

Har PISA tabt pusten?



Inge Henningsen, eXbus, DPU,
Aarhus Universitet

Abstract. For at kunne levere de rangordninger af lande som tydeligvis efterspørges hos beslutningstagerne og i offentligheden, har PISA valgt at basere sine analyser på et sæt af oversimplificerede modeller. I artiklen diskuteres nogle af de konsekvenser det får når de statistiske og datatekniske krav får forrang frem for de faglige samtidig med at der hos mange af deltagerne i de nationale PISA-konsortier synes at være en svigtende forståelse for de metodiske begrænsninger som de valgte analysemodeller medfører. Som eksempler ses der på undersøgelser af hvordan drenge og piger klarer forskellige opgavetyper, og på (mis)brug af kompetenceniveauerne til på papiret at identificere grupper af elever der ikke læser godt nok til at gennemføre en ungdomsuddannelse.

I kølvandet på offentliggørelsen af resultaterne fra PISA 2009 har resultaterne ikke givet de sædvanlige store overskrifter. Årsagerne kan være mange. S og SF har altid haft et ambivalent forhold til PISA. Efter ti år med regeringsmagten har de borgerlige svært ved at udnytte dårlige eller halvsloje resultater politisk. Samtidig sker der ikke rigtig noget nyt i PISA. Ranglisterne med lande, det primære formål med PISA, har vist sig alt for stationære til at de fortsat rigtig er interessante. At Danmark går fra en naturfagsscore på 496 i 2006 til 499 i 2009, eller at man flytter sig fra 15.- til 18.-pladsen i matematik, sætter ikke længere sindene i kog.

Trods dette spiller internationale læse/regneundersøgelser en væsentlig rolle i skoledebatten og har betydelig indflydelse på indretningen af de nationale skolesystemer. Det er derfor vigtigt at data fra sådanne undersøgelser bliver analyseret og fortolket professionelt og retvisende.

I forhold til PISA er der rejst mange spørgsmål vedrørende sammenligneligheden af testresultater på tværs af landegrænser og særligt om det er muligt at lave opgaver der er kulturelt og sprogligt neutrale, se fx Goldstein (2004) og Hopmann, Brinek & Retzl (2007). Der er imidlertid også problemer i meget af den nationale følgeforskning der udføres af PISA-konsortierne.

I takt med at interessen for de internationale rangeringer falder, rettes opmærksomheden andre steder hen. I Danmark i stigende grad mod udviklingen over tid i de

nationale scorer og undersøgelser af danske elevers præstationer ved forskellige opgavetyper. Men disse sammenligninger indeholder væsentlige metodologiske problemer som stammer direkte fra de greb som PISA har benyttet for at kunne retfærdiggøre udarbejdelsen af de nationale ranglister.

Raschmodellen

Problemet med at sikre sammenlignelighed på tværs af kulturer og uddannelsessystemer søges i PISA løst ved at udvælge opgavesæt inden for hvert af de faglige domæner matematik, læsning og naturvidenskab der på et empirisk grundlag (pilottestning) er valgt så de alle tester "de samme" færdigheder. Dette formaliseres ved at opgavebesvarelserne for hvert fagligt domæne skal kunne beskrives ved en endimensional item respons model (Raschmodel). Herved sikrer man entydigt bestemte nationale ranglister fordi PISA-scoren på meningsfyldt måde kan baseres på summen af rigtige besvarelser. Dette stiller imidlertid nogle helt særlige krav til de anvendte opgavesæt. Her skal alle opgaver nemlig "måle det samme". Det kan formaliseres i en række krav til opgaverne i sættet (se fx Kreiner, 2007):

- De målte færdigheder skal kunne repræsenteres ved en enkelt latent intervalskala.
- Ingen "differentiel item-funktion", dvs. at items (opgaver) skal fungere på samme måde for alle respondenter. Der må fx hverken være "pigeopgaver" eller "drengopgaver" i sættet, og items skal fungere på samme måde i forskellige lande.
- Sammenligninger mellem respondenter skal være principielt uafhængige af hvilke delsæt af opgaverne der inkluderes i prøven.

Vi skal se på hvad dette betyder for opgaverne. I PISA beskrives matematikdomænet eksempelvis ved *matematiske situationer, kontekster, idéområder, discipliner* og *kompetencer* (se Lindenskov & Weng, 2010, s. 85). Men uanset hvilke *idéområder* (rum og form, forandringer og sammenhænge, størrelser, usikkerhed) eller *matematiske discipliner* (tal, algebra, geometri, sandsynlighed, statistik) en opgave i PISA kommer fra, så skal den måle "dette samme" og fx være lige svær for to elever der er lige dygtige. Bemærk at dette naturligvis ikke gælder for alle matematikopgaver, men det er et krav til den samling af opgaver der sammen kan indgå i et prøvesæt hvis det skal analyseres med en Raschmodel.

I PISA forsøger man at opnå den homogenitet som Raschmodellen kræver, ved at lave en eller flere pilottest hvor man lader et antal elever regne opgaverne, og så fjerne de opgaver der ikke "passer sammen" med de andre. Hvis der er nogle opgaver som elever i de nordiske lande er særlig gode til at besvare, så kan de ikke være med

i sættet. Tilsvarende kan man hverken have “dreng-” eller “pigeopgaver” med i et opgavesæt. Man kan dog godt have opgaver der favoriserer elever fra et bestemt land. Men hvis man har det, så skal alle opgaverne favorisere elever fra dette land (oven i købet i samme grad alle sammen).

Ikke mange er i tvivl om at matematik og matematikkundskaber har mange forskellige dimensioner. Et forsøg på at beskrive dette er fx KOM-rapportens otte kompetencer (Undervisningsministeriet, 2002). De fleste vil også medgive at forskellige elever har deres styrke på forskellige områder, ligesom matematikundervisningen ikke lægger vægt på de samme ting i alle lande. Man kan også sige at eksistensen af alle de mange opgaver som ikke passer ind i PISA's Raschmodeller, er et empirisk “bevis” på at matematikkundskaber ikke kan beskrives med en “endimensionel latent variabel”. Raschanalysen benægter ikke dette. Den kræver bare at alle opgaver i et prøvesæt undersøger den samme latente dimension mens de lader alle de andre dimensioner ligge. Som det fremgår af det foregående, så lægger brug af Raschmodeller nogle betydelige begrænsninger på hvilke opgaver man kan inkludere i et prøvesæt. Brug af disse modeller har imidlertid også meget indsnævrende konsekvenser for hvilke forskningsspørgsmål man kan stille til data fra PISA-undersøgelserne.

Ikke-spørgsmål

Det følger af diskussionen i det foregående afsnit at visse forskningsspørgsmål ikke kan besvares når man analyserer data fra PISA, idet disse spørgsmål refererer til dimensioner som bevidst er fjernet ved testkonstruktionen. Eksempler på ikke-spørgsmål fra domænet matematik kunne være:

- Hvad karakteriserer items hvor piger scorer (relativt) højt?
- Hvad karakteriserer opgaver som er specielt vanskelige for de svage elever?
- Er der faglige områder hvor de danske elever er specielt stærke/svage?

Disse spørgsmål forudsætter nemlig implicit at matematik har flere dimensioner, og at det reflekteres i det anvendte prøvesæt. I modsætning hertil står konstruktionen af PISA-opgaverne hvor man gennem omfattende pilottestning forsøger at udelukke opgaver hvis de fungerer forskelligt fra land til land eller forskelligt mellem de to køn eller på anden vis er differentielle. Tilbage bliver et ensartet sæt af opgaver der tester en bestemt, men uspecificeret dimension i matematikkunnen.¹

1 Resultatet fremstilles imidlertid som et universelt gyldigt mål for matematikkunnen der kan bruges til rangordning af lande og forskellige grupper af elever.

“PISA fokuserer på matematik-kompetencer, som kan siges at være relevante for ethvert voksent menneske i et højt teknologisk demokratisk samfund.” (Lindenskov & Weng, 2004, s. 38)

Lykkes det for PISA?

Allerup (2005) har ved undersøgelse af læseresultater fra 2000 og 2003 vist at de brugte skalaer er skæve i forhold til køn. For nylig har Kreiner (2011) i en omfattende afprøvning af PISA's læseopgaver fra 2003 påvist eksistensen af differentiell item-funktion mellem lande. Han viser konkret at man kan finde to del sæt af de anvendte læseopgaver der placerer Danmark som henholdsvis nr. 3 og nr. 42 blandt alle lande, altså at landenes placering på PISA's rangliste er en funktion af hvilke opgaver der tages med i opgavesættet. Når PISA alligevel viser relativt konstante resultater fra år til år, kan det skyldes at opgavesættene i høj grad er de samme – og at nye opgaver vælges sådan at de passer med de gamle – men Kreiners resultater viser at et andet grundlæggende sæt af opgaver kunne have ført til en anden rangering.

Påvisningen af differentiell item-funktion i PISA's opgaver var egentlig ikke mit formål med at diskutere Raschmodellen. Derimod vil jeg beskrive min forundring over at PISA's forskere i mange af deres artikler forholder sig så afslappet til de grundlæggende modelkrav i PISA at de bygger deres forskning på at de ikke er opfyldt. Jeg har udvalgt nogle eksempler fra rapporten om PISA 2009.

I afsnittet om naturvidenskab diskuterer Sørensen & Davidsson forskelle mellem piger og drenge i PISA.

“Vores analyser i 2009 af de opgaver, som eleverne har løst, viser samme mønster som i 2006 i forhold til danske elevers præstationer. Piger og drenge klarer opgaver inden for Biologiske systemer på samme niveau, mens drenge er bedre inden for Teknologiske systemer, Fysiske systemer og Jordens og universets systemer. Piger klarer opgaver inden for Naturvidenskabelige undersøgelser bedre end drenge, men drenge er bedre til Naturvidenskabelige forklaringer. Dette sætter sig igennem i kompetenceområderne. Piger er bedre til området Anvende naturvidenskabelig evidens, men drenge er bedre til Identificere naturvidenskabelige spørgsmål og Forklare fænomener ud fra naturvidenskab.” (Sørensen & Davidsson, 2010, s. 121)

I dette citat giver forfatterne helt klart udtryk for at PISA-opgaverne i forhold til køn udviser differentiell item-funktion. Der er nogle opgaver der er lettere for piger, og nogle der er lettere for drenge. Hvis det er rigtigt, så holder Raschmodellen ikke, og så er det meningsløst at sammenligne pigers og drenges samlede scorer i naturvidenskab fordi den fundne forskel i drengenes favør udelukkende kan ses som et resultat af sammensætningen af prøvesættet. Hvis man havde valgt flere opgaver omhandlende pigeområderne *Naturvidenskabelige undersøgelser* og *Anvende naturvidenskabelig evidens*, så ville pigerne have klaret sig bedre, og vi havde fået et andet resultat i PISA 2009. Har man først konstateret at opgaverne fungerer forskelligt for piger og drenge, giver det ikke længere mening at sammenligne drenges og pigers totalscorer.

Man genfinder problemstillingen i undersøgelsen af matematik i PISA. Her skriver Lindenskov & Weng:

“Man kan her spørge, om der er særlige faglige områder, der motiverer og understøtter drengene, det være sig særlige kontekster eller opgaveformater. I PISA 2003 sås interessante variationer i præstationsforskellene mellem drenge og piger i Danmark. Præstationerne var svagest på området forandringer og sammenhænge, hvor der også var stor forskel på pigers og drenges præstationer. På området størrelser sås den mindste forskel på 9 scorepoint mellem piger og drenge, mens der på usikkerhed er en stor forskel på 22. Undersøgelser i danske klasserum vil sammen med nærmere analyser af datamateriale i PISA gennem årene være nødvendige for at forstå præstationsforskellene og deres variationer.” (Lindenskov & Weng, 2010, s.100)

I de to ovenfor omtalte situationer antages analyserne at svare på forskningsspørgsmål inden for henholdsvis naturvidenskab og matematik. Men det er kun meningsfyldt at spørge sådan på basis af PISA's materiale hvis man på forhånd går ud fra at opgaverne i PISA ikke opfylder det grundlæggende krav om fravær af differentiell item-funktion. Til det kan man sige at Allerups og Kreiners undersøgelser tyder på at der er differentiell item-funktion i PISA. Alligevel kan man undre sig over at centrale forskere i det danske PISA-projekt i deres forskning går ud fra at Raschmodellen ikke holder – uden at komme med nogen kommentarer til dette. Yderligere må man tage i betragtning at PISA-opgaverne ikke skal være repræsentative for problemerne i de forskellige sfærer, men tværtimod skal være tilpasset således at de trods forskellighed i testede områder, discipliner og kompetencer skal måle “det samme”. Derfor er PISA's opgaver på mange måder det dårligst mulige udgangspunkt for at undersøge forskelle fx mellem drenge og piger i matematik eller naturfag.

Opsummerende kan man sige at det at undersøge forskelle på piger og drenge på grundlag af PISA-opgaverne er at stille forskningsspørgsmål der er inkompatible med den grundmodel som PISA analyseres inden for, samtidig med at opgaverne i PISA næppe kan antages at være karakteristiske for forskellene mellem piger og drenge.

Kompetenceniveauer i PISA

Ud over testscorerne opererer PISA for hvert af domænerne med en række kompetenceniveauer. De angives at være beskrevet i form af de færdigheder som elever på det pågældende niveau antages at have. I det følgende beskrives niveau 5 i matematik.

“Elever, der præsterede på næsthøjeste niveau, 5, er karakteriserede ved at kunne arbejde med opstilling af modellering i komplekse matematikholdige situationer, for eksempel ved at kunne identificere muligheder og begrænsninger ved en model ud fra givne antagelser for opstillingen af modellen. Således kan eleven udvælge, sammenligne og vurdere,

hvilke strategier der er bedst egnede til at behandle et problem i relation til en eller flere mulige modeller. Elever på dette niveau udviser også indikationer på god forståelse og fortolkning af repræsentationer i problembehandling af de matematikholdige situationer.” (Lindenskov & Weng, 2010, s. 91)

Ovenstående beskrivelse forudsætter principielt en nøje sammenhæng mellem de krav opgaverne stiller, og de færdigheder eleverne har når de kan løse dem. Men i praksis er det sådan at opgaverne i PISA indplaceres på kompetenceniveauer alene ud fra elevscorerne, ikke ud fra deres indhold.

Peter Allerup har vist hvordan nogle opgaver skifter sværhedsgrad fra PISA 2000 til PISA 2003 (Allerup, 2005). Et eksempel er læseopgaven R067Q05. I PISA 2000 havde en gennemsnitlig OECD-elev en sandsynlighed på 38 % for at svare rigtigt på denne opgave. I PISA 2003 havde den tilsvarende elev 76 % chance for at svare rigtigt. Uanset at det er den samme opgave, så ligger den på to forskellige kompetenceniveauer de to år. Det er derfor svært at se hvordan det kan forenes med udsagn af typen “placering på “mathematical literacy” skalaen er foregået på baggrund af omhyggelige overvejelser af, hvilke typer af kompetencer der kan relateres til disse”. (Lindenskov & Weng, 2004, s. 48).

“Funktionelle analfabeter”

I Danmark har andelen af elever på det laveste PISA-niveau i læsning fået sit eget liv i uddannelsesdebatten. Her lyder det ofte at “17 % af de unge forlader folkeskolen uden at kunne læse og skrive”, til tider formuleret lidt mere forsigtigt som at “17 % ikke læser godt nok til at gennemføre en ungdomsuddannelse”, og i debatten er de blevet benævnt “funktionelle analfabeter”. Se mere i Henningsen (2008).

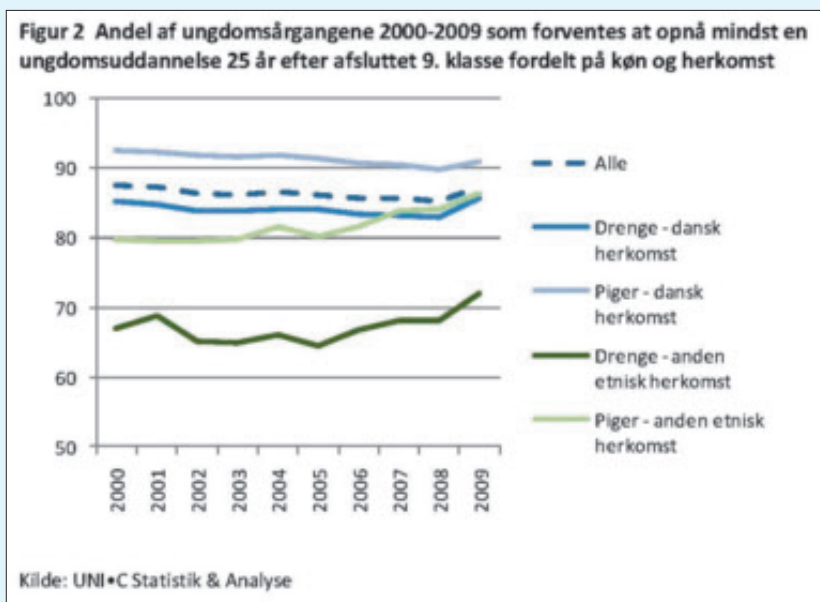
Udsagnet om “funktionelle analfabeter” stammer fra debatten om PISA 2003 hvor 17 % af de danske elever i læsning blev placeret på (eller under) det man betegner som laveste PISA-niveau. Disse elever fik den karakteristik at deres læsekundskaber ikke var gode nok til at de kunne gennemføre en ungdomsuddannelse. Problemet var bare at placeringen ikke sagde noget veldefineret om hvor godt eller dårligt eleverne læste. Sagen er nemlig at grænsen for niveau 1 i PISA fastlægges på grundlag af alle besvarelser fra OECD-landene, sådan at der altid skal være 18 % der ligger på eller under niveau 1. Om en elev falder over eller under niveau 1, siger altså kun noget om hvordan vedkommende læser i forhold til andre elever i OECD-landene, men ikke noget om hvordan eleven læser eksempelvis i forhold til kravene på en ungdomsuddannelse, der jo heller ikke er nogen entydigt bestemt størrelse – på tværs af lande og på tværs af uddannelser. Der foreligger heller ikke i dag – så vidt vides – systematiske danske eller internationale evidensbaserede undersøgelser af i hvilket omfang elever på de

forskellige PISA-niveauer faktisk kommer igennem ungdomsuddannelserne, eller hvilke konkrete krav disse egentlig stiller.

Det er fx betegnende at et af resultaterne fra PISA 2009 blev slået stort op: 50 % af alle elever med anden etnisk baggrund end dansk skulle have så dårlige skolekarakterer at de ikke kunne gennemføre en ungdomsuddannelse. Samtidig antages det i Undervisningsministeriets egne prognoser – baseret på gennemførelsestal fra tidligere år – at 77 % af disse unge i løbet af deres karriere vil gennemføre mindst en ungdomsuddannelse. Det er svært at se hvordan dette hænger sammen – og det gør det da heller ikke. I tekstboksen er de omtalte tal og argumenter vist i form af en fiktiv PISA-opgave.

En fiktiv PISA-opgave

Lykke og Asger diskuterer uddannelse. Lykke siger at der er en stigende tendens til at flere og flere drenge falder ud af uddannelsessystemet. Asger siger at han er uenig, og henviser til følgende tabel, der er udarbejdet for undervisningsministeriet af UNI-C Statistik og Analyse.



- Hvem mener du har ret ud fra tabellen?
- Hvad kunne Asger sige som belæg for sin påstand?

Sufian har hørt på deres diskussion. Han siger at det har stået i avisen at ifølge den nye PISA-undersøgelse forlader 50 % af de tosprogede elever folkeskolen uden at

beherske de basale færdigheder til at kunne fortsætte på en ungdomsuddannelse. Ud fra figuren mener han ikke at det kan passe.

c) Hvad kunne Sufian sige for at underbygge sit synspunkt?

Svar på fiktiv PISA-opgave

a) Asger har ret.

b) Figuren viser at andelen af drenge der har gennemført en ungdomsuddannelse, har ligget næsten konstant siden 2001, og at svingningerne ligger inden for hvad man ville forvente i en fremskrivningsmodel.

c) Figuren viser at i perioden efter 2000 forventes over 70 % af en årgang elever af anden etnisk herkomst at få mindst en ungdomsuddannelse. Dette modsiges PISA's påstand om de tosprogede elever.

Referencer:

Information. (marts 2011). *Tosprogede elever halter stadig bagefter*. www.information.dk/telegram/263612.

Lange, T. *Hvor stor en andel af en årgang forventes at få en uddannelse?*. <http://uvm.dk/service/Statistik/Tvaergaaende/Andel%20der%20faar%20uddannelse.aspx>.

Perspektiver

PISA er primært et politisk projekt der skal "forsyne uddannelsespolitikere, uddannelsesadministratorer og praktikere med en omfattende vurdering af læringsresultater ..." (Egelund, 2007, s.6). For at kunne levere de rangordninger af lande som så åbenbart efterspørges både hos beslutningstagere og i offentligheden, har PISA valgt at basere sine analyser på et sæt af oversimplificerede modeller. Men det har sin pris. Jeg har i artiklen forsøgt at diskutere nogle af de konsekvenser det får når man lader de statistiske og datatekniske krav få forrang frem for det faglige samtidig med at der hos mange af deltagerne i de nationale PISA-konsortier synes at være en svigtende forståelse for de metodiske begrænsninger som de valgte analysemodeller medfører.

Referencer

Allerup, P. (2005). PISA præstationer – målinger med skæve målestokke? *Dansk Pædagogisk Tidsskrift*, 53(1), s. 68-81.

Egelund, N. (2007). *PISA 2006 undersøgelsen – en sammenfatning*. København: Danmarks Pædagogiske Universitetsforlag.

- Goldstein, H. (2004). International Comparisons of Student Attainment: Some Issues Arising from the PISA Study. *Assessment in Education*, 11(3), s. 319-330.
- Henningsen, I. (2008). Holder PISA, hvad PISA lover? *Pædagogisk Psykologisk Tidsskrift*, 48(5-6), s. 380-394.
- Hopmann, S.T., Brinek, G. & Retzl, M. (red.). (2007). PISA zufolge PISA – PISA According to PISA. Berlin: LIT Verlag.
- Kreiner, S. (2007). Itemanalyse af matematikprøven. I: M. Hermansen (red.), *Skolens gode og onde cirkler*, s. 366-373. Frederiksberg: Samfundslitteratur.
- Kreiner, S. (2011). *Is the Foundation under PISA Solid? A Critical Look at the Scaling Model Underlying International Comparisons of Student Attainment*. (under udgivelse).
- Lange, T. (2011). *Hvor stor en andel af en årgang forventes at få en uddannelse?*. Lokaliseret 05.09.11 på: <http://uvm.dk/service/Statistik/Tvaergaaende/Andel%20der%20faar%20uddannelse.aspx>.
- Lindenskov, L. & Weng, P. (2004). Matematisk kompetence. I: J. Meiding (red.), *PISA 2003 – Danske unge i en international sammenligning* (s. 35-89). København: Danmarks Pædagogiske Universitets Forlag.
- Lindenskov, L. & Weng, P. (2010). Matematik. I: N. Egelund (red.), *PISA 2009. Danske unge i international sammenligning. Bind 1 – Resultatrapport* (s. 83-104). København: DPU, AKF, SFI.
- Sørensen, H. & Davidsson, E. (2010). Naturvidenskab. I: N. Egelund (red.), *PISA 2009. Danske unge i international sammenligning. Bind 1 – Resultatrapport* (s. 105-121). København: DPU, AKF, SFI.
- Undervisningsministeriet. (2002). *Kompetencer og matematiklæring*. København.