

Udviklingen i matematik målt i de danske PISA rapporter

Hans Bay

Professionshøjskolen UCC, E-mail: HAB@UCC.dk

Siden den første PISA rapport udkom i december 2001 har PISA undersøgelserne sat deres meget stærke præg på folkeskoledebatten. PISA rapporternes oprindelige intensi- on var at måle, hvor godt unge mennesker er forberedt til at møde udfordringer i et glo- baliseret samfund. Siden 2001 er der udkommet mindst 8 PISA rapporter, og disse PISA rapporter har fået stor opmærksomhed ikke mindst i pressen. Men PISA rapporterne sætter også deres spor i lovgivningen og rapporterne bruges i dag også til at evaluere mange centrale størrelser i folkeskolen. I den sidste nationale PISA rapport fremkom det meget overraskende resultat, at matematik-niveauet var faldet signifikant i forhold til de tre tidligere undersøgelser, dette på trods af at man med den seneste store ændring af folkeskoleloven har ønsket at styrke faget matematik. I denne artikel sættes spørgsmåls- tegn ved, om dette signifikante fald er udtryk for at eleverne er blevet dårligere til at løse matematik opgaver, eller om det skyldes en ændring i PISA undersøgelsernes stikprøve- design.

I foråret 2000 blev der gennemført en undersøgelse blandt 4.212 skolebørn fra 223 skoler. Eleverne var alle født i 1984, og de blev testet i læsning, matematik og natur- videnskab samt blev bedt om at udfylde et baggrundsskema. I december 2001 udkom rapporten: »Forventninger og færdigheder – danske unge i en international sammenligning«. Dette var den første danske PISA rapport. Siden da har der været afholdt tilsva- rende nationale test i 2003, 2006 og 2009. De efterfølgende danske PISA rapporter har alle fået stor opmærksomhed blandt forskere og i den offentlige debat, men PISA rap- porterne har også haft kraftigt indflydelse på lovgivningen inden for folkeskoleområdet.

Ved den sidste store revision af læreruddannelsen (ofte kaldet den nye læreruddan- nelse) i 2007, blev der netop fokuseret på fagligheden inden for dansk, matematik og na- turfag. Ydermere skal de lærerstuderende nu vælge mellem begynder- og sluttrin på lin- jefagene henholdsvis dansk og matematik. Netop opdelingen i begynder- og sluttrin vil mange opfatte som værende kraftigt inspireret af Finland. Finland har absolut klaret sig godt i disse internationale PISA undersøgelser. De er som regel blandt de tre første på ranglisterne og må klart siges at være det bedste europæiske land. Og når man har for- søgt at forklare Finlands høje placering, så har man ofte fokuseret på, at Finland har særligt uddannede lærere til at varetage undervisningen efter 7. klasse.

Det er samarbejdsorganisationen OECD, der har nedsat et internationalt PISA kon-

sortium, som har det overordnede ansvar for PISA testene. Det internationale PISA konsortium står for udvælgelse af testspørgsmålene samt konstruktionen af det baggrundsskema, som eleverne også skal udfylde. De deltagende lande skal sørge for oversættelse af testspørgsmål samt baggrundsskema. Tilsvarende skal de deltagende lande stå for den praktiske gennemførelse af testene. Et vigtigt karakteristikum – og til tider kraftigt diskuteret – ved PISA testene er, at det er de samme spørgsmål, der stilles i alle de deltagende lande. I og med at det er identiske spørgsmål, der stilles i alle lande, er det teknisk muligt at udarbejde en rangliste over, hvor godt de enkelte lande præsterer. Ranglisten og ikke mindst forudsætningen for at kunne udarbejde en valid rangliste har været voldsomt diskuteret, Kreiner (2011). I denne artikel vil der ikke blive fokuseret på Danmarks placering på denne rangliste, derimod vil udviklingen over tid inden for domænet matematik blive analyseret.

Også internationalt har PISA fået stor opmærksomhed. Ved den sidste PISA test i foråret 2009 deltog 65 lande, ved den første PISA test var der 32 lande, der deltog. En stigning som absolut kan fortolkes som en stor succes.

Nationalt har der også været stor fokus på PISA. Københavns kommune har således afholdt tre tilsvarende PISA test i 2004, 2007 og 2010. Disse PISA-K, som de ofte omtales, er en gentagelse af tidligere PISA test. På nationalt plan er PISA en stikprøve, mens PISA-K omfatter alle 9. klasser i alle Københavns folkeskoler samt de private skoler i København, som måtte ønske at deltage. I nedenstående tabel er vist en oversigt over afholdte PISA test i Danmark.

Man kan se af tabellen, at der har været 4 nationale PISA test og 3 københavnske PISA test. De nationale test er stikprøveundersøgelser, og designet er relativt kompliceret. Grundideen i et PISA design er at inddrage et mindre antal strata. Efterfølgende foretages en stikprøveudvælgelse af skoler, ofte efter skolestørrelse. Endelig udtrækkes tilfældige elever på skolen (uanset hvilken klasse de går i) dog maksimalt 28 elever pr. skoler. En passende betegnelse for designet er en tilnærmet stratificeret to-trins klyngeudvælgelse. Det komplicerede design har medført, at PISA datasættene er understøttet med vægtvariable, der skal anvendes når datasættene analyseres.

I den 4. nationale PISA test er designet ændret væsentligt for at få flere af anden etnisk herkomst med i undersøgelsen.

Når man skal følge en udvikling over tid baseret på stikprøver, er det meget væsentligt at sikre sig, at det ikke er ændringerne i stikprøvedesignet, der bevirker ændringer i interessevariablen. Interessevariablen skal helst måles på samme gentagne måde.

I dette tilfælde er det scoren i matematik, der er interessevariablen. Score i PISA undersøgelserne er konstrueret så gennemsnittet af OECD landene er på 500 (og den tilsvarende spredning er 100). I ovenstående fig. 1 er optegnet PISA scoren for de 7 afholdte PISA test i Danmark.

Tabel 1. Oversigt over PISA test i Danmark siden år 2000.

År	Test	Antal elever	Antal skoler	Rapport	Stikprøve Design	Deltagelse Elever	Deltagelse Skoler	Matematik Score
2000	1. PISA	4.212	223	Dec. 2001	Skoler inddelt i tre strata efter størrelse. Derefter klyngeudvælgelse. Elever skal være 15 år den 1. januar. Max 28 elever	92 %	95 %	514
2003	2. PISA	4.218	206	Dec. 2004	Som i 1. PISA	88,6 %	97,6 %	514
2004	1. PISA-K	2.352	59 folkeskoler 27 frie skoler		Alle 9. klasser, alle i klassen, alle folkeskoler. Samt de private der måtte ønske at deltage	86%	Alle folkeskoler; 27 ud af 44 frie	478
2006	3. PISA	4.532	209	Dec. 2007	Som i 1. PISA	89,1 %	95,9 %	513
2007	2. PISA-K	2.598	57 folkeskoler 28 frie skoler		Som i 1. PISA-K	82%	Alle folkeskoler; 28 ud af 45 frie	475
2007	2. PISA-K	2.598						
2009	4. PISA	5.924	285		Her sker en oversampling af skoler med elever af anden etnisk herkomst. Denne ekstra »skævhed« i design kompenseres ved at udarbejde nye vægte i datasættet	89,29 (vægtet)	90,75 (vægtet)	503
2010	3. PISA-K	3.196	28 frie skoler 55 folkeskoler	Dec. 2010	Som i 1. PISA-K	86%	Alle folkeskoler; 28 private ud af 45	482

Kilde: PISA rapporter. I oversigten er PISA etnisk fra 2005 udeladt.

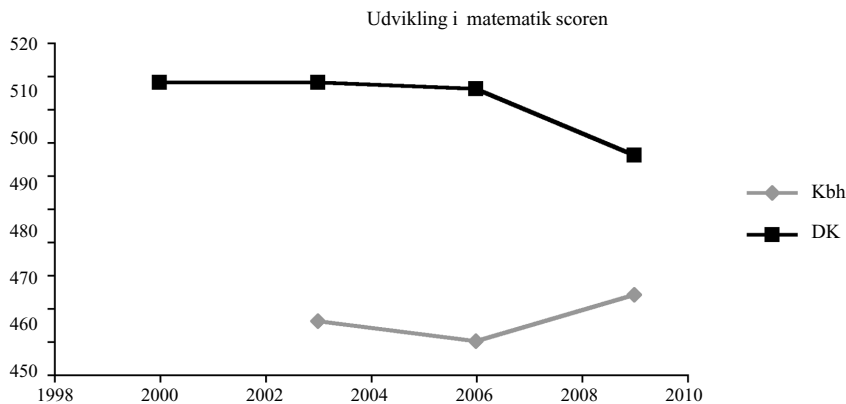


Fig. 1. Udvikling i matematikscoren i de nationale og københavnske PISA test, (rent visuelt er PISA-K skubbet et år tilbage).

Kilde: Diverse PISA rapporter.

Det fremgår af figuren, at de nationale PISA test er stabilt over de københavnske PISA test. Den helt afgørende årsag hertil er befolkningssammensætningen i København kontra resten af landet. Elever af anden etnisk herkomst scorer lavere end etniske danskere (et resultat som kan overføres til næsten alle OECD lande). Da København har relativt flere elever af anden etnisk herkomst er det helt forventeligt, at PISA København er mærkbart under de nationale test.

Det, der derimod er meget overraskende i figuren, er det dramatiske fald fra 2006 til 2009 i de nationale test (jvf. Bay 2011). Faldet er signifikant, som det konstateres i rapporten: »Der er tale om en signifikant tilbagegang i de danske elevers præstationer fra 2006 til 2009 på 11 point« (PISA2009 s. 88). Imens er udviklingen i de tilsvarende Københavnske test stabil. Det fremhæves i rapporten vedrørende matematik, at »For alle deltagende elever under ét er der ikke tale om signifikante ændringer fra 2007 til 2010 – ej heller når der deles op på køn eller skoletype.« (PISA-K(2010) s. 35).

Når der sker store ændringer i udviklinger, der er målt med stikprøver, kan dette skyldes at stikprøvedesignet er ændret. Og netop designet i PISA 2009 er kraftigt ændret i forhold til de tre første PISA runder. Stratifikationen af skoler er nu baseret på etnicitet frem for skolestørrelse, og stikprøven er øget for at få tilstrækkeligt mange elever med anden etnisk herkomst.

Kønsfordeling:

I første omgang skal det vurderes, hvor repræsentativt stikprøven er for 2009 undersøgelsen mht. fordelingen af drenge og piger. I tabel 2 er vist fordelingen af piger og drenge for den opnåede stikprøve.

Tabel 2. Fordeling på køn samt PISA score i matematik.

PISA 2009	køn	%	Fordeling af 15 årige pr. 1. januar 2009 (udtræk fra Statistikbanken, Danmarks Statistik)	PISA score i matematik (egne beregninger på PISA datasættet)
Piger	3038	51,3	48,7	495,3
Dreng	2886	48,7	51,3	511,4
I alt	5924	100,0	100,0	503,3

I den opnåede stikprøve for 2009 er der flere piger end drenge, der har deltaget i testen. Dette på trods af at der er flere drenge end piger i denne årgang. Udvalgelsen af elever er sket ved, at de skal være født i år 1990. Der er udført en test for, om fordelingen i stikprøven er lig fordelingen i hele landet. Testet bliver stærkt signifikant med en signifikanssandsynlighed, der er mindre end 0,0001. At den danske stikprøve er skæv mht. kønsfordelingen har ikke den store indflydelse på de nationale resultater. Men det er interessant, at stikprøven er blevet skæv, når designet er baseret på udtræk af skoler og efterfølgende tilfældigt udtræk af elever. Når det er vigtigt at undersøge, om stikprøven er repræsentativ på køn, så skyldes det, at det har vist sig at drenge klarer sig bedre end i piger i matematik. Dette gælder ikke alene i Danmark, men i stort set alle lande, der deltag i PISA undersøgelserne. Den umiddelbare konklusion vedr. stikprøvens repræsentativitet er, at drenge i højere grad end piger undgår at deltage. Deltagelsesprocenten for eleverne har siden den første undersøgelse i 2000 været under 90%. Her bør de danske PISA-aktører overveje, om ikke et nationalt krav på mindst 95% elevdeltagelse bør indføres. En opnåelse på 95% bør sikre, at stikprøven bliver repræsentativ mht. kønsfordeling. Trods alt er stikprøven baseret på elever, der går i skole, og prøven bliver gentaget senere for de elever, der har været syge på prøvedagen. En større opnåelse vil altid sikre en større repræsentativitet.

Det må dog også noteres, at kønsfordelingen altid har været lidt skæv, som det fremgår af tabel 3.

Det ses af tabellen, at man normalt i Danmark får flere piger end drenge til at deltage i PISA testene. Dette på trods af at der fødes flere drenge end piger. Den foreløbige konklusion er, at kønsfordelingen har været skæv i de sidste tre nationale PISA test, i den forstand, at flere piger end drenge deltag. Dette må så igen betyde, at den skæve kønsfordeling kun i begrænset omfang kan være årsag til den signifikante nedgang i matematik for 2009.

Table 3. Kønsfordeling for de sidste tre nationale PISA test i Danmark.

Antal elever	2003	2006	2009
Piger	2136	2331	3038
Drenge	2082	2201	2886
	4218	4532	5924

Kilde: Egne beregninger på PISA datasæt. Det har ikke muligt at se kønsfordelingen i datasættet fra 2000.

Socioøkonomisk baggrund

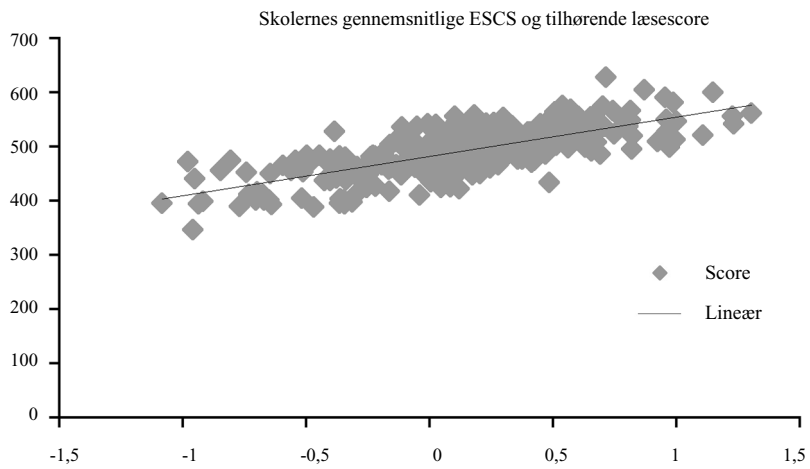
I forbindelse med PISA rapporter har det internationale PISA konsortium udarbejdet et indeks som er benævnt ESCS (economic, social and cultural status). »Dette indeks afspejler en række aspekter ved den enkelte elevs familie- og hjemmebaggrund, som kombinerer information om forældrenes uddannelsesniveau, erhvervs-mæssige stilling og forskellige typer af besiddelser i hjemmet« (PISA 2009 s. 130). Det er meget bemærkelsesværdigt, at indekset alene er udarbejdet ved at bruge elevernes egne svar på forældrenes uddannelse, stilling samt diverse besiddelser i hjemmet. Man skal altså ikke bruge registre til at beregne indekset men blot stille eleverne et antal spørgsmål, som de umiddelbart skal svare på. Det er faktisk ganske imponerende, at man med et begrænset antal spørgsmål til elever på ca. 15 år, kan konstruere et indeks, der virker i samtlige lande. Selve beregningen af ESCS ændrer sig lidt fra runde til runde, men et gennemgående træk er, at man altid har stillet spørgsmål om forældrenes uddannelse og beskæftigelse. Indekset ESCS er konstrueret så det har et gennemsnit på ca. nul og en spredning af størrelsesorden 1 for de deltagende OECD lande. Indekset ESCS har vist sig at være kraftigt korreleret med alle tre domæner, dvs. læsning, matematik og naturvidenskab. Der er brugt PISA scoren for læsning i nedenstående figurer. Men billedet er fuldstændig identisk, hvis man havde brugt scoren for matematik.

I PISA 2009 indgik 285 skoler, og for hver skole er gennemsnittet af elevernes læsescore og deres ESCS beregnet. Herved fås 285 punktpar, og disse punktpar er brugt til at illustrere sammenhængen mellem den socioøkonomiske baggrund og scoren for læsning. Scoren for læsning er tilsvarende fastlagt således, at gennemsnittet for OECD landene er ca. 500, og den tilhørende spredning er på 100.

Sammenhængen mellem skolernes gennemsnitlige socioøkonomiske baggrund og tilhørende læsescore er ganske tydelig. Har skolen en høj ESCS værdi vil den tilhørende score også være høj og vice versa. For fuldstændighedens skyld er den tilsvarende sammenhæng vist for 2006. (Figur 2a og 2b).

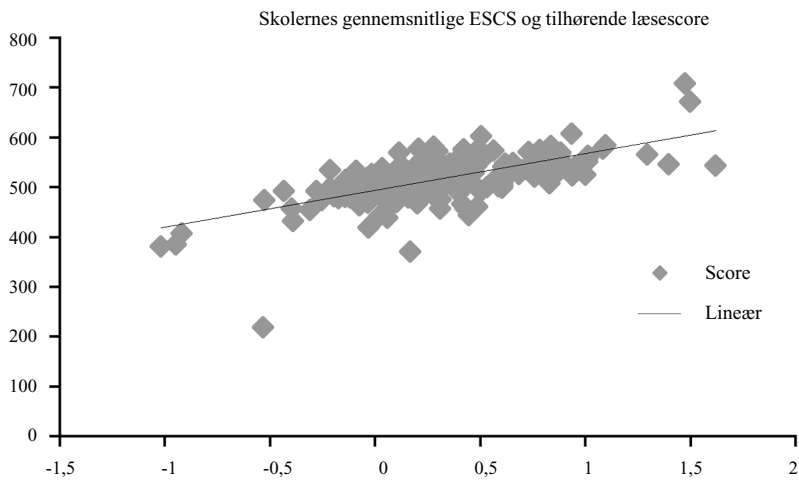
Igen er der en overbevisende sammenhæng mellem ESCS indekset og læsescoren.

Så i langt de fleste analyser på PISA data er det et uundgåeligt krav, at man skal inddrage ESCS indekset. Ovenstående figurer viser også meget klart, at hvis man sammen-



Figur 2a. Sammenhæng mellem ESCS og læsescore på skoleniveau, 2009.

Kilde: Egne beregninger på 2009 datasættet.



Figur 2b. Sammenhæng mellem ESCS og læsescore på skoleniveau, 2006.

Kilde: Egne beregninger på 2009 datasættet.

ligner skolernes gennemsnitskarakterer, så er det i praksis det samme som at sammenstille forældres socioøkonomiske placering. Så de bedste skoler (målt på karakterer) vil som hovedregel være i kommuner med en overvægt af veluddannede og højtlønnede personer.

Tabel 4. Udvikling i matematikscore og ESCS for årene 2006 og 2009.

2006	2009		Ændring i % fra 2006 til 2009
Mat. score			
Dansk	520,3	511,8	1,6
Etnisk	461,1	445,6	3,4
ESCS			
Dansk	0,376	0,367	2,4
Etnisk	-0,178	-0,221	24,2

Kilde: Egne beregninger på PISA datasættene.

Etnisk baggrund

I PISA rapporterne er det tilsvarende dokumenteret, at elever med anden etnisk baggrund som hovedregel får en lavere score. Også dette gælder i stort set alle de deltagende lande og også for alle tre domæner (læsning, matematik og naturvidenskab). I det følgende er eleverne opdelt i to grupper: danskere og etniske. Inddelingen er baseret på elevens besvarelse på to centrale spørgsmål: »I hvilket land blev du og dine forældre født?« og »Hvilket sprog taler du det meste af tiden derhjemme?«. Dermed bliver »dansker« defineret som en elev, der taler dansk i hjemmet og hvor mindst en af forældrene er født i Danmark. Resten bliver dermed defineret som etniske.

Udviklingen for de to grupper (danskere og personer af anden etnisk herkomst) mht. score i matematik og ESCS er vist i tabel 4.

De danske elever har haft et fald i matematikscore fra 520,3 til 511,8 i procent, som altså er signifikant. Elever af anden etnisk herkomst har haft et noget større fald. I 2006 scorede de i gennemsnit 461,1. Dette er faldet til 445,6 i 2009, og i procent er faldet ca. dobbelt så stort som hos de danske elever.

Når man ser på den tilsvarende udvikling for indekset for familiens økonomiske, sociale og kulturelle status, så er der også her tale om fald for begge grupper af elever. Dette leder uundgåeligt til, at man må opstille nogle hypoteser, som man burde undersøge nærmere.

Stikprøve designet i 2009 inddeler skoler efter etnicitet. Men skoler med mange etniske elever er også skoler med elever med generelt lave ESCS værdier. Det vil så betyde, at når man udtrækker skoler med mange etniske elever, så får man også generelt udtrukket elever med mindre ESCS værdier. Dette vil så igen betyde at den samlede stikprøve bliver skæv i forhold til den meget vigtige variabel ESCS. Denne hypotese er ikke mulig at teste på basis af de offentliggjorte datasæt med tilhørende rapporter. Det kræver en nøjere gennemgang af selve designet og den gennemførte stikprøveudtrækning.

Den næste hypotese man må overveje er, om opgaverne som stilles i matematik er skrevet i et så vanskeligt sprog, at især elever af anden etnisk herkomst har svært ved at læse opgaven. Hvis man først har svært ved at læse opgaven, så når man jo ikke at vise de matematiske kompetencer, som opgaven egentligt forsøger at teste i. Eksempelvis hvis opgaven omhandler, hvordan en tømrer skal løse en kompliceret opførelse af rækværk rundt om en bygning. Men de egentlige matematiske kompetencer omhandler beregning af en cirkels omfang. Det er derfor muligt, at det alene er den sproglige formulering der bevirker, at de etniske elever ikke løser opgaven. Dette kunne være årsagen til, at de etniske elever har et væsentligt større fald i matematik end de danske elever. Undersøgelsen af denne hypotese kræver detaljeret kendskab til de ca. 80 opgaver, som er stillet, og disse opgaver er forståeligt nok skjult for offentligheden. I den forbindelse bør man notere, at en nærmere analyse af den etniske del af PISA 2009 undersøgelsen viser, at tospørgede elever indhenter de danske elever, Tranæs 2011.

Opsummering

Den samlede vurdering af matematik niveauet i 2009 undersøgelse er:

- Der er et relativt stort frafald af elever på over 10%.
- Der er flere piger end drenge, der deltager i PISA testene
- Elever af anden etnisk herkomst har et større fald i matematikscore end de danske elever
- ESCS indekset er faldet i 2009 undersøgelsen i forhold til 2006 undersøgelsen
- Udviklingen i PISA-K er markant anderledes end udviklingen i de nationale undersøgelser.

Umiddelbart ledes man til at tro, at man må forkaste resultaterne i 2009 undersøgelsen på matematik domænet, da de alene afspejler, at designet er ændret.

Et godt princip, når man skal måle udviklinger over tid med stikprøver, er at bruge den »samme vægt«. Skal man følge ens egen udvikling over tid mht. om en motionsplan virker, så brug den samme vægt. At skifte en gammel badevægt ud med en digital vægt midt i forsøget med motion, vil uundgåeligt ødelægge muligheden for at vurdere udviklingen over tid. Et såkaldt trendbrud, som alle aktører inden for stikprøvedesign frygter. I PISA sammenhæng er »den samme vægt« udtryk for, at man bruger det samme stikprøvedesign over tid. Det må undre mange, at man har tilladt denne markante ændring af stikprøvedesignet, vel vidende at dette ofte bevirker trendbrud. Lige nu har ændringen i stikprøvedesignet rejst så mange spørgsmål om validiteten af udviklingen, at det i en årrække fremover vil være vanskeligt at foretage præcise og kvalificerede vurderinger af udviklingen i folkeskolen målt med PISA test.

Litteratur

- Bay, H. De danske PISA-rapporters håndtering af PISA-undersøgelserne. *MONA* 2011-3.
- Kreiner S. Is the foundation under PISA solid? A critical look at the scaling model underlying international comparisons of student attainment. *Research Report* 11/1 Biostatistiske institut, Københavns Universitet. På hjemmesiden www.pisa.oecd.org har Ray Adams kommenteret dette notat.
- PISA2000. *Forventninger og færdigheder – danske unge i en international sammenligning*. SFI 2001.
- PISA2003. PISA 2003. *Danske unge i en international sammenligning*. DPU 2004.
- PISA-K2004. *PISA-København – Kompetencer hos elever i 9. klasse i København*. AKF 2005.
- PISA-E2005. PISA etnisk 2005. *Kompetencer hos danske og etniske elever i 9. klasse i Danmark 2005*. Syddansk universitetsforlag 2007.
- PISA2006. *PISA 2006. Danske unge i en international sammenligning*. DPU 2007.
- PISA-K2007. *PISA København. Kompetencer hos elever i 9. klasse i København*. SFI 2008.
- PISA2009. PISA 2009. *Danske unge i en international sammenligning*. DPU 2010.
- PISA-K2010. *PISA København. 2010. Kompetencer hos elever i 9. klasse i København*. AKF 2011.
- Tranæs, T. PISA etnisk 2009. Store faglige fremskridt ledsaget af øget polarisering. *Notat fra Rockwool Fondens Forskningsenhed* april 2011.