

Validering af Læringsbarometer 2018

I dette notat gennemgås hovedresultaterne fra valideringen af Læringsbarometeret. Valideringen skal undersøge om de ni indikatorer i Læringsbarometeret korrekt afspejler de fænomener, som de repræsenterer. I dette notat gennemgås de overordnede resultater fra valideringen af Læringsbarometeret.

Den overordnede konklusion i valideringen af Læringsbarometeret er, at de ni teoretiske indikatorer, som Læringsbarometeret er bygget op omkring, er valide og kan genfindes i det danske data.

Valideringen benytter både traditionelle valideringsmetoder og mere avancerede metoder, som tester samvariation, intern konsistens og påvirkning fra udefrakommende faktorer inden for hver af de ni indikatorer i Læringsbarometeret.

I notatet gennemgås først resultaterne fra valideringen af pilottesten af Læringsbarometeret. Dernæst dykkes ned i selve Læringsbarometeret, som kontrolleres i forhold til syv krav om validitet.

I valideringen af Læringsbarometeret er der udført en række tests, og i dette notat er det kun de vigtigste pointer der fremhæves. Materialet fra valideringen kan findes i de vedlagte bilag: *IRT valideringsresultater* (for hver af de ni indikatorer) og *Item-item og item-skala validering*.

Validering af pilottesten

Uddannelses- og Forskningsministeriet valgte det finske spørgeskema til studerende, LEARN, som udgangspunkt for Læringsbarometeret efter enstemmig anbefaling fra et ekspertpanel og repræsentanter for uddannelsesinstitutioner og studerende. LEARN er tidligere valideret og oversat til en dansk kontekst jf. Hermann et al (2016).¹ De ni læringsindikatorer, som indgår i Læringsbarometeret udspringer af LEARN og var således allerede valideret på universitetsniveau i en række danske og internationale studier.

Konkret blev der taget udgangspunkt i Hermann et al's danske oversættelse af LEARN, som efterfølgende blev justeret efter anbefaling fra eksperterne.

For at teste det justerede spørgeskema i den samlede danske uddannelsessystem blev der gennemført en række pilottests.

Den samlede pilottest bestod af en kvalitativ test med semistrukturerede interviews blandt 29 studerende (heraf 6 engelsksprogede) og en kvantitativ test som blev sendt ud til studerende på 13 institutioner.

Forud for den kvalitative test havde ekspertpanelet, som har deltaget i udviklingen af Læringsbarometeret, og Uddannelses- og Forskningsministeriet (UFM) udpeget særlige fokusområder for interviewene. Resultaterne viste at de fleste studerende

¹Baggrunden for valg af LEARN som grundlag mv. er beskrevet i en række notater, som kan findes på Uddannelses- og Forskningsministeriets hjemmeside

20. december 2019

Uddannelses- og
Forskningsministeriet
Analyse

Børsgade 4
Postboks 2135
1015 København K
Tel. 3392 9700

www.ufm.dk

CVR-nr. 1680 5408

Ref.-nr.
90468

forstod spørgsmålene. Enkelte steder var der problemer med at forstå meningen med enkelte ord eller sætninger.

Danmarks Evalueringsinstitut (EVA) har udarbejdet en opsummering af den kvalitative pilottest: <https://ufm.dk/uddannelse/statistik-og-analyser/laeringsbarometer/filer/bilag-2-kvalitativ-pilottest-eva.pdf>

I maj 2018 blev spørgeskemaet testet i en kvantitativ pilottest. Epinion sendte spørgeskemaet ud til 16.200 studerende fra 13 institutioner. 3.500 studerende svarede på spørgeskemaet, hvilket giver en svarprocent på 21,6 procent. Resultaterne fra den kvantitative pilottest er brugt til efterfølgende valideringsanalyser.

Resultaterne fra den kvantitative pilottest er valideret på tre niveauer:

- Enkeltspørgsmål er testet i forhold til indikatorerne
- De ni indikatorer fra LEARN er testet
- Den samlede model er testet med ekstern validitet og faktoranalyse

Uddannelses- og
Forskningsministeriet

Resultaterne viser overordnet at spørgeskemaet empirisk passer til de teoretiske forventninger. På skalaniveau har syv ud af ni indikatorer tilfredsstillende værdier for intern konsistens. Resultaterne fra en konfirmativ faktoranalyse bekræfter i store træk, at de teoretiske dimensioner i LEARN kan genfindes empirisk i de indsamlede pilotdata, og at de enkelte spørgsmål passer til de tilhørende indikatorer.

Valideringsresultaterne blev opsummeret i et notat til ekspertpanelet og referencegruppen. Af notatet fremgår også mulige ændringsforslag til spørgeskemaet på baggrund af valideringerne: <https://ufm.dk/uddannelse/statistik-og-analyser/laeringsbarometer/filer/overordnede-valideringsresultater.pdf>.

Validering af Læringsbarometeret

Efter indsamlingen af Læringsbarometeret i efteråret 2018 er data valideret igen. Der er stadig fokus på, at genfinde de teoretiske indikatorer i data. I denne validering tages der udgangspunkt i RASCH-modellens krav til robuste tests, disse krav gennemgås nedenfor. En del af de tests, som blev gennemført i pilottesten, er også gentaget i Læringsbarometeret. I det nedenstående indgår de løbende som en del af gennemgangen af RASCH-modellen.

De gennemførte valideringer udgør et omfattende talmateriale, som findes vedlagt dette notat. I den nedenstående gennemgang af valideringen fremhæves relevante resultater, inkl. opmærksomhedspunkter i forhold til de enkelte indikatorer.

Syv krav til validitet

I RASCH modellen stilles der syv krav til validitet². De syv krav fungerer som en tjekliste, der bruges til at undersøge om Læringsbarometeret måler det, det er designet til at måle.

² Kreiner, 2009, *Om udvikling og afprøvning af pædagogiske tests*

Kravene til Læringsbarometeret er undersøgt med både traditionelle og mere avancerede valideringsmetoder. De forskellige metoder belyser kravene til validitet på forskellige måder. Der er ikke som sådan en mere korrekt valideringsmetode, men metoderne supplerer hinanden.

I det nedenstående gennemgås resultaterne af valideringen i forhold til de syv krav. I gennemgangen er kravene sat ind i Læringsbarometerets kontekst.

Unidimensionalitet

En indikator skal måle ét og kun ét fænomen. Eksempelvis skal konstruktiv feedback ikke også måle graden af støtte fra medstuderende.

Kravet om unidimensionalitet kan undersøges med flere forskellige metoder. I det nedenstående vises *eigenvalues* for alle indikatorerne i Læringsbarometeret. Eigenvalues viser, hvor meget information, der ligger i, at inkludere flere items (spørgsmål) i et indeks. For alle indikatorerne i Læringsbarometeret er eigenvalues markant højere for det første item end de efterfølgende items, jf. tabel 1. Det tyder på, at spørgsmålene i de enkelte indikatorer i høj grad måler det samme fænomen, og derfor er en unidimensional model et godt valg.

Uddannelses- og
Forskningsministeriet

Tabel 1

Eigenvalues for indikatorerne i Læringsbarometeret

	1. item	2. item	3. item	4. item	5. item
Overensstemmelse	2,61	0,59	0,46	0,35	
Konstruktivfeedback	3,06	0,44	0,29	0,21	
Interesse og motivation	2,80	0,52	0,40	0,29	
Støtte fra medstuderende	2,55	0,60	0,46	0,39	
Underviserinteraktion	2,97	0,70	0,50	0,48	0,34
Læring for forståelse	3,01	0,64	0,56	0,44	0,35
Organiseret læring	2,58	0,54	0,51	0,38	
Overfladelæring	2,54	0,67	0,51	0,28	
Dybdelæring	2,54	0,72	0,44	0,30	

Anm.: Indikatorernes eigenvalues er beregnet med en graded response model i SAS.

Kilde: Uddannelses- og Forskningsministeriet på baggrund af data fra Læringsbarometer (2018)

Unidimensionaliteten kan også undersøges med andre valideringsmetoder, f.eks. "item-item" og "item-skala" korrelationer.

Monotonicitet

Jo bedre en oplevelse en studerende har af et givet fænomen, jo højere skal den studerende vurdere fænomenet i spørgeskemaet. Eksempelvis skal studerende, der oplever at få god konstruktiv feedback, svare højere på spørgsmålene om konstruktiv feedback, end studerende, der oplever mindre god konstruktiv feedback.

Kravene om monotonicitet undersøges ved at måle samvariationen mellem det enkelte spørgsmål med de øvrige spørgsmål i indikatoren, en såkaldt item-skala test. Cronbachs Alpha, som er vist i tabel 2 nedenfor, er et mål for dette, som kombinerer item-skala tests fra alle spørgsmål i indikatorerne. En tommelfingerregel er, at hvis Cronbachs Alpha er over 0,70, udviser skalaen intern konsistens. Da det er tilfældet, vurderes det, at kravet om monotonicitet også er overholdt.

Tabel 2

Item-Skala analyse af Læringsbarometerets indikatorer

	Cronbachs Alpha
Overensstemmelse	0,77
Konstruktivfeedback	0,87
Interesse og motivation	0,81
Støtte fra medstuderende	0,75
Underviserinteraktion	0,78
Læring for forståelse	0,79
Organiseret læring	0,76
Overfladelæring	0,76
Dybdelæring	0,74

**Uddannelses- og
Forskningsministeriet**

Kilde: Uddannelses- og Forskningsministeriet på baggrund af data fra Læringsbarometer (2018)

Det skal bemærkes, at et enkelt underspørgsmål på indikatoren for *Underviserinteraktion* (spørgsmål 2: "Vi har i nogen grad indflydelse på, hvad underviserne lægger vægt på") udviser en forholdsvis lav sammenhæng med alle de øvrige spørgsmål i indikatoren. Spørgsmålet har gamma-koefficienter på under 0,30-0,37 i item-item analysen, mens gamma-koefficienterne i Læringsbarometeret ligger mellem 0,30 og 0,66). Spørgsmålene har dog stadig stor samvariation, og det vurderes derfor ikke at være et problem i forhold til konstruktionen af indikatoren.

Konsistens

Studerende, som svarer højt på ét spørgsmål inden for en indikator, skal også have tendens til at svare højere på andre spørgsmål inden for indikatoren. Dvs. besvarelserne inden for den enkelte indikator skal være positivt korrelerede.

Kravet om konsistens undersøges med både item-item og item-skala tests. Som vist under *monotonicitet* er der korrelation mellem de enkelte spørgsmål, og kravet om konsistens vurderes derfor overholdt, med samme opmærksomhedspunkt i forhold til spørgsmål 2 for *Underviserinteraktion*.

Kriterievaliditet

Hvis man ved, at resultaterne fra en undersøgelse skal hænge sammen med baggrundsvariable eller andre tidligere undersøgelsesresultater, bør det undersøges, at dette er tilfældet. Dvs. resultaterne fra Læringsbarometeret skal kunne spejles i tidligere resultater fra undersøgelser med LEARN spørgeskemaet.

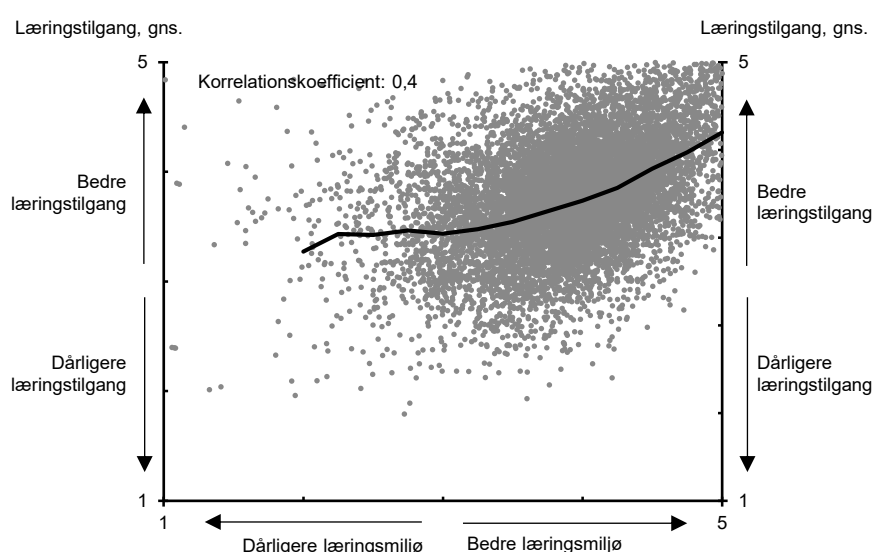
Læringsbarometeret kan især spejles i de konklusioner, der ligger i den forskning, som også har været en del af udviklingen af Læringsbarometeret. Læringsbarometeret ligger sig i høj grad også op af resultaterne fra forskningen. Det kommer

blandt andet til udtryk, når man ser på indikatorernes indbyrdes forhold. Det er f.eks. typisk de samme indikatorer, der scorer relativt højt på skalaen.

Et særligt fokuspunkt i forskningen, er sammenhængen mellem de indikatorer, der beskriver læringsmiljøet, og dem der beskriver de studerendes læringstilgange. Denne sammenhæng genfindes også i Læringsbarometeret, jf. figur 1. Disse sammenhænge undersøges i Uddannelses- og Forskningspolitisk Redegørelse 2019, hvor flere af forskningens resultater genfindes.

Figur 1

Sammenhæng mellem de studerendes vurdering af læringsmiljø og læringstilgang, gennemsnit, 2018



Uddannelses- og
Forskningsministeriet

Anm.: Hvert punkt i figuren angiver en persons gennemsnitsbesvarelse. Af diskretionshensyn er enkelte observationer ikke vist i figuren og de enkelte gennemsnit er forskudt marginalt og tilfældigt i visningen. Den sorte linje viser gennemsnittet på læringstilgangsindikatorerne, givet besvarelsen på læringsmiljøindikatorerne (f.eks. gennemsnittet på læringstilgange givet at man i gennemsnit svarer i intervallet 2,75-3,25 på læringsmiljøindikatorerne). Linjen starter ved 2 på læringsmiljøaksen grundet få observationer under 2.

Kilde: Uddannelses- og Forskningsministeriet på baggrund af data fra Læringsbarometer (2018).

Sammenhængen mellem de studerendes akademiske evner og deres vurdering på læringsindikatorerne viser også samme tendenser som tidligere undersøgelser³, nemlig at der er en positiv sammenhæng mellem karakterer og de studerendes vurdering på indikatorerne det beskriver Læringsmiljøet.

Homogenitet

Hvis studerende har lettere ved at erklære sig enig i ét spørgsmål inden for fænomenet frem for et andet, skal dette være tilfældet både for studerende, som har en god oplevelse af fænomenet og for studerende som har en dårlig oplevelse af fænomenet. Dvs. spørgsmål, som er 'svære' at erklære sig enige i, skal konsekvent vurderes lavere end spørgsmål, som er lette at erklære sig enig i.

³ Herrmann et al (2016): Measuring perceptions of the learning environment and approaches to learning validation of the learn questionnaire

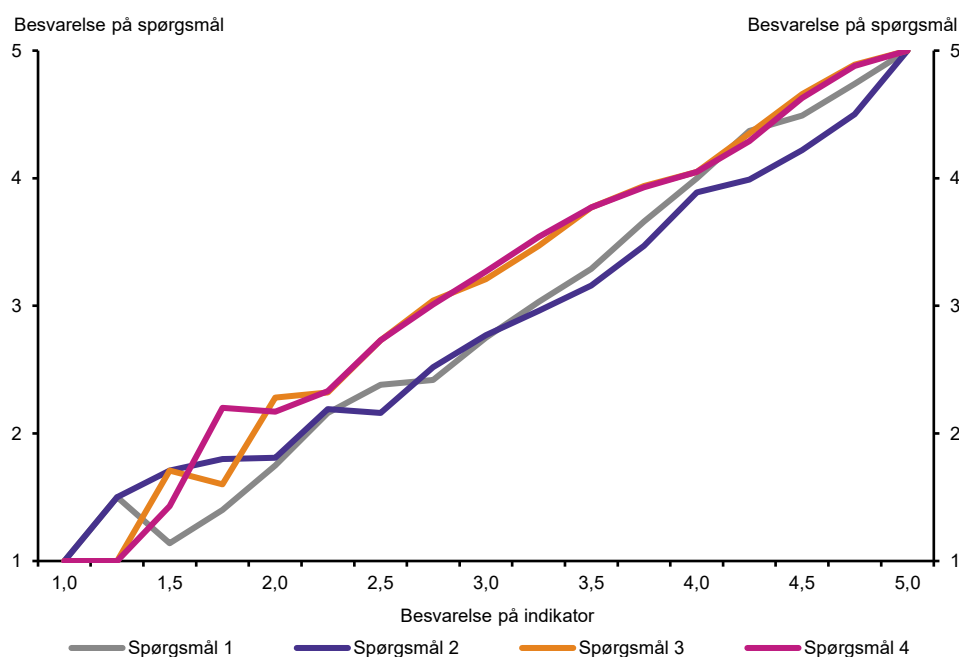
I testen for homogenitet sammenlignes besvarelserne på de enkelte spørgsmål med besvarelsen på indikatoren som helhed. Figur 2 viser testen for homogenitet for indikatoren *overensstemmelse*. For at en indikator er homogen, skal kurven for 'lette' spørgsmål ligge over kurven for 'sværere' spørgsmål, med forbehold for statistiske usikkerheder. Som figur 2 viser er spørgsmål 3 det sværeste, og det er sværest er erklære sig enig i, uanset hvordan man svarer på skalaen samlet set. Det er et udtryk for homogenitet i spørgsmålene. Det bemærkes desuden, at der er udsving i den lave ende af skalaen, hvor der også er relativt få observationer.

De øvrige indikatorer i Læringsbarometeret viser overordnet samme tegn på homogenitet som indikatoren for *overensstemmelse*. Der er enkelte undtagelser under indikatorerne *støtte fra medstuderende*, *læring for forståelse* og *overfladelæring*, som dog kan skyldes statistisk usikkerhed, og derfor vurderes det ikke til at være et problem i forhold til indikatorernes robusthed.

Uddannelses- og
Forskningsministeriet

Figur 2

Test af homogenitet for indikatoren *overensstemmelse*



Anm.: Evt. spørgsmålsformuleringer

Kilde: Uddannelses- og Forskningsministeriet på baggrund af data fra Læringsbarometer (2018).

Lokal uafhængighed

Evnen til at besvare et spørgsmål må ikke hænge sammen med formuleringen af eller evnen til at besvare andre spørgsmål.

Kravet om lokal uafhængighed er ikke relevant i forhold til Læringsbarometeret. Det skyldes, at de enkelte spørgsmålsformuleringer er selvstændige og også kunne fungere uden for Læringsbarometeret.

Ingen differentiell item-funktion (DIF)

Indikatorerne må ikke være afhængige af andet end det fænomen, man forsøger at måle. Dvs. personer, der har samme oplevelse af en uddannelse, må ikke svare forskelligt på Læringsbarometeret, fordi de tilhører forskellige grupper, f.eks. i forhold til køn eller alder. Kravet gælder også på de enkelte underspørgsmål.

Differentiell item-funktion (DIF) undersøges ved at opstille en række regressionsanalyser som har nedenstående form:

$$\text{underspørgsmål} = \text{indikatorværdi} + \text{baggrundsvARIABLE}$$

Uden DIF, forklarer indikatorværdien⁴ hele variationen i underspørgsmålet. Det vil sige, at hvis der er baggrundsvARIABLE som er signifikante i regressionsanalysen, er det et tegn på DIF.

Det skal bemærkes, at testen for DIF dermed er modsatrettet i forhold til normale regressionsanalyser, da vi søger at afvise, at baggrundsvARIABLE har betydning for besvarelsen på underspørgsmålene. Derved får antallet af observationer en betydning, fordi flere observationer giver mere statistisk power. Figur 2 viser, at antallet af p-værdier, på tværs af alle baggrundsvARIABLE og alle underspørgsmål, som ligger under 5 procent signifikansniveauet stiger med antallet af observationer.

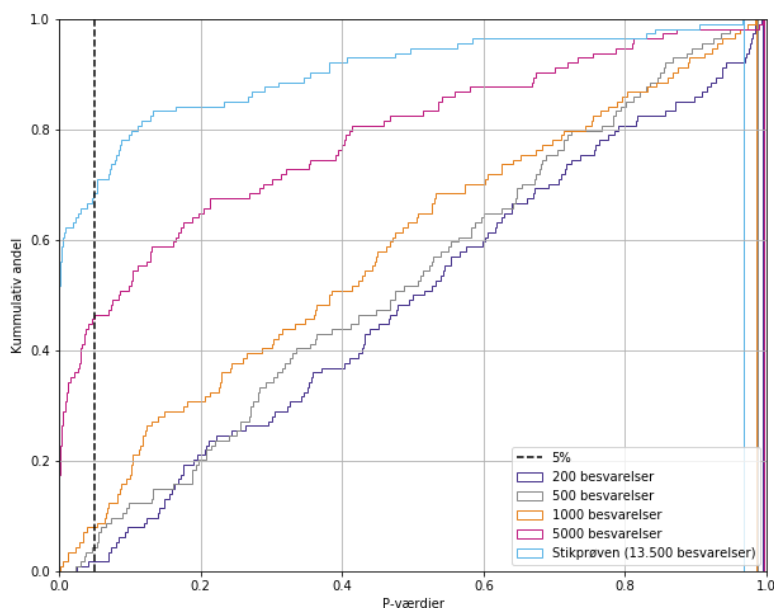
Der er ikke en fast regel for, hvor mange observationer, der skal inkluderes i en test af DIF. I valideringen af Læringsbarometeret tages der udgangspunkt i stikprøver med 500 observationer.

Uddannelses- og
Forskningsministeriet

⁴ Indikatorværdien er beregnet i en graded response model, og er et estimat af den enkelte persons latente værdi på indikatoren.

Figur 3

Fordeling af p-værdier i test af differentiell item-funktion



Anm.: Den stiplede linje markerer 5 procent signifikansniveauet

Kilde: Uddannelses- og Forskningsministeriet på baggrund af data fra Læringsbarometer (2018).

Uddannelses- og
Forskningsministeriet

Der er i alt taget 200 stikprøver af 500 observationer (figur 2 viser kun en enkelt af disse). På tværs af 38 spørgsmål i Læringsbarometeret og 7 kategorier af baggrundsvariable giver det 53.200 p-værdier, som alle er kontrolleret i forhold til DIF. I alt er der undersøgelsen tegn på DIF i godt 10 procent af tilfældene. Dette vurderes ikke til at være et problem for indikatorerne i Læringsbarometeret.

Det er også undersøgt, om der er systematik i, hvilke spørgsmål eller baggrundsvariable der typisk udviser tegn på DIF. Her er der to spørgsmål under indikatoren *læring for forståelse*⁵, som skiller sig ud. Spørgsmålene udviser tegn på DIF for erhvervsakademierne.

⁵ Spørgsmålene er 'Undervisningen hjælper mig til at kunne forholde mig kritisk til ny viden eller forskning, jeg bliver præsenteret for' og 'Underviserne lægger vægt på, at vi kan anvende det lærte i en ny kontekst eller på praktiske problemstillinger'. Svarmuligheder: meget enig, enig, hverken enig eller uenig, uenig, meget uenig.