dansk kontekst. Udvalg for Kvalitet og Relevans i de Videregående Uddannelser (Kvalitetsudvalget) har tidligere anvendt skalaer for *effektiv undervisning* og *dybdelæring*, der ligeledes var baseret på studerendes oplevelse af undervisningen, ligesom Uddannelses- og Forskningsministeriet har spurgt ind til studerendes oplevelse af *underviserinteraktion* og *læring for forståelse* i Læringsbarometeret[1], i begge tilfælde dog kun som delelement i en større undersøgelse, der også fokuserer på mange andre ting. Denne undersøgelse er delvist baseret på de to skalaer, Kvalitetsudvalget brugte. Derudover er undersøgelsen inspireret af det amerikansk udviklede surveyredskab *Student Evaluation of Educational Quality* samt nogle principper for god vurdering og feedback fra litteraturen (Nicol, 2007).

Målgruppe

Rapporten er særlig relevant for ledere og medarbejdere på de videregående uddannelser, der arbejder med at bruge data til udvikling af undervisning, men rapporten kan også læses af andre, der

1 Se https://ufm.dk/uddannelse/statistik-og-analyser/laeringsbarometer/malingen-2018 for nærmere information.

interesserer sig bredere for udviklingen af undervisningen på videregående uddannelser, herunder politikere, uddannelsesinstitutioner og de studerendes organisationer.

Resultater

Vi måler i undersøgelsen syv aspekter af undervisning med skalaer, som består af tre til fem spørgeskemaspørgsmål. De gennemførte valideringsanalyser viser, at disse skalaer fungerer med tilfredsstillende præcision og ensartethed på tværs af forskellige grupper af studerende.[2] Vi konkluderer på den baggrund, at skalaerne kan bruges til at gennemføre valide målinger af de studerendes oplevelse af de forskellige aspekter af undervisning.

Tabel 2.1 viser de syv aspekter af studerendes oplevelse af undervisning, som rapporten adresserer. Fem aspekter vedrører undervisningen/underviserne, mens to aspekter vedrører den feedback, de studerende modtager fra underviserne.

TABEL 2.1

Aspekter af undervisningskvalitet der indgår i undersøgelsen

Undervisning/undervisere	Feedback
Tydelig undervisning	Tydelig feedback
Fokus på deltagelse	Feedforward
Imødekommenhed og tilgængelighed	
Vægt på elaborering	
Vægt på selvstændighed	

Kilde: Danmarks Evalueringsinstitut, 2021.

Perspektivering

Denne undersøgelse viser, at det er muligt at måle de studerendes oplevelse af en række centrale aspekter af undervisning med syv korte skalaer.

På den baggrund kan EVA stille nogle konkrete måleredskaber til rådighed for institutioner, der ønsker at følge de studerendes oplevelse af kvaliteten af undervisningen på deres uddannelser. De syv skalaer kan fx anvendes til at databasere arbejdet med kvalitetsudvikling af uddannelserne. Valide mål for centrale aspekter af undervisningen kan forhåbentlig bidrage til udviklingen af god undervisning og hermed øge de studerendes læringsudbytte.

Selvom skalaerne kan bruges til at gennemføre valide målinger, kan den teoretiske forståelse af aspekterne og de specifikke spørgeskemaspørgsmål fortsat videreudvikles. Både hvad angår teori og konkrete spørgsmål er vores måling inspireret af tidligere målinger af undervisningskvalitet. En mere udfoldet teoretisk forståelse af de forskellige aspekter kunne dog styrke indsigten i, hvilken betydning de forskellige aspekter mere præcist har for fx de studerendes læringsudbytte eller fastholdelse, og ville formentlig kunne udpege andre relevante aspekter af undervisningskvalitet i

2 Der er gennemført Rasch-analyser, som viser, at det for alle syv skalaer er muligt at fitte data til enten en Rasch-model eller en udvidet Rasch-model (en såkaldt grafisk log-lineær Rasch-model).

fremtidige målinger. Derudover er der også potentiale til at videreudvikle spørgsmål og skalaer gennem en øget tilpasning til konkrete undervisningskontekster med henblik på at styrke praksisrelevansen af de data, der indsamles.

Datagrundlag

Datagrundlaget er spørgeskemabesvarelser fra studerende på de videregående uddannelser. Populationen omfatter studerende fra samtlige videregående uddannelser, der optog studerende via Den Koordinerede Tilmelding (KOT) i sommeren 2018 inden for de tre store uddannelsestyper: erhvervsakademi, professionsbachelor og universitetsuddannelser. Undersøgelsen er baseret på et spørgeskema, som er udsendt i marts 2019, dvs. et halvt år efter studiestart.

2 Indledning

Denne rapport præsenterer konkrete bud på, hvordan man kan måle studerendes oplevelse af syv centrale aspekter af undervisning på de videregående uddannelser.

2.1 Formål og fokus

Formålet med denne undersøgelse er at udvikle vidensbaserede og validerede mål for centrale aspekter af kvalitet af undervisning, som gør det muligt for institutioner eller andre aktører at undersøge og følge centrale aspekter af undervisningskvalitet på de videregående uddannelser.[3] Dermed undersøger vi i rapporten, hvordan syv udvalgte aspekter af undervisning kan defineres og måles.

En væsentlig afgrænsning er, at undersøgelsen fokuserer på generiske aspekter af undervisning, som kan observeres på tværs af forskellige uddannelser og sektorer. Det betyder, at der fx ikke kan være fokus på kvaliteten af det faglige indhold af undervisningen, selvom den udgør et meget vigtigt aspekt af undervisningens kvalitet. Fokus er i stedet på nogle udvalgte, generiske aspekter af undervisningen, som er inspireret af tidligere mål for undervisningskvalitet, og som formodes at udgøre væsentlige aspekter.

2.2 Baggrund

Selvom de studerendes oplevelse af undervisningen må betragtes som et indirekte mål for undervisningskvalitet, udgør det samtidig den mest oplagte måde at forsøge at måle undervisningskvalitet på. De studerende er som modtagere af undervisningen de vigtigste respondenter til at kunne vurdere undervisningen, og hvordan undervisningen påvirker dem. Selvom de studerendes oplevelse ikke er et objektivt mål for undervisningskvalitet, er det uomtvisteligt, at deres oplevelse af undervisningen udgør et centralt perspektiv i vurderingen af undervisningskvalitet. Dertil kommer, at der næppe findes alternative direkte eller objektive mål for undervisningskvalitet.

3 Formålet er dermed ikke at udvikle et kompleks af mål, der tilsammen udgør et dækkende mål for undervisningskvalitet.

2.3 Syv korte skalaer om oplevelser af undervisning på uddannelsesniveau

Vi måler de studerendes oplevelser af undervisningen gennem syv aspekter med skalaer, som består af tre til fem spørgeskemaspørgsmål. Til forskel fra undervisningsevalueringer, som uddannelsesinstitutionerne skal gennemføre på kursusniveau, er de anvendte spørgsmål rettet mod de studerendes generelle oplevelser med undervisning på den uddannelse, de går på. Målene for de syv aspekter er udviklet med afsæt i tidligere mål for studerendes oplevelse af undervisningskvalitet.[4]

2.4 Fra begreber til måling af undervisningskvalitet

Vi gennemfører avancerede valideringsanalyser med henblik på at afdække, om de anvendte spørgeskemaspørgsmål måler de aspekter, vi ønsker at måle med tilstrækkelig præcision og ensartethed på tværs af forskellige grupper af studerende. Hvis de gør det, betragtes de anvendte spørgeskemaspørgsmål og skalaer som valide mål for de undersøgte aspekter.

I rapporten præsenterer vi de undersøgte aspekter af undervisning, operationaliseringen af begreber i form af en række spørgeskemaspørgsmål og skalaer samt dokumentation for de gennemførte valideringsanalyser. Rapporten dokumenterer således forbindelsen mellem begreberne for forskellige aspekter af undervisning, de anvendte skalaer og de indsamlede observationer.[5]

2.5 Rapportens opbygning

Vi præsenterer i kapitel 3 de forskellige aspekter af undervisning, som indgår i undersøgelsen, forbindelsen til tidligere mål for undervisningskvalitet og de skalaer, vi i undersøgelsen har brugt til at måle de forskellige aspekter. Kapitel 4 viser fordelingen på de anvendte spørgsmål og de konstruerede skalaer og dokumenterer de gennemførte Rasch-analyser, som er en avanceret form for valideringsanalyse.[6]

4 *Mål* refererer i denne sammenhæng til *måling* og ikke til formål med undervisning.

5 Dette er ifølge Adcock og Collier (2001) essensen i undersøgelser af målingsvaliditet.

6 Vi anvender Rasch-analyse til at undersøge validiteten af målene for studerendes oplevelse af undervisningskvalitet.

3 Centrale aspekter af undervisning på de videregående uddannelser

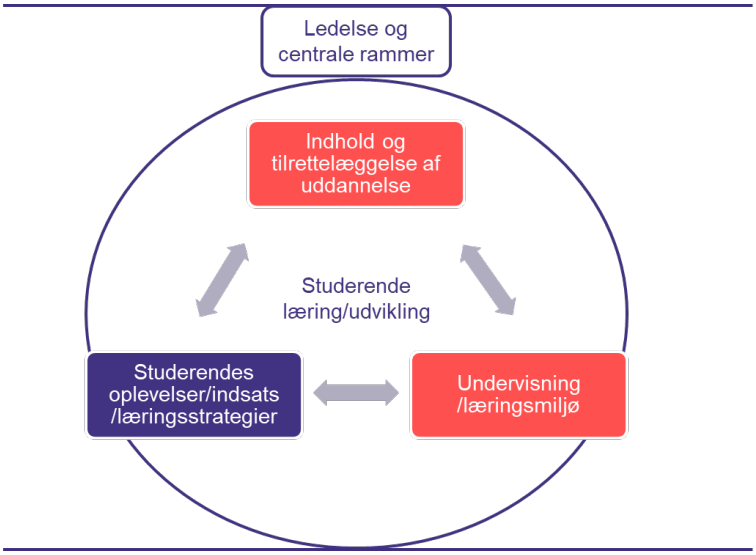
I dette kapitel præsenterer vi de syv aspekter af undervisning, som indgår i undersøgelsen. Udvælgelsen af de syv aspekter er baseret på en bred forståelse af undervisningskvalitet, hvor fem aspekter handler om selve undervisningen/underviseren, mens to aspekter handler om undervisernes feedback. Vi gennemgår nedenfor, hvordan vi måler disse aspekter ved hjælp af en række skalaer baseret på data fra spørgeskemaer, hvori vi spørger ind til de studerendes oplevelse af de syv aspekter. Efterfølgende giver vi en kort introduktion til de mål for undervisningskvalitet, som vores mål er udviklet fra.

Indledningsvist præsenteres den konceptuelle ramme for undersøgelsen. Kvalitet af undervisningen betragtes her som et centralt element af uddannelseskvalitet, jf. figur 3.1. Og studerendes *oplevelse* af undervisningen betragtes som et vigtigt perspektiv på kvaliteten af undervisningen i overensstemmelse med Chen & Hoshower (2003):

> Since student ratings are the most, if not the only, influential measure of teaching effectiveness, active participation by and meaningful input from students can be critical in the success of such teaching evaluation systems.

FIGUR 3.1

Konceptuel ramme for undersøgelsen: Centrale elementer af uddannelseskvalitet på de videregående uddannelser



Kilde: Danmarks Evalueringsinstitut, 2021.

Indholdet og tilrettelæggelsen af uddannelsen samt undervisningen og læringsmiljøet kan betragtes som centrale elementer af uddannelseskvalitet, der både har en direkte påvirkning på de studerendes læringsudbytte, og som uddannelsesinstitutionerne i høj grad selv er med til at definere. Vores måling af undervisningskvalitet betragtes dog som et *indirekte* mål, eftersom vi baserer målingerne på de studerendes *oplevelse* af undervisningen.

De studerendes egen indsats, oplevelser og læringsstrategier kan jf. figur 3.1 også betragtes som centrale elementer af uddannelseskvalitet. Disse forhold kan i sig selv formodes at have betydning for de studerendes læringsudbytte. Til gengæld er det ikke forhold, uddannelserne selv kan tilrettelægge, men i stedet noget de kan forsøge at fremme indirekte ved at påvirke de studerende. Andre har på baggrund af denne tilgang betragtet målinger af de studerendes oplevelser, indsats og læringsstrategier som mål for uddannelseskvalitet.[7]

Ledelsen af og de centrale rammer for uddannelserne kan ligeledes betragtes som centrale elementer af uddannelseskvalitet. De er vigtige betingelser for uddannelsernes indhold og tilrettelæggelse samt undervisningen og læringsmiljøet. Ud fra dette perspektiv betragtes de som forhold, der indirekte påvirker de studerendes læring og udvikling, jf. figur 3.1.

3.1 EVA's mål for undervisningskvalitet på de videregående uddannelser

3.1.1 Aspekter vedrørende undervisningen/underviserne på uddannelsen

Undersøgelsen omfatter fem aspekter, der vedrører undervisningen/underviserne. To af disse vedrører underviserens didaktiske praksis: *Tydelig undervisning* og *Fokus på deltagelse*. *Tydelig undervisning* handler om, i hvilken grad underviserne er tydelige, både hvad angår fagenes formål og krav og i deres formidling af det faglige indhold, mens *Fokus på deltagelse* handler om undervisernes fokus på at inddrage og aktivere de studerende i forbindelse med undervisningen.

To af de målte aspekter vedrører undervisernes fokus på, i hvilken grad de studerende engagerer sig i dybdelæringsprocesser: *Vægt på elaborering* og *Vægt på selvstændighed*. *Vægt på elaborering* handler om undervisernes vægtning af de studerendes arbejde med at skabe forbindelser mellem forskellige typer af viden. *Vægt på selvstændighed* handler om undervisernes vægtning af de studerendes arbejde med selvstændigt at vurdere, tage kritisk stilling og bidrage med nye perspektiver.

Endelig vedrører det femte aspekt vedrørende undervisningen/underviserne mødet mellem de studerende og underviserne. *Imødekommenhed og tilgængelighed* handler om undervisernes imødekommenhed, interesse og tilgængelighed i dette møde.

3.1.2 Aspekter vedrørende feedback på uddannelsen

Undersøgelsen omfatter to aspekter, der vedrører feedback på uddannelsen. Disse aspekter vedrører ligeledes undervisernes didaktiske praksis: *Tydelig feedback* og *Feedforward*. *Tydelig feedback* handler om, hvorvidt den modtagne feedback er med til at tydeliggøre styrker og svagheder i de

7 Se fx https://ufm.dk/uddannelse/statistik-og-analyser/laeringsbarometer/om-malingen.

studerendes opgaver og dermed også tydeliggøre for de studerende, hvad en god opgave er. *Feedforward* handler om, hvorvidt den modtagne feedback understøtter de studerendes faglige udvikling fremadrettet, fx gennem hjælp til prioritering og konkret anvisning af, hvordan de studerende kan forbedre sig.

Tabel 3.1 viser de forskellige aspekter af undervisningskvalitet, eksempler på spørgsmål samt den teoretiske forståelse af aspekterne.

TABEL 3.1

EVA's mål for studerendes oplevelse af undervisningskvalitet

	Eksempel på spørgsmål	Teoretisk forståelse
Undervisning		
Tydelig undervisning	• I hvilken grad oplever du, at underviserne i dine fag er gode til at forklare faglige pointer tydeligt?	*Tydelig undervisning* handler om undervisernes tydelighed med hensyn til formål og krav og tydeligheden i deres formidling af det faglige indhold.
Fokus på deltagelse	• I hvilken grad oplever du, at underviserne i dine fag tilskynder de studerende til at deltage i faglige diskussioner i undervisningen?	*Fokus på deltagelse* handler om undervisernes tilbøjelighed til at inddrage de studerende og tilskynde dem til at deltage aktivt i forbindelse med undervisningen.
Vægt på elaborering	• I hvilken grad oplever du, at underviserne i dine fag lægger vægt på, at de studerende kobler ny viden med det, de allerede ved om emnet?	*Vægt på elaborering* handler om undervisernes vægtning af de studerendes arbejde med at skabe forbindelser mellem forskellige typer af viden, herunder mellem ny og eksisterende viden.
Vægt på selvstændighed	• I hvilken grad oplever du, at underviserne i dine fag lægger vægt på, at de studerende kan vurdere et synspunkt, en afgørelse eller en informationskilde?	*Vægt på selvstændighed* handler om undervisernes vægtning af de studerendes arbejde med selvstændigt at vurdere, tage kritisk stilling og bidrage med nye perspektiver.
Imødekommenhed og tilgængelighed	• I hvilken grad oplever du, at underviserne i dine fag er imødekommende over for studerende, der søger hjælp eller råd både i og uden for undervisningen?	*Imødekommenhed og tilgængelighed* handler om undervisernes imødekommenhed, interesse og tilgængelighed i mødet med de studerende.
Feedback		
Tydelig feedback	Hvor ofte gør dine undervisere følgende i forbindelse med feedback på en opgave på studiet? • Gør det tydeligt for mig, hvad jeg ikke har forstået.	Tydelig feedback handler om, hvorvidt den modtagne feedback er med til at tydeliggøre styrker og svagheder i de studerendes opgaver.
Feedforward	Hvor ofte gør dine undervisere følgende i forbindelse med feedback på en opgave på studiet? • Siger hvad jeg bør gøre anderledes fremover.	Feedforward handler om, hvorvidt den modtagne feedback understøtter de studerendes faglige udvikling fremadrettet.

Kilde: Danmarks Evalueringsinstitut, 2021.
Note: Se oversigt over alle items, der indgår i hvert af aspekterne, i tabel 3.5 i slutningen af kapitlet.

3.2 Udviklingen af skalaer om kvalitet af undervisning

Udviklingen af skalaerne for de syv aspekter af undervisning og feedback, som fremgår af tabel 3.1, er dels baseret på EVA's egne valideringsanalyser af tidligere skalaer om undervisning og dels med inspiration fra:

- Spørgeskemaredskabet *Student Evaluation of Educational Quality* (SEEQ)
- Kvalitetsudvalgets spørgeskema til studerende på de videregående uddannelser[8]
- Principper for god vurdering og feedback fra Nicol (2007).

I de følgende tre afsnit præsenteres hver enkel af ovenstående inspirationskilder og deres relation til EVA's skalaer, inden de spørgsmål, der indgår i EVA's mål for de syv aspekter af undervisning og feedback, præsenteres.

3.2.1 *Student Evaluation of Educational Quality*: Et spørgeskemabaseret mål for undervisningskvalitet

Student Evaluation of Educational Quality (SEEQ) er et amerikansk udviklet spørgeskemainstrument, der er designet til at måle studerendes evaluering af undervisningens kvalitet. SEEQ er baseret på et større udviklingsarbejde med udarbejdelsen af en pulje spørgeskemaspørgsmål, som efterfølgende er sorteret ud fra følgende kriterier: 1) studerendes angivelse af spørgsmålenes vigtighed, 2) underviseres angivelse af spørgsmålenes brugbarhed, 3) faktoranalyse, samt 4) spørgsmålenes reliabilitet (Marsh, 1982, s. 77). De forskellige aspekter af studerendes oplevelse af undervisning har siden vist sig at være stabile på tværs af tid og forskellige populationer i forskellige kulturer (Coffey & Gibbs, 2010).[9]

SEEQ omfatter skalaer for ni forskellige aspekter af uddannelseskvalitet, der hver måles med mellem to og fire spørgsmål (Catano & Harvey, 2011). Vi har i vores arbejde ladet os inspirere af de fire af disse aspekter, der mest direkte adresserer undervisning, hvilket fremgår af tabel 3.2 nedenfor:

8 Udvalg for Kvalitet og Relevans (2014)

9 Dette er undersøgt ved hjælp af faktoranalyse.

TABEL 3.2

Oversigt over dimensioner fra SEEQ og relation til EVA-skalaer

Aspekt af undervisningskvalitet i SEEQ	EVA-aspekt
Learning	-
Individual rapport	Imødekommenhed og tilgængelighed
Enthusiasm	-
Examinations	-
Organisation	Tydelig undervisning
Breadth	Fokus på elaborering
Group interaction	Fokus på deltagelse
Assignments	-
Workload/difficulty	-

Kilde: Danmarks Evalueringsinstitut, 2021.
Note: Baseret på Catano & Harvey (2011) og Coffey & Gibbs (2001).

3.2.2 Kvalitetsudvalgets skala om undervisning

Kvalitetsudvalget gennemførte i 2014 en spørgeskemaundersøgelse på de videregående uddannelser i Danmark, og denne var en oversættelse af det amerikansk udviklede *National Survey of Student Engagement* (NSSE). Spørgeskemaet omfatter spørgsmål angående de studerendes deltagelse i en række uddannelsesaktiviteter, institutionernes og fagenes faglige krav, opfattelse af læringsmiljøet, undervisningen på uddannelserne med videre (Udvalg for Kvalitet og Relevans i de Videregående Uddannelser, 2014).[10]

Blandt de skalaer, der indgår i Kvalitetsudvalgets spørgeskema vedrører to undervisningen på uddannelserne. Det gælder skalaen om Dybdelæring (fire spørgsmål) og skalaen om Effektiv undervisning (syv spørgsmål). Tematisk omhandler skalaen for Dybdelæring, hvor meget underviserne har lagt vægt på henholdsvis elaborering (at forbinde forskellige typer af information), dybdelæring, selvstændighed og nye idéer/perspektiver. Tematisk omhandler skalaen for Effektiv undervisning undervisernes tydelighed (tre spørgsmål), deres struktur (ét spørgsmål) samt i hvilket omfang underviserne giver studerende forskellige typer af feedback (tre spørgsmål).

Set i forhold de aspekter, der undersøges i denne rapport (EVA-aspekter), synes begge skalaer tematisk at måle flere forskellige aspekter. Særligt med hensyn til skalaen for Dybdelæring synes spørgsmålene jf. tabel 3.3 at dække over forskellige EVA-aspekter (*Vægt på elaborering* og *Vægt på selvstændighed*). EVA har indledningsvist gennemført valideringsanalyser af kvalitetsvalgets skalaer, som ligeledes pegede på, at de indeholder flere dimensioner. Det er baggrunden for, at skalaerne ikke er anvendt i deres oprindelige form. Tabel 3.3 viser en oversigt over relevansen af de to skalaer for EVA's skalaer.

10 Se også http://nsse.indiana.edu/.

TABEL 3.3

Oversigt over Kvalitetsudvalgets skalaer og beslægtede EVA-aspekter

Aspekt af undervisningskvalitet i Kvalitetsudvalgets skalaer	EVA-aspekt
Dybdelæring	Vægt på elaborering
Dybdelæring	Vægt på selvstændighed
Effektiv undervisning	Tydelighed

Kilde: Danmarks Evalueringsinstitut, 2021.
Note: Baseret på Udvalg for Kvalitet og Relevans i de Videregående Uddannelser.

3.2.3 Nicols principper for god vurdering og feedback

EVA's skalaer for forskellige aspekter af feedback tager udgangspunkt i *Principles of good assessment and feedback: Theory and practice* (Nicol, 2007), der identificerer ti principper for god vurderings- og feedbackpraksis. Principperne er udviklet med henblik på at styrke studerendes ansvar for og kontrol over deres læringssituation med reference til litteraturen om selv-reguleret læring (Nicol, 2007), hvor kontrol med egen læring også formodes at styrke de studerendes motivation for at studere (Pintrich, 2003; EVA, 2020). EVA har med afsæt i de ti principper for god vurderings- og feedbackpraksis udviklet to skalaer for feedback. De to skalaer er udvalgt, fordi det er de aspekter, der mest direkte vedrører undervisernes didaktiske praksis. Tabel 3.4 viser en oversigt over de ti principper og de to aspekter af feedback, som EVA har udviklet mål for.

TABEL 3.4

Oversigt over principper for god vurderings- og feedbackpraksis samt beslægtede EVA-aspekter

Princip for god vurderings- og feedbackpraksis	EVA-aspekt
Help clarify what good performance is (goals, criteria, standards)	Tydelig feedback
Deliver high quality feedback information that helps learners self-correct	Tydelig feedback Feedforward
Encourage "time and effort" on challenging learning tasks	
Encourage positive motivational beliefs and self-esteem	
Encourage interaction and dialogue around learning (peer and teacherstudent	
Facilitate the development of self-assessment and reflection in learning.	
Give learners choice in assessment – content and processes	
Involve students in decision-making about assessment policy and practice	
Support the development of learning communities	
Help teachers adapt teaching to student needs	

Kilde: Danmarks Evalueringsinstitut, 2021.
Note: Principper og reflektionsspørgsmål er baseret på Nicol (2007).

3.2.4 EVA's mål for studerendes oplevelse af undervisningskvalitet

EVA's skalaer for studerendes oplevelse af undervisning er som nævnt inspireret af ovenstående skalaer og teoretiske forståelser. Skalaerne er på den baggrund udviklet af EVA selv. Tabel 3.5 viser de forskellige aspekter for feedback, undervisning samt undervisernes imødekommenhed og tilgængelighed, og de spørgeskemaspørgsmål der indgår i skalaen for hvert aspekt. Ud over den konceptuelle ramme, som er præsenteret overfor, er skalaerne altså ikke udviklet på baggrund af en samlet teori for, hvordan de forskellige aspekter påvirker hinanden eller deres præcise betydning for de studerendes fastholdelse og læringsudbytte. EVA har dog udvalgt de forskellige aspekter på baggrund af en vurdering af, at der er tale om centrale aspekter af undervisningskvalitet.

TABEL 3.5

EVA's mål for studerendes oplevelse af undervisningskvalitet

	Måling af aspekt
Undervisning	
Tydelig undervisning	hvilken grad oplever du, at underviserne i dine fag: • … er gode til at forklare fagenes formål og krav tydeligt? • … er gode til at vise, hvordan fagene kan anvendes i virkelige situationer? • … ofte bruger eksempler eller illustrationer til at forklare svære pointer? • … er gode til at forklare faglige pointer tydeligt?
Fokus på deltagelse	I hvilken grad oplever du, at underviserne i dine fag: • … tilskynder de studerende til at deltage i faglige diskussioner i undervisningen? • … tilskynder de studerende til at dele deres idéer og viden i undervisningen? • … tilskynder de studerende til at give udtryk for egne holdning og/eller forholde sig kritisk til undervisningen? • … er gode til at inddrage de studerende i undervisningen?
Imødekommenhed og	I hvilken grad oplever du, at underviserne i dine fag: • … er venlige overfor hver enkelt studerende? • … er imødekommende over for studerende, der søger hjælp eller råd både i og uden for undervisningen? • … er oprigtigt interesserede i hver enkelt studerende? • … er tilgængelige for dialog uden for undervisningen? • … svarer på skriftlige henvendelser?
tilgængelighed	I hvilken grad oplever du, at underviserne i dine fag: • … lægger vægt på, at de studerende kan anvende det lærte på praktiske problemstillinger eller i en ny kontekst? • … lægger vægt på, at de studerende kobler ny viden med det, de allerede ved om emnet? • … lægger vægt på, at de studerende gør deres bedste for at forbinde det, de lærer på de forskellige dele af uddannelsen?
Vægt på elaborering	I hvilken grad oplever du, at underviserne i dine fag: • … lægger vægt på, at de studerende kan vurdere et synspunkt, en afgørelse eller en informationskilde? • … lægger vægt på, at de studerende kommer med nye idéer eller perspektiver i relation til det faglige stof? • … lægger vægt på, at de studerende forholder sig kritisk til det faglige stof, de præsenterer?
Feedback	
Tydelig feedback	Hvor ofte gør dine undervisere følgende i forbindelse med feedback på en opgave på studiet? • Udpeger fejl eller svagheder ved det, jeg har lavet • Fremhæver de gode aspekter ved mit arbejde • Gør det tydeligt for mig hvad jeg ikke har forstået

	Måling af aspekt
Feedforward	Hvor ofte gør dine undervisere følgende i forbindelse med feedback på en opgave på studiet? • Siger hvad jeg bør gøre anderledes fremover • Hjælper mig med at prioritere, hvad det er særligt vigtigt, at jeg arbejder videre med • Beskriver hvordan jeg kommer op på et højere fagligt niveau

Kilde: Danmarks Evalueringsinstitut, 2021.
Note: Svarkategorier for de to skalaer om feedback: *Altid*, *Ofte*, *Nogle gange*, *Sjældent*, *Aldrig*, *Ikke relevant*. Svarkategorier for de fem skalaer om undervisning og imødekommenhed: *I meget høj grad*, *I høj grad*, *I nogen grad*, *I mindre grad*, *Slet ikke*.

4 Valideringsanalyse

4.1 Intro til valideringsanalyser

For at undersøge validiteten af de anvendte skalaer så grundigt som muligt, har vi gennemført detaljerede *item respons*-analyser. Vi har benyttet Rasch-modellen, fordi en skala, som passer til Rasch-modellen, har optimale måleegenskaber og har en statistisk sufficient sumscore.[11] Dette betyder, at sumscoren indeholder al den nødvendige information for at kunne vurdere personers niveau for det, der måles, og at vi derfor kan benytte sumscoren som skalascore i de videre analyser. I så fald vil det kun være niveauet af det, vi måler, der er bestemmende for, hvordan der svares på hvert enkelt spørgsmål i skalaen. Den samlede score på den pågældende skala vil derfor i sig selv indeholde tilstrækkelig information om de studerendes oplevelse af fx undervisernes tydelighed.

Som en del af analyserne undersøger vi, om der er systematiske skævheder i scoren, som er forårsaget af, at ét eller flere spørgsmål ikke fungerer på samme måde for forskellige grupper af studerende – det, som kaldes *Differentiel Item Funktion* (DIF). Hvis der er DIF, betyder det, at studerende, der fx scorer lige højt på skalaen, der måler elaborering, ikke kun har opnået denne score som følge af deres niveau af elaborering, men også delvist pga. andre karakteristika, såsom køn eller alder. Den skævhed i scoren, der opstår på grund af DIF, kan betyde, at efterfølgende statistiske analyser, hvor grupper af studerende sammenlignes, også bliver skævvredet. I de tilfælde, hvor der er DIF, har vi benyttet en udvidet udgave af Rasch-modellen (de såkaldte grafiske log-lineære Rasch-modeller), som kan justere den samlede score for skævheder forårsaget af DIF, og derved opnås et mere retvisende og validt mål for skalaen, der kan bruges på tværs af undergrupper.[12] Vi har undersøgt DIF ud fra fem baggrundsvariable: køn, alder, optagelsesprioritet, type af gymnasial uddannelse samt hovedområde.

Som en anden del af analyserne har vi undersøgt, om nogle spørgsmål i skalaen hænger stærkere sammen, end hvad der kan forklares af det, skalaen måler. Vi kan altså finde ud af, om det blot er de studerendes oplevelse af fx undervisernes tydelighed, der afgør, hvad der svares på et givent spørgsmål, eller om det også afhænger af, hvad der er svaret på ét eller flere af de andre spørgsmål i skalaen. Dette kaldes lokal afhængighed og er et krav i alle almindelige målingsmodeller. Hvis der ikke tages højde for lokal afhængighed mellem spørgsmål, når reliabiliteten beregnes (for eksempel med *Cronbachs alpha*), vil reliabiliteten fremstå som højere, end den egentlig er. Vi har derfor i forbindelse med de udvidede Rasch-modeller (Kreiner & Christensen, 2007) anvendt en metode til beregning af reliabiliteten, som tager højde for lokal afhængighed, og som giver et mere retvisende

11 For en letlæselig og detaljeret gennemgang af Rasch-modellen, se Kreiner (2009). For mere om Rasch-modellen, se Kreiner (2007).
12 For information om de grafiske loglineære Rasch-modeller, se Kreiner & Christensen (2007).

billede af skalaernes reliabilitet. Rasch-modellerne kan altså ikke blot validere skalaer af den karakter, som vi her arbejder med, men de kan også løse nogle af de fundamentale problemer, der kan være med sådanne skalaer.

4.2 Analysestrategi

De gennemførte valideringsanalyser er baseret på tilfældigt udtrukne stikprøver fra den samlede population. Dette skyldes, at valideringsanalyserne anvender en lang række test, der er baseret på at identificere problemer som signifikante forskelle, baseret på forskellige kriterier. Når vi har store datasæt, hvilket er tilfældet her, medfører det problemer med, at selv små forskelle uden substantiel betydning ofte bliver signifikante.

Derfor er der udtrukket stikprøver af en størrelse, som erfaringsmæssigt er velegnet til at besvare de undersøgelsesspørgsmål, vi stiller i valideringsanalyserne. Vi har i valideringsanalyserne især interesseret os for 1) om der samlet set er en god overensstemmelse mellem den specificerede model og observerede data, og 2) om der kan påvises systematiske forskelle i den måde, hvorpå spørgsmålene fungerer mellem forskellige hovedområder.

På den baggrund er der udtrukket tilfældige stikprøver for hver af de tre sektorer (erhvervsakademiuddannelser, professionsbacheloruddannelser og universitetsuddannelser) fra den samlede population af respondenter på 8.153, så der er 100 tilfældigt udtrukne respondenter inden for hvert hovedområde. Det betyder, at de samlede stikprøver for hver sektor er på 400-600. Dette vurderes at være et passende datagrundlag til at undersøge de spørgsmål, vi interesserer os for, i valideringsanalyserne.

4.3 Resultater af valideringsanalyser

På baggrund af valideringsanalyserne finder vi, at vi kan måle de syv aspekter af de studerendes oplevelse af undervisningskvalitet, som der er gennemført valideringsanalyser af, med tilstrækkelig præcision og ensartethed på tværs af forskellige grupper af studerende. For to af skalaerne (*Vægt på elaborering* og *Vægt på selvstændighed*) er det nødvendigt at justere skalascoren for at kunne foretage præcise sammenligninger mellem henholdsvis forskellige aldersgrupper (*Vægt på elaborering*) og mellem mænd og kvinder (*Vægt på selvstændighed*). Valideringsanalyserne dokumenterer således, at de anvendte skalaer udgør valide mål for de syv aspekter af undervisningskvalitet.

4.4 Skalakonstruktion

Skalaerne baseres på tre-fem spørgsmål. Skalaerne er konstrueret ved at lægge værdien af besvarelserne på de pågældende spørgsmål sammen. Herefter er skalaerne omskaleret, så minimum er 0 og maksimum er 1 for derved at standardisere skalaerne, så de har samme variationsbredde. Skalaerne er justeret for såkaldt *Differentiel Item Funktion* (DIF) i de tilfælde, hvor det er relevant (*Vægt på elaborering* og *Vægt på selvstændighed*).

4.5 Dokumentation for de enkelte skalaer

4.5.1 Præsentation af skala for tydelig undervisning

Skalaen for tydelig og velorganiseret undervisning er dannet på baggrund af nedenstående fire items:
" I hvilken grad oplever du, at underviserne i dine fag:"

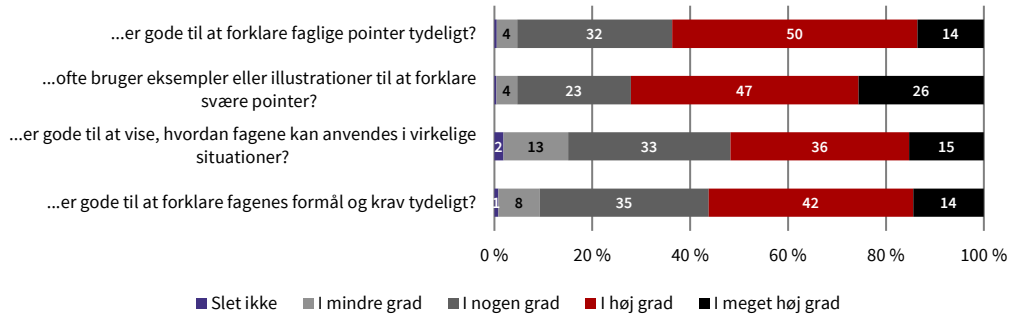
[I meget høj grad, I høj grad, I nogen grad, I mindre grad, Slet ikke]

1. … er gode til at forklare fagenes formål og krav tydeligt?
2. … er gode til at vise, hvordan fagene kan anvendes i virkelige situationer?
3. … ofte bruger eksempler eller illustrationer til at forklare svære pointer?
4. … er gode til at forklare faglige pointer tydeligt?

FIGUR 4.1

Fordeling af enkeltspørgsmål, der indgår i skala for tydelig undervisning

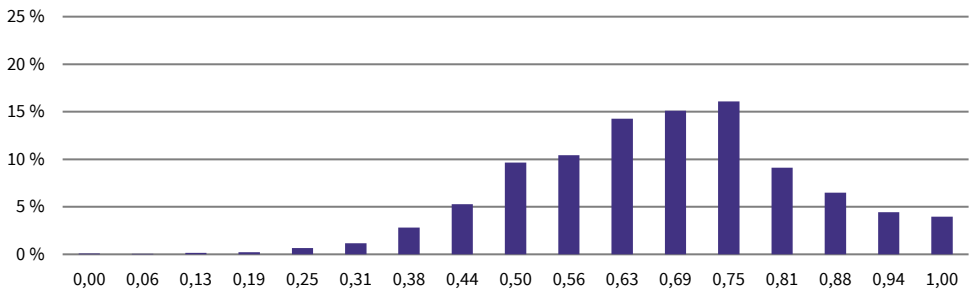
I hvilken grad oplever du, at underviserne i dine fag:



Kilde: Danmarks Evalueringsinstitut, 2021.
Note: n = 8.153.

FIGUR 4.2

Fordeling af skala for tydelig undervisning



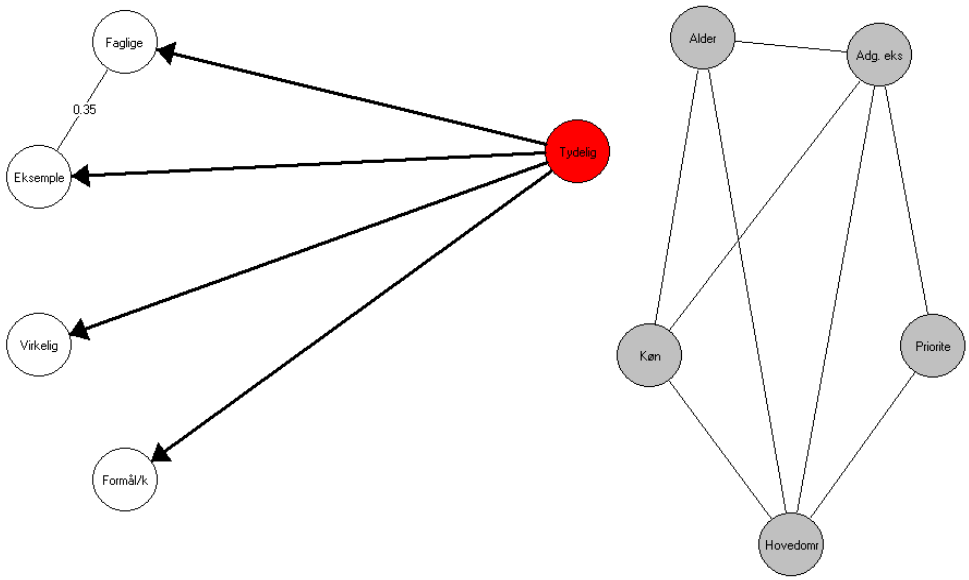
Kilde: Danmarks Evalueringsinstitut , 2021.
Note: n = 8.153.

Resultater for erhvervsakademier

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for lokal afhængighed mellem spørgsmål 3 og 4. Der er ingen DIF mellem spørgsmål og baggrundsvariable. Reliabiliteten for skalaen er på 0,74. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.3

Grafisk fremstilling af tydelig undervisning (for erhvervsakademier)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 400. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.1

Global homogenitetstest samt globale test for DIF (for erhvervsakademier)

Test	CLR	DF	p-værdi
Global homogenitetstest	14,9	21	0,829
Globale test for DIF:			
Køn	21,8	21	0,412
Alder	35,7	42	0,742
Adgangseksamen	28,5	21	0,126
Prioritet	18,7	21	0,603
Hovedområde	58,1	63	0,652

Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

Fortolkning af grafiske fremstillinger

Figur 4.3 er et eksempel på en grafisk fremstilling af de gennemførte Rasch-analyser. Pilene mellem den røde cirkel (den latente variabel, vi ønsker at måle) og cirklerne (de spørgsmål, der indgår i skalaen) i venstre side viser, at spørgsmålene forventes at være udtryk for (og derfor kausalt bestemt af) den latente variabel.

Hvis data passer til en Rasch-model, vil spørgsmålene alene være bestemt af den latente variabel. Det er dog ikke altid tilfældet. Derfor undersøges det for det første, om spørgsmålene hænger sammen med bagvedliggende variable, når der er taget højde for forskelle i den latente variabel (ovenfor omtalt om *Differentiel Item Funktion* (DIF)). Hvis det er tilfældet vil der være en streg, som illustrerer denne sammenhæng fra den bagvedliggende variabel til det spørgsmål, der er tale om. Det fremgår af figur 4.3, at der ikke er problemer med DIF i dette tilfælde.

For det andet undersøges det ligeledes, om der er to spørgsmål, der har en stærkere indbyrdes sammenhæng, end hvad vores bud på den latente variabel kan forklare (ovenfor omtalt som *lokal afhængighed*). Hvis det er tilfældet, er der en streg mellem de to spørgsmål, hvor styrken af denne sammenhæng også er angivet med en gammakoefficient. Af figur 4.3 fremgår det, at der er lokal afhængighed mellem de to øverste items, og vi kan se, at gammakoefficienten er 0,35, hvilket der derfor tages højde for i modellen.

Endelig kan der være sammenhæng mellem de bagvedliggende variable og den latente variabel, hvis en bestemt gruppe af studerende ser ud til konsekvent at score højere eller lavere på det, vi forsøger at måle. I dette tilfælde vil der være en streg mellem den bagvedliggende variabel og den latente variabel. Dette er dog ikke et problem for valideringen af skalaen. Af figur 4.3 ser vi, at der i dette tilfælde ikke er nogen grupper, som systematisk scorer højere eller lavere på vores latente variabel.

Hvis der er problemer med DIF eller lokal afhængighed kan der ofte tages højde for det i en såkaldt Grafisk Log-lineær Rasch-Model (GLLRM). Hvis der er pile, som angiver DIF eller lokal afhængighed i modellen, betyder det, at det har været muligt at tage højde for disse udfordringer og finde et acceptabelt fit til en GLLRM.

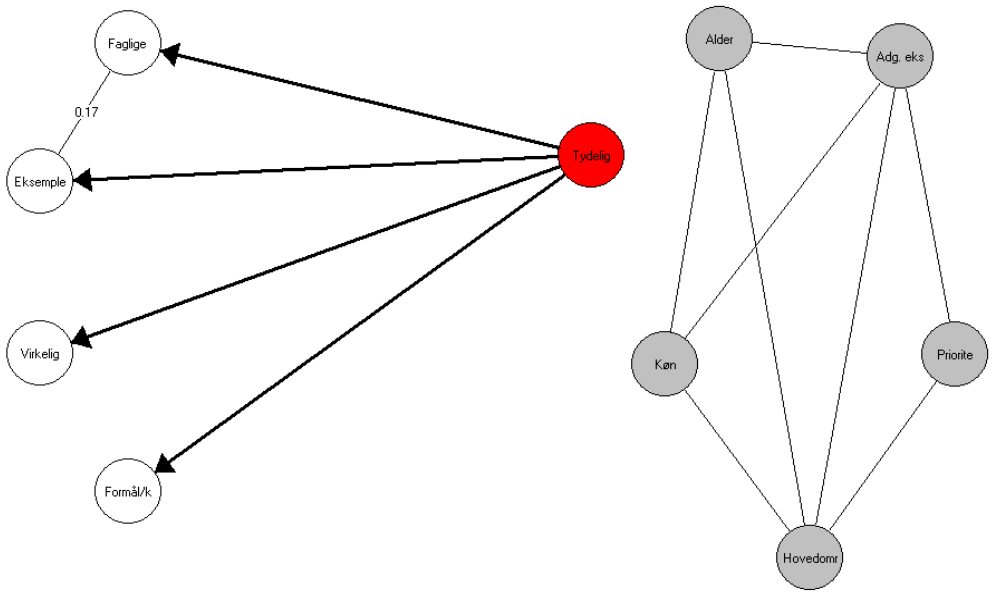
En atypisk ting ved disse analyser er, at vi identificerer problemer ved at påvise, at ovenstående sammenhænge er signifikante. Det skaber den udfordring, at man finder mange problemer, når man arbejder med store datasæt, hvilket der dog er taget hånd om i analysestrategien (jf. ovenfor).

Resultater for professionshøjskoler

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for lokal afhængighed mellem spørgsmål 3 og 4. Der er ingen DIF mellem spørgsmål og baggrundsvariable. Reliabiliteten for skalaen er på 0,72. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.4

Grafisk fremstilling af tydelig undervisning (for professionshøjskoler)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 600. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.2

Global homogenitetstest samt globale test for DIF (for professionshøjskoler)

Test	CLR	DF	p-værdi
Global homogenitetstest	20,4	25	0,727
Globale test for DIF:			
Køn	14,2	25	0,958
Alder	55,7	50	0,269
Adgangseksamen	24,6	25	0,484
Prioritet	22,1	25	0,629
Hovedområde	117,3	125	0,675

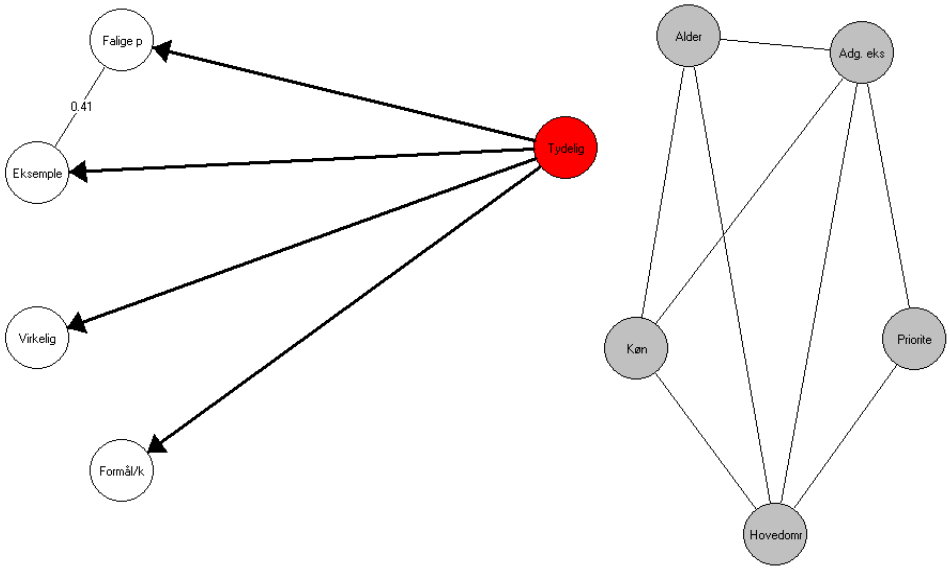
Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

Resultater for universiteter

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for lokal afhængighed mellem spørgsmål 3 og 4. Der er ingen DIF mellem spørgsmål og baggrundsvariable. Reliabiliteten for skalaen er på 0,75. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.5

Grafisk fremstilling af tydelig undervisning (for universiteter)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 500. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.3

Global homogenitetstest samt globale test for DIF (for universiteter)

Test	CLR	DF	p-værdi
Global homogenitetstest	19,7	24	0,714
Globale test for DIF:			
Køn	24,6	24	0,43
Alder	40,4	48	0,775
Adgangseksamen	20,4	24	0,672
Prioritet	34,9	24	0,07
Hovedområde	84,7	96	0,789

Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

4.5.2 Præsentation af skala for imødekommenhed og tilgængelighed

Skalaen for imødekommenhed og tilgængelighed er dannet på baggrund af fem items:
I hvilken grad oplever du, at underviserne i dine fag:

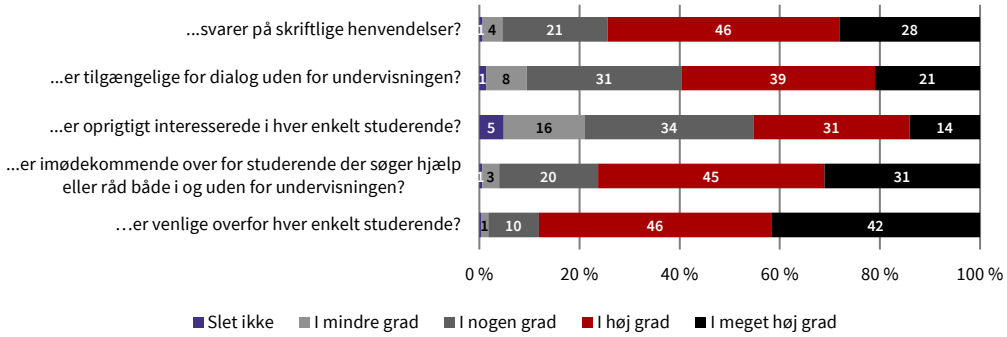
[I meget høj grad, I høj grad, I nogen grad, I mindre grad, Slet ikke]

1. … er venlige overfor hver enkelt studerende?
2. … er imødekommende over for studerende der søger hjælp eller råd både i og uden for undervisningen?
3. … er oprigtigt interesserede i hver enkelt studerende?
4. … er tilgængelige for dialog uden for undervisningen?
5. … svarer på skriftlige henvendelser?

FIGUR 4.6

Fordeling af enkeltspørgsmål, der indgår i skala for imødekommenhed og tilgængelighed

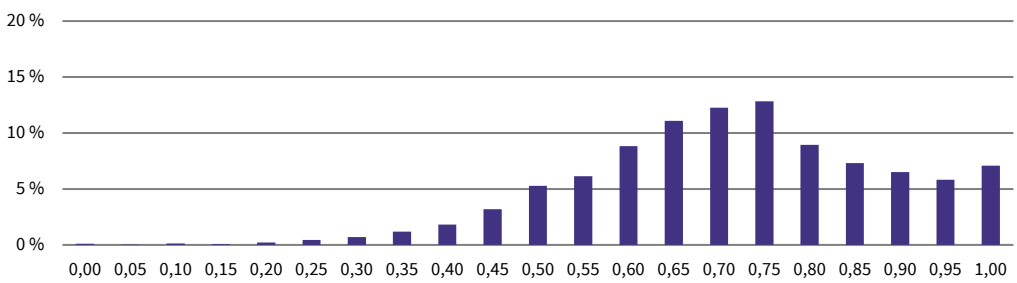
I hvilken grad oplever du, at underviserne i dine fag:



Kilde: Danmarks Evalueringsinstitut, 2021.
Note: n = 8.153.

FIGUR 4.7

Fordeling af skala for imødekommenhed og tilgængelighed



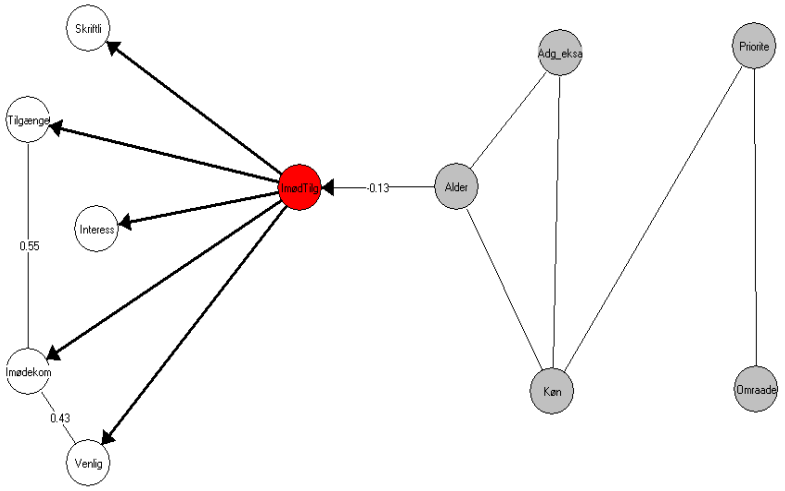
Kilde: Danmarks Evalueringsinstitut, 2021.
Note: n = 8.153.

Resultater for erhvervsakademier

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for lokal afhængighed mellem spørgsmål 1 og 2 samt 2 og 4. Der er ingen DIF mellem spørgsmål og baggrundsvariable. Reliabiliteten for skalaen er på 0,77. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.8

Grafisk fremstilling af imødekommenhed og tilgængelighed (for erhvervsakademier)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 400. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.4

Global homogenitetstest samt globale test for DIF (for erhvervsakademier)

Test	CLR	DF	p-værdi
Global homogenitetstest	23,8	29	0,74
Globale test for DIF:			
Køn	26,4	29	0,6
Alder	61,9	58	0,34
Adgangseksamen	24,5	29	0,7
Prioritet	30,9	29	0,37
Hovedområde	101,2	87	0,14

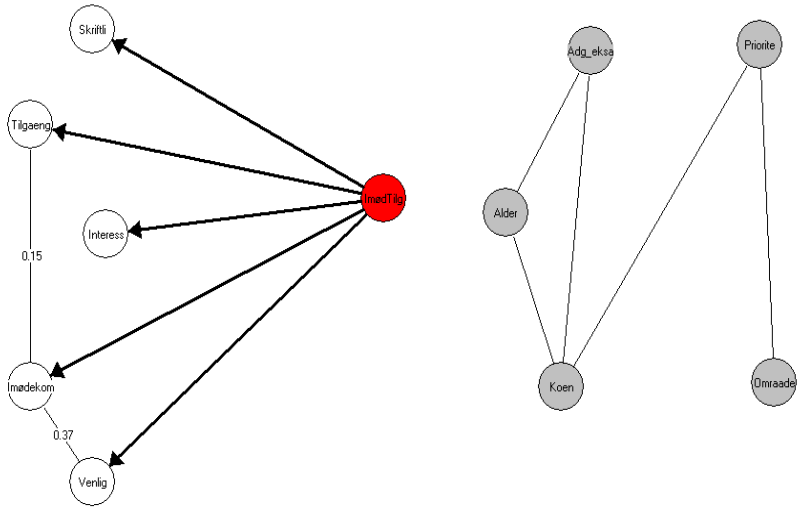
Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

Resultater for professionshøjskoler

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for lokal afhængighed mellem spørgsmål 1 og 2 samt 2 og 4. Der er ingen DIF mellem spørgsmål og baggrundsvariable. Reliabiliteten for skalaen er på 0,81. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.9

Grafisk fremstilling af imødekommenhed og tilgængelighed (for professionshøjskoler)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 600. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.5

Globale homogenitetstest samt globale test for DIF (for professionshøjskoler)

Test	CLR	DF	p-værdi
Global homogenitetstest	38,8	30	0,131
Globale test for DIF:			
Køn	40,8	30	0,9
Alder	60,3	60	0,46
Adgangseksamen	30,9	30	0,42
Prioritet	35,6	30	0,22
Hovedområde	139,8	150	0,713

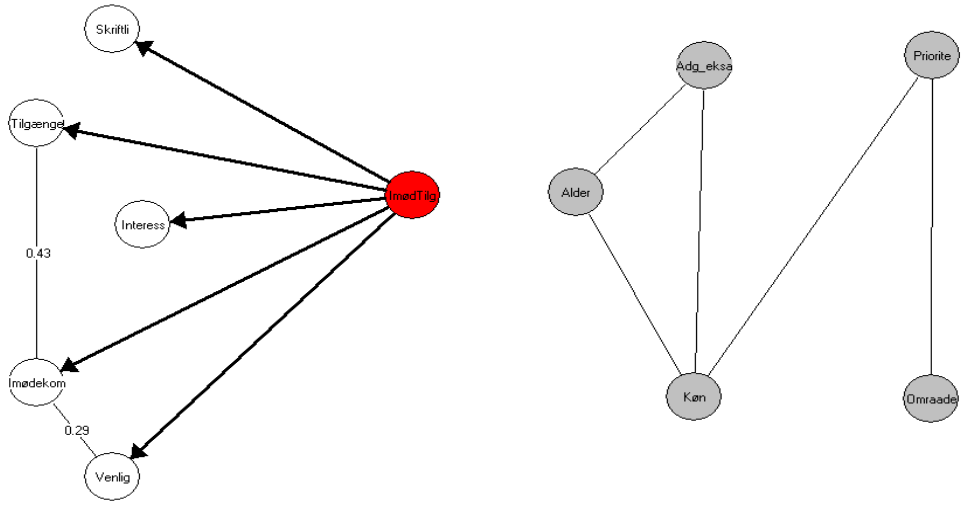
Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

Resultater for universiteter

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for lokal afhængighed mellem spørgsmål 1 og 2 samt 2 og 4. Der er ingen DIF mellem spørgsmål og baggrundsvariable. Reliabiliteten for skalaen er på 0,81. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.10

Grafisk fremstilling af imødekommenhed og tilgængelighed (for universiteter)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 400. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.6

Global homogenitetstest samt globale test for DIF (for universitetsuddannelser)

Test	CLR	DF	p-værdi
Global homogenitetstest	40,7	31	0,114
Globale test for DIF:			
Køn	33,8	31	0,332
Alder	57,8	62	0,627
Adgangseksamen	28,3	31	0,607
Prioritet	39,9	31	0,132
Hovedområde	143,6	124	0,110

Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

4.5.3 Præsentation af skala for fokus på deltagelse

Skalaen for fokus på deltagelse er dannet på baggrund af nedenstående fire items:
"I hvilken grad oplever du, at underviserne i dine fag:"

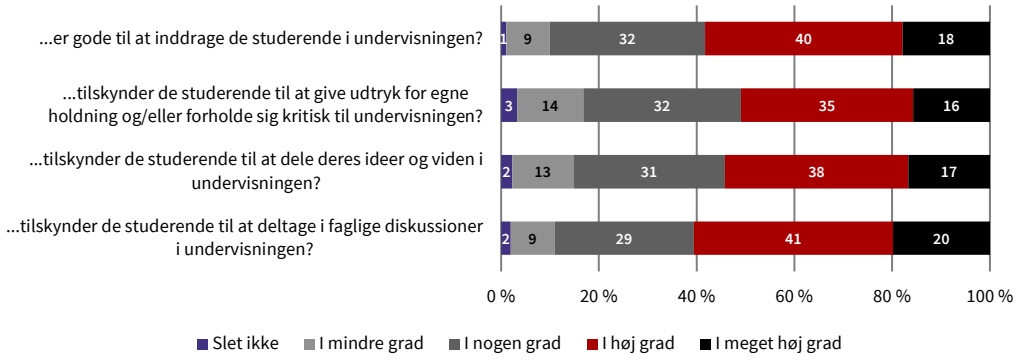
[I meget høj grad, I høj grad, I nogen grad, I mindre grad, Slet ikke]

1. … tilskynder de studerende til at deltage i faglige diskussioner i undervisningen?
2. … tilskynder de studerende til at dele deres ideer og viden i undervisningen?
3. … tilskynder de studerende til at give udtryk for egne holdning og/eller forholde sig kritisk til undervisningen?
4. … er gode til at inddrage de studerende i undervisningen?

FIGUR 4.11

Fordeling af enkeltspørgsmål, der indgår i skala for fokus på deltagelse

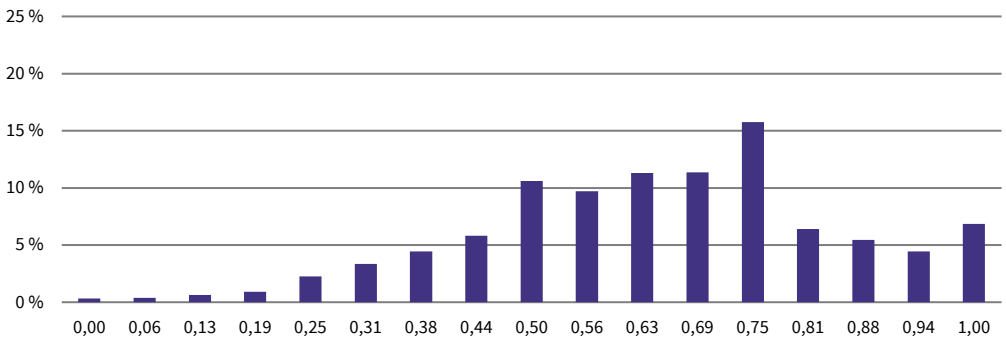
I hvilken grad oplever du, at underviserne i dine fag:



Kilde: Danmarks Evalueringsinstitut, 2021.
Note: n = 8.153.

FIGUR 4.12

Fordeling af skala for fokus på deltagelse



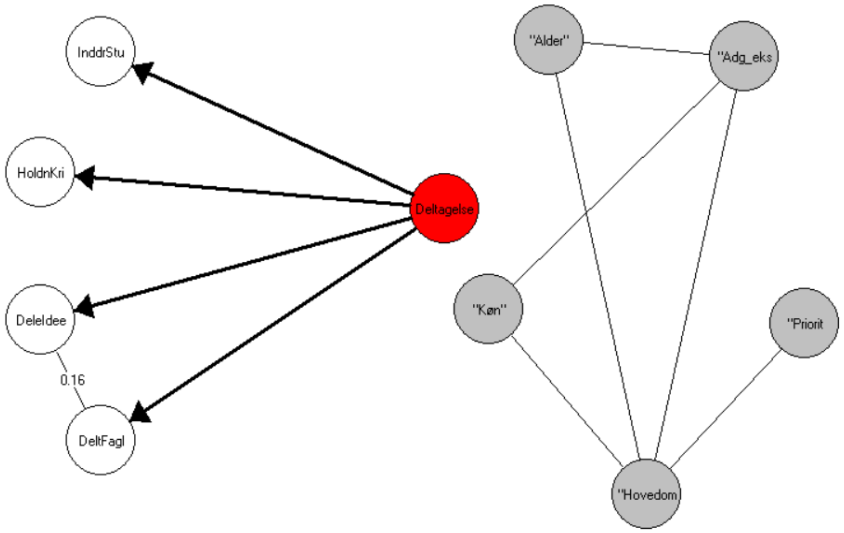
Kilde: Danmarks Evalueringsinstitut, 2021.
Note: n = 8.153.

Resultater for erhvervsakademier

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for lokal afhængighed mellem spørgsmål 1 og 2. Der er ingen DIF mellem spørgsmål og baggrundsvariable. Reliabiliteten for skalaen er på 0,86. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.13

Grafisk fremstilling af fokus på deltagelse (for erhvervsakademier)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 400. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.7

Global homogenitetstest samt globale test for DIF (for erhvervsakademier)

Test	CLR	DF	p-værdi
Global homogenitetstest	28,2	26	0,35
Globale test for DIF:			
Køn	24,2	26	0,57
Alder	41,8	52	0,84
Adgangseksamen	20,7	26	0,76
Prioritet	15,0	26	0,96
Hovedområde	85,7	78	0,26

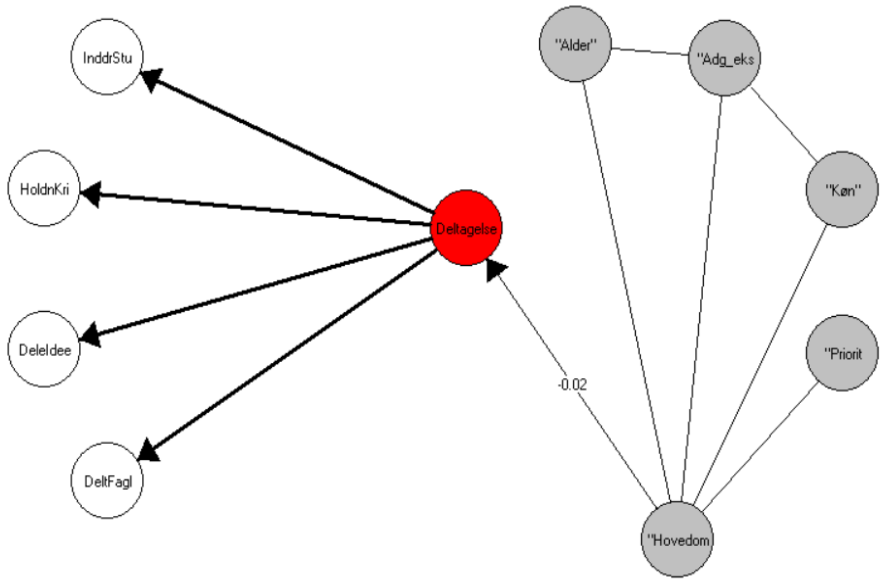
Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

Resultater for professionshøjskoler

Skalaen passer til en Rasch-model. Reliabiliteten for skalaen er på 0,85. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.14

Grafisk fremstilling af fokus på deltagelse (for professionshøjskoler)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 600. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.8

Global homogenitetstest samt globale test for DIF (for professionshøjskoler)

Test	CLR	DF	p-værdi
Global homogenitetstest	17,0	15	0,32
Globale test for DIF:			
Køn	6,7	15	0,97
Alder	25,7	30	0,69
Adgangseksamen	11,8	15	0,694
Prioritet	13,6	15	0,557
Hovedområde	60,3	75	0,89

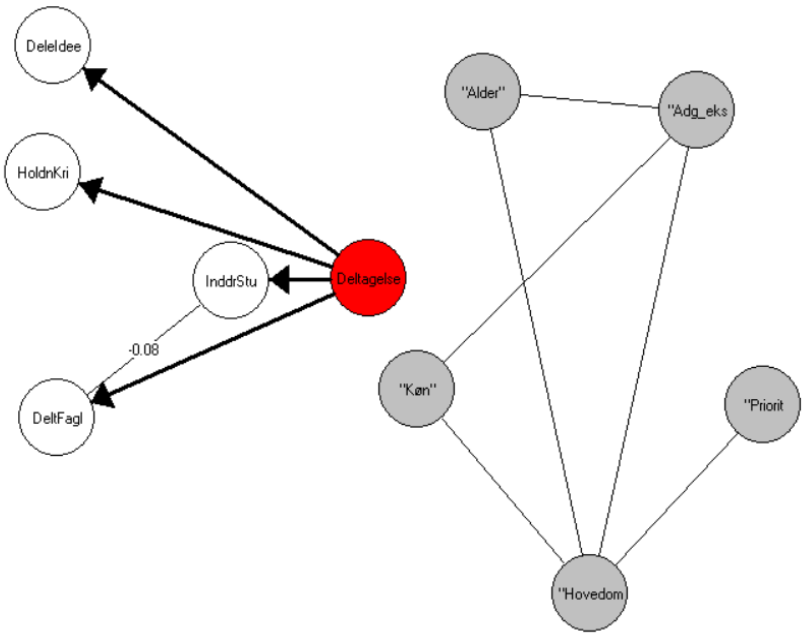
Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

Resultater for universiteter

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for lokal afhængighed mellem spørgsmål 1 og 4. Der er ingen DIF mellem spørgsmål og baggrundsvariable. Reliabiliteten for skalaen er på 0,84. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.15

Grafisk fremstilling af fokus på deltagelse (for universiteter)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 500. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.9

Global homogenitetstest samt globale test for DIF (for universiteter)

Test	CLR	DF	p-værdi
Global homogenitetstest	30,9	29	0,37
Globale test for DIF:			
Køn	41,2	29	0,07
Alder	57,2	58	0,51
Adgangseksamen	45,0	29	0,03
Prioritet	23,2	29	0,77
Hovedområde	112,7	116	0,571

Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester om, spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

4.5.4 Præsentation af skala for vægt på elaborering

Skalaen for vægt på elaborering er dannet på baggrund af nedenstående tre items:
"I hvilken grad oplever du, at underviserne i dine fag:"

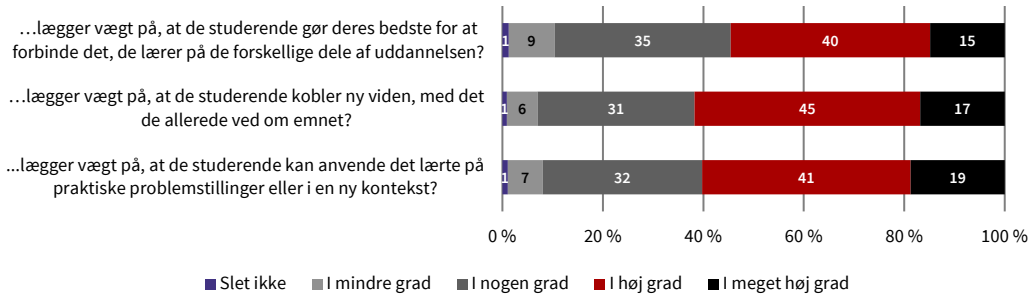
[I meget høj grad, I høj grad, I nogen grad, I mindre grad, Slet ikke]

1. … lægger vægt på, at de studerende kan anvende det lærte på praktiske problemstillinger eller i en ny kontekst?
2. … lægger vægt på, at de studerende kobler ny viden, med det de allerede ved om emnet?
3. … lægger vægt på, at de studerende gør deres bedste for at forbinde det, de lærer på de forskellige dele af uddannelsen?

FIGUR 4.16

Fordeling af enkeltspørgsmål, der indgår i skala for vægt på elaborering

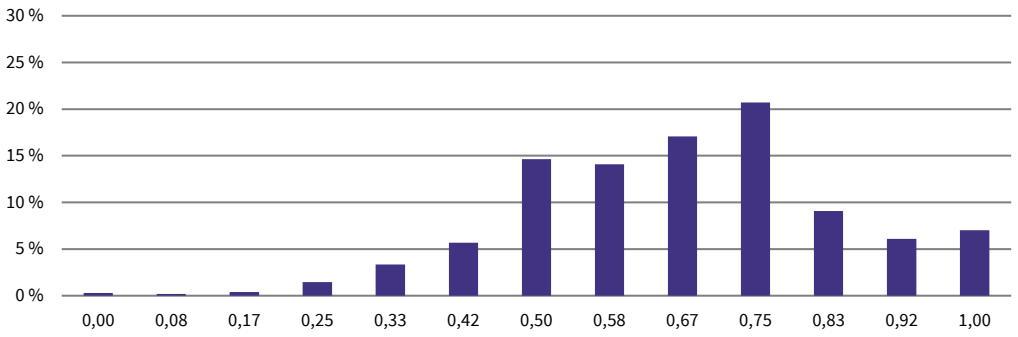
I hvilken grad oplever du, at underviserne i dine fag:



Kilde: Danmarks Evalueringsinstitut, 2021.
Note: n = 8.153.

FIGUR 4.17

Fordeling af skala for vægt på elaborering



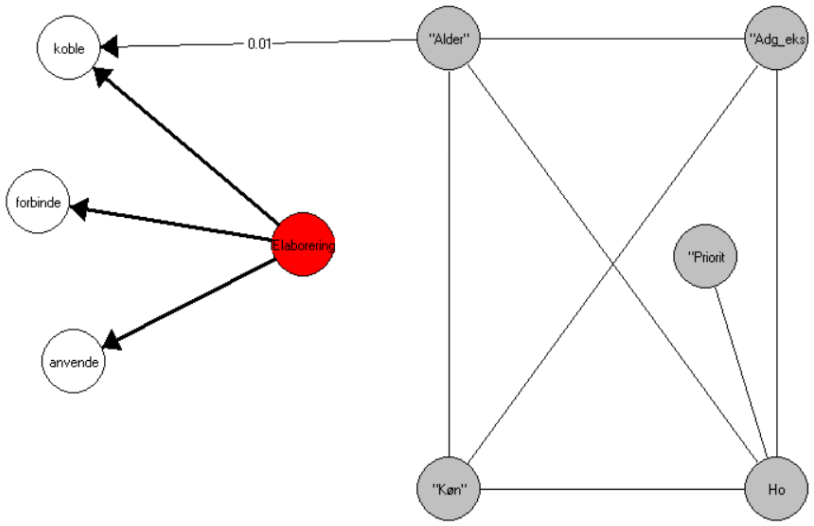
Kilde: Danmarks Evalueringsinstitut, 2021.
Note: n = 8.153. Den viste skala er ikke justeret for DIF.

Resultater for erhvervsakademier

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for DIF mellem spørgsmål 2 og baggrundsvariablen *Alder*. Der er ikke lokal afhængighed mellem spørgsmålene. Reliabiliteten for skalaen er på 0,79-0,81. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.18

Grafisk fremstilling af vægt på elaborering (for erhvervsakademier)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 400. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.10

Global homogenitetstest samt globale test for DIF (for erhvervsakademier)

Test	CLR	DF	p-værdi
Global homogenitetstest	21,1	19	0,33
Globale test for DIF:			
Køn	31,4	19	0,04
Alder	9,6	16	0,89
Adgangseksamen	12,5	19	0,86
Prioritet	15,3	19	0,71
Hovedområde	61,6	57	0,31

Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

TABEL 4.11

Score equation af skalaen *Vægt på elaborering* mht. alder (for erhvervsakademier), justering af sumscoren

17-20 år	21-25 år	26-70 år
1.00	0.75	0.06
2.00	1.73	1.16
3.00	2.85	2.39
4.00	4.00	3.63
5.00	5.08	4.81
6.00	6.11	5.91
7.00	7.11	6.94
8.00	8.11	7.97
9.00	9.14	9.00
10.00	10.19	10.01
11.00	11.15	11.01

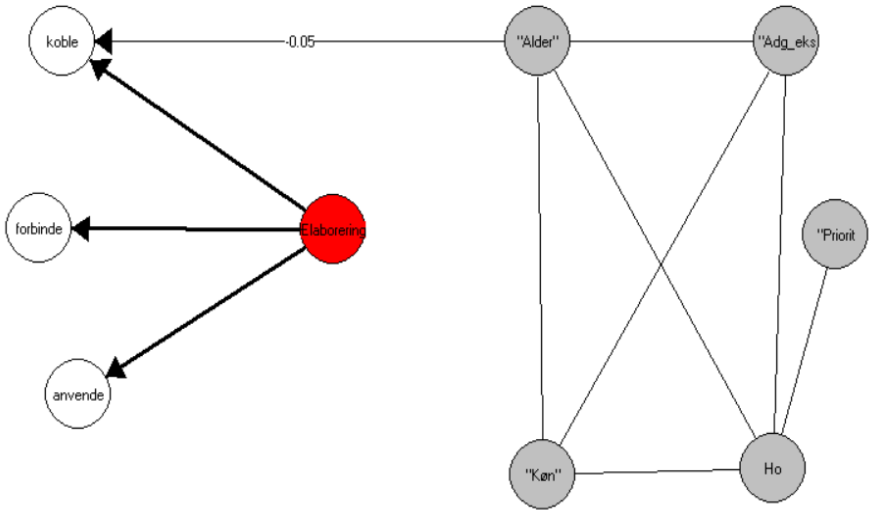
Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: Tabellen viser, hvordan sumscoren justeres på baggrund af for den identificerede DIF. Respondenter i kolonnen længst til venstre er referencegruppen, ift. hvilken de øvrige justeres.

Resultater for professionshøjskoler

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for DIF mellem spørgsmål 3 og baggrundsvariablen *Alder*. Der er ikke lokal afhængighed mellem spørgsmål. Reliabiliteten for skalaen er på 0,75-0,77. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.19

Grafisk fremstilling af vægt på elaborering (for professionshøjskoler)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 600. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.12

Global homogenitetstest samt globale test for DIF (for professionshøjskoler)

Test	CLR	DF	p-værdi
Global homogenitetstest	25,8	18	0,1
Globale test for DIF:			
Køn	18,3	18	0,44
Alder	10,8	16	0,82
Adgangseksamen	24,3	18	0,14
Prioritet	20,5	18	0,3
Hovedområde	81,8	90	0,72

Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

TABEL 4.13

Score equation af skalaen *Vægt på elaborering* mht. alder (for professionshøjskoler), justering af sumscoren

17-20 år	21-25 år	26-70 år
1.00	1.77	1.02
2.00	2.48	1.83
3.00	3.19	2.74
4.00	3.95	3.70
5.00	4.82	4.73
6.00	5.83	5.85
7.00	6.96	7.00
8.00	8.08	8.10
9.00	9.18	9.14
10.00	10.23	10.14
11.00	11.18	11.09

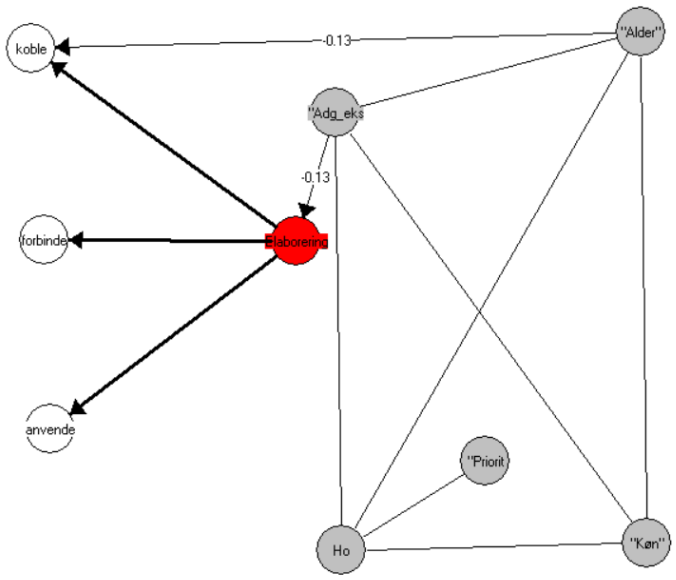
Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: Tabellen viser, hvordan sumscoren justeres på baggrund af for den identificerede DIF. Respondenter i kolonnen længst til venstre er referencegruppen, ift. hvilken de øvrige justeres.

Resultater for universiteter

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for DIF mellem spørgsmål 3 og baggrundsvariablen *Alder*. Der er ikke lokal afhængighed mellem spørgsmålene. Reliabiliteten for skalaen er på 0,67-0,79. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.20

Grafisk fremstilling af vægt på elaborering (for universiteter)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 500. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.14

Global homogenitetstest samt globale test for DIF (for universiteter)

Test	CLR	DF	p-værdi
Global homogenitetstest	23,2	19	0,23
Globale test for DIF:			
Køn	19,6	19	0,42
Alder	12,2	16	0,73
Adgangseksamen	27,8	19	0,09
Prioritet	13,2	19	0,83
Hovedområde	71,8	76	0,62

Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

TABEL 4.15

Score equation af skalaen *Vægt på elaborering* mht. alder (for universiteter), justering af sumscoren

17-20 år	21-25 år	26-70 år
1.00	0.79	0.07
2.00	1.59	0.91
3.00	2.47	1.84
4.00	3.49	2.90
5.00	4.70	4.24
6.00	5.93	5.78
7.00	7.01	7.07
8.00	8.01	9.03
9.00	9.02	9.03
10.00	10.04	9.94
11.00	11.04	10.91

Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: Tabellen viser, hvordan sumscoren justeres på baggrund af for den identificerede DIF. Respondenter i kolonnen længst til venstre er referencegruppen, ift. hvilken de øvrige justeres.

4.5.5 Præsentation af skala for vægt på selvstændighed

Skalaen vægt på selvstændighed er dannet på baggrund af nedenstående tre items:
"I hvilken grad oplever du, at underviserne i dine fag:"

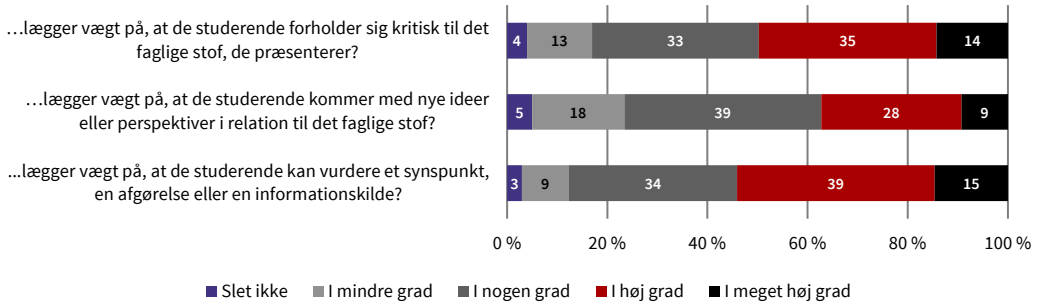
[I meget høj grad, I høj grad, I nogen grad, I mindre grad, Slet ikke]

1. ... lægger vægt på, at de studerende kan vurdere et synspunkt, en afgørelse eller en informationskilde?
2. … lægger vægt på, at de studerende kommer med nye ideer eller perspektiver i relation til det faglige stof?
3. … lægger vægt på, at de studerende forholder sig kritisk til det faglige stof, de præsenterer?

FIGUR 4.21

Fordeling af enkeltspørgsmål, der indgår i skala for vægt på selvstændighed

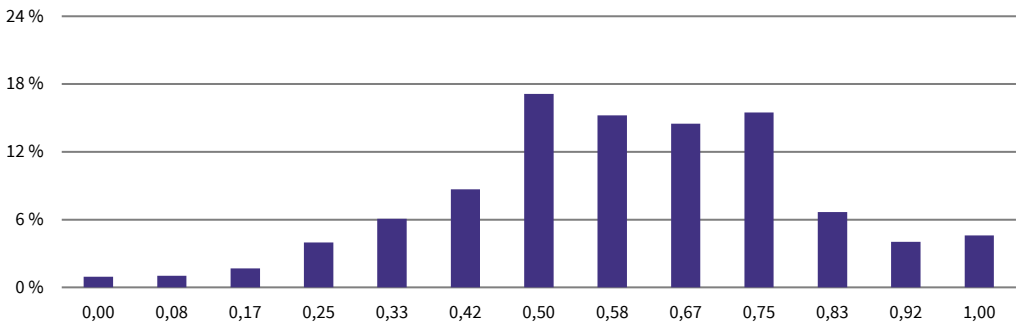
I hvilken grad oplever du, at underviserne i dine fag:



Kilde: Danmarks Evalueringsinstitut, 2021.
Note: n = 8.153.

FIGUR 4.22

Fordeling af skala for vægt på selvstændighed



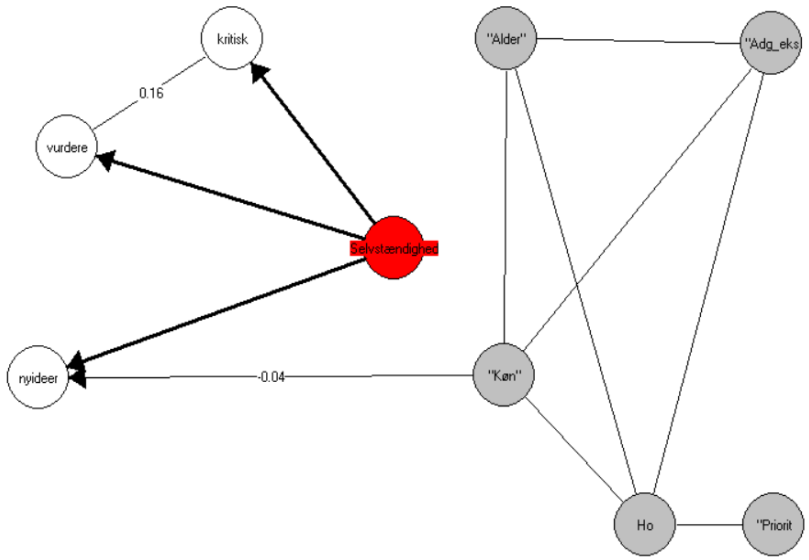
Kilde: Danmarks Evalueringsinstitut, 2021.
Note: n = 8.153. Den viste skala er ikke justeret for DIF.

Resultater for erhvervsakademier

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for DIF mellem spørgsmål 1 og baggrundsvariablen *Køn*. Der er lokal afhængighed mellem spørgsmål 1 og 3. Reliabiliteten for skalaen er på 0,79-0,80. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.23

Grafisk fremstilling af vægt på selvstændighed (for erhvervsakademier)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 400. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.16

Global homogenitetstest samt globale test for DIF (for erhvervsakademier)

Test	CLR	DF	p-værdi
Global homogenitetstest	7,5	28	1
Globale test for DIF:			
Køn	23,3	20	0,03
Alder	75,9	56	0,04
Adgangseksamen	20,9	28	0,83
Prioritet	23,6	28	0,7
Hovedområde	100,1	84	0,11

Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

TABEL 4.17

Score equation af skalaen *Vægt på selvstændighed* mht. køn (for erhvervsakademier), justering af sumscoren

Mænd	Kvinder
1.00	0.88
2.00	1.80
3.00	2.81
4.00	3.88
5.00	4.94
6.00	6.00
7.00	7.05
8.00	8.09
9.00	9.11
10.00	10.10
11.00	11.07

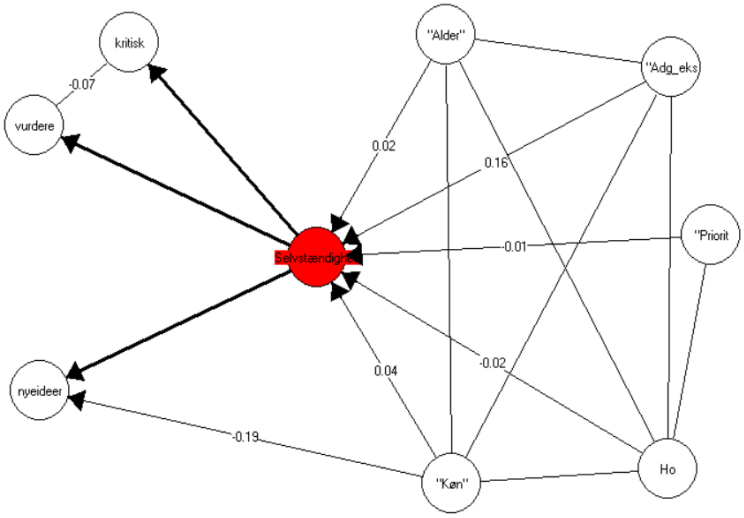
Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: Tabellen viser, hvordan sumscoren justeres på baggrund af for den identificerede DIF. Respondenter i kolonnen længst til venstre er referencegruppen, ift. hvilken de øvrige justeres.

Resultater for professionshøjskoler

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for DIF mellem spørgsmål 1 og baggrundsvariablen *Køn*. Der er lokal afhængighed mellem spørgsmål 1 og 3. Reliabiliteten for skalaen er på 0,76-0,78. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.24

Grafisk fremstilling af vægt på selvstændighed (for professionshøjskoler)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 600. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.18

Global homogenitetstest samt globale test for DIF (for professionshøjskoler)

Test	CLR	DF	p-værdi
Global homogenitetstest	16,5	30	0,98
Globale test for DIF:			
Køn	23,4	22	0,38
Alder	65,5	60	0,29
Adgangseksamen	31,5	30	0,39
Prioritet	45,8	30	0,03
Hovedområde	178,6	150	0,06

Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

TABEL 4.19

Score equation af skalaen *Vægt på selvstændighed* mht. køn (for professionshøjskoler), justering af sumscoren

Mænd	Kvinder
1.00	0.84
2.00	1.79
3.00	2.79
4.00	3.82
5.00	4.88
6.00	6.01
7.00	7.17
8.00	8.24
9.00	9.15
10.00	9.98
11.00	10.85

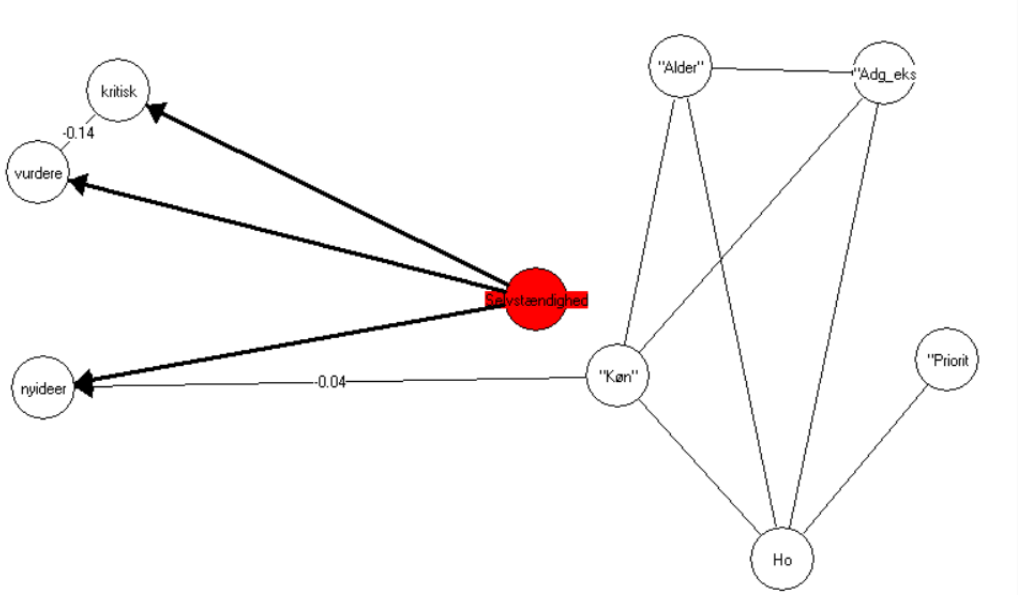
Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: Tabellen viser, hvordan sumscoren justeres på baggrund af for den identificerede DIF. Respondenter i kolonnen længst til venstre er referencegruppen, ift. hvilken de øvrige justeres.

Resultater for universiteter

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for DIF mellem spørgsmål 1 og baggrundsvariablen *Køn*. Der er lokal afhængighed mellem spørgsmål 1 og 3. Reliabiliteten for skalaen er på 0,74-0,79. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.25

Grafisk fremstilling af vægt på selvstændighed (for universiteter)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 500. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.20

Global homogenitetstest samt globale test for DIF (for universiteter)

Test	CLR	DF	p-værdi
Global homogenitetstest	38,5	31	0,17
Globale test for DIF:			
Køn	28,6	23	0,19
Alder	71,6	62	0,19
Adgangseksamen	37,1	31	0,21
Prioritet	29,9	31	0,52
Hovedområde	125,0	124	0,46

Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

TABEL 4.21

Score equation af skalaen *Vægt på selvstændighed* mht. køn (for universiteter), justering af sumscoren

Mænd	Kvinder
1.00	0.95
2.00	1.91
3.00	2.90
4.00	3.93
5.00	4.98
6.00	6.01
7.00	7.04
8.00	8.06
9.00	9.06
10.00	10.05
11.00	11.03

Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: Tabellen viser, hvordan sumscoren justeres på baggrund af for den identificerede DIF. Respondenter i kolonnen længst til venstre er referencegruppen, ift. hvilken de øvrige justeres.

4.5.6 Præsentation af skala for tydelig feedback

Skalaen for tydelig og velorganiseret undervisning er dannet på baggrund af nedenstående tre items:

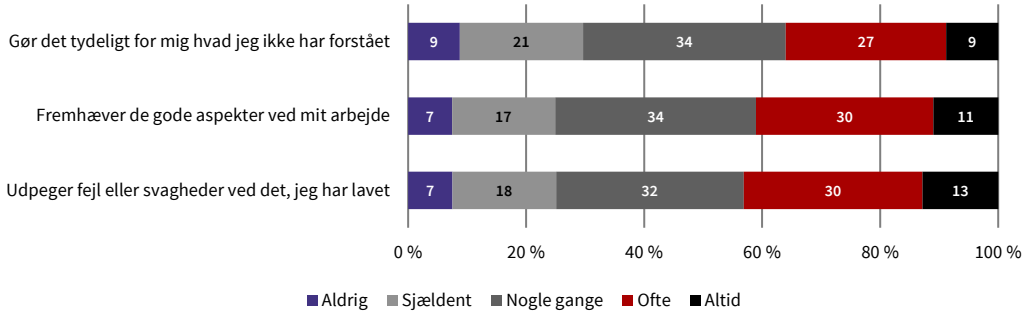
"Hvor ofte gør dine undervisere følgende i forbindelse med feedback på en opgave på studiet?" [Altid, Ofte, Nogle gange, Sjældent, Aldrig]

1. … udpeger fejl eller svagheder ved det, jeg har lavet
2. … fremhæver de gode aspekter ved mit arbejde
3. … før det tydeligt for mig hvad jeg ikke har forstået

FIGUR 4.26

Fordeling af enkeltspørgsmål, der indgår i skala for tydelig feedback

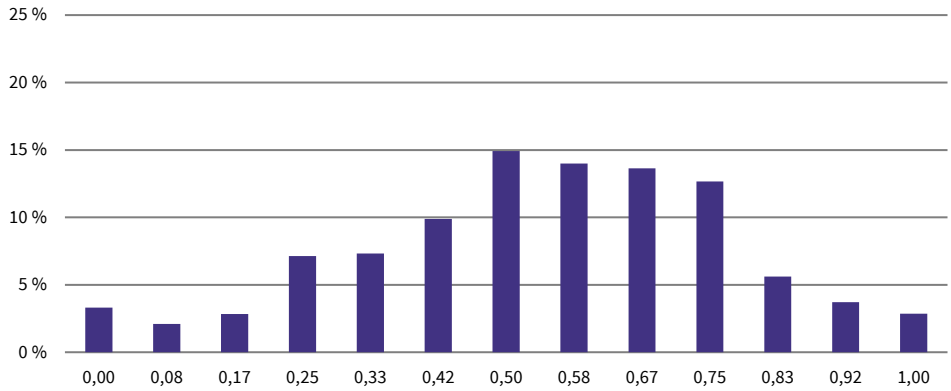
Hvor ofte gør dine undervisere følgende i forbindelse med feedback på en opgave på studiet?



Kilde: Danmarks Evalueringsinstitut, 2021.
Note: n = 7.382.

FIGUR 4.27

Fordeling af skala for tydelig feedback



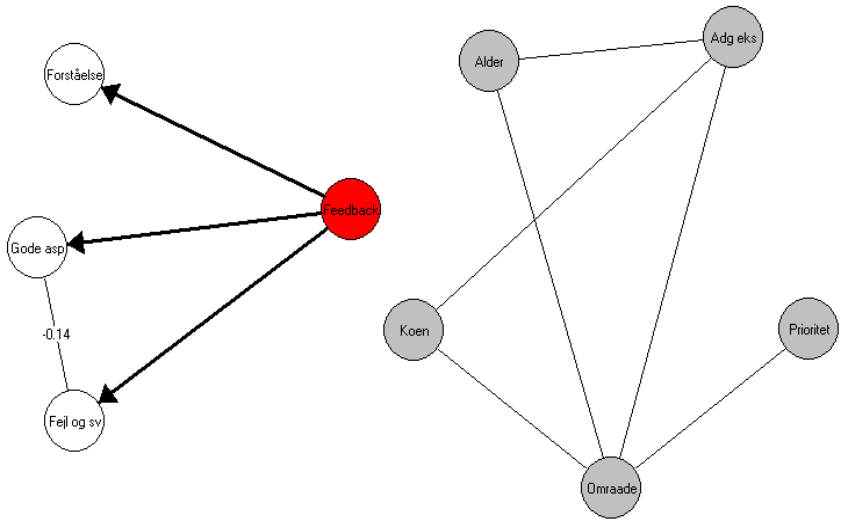
Kilde: Danmarks Evalueringsinstitut, 2021.
Note: n = 7.382.

Resultater for erhvervsakademier

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for lokal afhængighed mellem spørgsmål 1 og 2. Der er ingen DIF mellem spørgsmål og baggrundsvariable. Reliabiliteten for skalaen er på 0,82. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.28

Grafisk fremstilling af tydelig feedback (for erhvervsakademier)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 400. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.22

Global homogenitetstest samt globale test for DIF (for erhvervsakademier)

Test	CLR	DF	p-værdi
Global homogenitetstest	32,6	25	0,14
Globale test for DIF:			
Køn	21,8	25	0,65
Alder	56,7	50	0,24
Adgangseksamen	22,3	25	0,62
Prioritet	21,3	25	0,68
Hovedområde	93,8	75	0,07

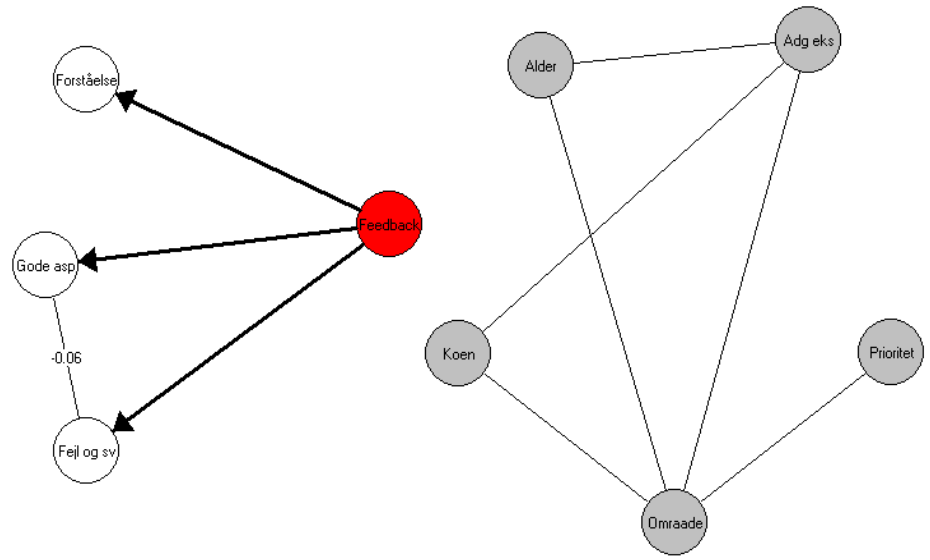
Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

Resultater for professionshøjskoler

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for lokal afhængighed mellem spørgsmål 1 og 2. Der er ingen DIF mellem spørgsmål og baggrundsvariable. Reliabiliteten for skalaen er på 0,79. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.29

Grafisk fremstilling af tydelig feedback (for professionshøjskoler)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 600. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.23

Global homogenitetstest samt globale test for DIF (for professionshøjskoler)

Test	CLR	DF	p-værdi
Global homogenitetstest	37,2	27	0,09
Globale test for DIF:			
Køn	27,2	27	0,46
Alder	81,7	54	0,01
Adgangseksamen	33,7	27	0,17
Prioritet	32,8	27	0,2
Hovedområde	168,5	135	0,03

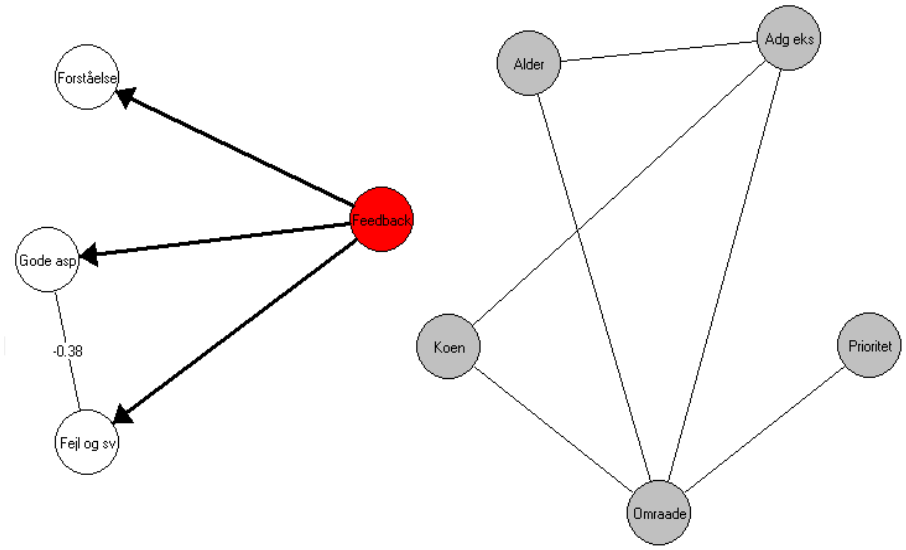
Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

Resultater for universiteter

Skalaen passer til en grafisk log-lineær Rasch-model, hvor der tages højde for lokal afhængighed mellem spørgsmål 1 og 2. Der er ingen DIF mellem spørgsmål og baggrundsvariable. Reliabiliteten for skalaen er på 0,79. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.30

Grafisk fremstilling af tydelig feedback (for universiteter)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 500. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.24

Global homogenitetstest samt globale test for DIF (for universiteter)

Test	CLR	DF	p-værdi
Global homogenitetstest	25,9	30	0,6
Globale test for DIF:			
Køn	34,9	30	0,25
Alder	64,9	60	0,31
Adgangseksamen	24,3	30	0,76
Prioritet	25,5	26	0,49
Hovedområde	130,0	120	0,25

Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

4.5.7 Præsentation af skala for feedforward

Skalaen for tydelig og velorganiseret undervisning er dannet på baggrund af nedenstående tre items:

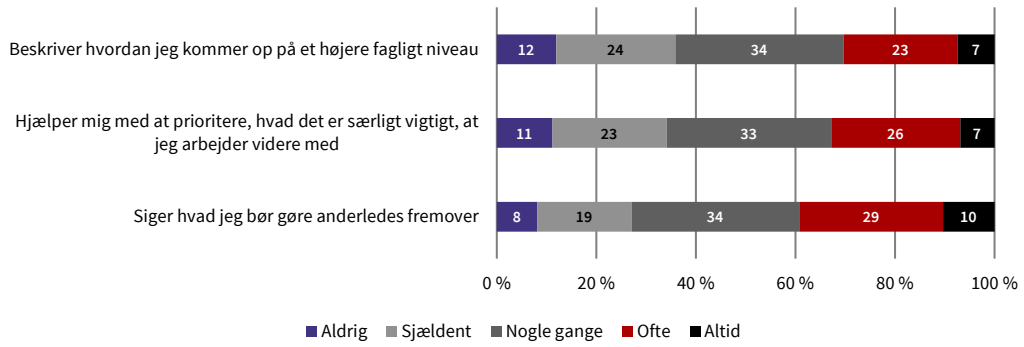
"Hvor ofte gør dine undervisere følgende i forbindelse med feedback på en opgave på studiet?" [Altid, Ofte, Nogle gange, Sjældent, Aldrig]

1. … siger hvad jeg bør gøre anderledes fremover
2. … hjælper mig med at prioritere, hvad det er særligt vigtigt, at jeg arbejder videre med
3. … beskriver hvordan jeg kommer op på et højere fagligt niveau

FIGUR 4.31

Fordeling af enkeltspørgsmål, der indgår i skala for feedforward

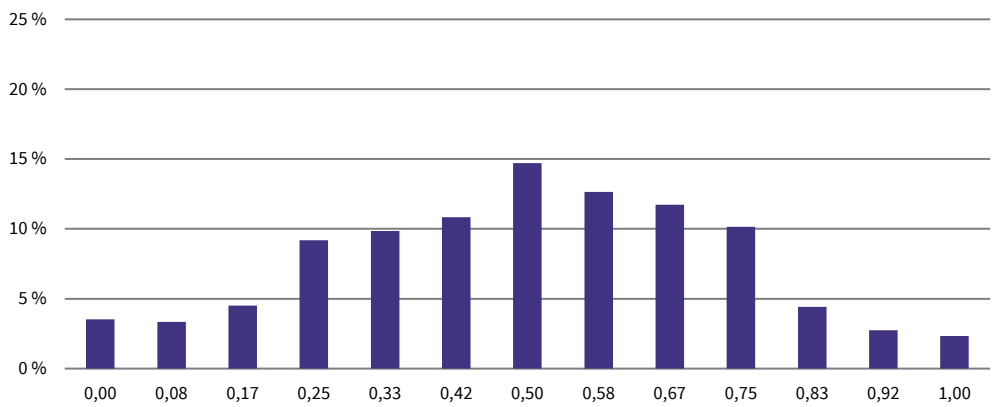
Hvor ofte gør dine undervisere følgende i forbindelse med feedback på en opgave på studiet?



Kilde: Danmarks Evalueringsinstitut, 2021.
Note: n = 7.407.

FIGUR 4.32

Fordeling af skala for feedforward



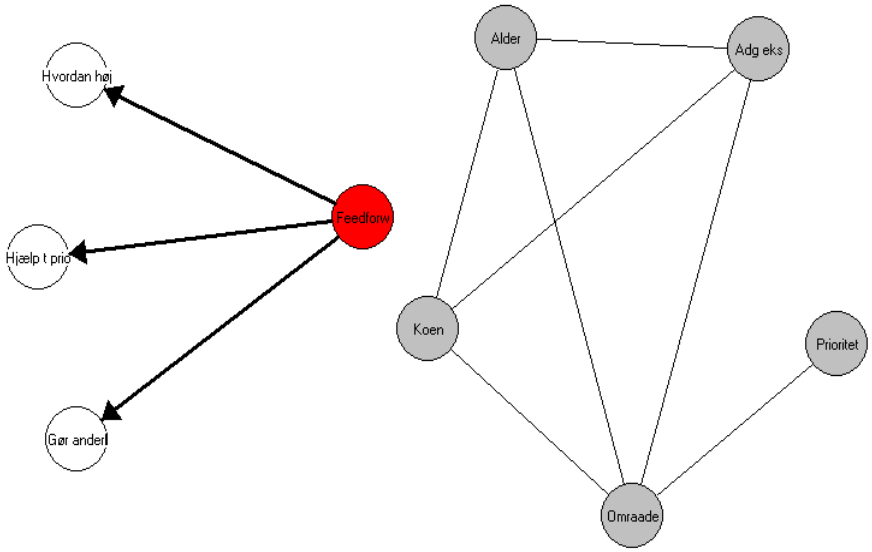
Kilde: Danmarks Evalueringsinstitut, 2021.
Note: n = 7.407.

Resultater for erhvervsakademier

Skalaen passer til en Rasch-model, da der hverken er DIF mellem spørgsmål eller lokal afhængighed mellem spørgsmål og baggrundsvariable. Reliabiliteten for skalaen er på 0,82. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.33

Grafisk fremstilling af feedforward (for erhvervsakademier)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 400. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.25

Global homogenitetstest samt globale test for DIF (for erhvervsakademier)

Test	CLR	DF	p-værdi
Global homogenitetstest	10,7	11	0,47
Globale test for DIF:			
Køn	8,8	11	0,64
Alder	6,2	22	1
Adgangseksamen	13,2	11	0,28
Prioritet	8,4	11	0,068
Hovedområde	21,1	33	0,95

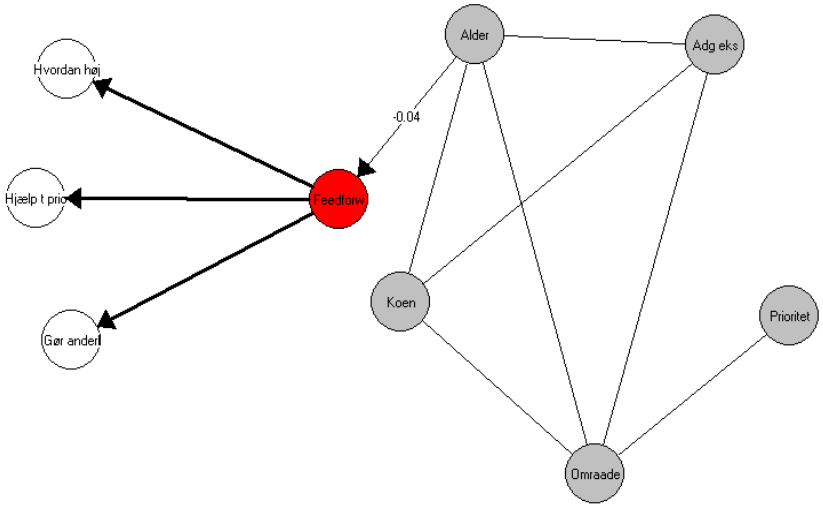
Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

Resultater for professionshøjskoler

Skalaen passer til en Rasch-model, da der hverken er DIF mellem spørgsmål eller lokal afhængighed mellem spørgsmål og baggrundsvariable. Reliabiliteten for skalaen er på 0,79-0,84. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.34

Grafisk fremstilling af feedforward (for professionshøjskoler)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 600. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.26

Global homogenitetstest samt globale test for DIF (for professionshøjskoler)

Test	CLR	DF	p-værdi
Global homogenitetstest	7,3	11	0,77
Globale test for DIF:			
Køn	5,5	11	0,9
Alder	12,3	22	0,95
Adgangseksamen	4,9	11	0,94
Prioritet	12,6	11	0,32
Hovedområde	56,7	55	0,41

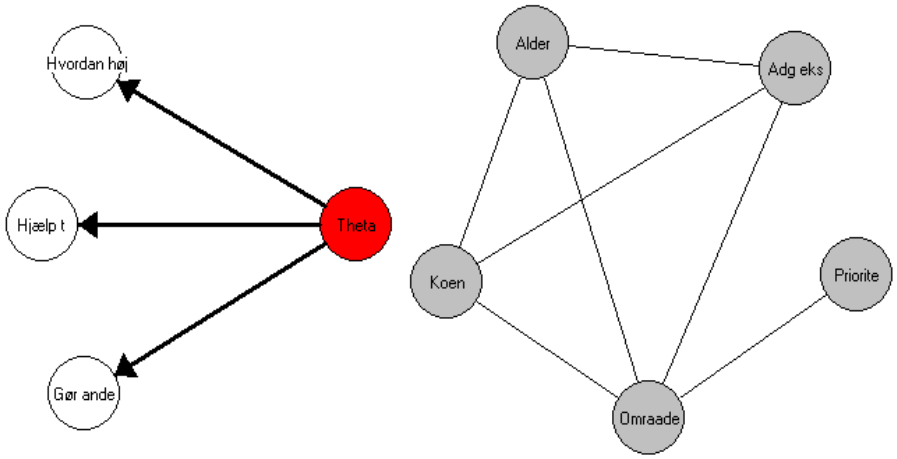
Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

Resultater for universiteter

Skalaen passer til en Rasch-model, da der hverken er DIF mellem spørgsmål eller lokal afhængighed mellem spørgsmål og baggrundsvariable. Reliabiliteten for skalaen er på 0,79. I nedenstående figur ses en grafisk fremstilling af skalaen.

FIGUR 4.35

Grafisk fremstilling af feedforward (for universiteter)



Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: n = 500. Tallene mellem spørgsmålene beskriver korrelationen vha. gammakoefficienter. Korrelationen med *Hovedområde* beskrives ved en chi^2-værdi, da variablen *Hovedområde* er en nominel variabel.

TABEL 4.27

Global homogenitetstest samt globale test for DIF (for universiteter)

Test	CLR	DF	p-værdi
Global homogenitetstest	15,2	11	0,17
Globale test for DIF:			
Køn	9,6	11	0,57
Alder	22,3	22	0,44
Adgangseksamen	11,9	11	0,37
Prioritet	20,2	11	0,04
Hovedområde	44	44	0,47

Kilde: Rasch-analyse af spørgeskemadata fra W3 2018.
Note: En Conditional Likelihood Ratio-test (CLR) tester, om spørgsmåls-parametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmåls-parametre for to lige store grupper af studerende, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Alle test er foretaget, efter der er taget højde for DIF og lokal afhængighed.

Appendiks A – Litteraturliste

Adcock, R., & Collier, D. (2001). Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *The American Political Science Review*, 95(3).

Catano, Victor M. and Harvey, Steve (2011). Student perception of teaching effectiveness: development and validation of the Evaluation of Teaching Competencies Scale (ETCS). *Assessment & Evaluation in Higher Education*, 36: 6, 701 — 717.

Chen, Y., & Hoshower, L. (2003). Student Evaluation of Teaching Effectiveness: An assessment of student perception and motivation. *Assessment & Evaluation in Higher Education, 28*(1), 71 -88.

Coffey, M., & Gibbs, G. (2001). The Evaluation of the Student Evaluation of Educational Quality Questionnaire (SEEQ) in UK Higher Education. *Assessment & Evaluation in Higher Education, 26*(1), 89 -93.

EVA (2020). *Fagligt engagement og social tilknytning - fra teoretisk begreb til empirisk måling*. København: Danmarks Evalueringsinstitut.

Kreiner, S., & Christensen, K. B. (2007). Validity and Objectivity in Health-Related Scales: Analysis by Graphical Loglinear Rasch Models. I: von Davier, M., & Carstensen, C. H. (red.), *Multivariate and Mixture Distribution Rasch Models. Extensions and Applications*. (s. 329-346). New York: Springer.

Kreiner, S. (2007). Validity and objectivity: Reflections on the role and nature of Rasch models. *Nordic Psychology, 59*(3), 268-298.

Kreiner, S. (2009). Om udvikling og afprøvning af pædagogiske test. I: C. Bendixen, & S. Kreiner (red.), *Test i folkeskolen*. (s. 43-82). København: Hans Reitzels Forlag.

Loukkola, Tia et al (2020). *Exploring higher education indicators*. European University Association.

Marsh, H. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology, 52*(1), 77-95.

Nicol, D. (2007, 29.-31. maj). *Principles of good assessment and feedback: Theory and practice* [keynote-præsentation]. REAP International Online-konference omhandlende *Assessment Design for Learner Responsibility*.

Pintrich, P. R. (2003). A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Contexts. *Journal of Educational Psychology*, *95*(4), 667-686.

Udvalget for Kvalitet og Relevans I de Videregående Uddannelser (2014). *Spørgeskemaundersøgelse blandt studerende og undervisere på de videregående uddannelser. Metodebilag.* København: Udvalget for Kvalitet og Relevans I de Videregående Uddannelser.

Måling af undervisningskvalitet på de videregående uddannelser

© 2021 Danmarks Evalueringsinstitut

Citat med kildeangivelse er tilladt

Publikationen er kun udgivet i elektronisk form på: www.eva.dk

Foto: Thomas Søndergaard

ISBN (www) 978-87-7182-481-0

Danmarks Evalueringsinstitut (EVA) gør uddannelse og dagtilbud bedre. Vi leverer viden, der bruges på alle niveauer – fra institutioner og skoler til kommuner og ministerier.

DANMARKS EVALUERINGSINSTITUT

T 3555 0101
E eva@eva.dk
H www.eva.dk