

Den 2/6 - 2015

Til

Kontorchef Peter Linde DST Survey ([pli@dst.dk](mailto:pli@dst.dk))

Tak for svar på mine spørgsmål til ministeren angående PISA - stikprøve, stratifikation, vægtning, repræsentation mv. Jeg vil benytte mig af tilbuddet om at stille yderligere spørgsmål til den danske undersøgelse - se i øvrigt indsat kommunikation med ministeren.

Der er både nye og gamle pkt. i ministerens svar. Jeg vil kommentere de fleste og stille en række spørgsmål. Hm, jeg er vist også kommet til at konkludere på punkterne, så lad mig tage konklusionerne først.

#### **Færøerne – eller rettere - de færøske elever fra stratum 5 (immigrant level).**

**Jeg mener, det er klart dokumenteret, at de færøske elevers resultater er med i det samlede danske resultat i PISA 2012.** Det gælder både de tre hoveddomæner (læsning, matematik og naturfag) og de computer baserede test (CBA Læsning, matematik og problemlæsning). Det er endvidere klart, at de færøske elever er med i PISA etnisk 2012, de færøske skoler befinder sig stratum 5, der er et immigrant level.

#### **KeyQuest - kontrollanter.**

KeyQuest bruges af det internationale konsortium til udvælgelse af skolerne. Udvælgelsen sker i forhold til den stratificering, det danske konsortium har foretaget på grundlag af skolelisten. I både 09 og 12 glemmer det danske konsortium en række internationale skoler, hvilket først opdages, da man finder ud af Danmarks eksklusionsprocent er større end 5%. Dette sker både i 09 og 12. Så kontrollen fra det internationale konsortiums side er en slags efterkontrol, hvilket er i overensstemmelse med standarderne. KeyQuest bruges af de danske konsortium ved udvælgelsen af elever fra de udtrukne skoler. Denne udvælgelse sker på grundlag af lister indsendt af skolerne og her ser det ud til, at man bare tager alle med fra listerne i stratum 2, 3, 4 (og samtlige elever med anden etnisk baggrund end dansk blev tilføjet udtrækket).

Der findes et korps af kontrollanter, der rejser rundt og overværer testningen. Der er vist ikke nogen, der rejser rundt og kigger det nationale center over skulderen. **Det er i hovedsagen først, når oplysninger er indsendt, man stiller spørgsmål for at få opklaret** uoverensstemmelser mv – jævnfør fx vores overskridelse af 5 % grænsen for eksklusion af elev pga. SEN.

#### **Stikprøvens størrelse**

**Resultaterne bliver ikke væsentligt mere sikre fordi man tagen er lidt større stikprøve.** En fire gange større stikprøve halverer kun usikkerheden. I PISA etnisk 2005 var den etniske stikprøve 1200 elever i 2012 var den 1800 elever, altså kun 1,5 gange større.

#### **To stikprøver**

**Der er to stikprøver i PISA 2012.** Den ene – med 82 UH elever – giver et resultat, der er lavere end stikprøven uden de 82 UH elever. Forskellen er 2-3 PISA point.

#### **Repræsentativitet – vægtning**

**I PISA er det en standard, at der skal udtrækkes en repræsentativ stikprøve, hvis man skal drage konklusion om alle 15 årige i uddannelsessystemet. Den danske stikprøve er ikke repræsentativ (kilde: Danske PISA).** Vægtningen i PISA "genskaber" ikke repræsentativiteten, da man ikke bruger registerdata, når man analyserer stikprøven. Desuden

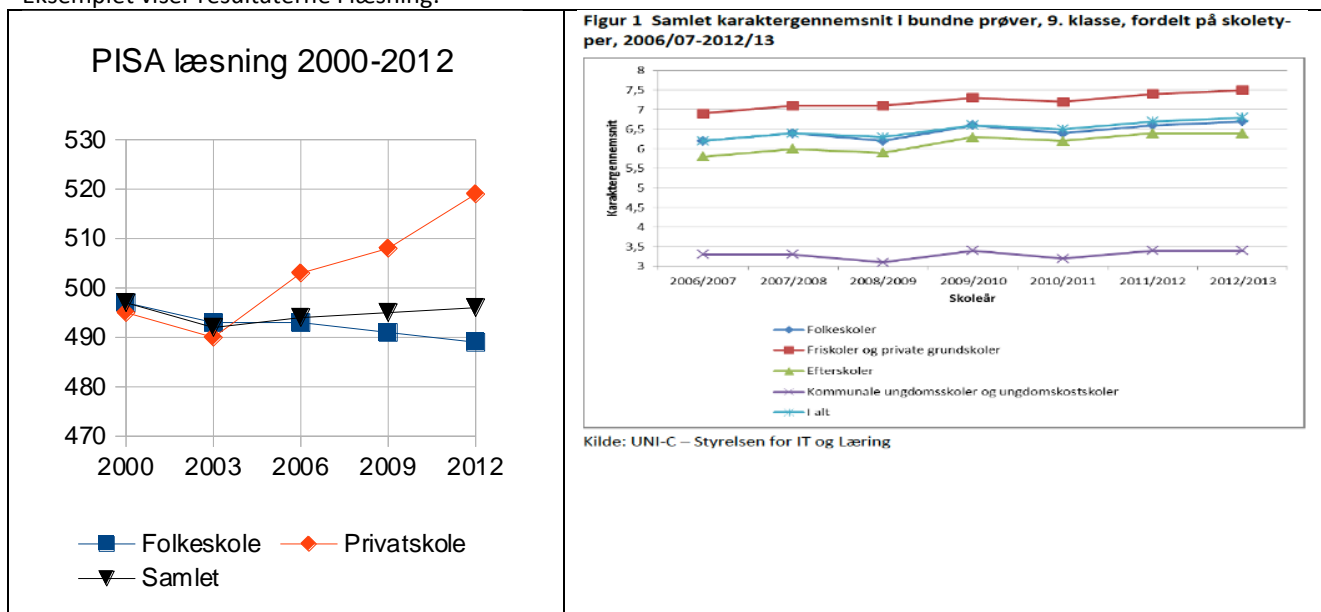
opgøres resultatet ikke skole for skole, men elev for elev. Repræsentativiteten skal sikres via stratificeringen og udtrukket af elever (det nationale konsortiums ansvar). Den manglende repræsentativitet i den danske stikprøve handler i hovedsagen om følgende:

- **Socioøkonomisk status** – for mange elever med svag status
- **To-sprogede elever** – der er udtrukket for mange tosprogede elever/indvandrerelver selv om de ikke eksisterer, når man ser på dem, med den definition af indvandrerelver PISA bruger
- Der er færøske elever med i stikprøven – stratum 5.
- Det faglige niveau blandt de kommunale elever i stikprøven ligger pænt under det faglige niveau blandt de kommunale elever på landsplan
- Der mangler elever fra nogle internationale skoler. Skolerne var ikke med på den liste det internationale konsortium brugte ved udvælgelsen. Det virker også som der mangler skoler fra den private skolesektor

### Udvikling i de danske PISA resultater fra 2000 til 2012.

Ser man på udviklingen i de danske PISA resultater fra 2000 til 2012, så ser det ud til, at der sker ændringer, når der ændres ved stratificering og elevudvælgelse. Ændringer påvirker den kommunale og den private skolesektor forskelligt - de private skolers score stiger, de kommunale skolers score falder. Udvikling afspejles dog ikke i FSA karaktererne fra 2007-2013.

Eksemplet viser resultaterne i læsning.



**2003:** Fejl i fordeling af prøvehæfter.

**2006:** Færøerne deltager som dansk undergruppe, ændring i implicite variable - amterne nedlægges først i 2007 et år efter PISA 2006 afvikles.

**2009:** Oversampling, færøske elever.

**2012:** Oversampling, Færøerne får tildelt immigrat level stratum 5.

**Konklusionen er derfor, at de danske PISA resultater fra 2006 og frem ikke kan bruges til at drage konklusioner om alle 15 årige i skolesystemet.**

## Uddybning

### Pkt. 1. Færøske elever - PISA 2012

Danmark deltager i Computer-based assessment of mathematics and reading samt problem solving. Der udtrækkes 4149 elever og 3332 bliver ikke udtrukket ( side 220 i den internationale tekniske rapport PISA 2012). Summen af udtrukne og ikke udtrukne er 7481, lig med det samme antal elever der deltog fra Danmark.

Denmark	PS+CBA	15, but 0 in SEN schools, and all students in stratum 5 schools (Faroes)	28 for strata 1, and 4; 20 for strata 2, and 3, all for stratum 5
---------	--------	--	---

Altså, alle elever fra stratum 5 skolerne - Færøerne - deltager i CBA.

Det var i 2012 ikke et krav om, at deltagerne i CBA også deltog i den papirbaserede test. Men da summen af de elever der deltog - inklusive de Færøske elever - og de elever der ikke deltog er lig med de elever, der deltog i de papirbaserede test+UH, så må alle stratum 5 elever jo have deltaget i de papirbaserede test. Hvilket så åbenbart bekræftes af følgende:

Side 400-401 i den internationale tekniske rapport:

"With the exception of Brazil, Italy and Spain the number of cases included in the computer-based assessment (CBA) database is the same as the number of cases in the PISA 2012 international database. Brazil, Italy and Spain chose to subsample schools from their large national school sample — see Chapter 4 for details of CBA sampling. The weight and replicate weight variables for these three countries have been adjusted in the CBA database to reflect this subsampling. **For all other countries, the CBA weights and paper-based weights are identical.**"

Side 157 i den tekniske rapport:

All students from schools that are sampled for computer-based assessment **received plausible values for paper-based PISA** and plausible values for computerbased assessment.

Fra Nationaløkonomisk Tidsskrift 151 (2013): 247-258 (Hans Bay) haves følgende:

"Når elevernes resultater bliver beregnet, sker dette ved, at den enkelte elev får tillagt 5 »plausible values« benævnt som PV. Gennemsnittet af disse 5 værdier vil som regel blive betragtet som denne elevs score for det pågældende domæne, som enten er læsning, matematik eller naturfag. Og hvis man tager gennemsnittet af de 5 PV'er og derefter gennemsnittet af alle elever, får man landets resultat for det pågældende domæne."

Altså opregningen foregår ikke skole for skole, men elev for elev. Og da de Færøske elever har PV'er i alle domæner, er de altså med i det danske resultat både med hensyn CBA og den papirbaserede test.

### Pkt. 2. KeyQuest - kontrollanter.

KeyQuest kan åbenbart indstilles til at tage højde for en aftalt oversampling. Spanien har registreret en oversampling (side 221 i den internationale rapport) og i Columbia blev der undervejs i processen identificeret en nødvendighed for oversampling i en særlig region (side 79 i den internationale rapport) og man måtte foretage en ny stratifikation og udvælgelse.

I Danmark har man i stratum 2, 3, 4 først udtaget TCS danske elever og derpå tilføjet samtlige etniske elever til udtrækket. At man "tilføjer" de etniske elever tyder på, at KeyQuest ikke har været indstillet til oversampling af de indvandrere elever, som PISA etnisk tidligere har beskæftiget sig med. Der er jo forskel på definitionerne af indvandrer i DST og PISA og forskellen på definitionerne vil give to forskellige stratificeringer og dermed også forskellige vægte og slutresultat (Side 70 i den internationale tekniske rapport: **The quality of the sampling frame directly affects the survey results through the schools' probabilities of selection and therefore their weights and the final survey estimates.** NPMs were therefore advised to be diligent and thorough in constructing their school sampling frames).

PISA's definition af indvandrer kan læses på side 307 i den internationale tekniske rapport. I PISA er man indfødt, hvis bare en af forældrene er født i landet (native students (those students who had at least one parent born in the country)). DST: Personer med dansk oprindelse er personer – uanset fødested – der har mindst én forælder, der både er **dansk statsborger** og født i Danmark). I lyset af de sidste 10-15 års stramme udlændingepolitik, er der formentlig ikke ret mange af vores to-sprogede (oversamplede) elever, der er indvandrere i den betydning, der ligger i PISA's definition. I PISA etnisk regnes elever, der taler et andet sprog end dansk i hjemmet også med som indvandrer. I PISA er sprog talt i hjemmet ikke en indvandrer variabel (s. 307). Så alt i alt er de oversamplede elever i det store og hele nok indfødte målt med PISA's alen (de Færøske elever må vel betragtes som indfødte på Færøerne).

### Kontrol

PISA kontrollerer en mængde ting, og kontrol er da også bygget ind i KeyQuest. Der er et korps af kontrollanter (lønnet og trænet af OECD) som tager rundt på uanmeldt besøg, når testen afvikles, men bortset fra disse kontrollanter er der næppe kontrollanter, der rejser rundt til de deltagende lande og kigger det nationale center over skulderen. Formentlig afklares de fleste spørgsmål via elektronisk kommunikation. Og så er det jo også sådan, at PISA ikke har kapacitet til at kontrollere og senere indarbejde alle lokale "løsninger". Man skal holde sig til standarderne - se side 447 Timeliness, samt følgende:

#### The Selection of the Student Sample

Following the selection of the school sample by the international contractor, the list of sampled schools was returned to National Centres. NPMs then contacted these schools and requested a list of all PISA-eligible students from each school.

This was provided on the *List of Students*, and was used by NPMs to select the student sample. NPMs were required to select the student sample using *KeyQuest*, the PISA student sampling and data entry software prepared by the international contractor. *KeyQuest* generated the list of sampled students for each school, known as the *Student Tracking Form* and the *Session Attendance Form* that served as the central administration documents for the study and linked students, test booklets and Student Questionnaires - side 111 i den internationale tekniske rapport.

Så listen over elever stammer fra skolerne og den skal åbenbart indtastes i KeyQuest af personale på det nationale center. En væsentlig kontrolfunktion i KeyQuest ser ud til at være at fange taste fejl.

Bundlinjen er derfor, at det nationale center er ansvarlig for stratificering, udtræk af elever (ikke skoler) og aftalte modificeringer af KeyQuest. Det er også det nationale centers ansvar, at man vælger at stratificere med udgangspunkt i en variabel, som ikke er defineret ens i PISA og DNK. At der er et problem fremgår af, at man i 2009 kalder stratificeringen *Minority Enrollment (4)*; og i 2012 *"Immigrant Levels (5)"* selv om det i praksis - vurderet i forhold til de danske rapporter - ser ud til at være den samme stratificering i de to undersøgelser.

### **Pkt. 3. Størrelse af stikprøve.**

PISA kræver, at man stiller med mindst 4500 elever og mindst 150 skoler til hovedundersøgelsen - 1500 pr. domæne. Danske PISA ønskede en oversampling af etniske elever, fordi det ville give et mere sikkert resultat. I 2005 lavede man en PISA etnisk i forbindelse pilotundersøgelse til PISA 2006. Der deltog ca 1200 etniske elever i 2005 undersøgelsen (plus færøske elever). I 2012 deltog ca. 1800 etniske elever i den almindelige PISA undersøgelse altså ca 1,5 gange flere elever i 12 end i 06. Hvis usikkerheden i 2005 var fx 5% så ville usikkerheden med 1800 elever være  $5/\sqrt{1,5} = 4\%$ . Det virker ikke umiddelbart som om, at sikkerheden forøges så meget, at oversamplingen var "nødvendig", især i betragtning af PISA's definition af indvandrere ikke svarer til den definition DST bruger og PISA bruger heller ikke sprog i hjemmet som indvandrer variabel. Det er formentlig god skik, at man i sammenlignende undersøgelser benytter de samme definitioner ellers bliver sammenligningen vel ret umulig.

### **Pkt. 4. To stikprøver - to resultater**

Der ser faktisk ud til at være to forskellige resultater i 2012. Booklet mean er jo større end det resultat, der fremgår af den danske rapport (2012).

Estimated booklet effects on the PISA scale - tabel 12.26 til 12.28 i den internationale tekniske rapport PISA 2012 - side 245 ff.

Matematik: Expected mean 502 (503). Resultat rapporteret i den danske rapport: 500 (503)

Læsning: Expected mean 499 (495). Resultat rapporteret i den danske rapport: 496 (495)

Naturfag: Expected mean 501 (499). Resultat rapporteret i den danske rapport: 498 (499)

(Der er tale om standard Booklets. Tallene i parentes stammer fra 2009 undersøgelsen)

Det vil sige, at vurderet ud fra besvarelser af opgaverne i standard prøvehæfterne ville vi ligge 2-3 point højere i 2012 end det fremgår af den Danske rapport.

Den lille stikprøve må i denne forbindelse være de elever, som får tildelt standard booklets og den store stikprøve må være den lille stikprøve plus de 82 UH elever - side 218. Måske er det pga. af SEN skolernes deltagelse (se side 221) at scoren falder 2-3 point i forhold til "expected mean" - da SEN elevernes resultater indgår på lige fod med de øvrige elevers.

SEN skolerne/eleverne er taget med i 2012 for at bringe eksklusionsprocenten tættere på kravet om højst 5% - se side 158 i den danske rapport.

Imidlertid så skyldes den høje danske eksklusionsprocent - i både 09 og 12 - ikke, at vi har ekskluderet for mange svage elever, men at vi **ikke** taget en række internationale skoler med på skolelisten.

"Overall exclusions were greater than 5% (6.18%). Data were fully explained – there was a difficulty in defining the school population – some international schools were not included when they should have been" - side 284 i den internationale rapport PISA 2012.

Så vi har altså "ofret" 2-3 PISA point i 2012, fordi det var vanskeligt at definere skolepopulationen. Vi havde den samme vanskelighed i 2009 - bemærkelsesværdigt.

## Pkt. 5. Repræsentativitet - vægtning

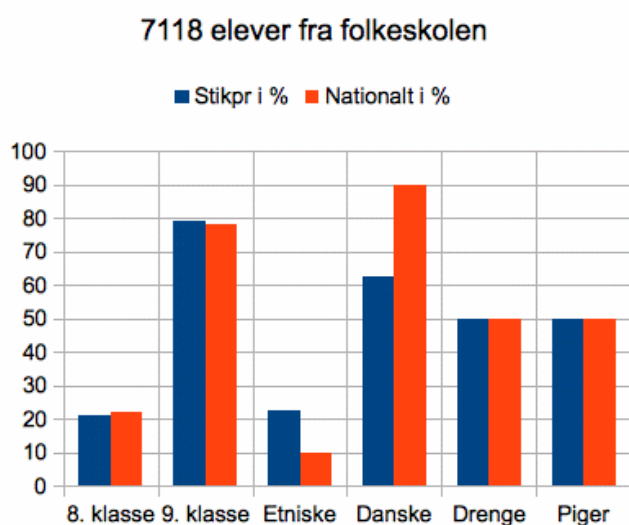
Standarden for målpopulation og sampling siger blandt andet: "Furthermore, to be able to draw conclusions that are valid for the entire population of fifteen-year-old students, **a representative sample shall be selected for participation in the test**" - side 448 i den internationale rapport (min understregning/fremhævning).

Vægtningen i PISA forudsætter denne standard og den genskaber derfor ikke repræsentativitet via vægtningen. Man vægter for at estimere stikprøvevariansen til brug i analyserne. Og vægtningen foregår som beskrevet i kapitel 8 i de internationale tekniske rapporter og ikke som beskrevet i de danske rapporter. Elever på samme skole har samme vægt og der er intet i justeringsfaktorerne der handler om, at der vægtes i forhold til særlige elevkategorier fx tosprogede. Og som nævnt foregår opregningen ikke skole for skole, men elev for elev - Nationaløkonomisk Tidsskrift 151 (2013): 247-258 (Hans Bay).

Repræsentativitet på elevniveau i den danske stikprøve - 2012 - kan illustreres ved hjælp af oplysningerne fra de to rapporter om kriteriebaseret af de nationale test, som UVM lod udarbejde i forbindelse med de nationale tests ændrede rolle i skolereformen.

Rapporterne beskæftiger sig med elever, der deltog i de nationale test i matematik 2010, læsning 2011/12 og PISA 2012. De fleste elever kommer fra kommunale skoler da testen er obligatorisk for disse elever, der kan dog godt være elever fra private skoler i materialet. Der er i så fald nok tale om elever, som er flyttet fra en kommunal skole til en privat skole inden PISA blev afviklet.

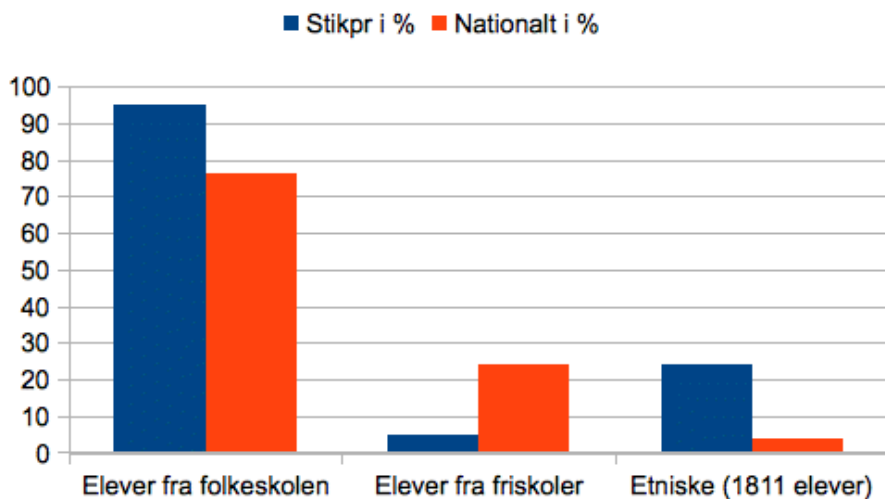
### Klassetrin, etnicitet, køn



Note: "Etniske" repræsenterer formentlig DST's definition: Indvandrere + efterkommere og Danske PISA's opfattelse af, at sprog talt i hjemmet er en indvandrer variabel.

## Offentlige skoler private skoler – indvandrer elever

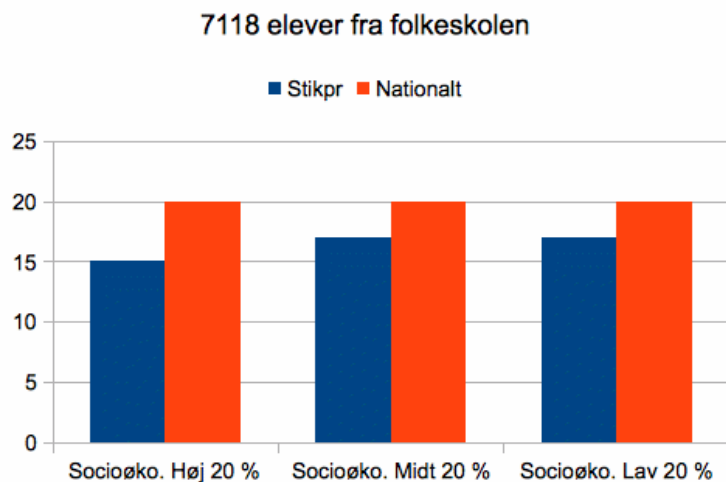
Den samlede stikprøve 7481 elever



**Note:** Den røde søjle - "Etniske" - repræsenterer de ca. 2303 (4%) 15 årige elever der gik i 7.-10. klasse i 2012 og som er registreret som indvandrere af DST. Efterkommere er lig med 4578, og de vil formentlig alle blive betegnet som indfødte i PISA, da PISA ikke har statsborgerskab med i sin definition (PISA native: Den ene forældre er født i landet. DST: Personer med dansk oprindelse er personer – uanset fødested – der har mindst én forælder, der både er dansk statsborger og født i Danmark.)

De manglende friskoleelever har formodentlig deltaget i de nationale test (2010, 2011, 2012) i folkeskolen og så siden flyttet til friskole eller efterskole. Det virker dog som om, der lidt for få elever fra denne skolesektor - de færøske elever er med i "friskoler"- der burde nok have stået "andre skoler".

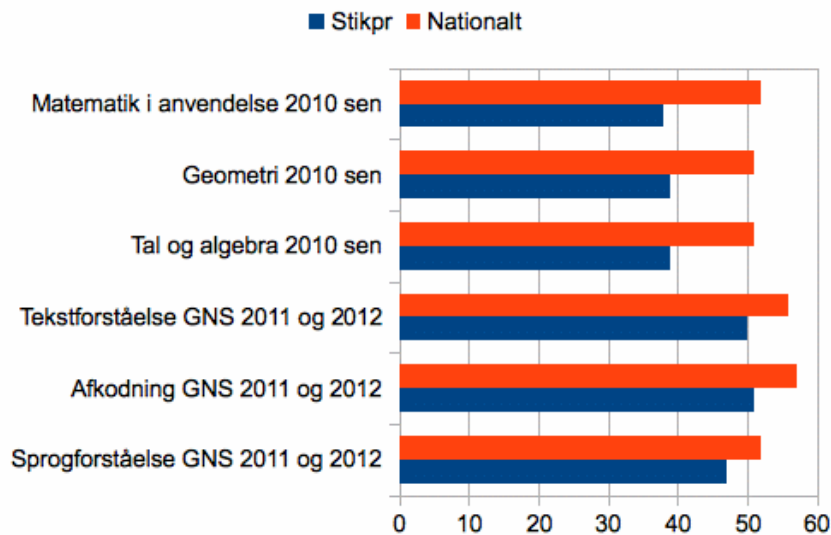
## Socioøkonomisk status



Note: Pointen er, at der er overvægt af elever med lav socioøkonomisk status. Hvilket øjensynligt i PISA slår igennem, som lavere score på opgaver man ikke har løst/fået tildelt.

## Fagligt niveau målt med de nationale test

### Niveau i læsning og matematik - de nationale test



Note: "sen" betyder elever med sen skolestart. De gik i 8. klasse, da de deltog i PISA 2012.  
Kilde tabel 3.7 til 3.12 i rapport 1 om kriteriebasering af de nationale test.

De skævheder, der er i stikprøven rettes ikke op ved den vægtning, der foregår i PISA. Det er en forudsætning, at stikprøven er repræsentativt udtrukket. I en undersøgelse af fagligt niveau (PISAs definition) må det vel anses for ret vigtigt, at stikprøven er repræsentativ, målt med danske alen også på dette punkt. Kalibreringen af testens opgaver

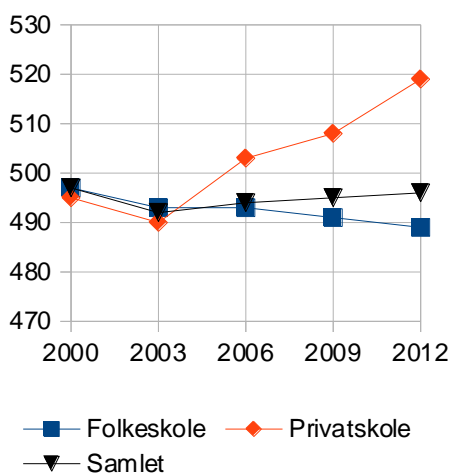


foretages jo med uvægtede data både nationalt og internationalt. De justeringer af opgaveparametrene kalibreringen giver anledning til er ens for alle lande for ikke at forstyrre rangordningen.

DNK stiller med en stikprøve fra folkeskolen med et fagligt niveau (matematik og læsning i de nationale test) der ligger pænt under landsgennemsnittet plus omkring 700 færøske elever, der åbenbart fagligt ligger på linje med vores tosprogede elever, når de måles med opgaver beregnet på danske elever. Det betyder formentlig flere forkerte svar på opgaverne end forventet og dermed kommer opgaverne til at se vanskeligere ud for danske elever end forventet. Vi får muligvis lidt hjælp af den internationale kalibrering, men ikke noget der flytter på rangordnen - se kapitel 9 i den internationale tekniske rapport.

**Set over tid ser det ud til, at der er et mønster i forhold til "stikprøven" og PISA resultaterne – tallene stammer fra Interactive Data Selection, fx <http://pisa2012.acer.edu.au/interactive.php>**

### PISA læsning 2000-2012



Ændringen fra 2000 til 2003 blev forårsaget af en fejl ved fordeling af prøvehæfterne i 2003. Er beskrevet i den internationale tekniske rapport 2003 side 211: In denmark .... It has been noted... Begge skoleformer reagerede ens på fejlen.

2006: De private skoler tager et pænt hop i scoren. De kommunale skoler er uændret - falder lidt. Der intet stratum for skolerne på Færøerne. Skolerne opdeles efter størrelse.

2009: Oversampling af tosprogede elever.

Skolerne opdeles nu efter antallet af tosprogede elever - Minority Enrollment. Færøerne får øjensynligt deres eget stratum - Certainty Selections.

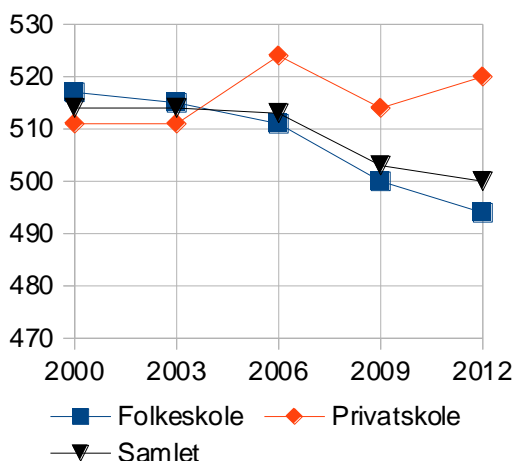
De private skolers score stiger.

De kommunale skolers score falder.

2012: Oversampling af tosprogede elever. Navnet på opdelingen af skoler ændres til Immigrant Levels, der er nu 5 immigrant levels mod 4 i 2009. Færøerne har deres eget stratum 5, som er et "**Immigrant level**" stratum. Vi deltager i de computerbaserede test, alle de færøske elever deltager. Da vægtene i CBA er identiske med vægtene i de papirbaserede test, kan man slutte, at de færøske elever også deltager i de papirbaserede test under dansk flag, de får tildelt PV-er i de papirbaserede test - se side 157 (Int. tek. rapport)

De private skolers score stiger.  
De kommunale skolers score falder.

## PISA matematik 2000-2012



Der er ingen ændring i den samlede score fra 2000-2003. De kommunale skoler taber dog lidt i forhold til 2000. De private skoler er uændret.

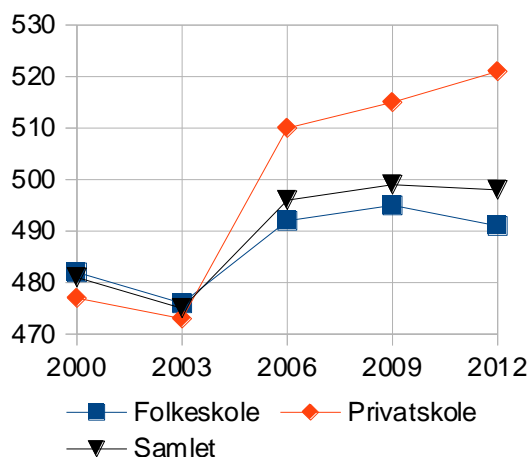
2006: De private skoler tager et pænt hop i scoren. De kommunale skoler falder lidt. Færøske skoler deltager uden eget stratum. Skolerne opdeles efter størrelse.

2009: Oversampling af tosprogede elever. Skolerne opdeles nu efter antallet af tosprogede elever - Minority Enrollment. Færøerne får øjensynligt deres eget stratum - Certainty Selections. Både kommunale skoler og de private skolers score falder.  
Det samlede resultat falder.

2012: Oversampling af tosprogede elever. Navnet på opdelingen af skoler ændres til Immigrant Levels, der er nu 5 immigrant levels mod 4 i 2009. Færøerne har deres eget stratum 5, som er et "**Immigrant level**" stratum. Vi deltager i de computerbaserede test, alle de færøske elever deltager. Da vægtene i CBA er identiske med vægtene i de papirbaserede test, kan man slutte, at de færøske elever også deltager i de papirbaserede test under dansk flag, de får tildelt PV- i de papirbaserede test - se side 157 (Int. tek. rapport)

De private skolers score stiger.  
De kommunale skolers score falder.

## PISA naturfag 2000 - 2012



Ændringen fra 2000 til 2003 blev forårsaget af en fejl ved fordeling af prøvehæfterne i 2003. Er beskrevet i den internationale tekniske rapport 2003 side 211: In Denmark .... It has been noted...

Begge skoleformer reagerer ens på fejlen.

2006: De private og kommunale skoler tager et pænt hop i scoren. Stigningen er størst for de private skoler. Der intet stratum for Færøerne. Skolerne opdeles efter størrelse.

2009: Oversampling af tosprogede elever. Skolerne opdeles nu efter antallet af tosprogede elever - Minority Enrollment. Færøerne får øjensynligt deres eget stratum - Certainty Selections.

De private skolers score stiger.

De kommunale skolers score stiger også, men dog er stigningen mindre.

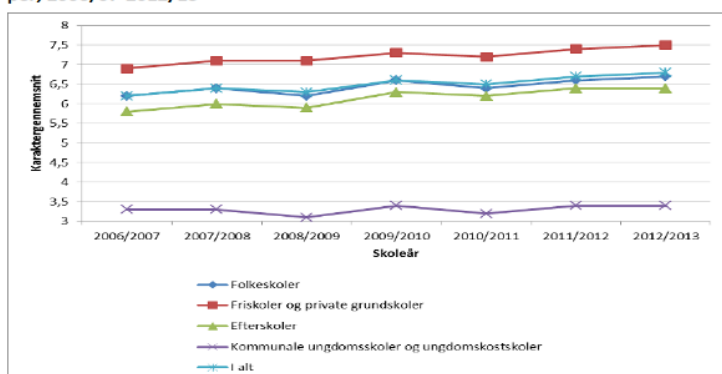
2012: Oversampling af tosprogede elever. Navnet på opdelingen af skoler ændres til Immigrant Levels, der er nu 5 immigrant levels mod 4 i 2009. Færøerne har deres eget stratum 5, som er et "**Immigrant levels**" stratum. Vi deltager i de computerbaserede test, alle de færøske elever deltager. Da vægtene i CBA er identiske med vægtene i de papirbaserede test, kan man slutte, at de færøske elever også deltager i de papirbaserede test under dansk flag, de får tildelt PV- i de papirbaserede test - se side 157 (Int. tek. rapport).

De private skolers score stiger.

De kommunale skolers score falder.

Til sammenligning er her vist udviklingen den samlede karakter ved FSA 2007- 2013.

**Figur 1 Samlet karaktergennemsnit i bundne prøver, 9. klasse, fordelt på skoletyper, 2006/07-2012/13**



Kilde: UNI-C – Styrelsen for IT og Læring

Hvis den divergerende udvikling i PISA scoren for private og kommunale skoler skyldes, at de to systemer har udviklet en forskellig faglighed, så burde man også kunne se denne udvikling i resultaterne ved FSA. Det kan man åbenbart ikke. Det samlede resultat af FSA for de to skoleformer udvikler sig ensartet fra 2007 til 2013.

Min forklaring ville umiddelbart være:

#### **2006:**

De færøske resultater er ikke blevet sorteret fra i det danske resultat - de har ikke deres eget stratum. Da de færøske skoler er kommunale ses virkningen først og fremmest på de kommunale skolars resultater.

#### **2009 og 2012.**

Oversamlingen i 09 og 12 bevirker åbenbart, at man får en stikprøve, hvor de deltagende elever fra de kommunale skoler ligger fagligt langt under landsgennemsnittet i læsning og matematik målt med de nationale test samtidig med, at der er en klar overvægt af elever med svag socioøkonomisk status. Det lave snit i DNT må formodes at føre til færre rigtige svar i PISA og overvægten af elever med svag socioøkonomisk status har ligeledes en negativ effekt på resultatet. Desuden deltager eleverne fra Færøerne også i de papirbaserede test i PISA 2012 og får tildelt PV-erne i de papirbaserede domæner åbenbart som "elever med anden etnisk baggrund end dansk" - stratum 5. Og da opregningen foregår elev for elev, så påvirkes det samlede danske resultat nedad, de færøske elever scorer nogenlunde på linje med de danske tosprøgede elever.

Jævnfør også side 211 i den tekniske rapport for PISA 2003.

#### **Sammenfattende**

##### **Repræsentativitet**

Den danske stikprøve er ikke repræsentativt udtrukket 09 og 12 - kilde danske PISA.

Standarden er, at der skal udtrækkes en repræsentativ stikprøve for at kunne drage konklusioner om alle 15 årige under uddannelse.

Det nationale center står for stratificeringen. Det internationale center for udvælgelse af skolerne. De nationale center står udvælgelse af eleverne.

Vægtningen i PISA genskaber ikke repræsentativiteten, da det jo er forudsat, at stikprøven er repræsentativ udtrukket. Opregningen foregår ikke skole for skole, men elev for elev.

Der mangler en række internationale skoler på skolelisten.

Færøske elever tæller med, da de får tildelt PV-værdier også i de papirbaserede test, hvor de så også deltager med et antal elever.

Landets resultat er gennemsnittet af PV-erne.

De valgte skoler (internationalt konsortium) sender elevlister til det nationale center, der så står for udtrækningen af elever via KeyQuest og "tilføjelse af alle elever med anden etnisk baggrund end dansk".

## **Færøerne**

Alle færøske elever deltager i CBA og får derfor også tildelt PV'er i de papirbaserede test, hvor de så også deltager. De færøske elevers resultat er derfor med i beregningen af det danske resultat.

## **Oversampling**

Oversamplingen af eleverne foregår øjensynlig manuelt (og alle elever med anden etnisk baggrund tilføjes udtrækket). KeyQest kan øjensynlig indstilles til en aftalt oversampling - jævnfør Columbia - men det sker øjensynlig ikke i Danmark. Oversamplingen bevirker, at man får en stikprøve med kommunale elever, der ligger pænt under det nationale niveau i læsning og matematik målt med de nationale test og samtidig indeholder stikprøven en overvægt af elever med svag socioøkonomisk status. Målt med PISA's indvander definition er oversamplingen formentlig beskeden.

## **Kontrol**

PISA har indbygget kontroller i KeyQest også kontrol for tastefejl. Det virker ikke sandsynligt, at man har en hær af kontrollanter, der rejser rundt og kigger det nationale center over skulderen. PISA har ikke kapacitet til at sende folk rundt og kigge de nationale centre over skulderen med henblik på senere at indarbejde "lokale" løsninger i undersøgelsen – Timeliness. Man skal holde sig til standarder.

Man foretager dog kontrol af selve afviklingen på skolerne via uanmeldte besøg af kontrollanter udannet og lønnet af OECD.

## **To stikprøver**

Der er to stikprøver i PISA 2012. En med UH elever og en uden. Der er også to resultater.

## **Sikkerhed**

Sikkerheden øges ikke væsentligt ved at medtage lidt flere elever.

## **Spørgsmål**

**Er de danske PISA resultater i 2012 repræsentative for alle 15 årige under uddannelse i den danske del af rigsfællesskabet?**

1. Vi har 5 eksplicite strata i 2009, men 6 i 2012. Certainty Selections er med begge gange. Hvilke skoler er der tale om i 09 og 12?
2. Er der også foretaget en oversampling i PISA 2015 af samme beskaffenhed som i 09 og 12?
3. Hvor mange elever fra Færøerne deltog i PISA 2012?
4. Er de færøske elever med i PISA 2015?
5. I 2009 er der 4 "Minority Enrollment levels", i 2012 er der 5 "Immigrant levels", i de danske rapporter er der kun nævnt 4 levels i både 09 og 12. De færøske skoler er stratum 5 i 2012. Er de færøske elever blevet til "elever med anden etnisk baggrund end dansk" i 2012 undersøgelsen?

Med venlig hilsen Niels Christoffersen

Rorupvej 9

4320 Lejre

Kopi til ministeren og uddannelsesudvalget.

Tidligere kommunikation med ministeren – indsat herunder:



Den 09-03-2015

Til Undervisningsminister Christine Antorini

Undertegnede skal hermed gøre opmærksom på, at det ser ud til, at ministeren har fremlagt ukorrekte oplysninger i forbindelse svar på spørgsmål 74 i Børne- og Undervisningsudvalget den 03-02-2014. Jeg ønsker samtidig at spørge ministeren om, det stadig er ministerens opfattelse, at konsortiet bag undersøgelsen sikrer, at PISA resultaterne i 2009 og 2012 er repræsentative for 15 årige under uddannelse i Danmark?

Link: <http://www.ft.dk/samling/20131/almdel/buu/spm/74/svar/1109064/1329687/index.htm>

Spørgsmålet (74) lød:

Kan ministeren i fortsættelse af samrådet 14/1-14 om PISA 2012 bekræfte, at der er en overvægt af skoler med mange 2-sprogede elever, der ligger til grund for undersøgelsen, og vil ministeren i bekræftende fald oplyse hvorfor?

Omtrent midtvejs nede i teksten står følgende:

Gennem en særlig udtræksmetode sikrer konsortiet bag undersøgelsen, at PISA er repræsentativ til trods for, at der er en overrepræsentation af elever med anden etnisk baggrund.

I slutningen af svaret står:

På skoler med mange elever med anden etnisk baggrund foretages udtrækket som en tilfældig stikprøve. På de øvrige skoler – med nogle, få eller ingen elever med anden etnisk baggrund – udtrækkes først en fuld stikprøve af etnisk danske elever. Efterfølgende bliver samtlige elever med anden etnisk baggrund udtrukket til at deltage. Som følge af denne metode til udtræk opnås en markant overrepræsentation af elever med anden etnisk baggrund end dansk. Vægtningen indebærer, at testbesvarelserne fra eleverne indgår med forskellig vægt i udregningen af de danske PISA-resultater. Testresultater fra etnisk danske elever på de udtrukne skoler med nogle, få eller ingen elever med anden etnisk baggrund tildeles en større vægt end både resultaterne fra deres klassekammerater med anden etnisk baggrund og alle elevresultaterne fra skoler med mange elever med anden etnisk baggrund. Metoden bag udtrækket samt principperne for vægtningen af besvarelserne er nærmere beskrevet i kapitel 6 i *PISA 2009 – teknisk rapport*.

Det er rigtigt, at metoden bag udtrækket samt principperne for vægtningen af besvarelserne er nærmere beskrevet i kapitel 6 i PISA 2009 - teknisk rapport.

**Men beskrivelsen er forkert.**

Det er rigtigt, at testbesvarelserne indgår med forskellig vægt i udregningen af de danske resultater, men fordi beskrivelsen i kapitel 6 er forkert, så ser det ud som om testresultater fra etnisk danske elever på de udtrukne skoler med nogle, få eller ingen elever med anden etnisk baggrund tildeles en større vægt end både resultaterne fra deres klassekammerater med anden etnisk baggrund og alle elevresultaterne fra skoler med mange elever med anden etnisk baggrund.

**Det forholder sig faktisk omvendt.**

Der gælder altså, at testresultater fra etnisk danske elever (faktisk alle elever) på de udtrukne skoler med nogle, få eller ingen elever med anden etnisk baggrund tildeles en **MINDRE** vægt end både resultaterne fra deres klassekammerater med anden etnisk baggrund og alle elevresultaterne fra skoler med mange elever med anden etnisk

baggrund.

**Det er ligeledes forkert**, at der på skoler med mange elever med anden etnisk baggrund foretages udtrækket som en tilfældig stikprøve.

Det der er tilfældet er, at der på skoler, hvor der kun er danske elever, foretages udtrækket som en tilfældig stikprøve.

**Her kommer begrundelserne – se også bilag 1 og 2.**

På side 168 i PISA 2009 – teknisk rapport er stratificeringen beskrevet:

**Høj:** Skoler med mindst 4 og mindst 33 pct. elever med anden etnisk baggrund end dansk.

**Mellem:** Skoler med mindst 4 og mindst 10 pct., men færre end 33 pct. elever med anden etnisk baggrund end dansk

**Lav:** Skoler, hvor antallet elever med anden etnisk baggrund end dansk var mellem 1 og 3 pct., og/eller hvor andelen elever med anden etnisk baggrund end dansk udgjorde mere end 0, men under 10 pct., og

**Ingen:** Skoler uden nogen elever med anden etnisk baggrund end dansk.

Ser man derpå på udtræksproceduren, så står der:

På de udtrukne skoler i stratum 01 (HØJ) blev der lavet et tilfældigt udtræk med 28 elever.

På de udtrukne skoler i stratum 02 (MELLEM), 03 (LAV) og 04 (INGEN) blev eleverne delt i to grupper, hvorfra der blev udtrukket elever.

I stratum 02 (MELLEM) blev der udtrukket 16 elever med dansk etnisk baggrund, og alle eleverne med anden etnisk baggrund end dansk blev tilføjet til udtrækket.

I stratum 03 (LAV) blev der udtrukket 24 elever med dansk etnisk baggrund, og alle eleverne med anden etnisk baggrund end dansk blev tilføjet til udtrækket.

I stratum 04 (HØJ) blev der udtrukket 28 elever med dansk etnisk baggrund, og alle eleverne med anden etnisk baggrund end dansk blev tilføjet til udtrækket.

Det fremgår af tabel 6.2 at stratum 1 = HØJ, stratum 2 = MELLEM, stratum 3 = LAV og stratum 4 = INGEN.

Så i forhold til stratificeringen vil den beskrevne udtræksprocedure føre til, at der på indvandrerunge skoler i stratum 1 (HØJ) ikke foretages nogen oversampling af indvandrere elever, og der på skolerne i stratum 4, hvor der ikke er nogen indvandrere elever foretages en oversampling af elever som ikke findes på skolerne.

Så udtræksproceduren og stratificeringen passer ikke sammen.

Udtræksproceduren og stratificering i hovedrapporten PISA - 2009 (side 118) passer til gengæld fint sammen. Der er bare et problem omkring logikken i betegnelserne HØJ, MELLEM, LAV og INGEN.

I hovedrapporten (2009) er

**HØJ beskrevet sådan her:**

**Høj:** Skoler uden nogen elever med anden etnisk baggrund end dansk. HØJ henviser åbenbart til andelen af danske elever, og det giver derfor god mening ikke at oversample etniske elever, der er jo ikke nogen.



**INGEN er beskrevet sådan her:**

**Ingen:** Skoler med mindst 4 og mindst 33 pct. elever med anden etnisk baggrund end dansk. INGEN henviser ikke til noget, da begge elevgrupper er repræsenteret, men det giver god mening at oversample etniske elever, da der er mindst er 33 % etniske elever.

Fejlene går igen i hovedrapporten fra PISA 2012. Udtræksproceduren i 2009 og 2012 er ens, så det må være stratificeringen i hovedrapporten fra 2009, der også er brugt i 2012.

Retter man fejlene, så ser resultatet sådan her ud - tal fra 2012 rapporten. Vægtene fremgår af side 161 i PISA 2012 resultatrapport.

**INGEN (Stratum 01 og 06) - etniske elever.**

Der udtrækkes 28 elever (danske).

Vægt: 1,83 og 2,09. Stratummet er delt i to dele

**LAV (Stratum 02) - under 10 % etniske elever.**

Der udtrækkes 25 danske elever og alle etniske elever tilføjes udtrækket.

Vægt: 2,36

**MELLEEM (Stratum 03) - 10 % og 33 % etniske elever.**

Der udtrækkes 20 danske elever og alle etniske elever tilføjes udtrækket.

Vægt: 14,17

**HØJ (Stratum 04) - mere end 33 % etniske elever.**

Der udtrækkes 28 danske elever og alle etniske elever tilføjes udtrækket.

Vægt: 16,77

## Konklusion:

1. Jo større andel af etniske elever på skolen, jo højere vægt får alle elevernes resultat. Årsagen hænger sammen med skolestørrelsen og har altså intet at gøre med hverken etnicitet, hårfarve, skolestørrelse mv.
2. Det er på skoler med udelukkende danske elever, der foretages et tilfældigt udtræk på 28 elever. En del af disse elever kan godt være to-sprogede, da sprogethed ikke er en indvandrer kategori i PISA (se den internationale tekniske rapport 2012 side 307)

## Spørgsmålet (for nu ikke at glemme det):

Er det stadig ministerens opfattelse, at konsortiet bag undersøgelsen sikrer, at PISA resultaterne i 2009 og 2012 er repræsentative for 15 årige under uddannelse i Danmark?

Med venlig hilsen  
Niels Christoffersen  
Rorup vej 9  
4320 Lejre

Kopi til Børne- og Undervisningsudvalget.

Vedhæftede filer

Bilag 1: Analyse af stratificering og udtræksprocedure som beskrevet i de danske rapporter - 2009 og 2012

Bilag 2: Vægtberegninger i PISA - kapitel 8 i den internationale tekniske rapport

Bilag 3: Repræsentation og andre uregelmæssigheder (mærkværdigheder) i danske PISA

Nationaløkonomisk Tidsskrift 151 (2013): 247-258: Udviklingen i PISA for de tre skandinaviske lande af Hans Bay

Link:

International teknisk rapport:

<http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>

Kriterie basering af de nationale test:

<http://www.uvm.dk/~media/UVM/Filer/Udd/Folke/PDF14/Okt/141008%20Kriteriebaserede%20test%20delrapport%201.pdf>

<http://www.uvm.dk/~media/UVM/Filer/Udd/Folke/PDF14/Okt/141008%20Kriteriebaserede%20test%20delrapport%202.pdf>

## Analyse af

### Stratificering og udvælgelsesprocedure

2009 – Hovedrapport side 118

Stratificering

**Høj:** Skoler uden nogen elever med anden etnisk baggrund end dansk,

**Mellem:** Skoler med mindst en og højst tre elever med anden etnisk baggrund end dansk, og hvor andelen af elever med anden etnisk baggrund end dansk er under 10 pct.

**Lav:** Skoler med mindst 4 og mindst 10 pct., men færre end 33 pct. elever med anden etnisk baggrund end dansk.

**Ingen:** Skoler med mindst 4 og mindst 33 pct. elever med anden etnisk baggrund end dansk.

Stratificering

**Høj:** Skoler med mindst fire elever med anden etnisk baggrund end dansk, og hvor andelen af elever med anden etnisk baggrund end dansk udgør mindst 33 pct.

**Mellem:** Skoler med mindst fire elever med anden etnisk baggrund end dansk, og hvor andelen af elever med anden etnisk baggrund end dansk udgør mindst 10 pct., men færre end 33 pct.

**Lav:** Skoler hvor antallet af elever med anden etnisk baggrund end dansk var mellem en og tre, og/eller hvor andelen af elever med anden etnisk baggrund end dansk udgjorde mere end 0, men under 10 pct., og

**Ingen:** Skoler uden elever med anden etnisk baggrund end dansk.

Skematiseret – stratificering 2009 hovedrapport og 2012 hovedrapport.

<b>2009 stratificering</b>	<b>Etniske</b>	<b>Danske</b>	<b>2012 stratificering</b>	<b>Etniske</b>	<b>Danske</b>
Str 1: HØJ	0	Alle	Str 1: HØJ	Mindst 33 %	Højst 67 %
Str 2: MELLEME	Under 10 %	Mindst 90 %	Str 2: MELLEME	Mellem 10 % og 33 %	Mellem 67 % og 90 %
Str. 3: LAV	Mellem 10 % og 33 %	Mellem 67 % og 90 %	Str. 3: LAV	Under 10 %	Mindst 90 %
Str 4: INGEN	Mindst 33 %	Højst 67 %	Str 4: INGEN	0	Alle

Som det ses, er der byttet om betegnelserne i 2012 i forhold til 2009. Så skoler, der i 2009 var HØJ i 2012 er INGEN.

For at finde ud af hvad der er op og ned i forhold til betegnelserne "HØJ,....INGEN" kan man se på udtræksproceduren de to år.

**2009 - teknisk rapport:**

På de udtrukne skoler i stratum 01 blev der lavet et tilfældigt udtræk med 28 elever.

På de udtrukne skoler i stratum 02, 03 og 04 blev eleverne delt i to grupper, hvorfra der blev udtrukket elever.

I stratum 02 blev der udtrukket 16 elever med dansk etnisk baggrund, og alle eleverne med anden etnisk baggrund end dansk blev tilføjet til udtrækket.

I stratum 03 blev der udtrukket 24 elever med dansk etnisk baggrund, og alle eleverne med anden etnisk baggrund end dansk blev tilføjet til udtrækket.

I stratum 04 blev der udtrukket 28 elever med dansk etnisk baggrund, og alle eleverne med anden etnisk baggrund end dansk blev tilføjet til udtrækket.

**2012 - resultat rapport:**

På de udtrukne skoler i stratum 01 (og stratum 06, - det er det samme, men opdelt i to i forhold til vægtning) blev der lavet et tilfældigt udtræk på 28 elever (TCS=28)

På de udtrukne skoler i stratum 02, 03 og 04 blev eleverne delt i to grupper, hvorfra der blev udtrukket elever.

I stratum 02 blev der udtrukket 20 elever (TCS=20) med dansk etnisk baggrund, og alle eleverne med anden etnisk baggrund end dansk blev tilføjet til udtrækket.

I stratum 03 blev der udtrukket 25 elever (TCS=25) med dansk etnisk baggrund, og alle eleverne med anden etnisk baggrund end dansk blev tilføjet til udtrækket.

I stratum 04 blev der udtrukket 28 elever (TCS=28) med dansk etnisk baggrund, og alle eleverne med anden etnisk baggrund end dansk blev tilføjet til udtrækket.

### Skematiseret - udtræksprocedure

2009	Danske	Etniske	2012	Danske	Etniske
Str. 1 HØJ	28		Str. 1 HØJ	28	
Str. 2 MELLEME	16	Tilføjet	Str. 2 MELLEME	20	Tilføjet
Str. 3 LAV	24	Tilføjet	Str. 3 LAV	25	Tilføjet
Str. 4 INGEN	28	Tilføjet	Str. 4 INGEN	28	Tilføjet

Da udtræksproceduren de to år er ens, må man konkludere, at det er stratificeringen fra 2009 - hovedrapporten - der også er brugt i 2012. Fordi stratificeringen i 2009 hovedrapporten stemmer med udtræksproceduren i både 09 og 12.

Bemærk at betegnelserne ikke er logiske, hvis der fx INGEN etniske elever er, så kan man jo ikke tilføje dem til udtrækket.

I nedenstående skema er beskrivelsen af stratificeringen i 2012 bragt i overensstemmelse med udtræksproceduren. Jeg har endvidere flyttet på betegnelserne, så de logisk henviser til andelen af etniske elever. Nummereringen er dog uændret.

**Skematisk gengivelse af stratificering 2009 og 2012 – så stratificering og udtræksproceduren passer sammen. Betegnelserne INGEN etc. henviser til andelen af etniske elever.**

2009 stratificering	Etniske	Danske	2012 stratificering	Etniske	Danske
Str. 1: INGEN	0	Alle	Str. 1: INGEN	0	Alle
Str. 2: LAV	Under 10 %	Mindst 90 %	Str. 2: LAV	Under 10 %	Mindst 90 %
Str. 3: MELLEME	Mellem 10 % og 33 %	Mellem 67 % og 90 %	Str. 3: MELLEME	Mellem 10 % og 33 %	Mellem 67 % og 90 %
Str. 4: HØJ	Mindst 33 %	Højst 67 %	Str. 4: HØJ	Mindst 33 %	Højst 67 %

### Udtræksprocedure - 2012:

På de udtrukne skoler i stratum 01 (og stratum 06, - det er det samme, men opdelt i to i forhold til vægtning) blev der lavet et tilfældigt udtræk på 28 elever (TCS=28)

På de udtrukne skoler i stratum 02, 03 og 04 blev eleverne delt i to grupper, hvorfra der blev udtrukket elever.

I stratum 02 blev der udtrukket 20 elever (TCS=20) med dansk etnisk baggrund, og alle eleverne med anden etnisk baggrund end dansk blev tilføjet til udtrækket.

I stratum 03 blev der udtrukket 25 elever (TCS=25) med dansk etnisk baggrund, og alle eleverne med anden etnisk baggrund end dansk blev tilføjet til udtrækket.

I stratum 04 blev der udtrukket 28 elever (TCS=28) med dansk etnisk baggrund, og alle eleverne med anden etnisk baggrund end dansk blev tilføjet til udtrækket.

Her giver det god mening ikke at tilføje etniske elever til udtrækket i stratum 01 - der er jo ikke nogen - og det giver også god mening at tilføje etniske elever til udtrækket i de andre strata, der er jo nogen at tilføje.

### Niels Christoffersen

Om repræsentation og andre uregelmæssigheder i danske PISA 2012 mv

Liste over "mærkværdigheder" og lignende i Danske PISA

- PISAs har en standard om at udtage en repræsentativ stikprøve. Citat fra den internationale tekniske rapport side 447:  
**Generalisability:** Data are collected from specific individuals, in a specific situation, and at a certain point in time. Individuals to be tested, test materials and tasks, etc. should be selected in a way that will ensure that the conclusions reached from a given set of data do not simply reflect the setting in which the data were collected but hold for a variety of settings and are valid in the target population at large. **Thus, collecting data from a representative sample of the population, for example, will lead to results that accurately reflect the level of literacy of fifteen-year-old students in a country.**  
**(Min fremhævnin g)**  
Det fremgår af rapporterne om kriteriebasering af de nationale test, at de elever fra stikprøven, som både har deltaget i de nationale test i matematik i 2010 og læsning 2011 og 2012 lå pænt under det nationale gennemsnit i de nævnte test. Det fremgår også, at elevernes socioøkonomiske status heller ikke er repræsentativ for hele populationen. Med hensyn til fagligheden målt med de nationale test, så er der flere elever som ligger under det nationale gennemsnit og færre elever som ligger over. Det samme billede viser tallene for socioøkonomisk status. Fordelingen på køn og 8-9 klasse ser ud til at passe.  
(Test resultater: Se tabel 3.7 til 3.12 på side 21 i rapport 2. Socioøkonomisk status: se tabel 3.2 side 14 i rapport 1). På side  
**Konklusion:** Stikprøven for de 7118 folkeskoleelever er ikke repræsentativ med hensyn til fagligt niveau (målt med de nationale test) og socioøkonomisk status.
- PISA foretager en kalibrering af testen på både nationalt niveau og internationalt niveau. Begge disse kalibreringer foretages med uvægtede data - se den internationale rapport side 148 og side 155.

Ministeriet har fået foretaget en kriteriebaseret (PISA) af de nationale test i læsning matematik. I rapport 1 side 13 finder man en score i læsning og matematik for stikprøvens elever på (Gennemsnitlig PISA-score i datasættet) på henholdsvis 487 p. og 489 p. Til sammenligning angiver PISA databasen en score i læsning og matematik på henholdsvis 489 p. og 494 p. Når man tager højde for den statistiske usikkerhed er der sikkert tale om samme score.

På side 51 i rapport 2 om kriteriebaseret af de nationale test står følgende:

Alderen for elever med normal skolestart ligger i intervallet fra 13,8 til 15,3 år med et gennemsnit på 14,8 år.

Alderen for elever med sen skolestart ligger i intervallet fra 15,3 til 17,2 år med et gennemsnit på 15,6.

PISA handler om 15 årige med en alder på 15 år og 3 mdr. til 16 år og 2 mdr. på testdagen. Den danske stikprøve medtager en betydelig del elever, som ikke burde være med, fordi de enten er for gamle eller for unge. Så heller ikke her er stikprøven repræsentativ og den manglende repræsentation slår formentlig igennem ved kalibrering af opgavernes parametre – opgaver kommer til at se sværere ud end de ville gøre med en fagligt repræsentativ stikprøve.

Fra 2006 indgår Færøske elever (9. klasse 4-500 elever) i den danske undersøgelse. Det påstås, at de færøske resultater ikke indgår de danske resultater. Det virker så ret underligt, at alle færøske skoler indgår i problemløsning og skærmlæsning - 2012. Mon de så ikke også indgår i de andre domæner - se side 221 i den internationale rapport - sandsynligvis igen via deres bidrag til kalibreringen af testen?

- Der er en række krav til stikprøveudtrækket. Eksempelvis skal det omfatte mindst 150 skoler fra det pågældende land, og **maksimalt 28 (TCS=28) elever fra den udvalgte skole må deltage - Se Hans Bay vedhæftet.**  
Man udtager TCS danske elever og tilføjer samtlige etniske elever til udtrækket. Det ser således ud til, at man i Danmark overskrider det maksimale antal elever ved udtrækket.
- Eksklusionsprocenten er for stor, fordi nogle internationale skoler mangler i skolelisten; ikke fordi der er mange svage elever i populationen, men der er det selvfølgelig i stikprøven – se side 284 i den internationale rapport.
- Der optræder en del meget små skoler i den danske stikprøve, tilsyneladende uden der taget højde for det i stratificeringen/vægtningen – TCS er alt for stor. I stratum 4 er TCS fx lig med 28, men i GNS er der omkring 10 elever pr skole i stratum 4 - se også side 187 i den internationale rapport (søg på Demark).
- Med udgangspunkt i formlerne om kriteriebaseret af de nationale test kan man lave en beregning af scoren i matematik og læsning i 2012 under forudsætning af, at der var tale om repræsentativ stikprøve og målene for elevernes kunnen repræsenteret ved kravene til FSA svarer til det PISA opgaverne måler. Resultatet ligger pænt over 550 p. for begge domæners vedkommende.

Formlerne er

Matematik:  $PISA\text{-score} = 515,188 + 68,185 * \text{logit} - 2,155 * \text{logit} * \text{logit}$

Læsning:  $PISA\text{-score} = 471,374 + 79,663 * \text{logit} - 5,719 * \text{logit} * \text{logit}$

Logit er et tal, der fremkommer, når eleven løser testopgaverne.

<http://uvm.dk/~media/UVM/Filer/Udd...>

PISA har nogle tilsvarende formler

Matematik:  $Score = ((\text{logit} + 0,0981) / 1,2838) * 100 + 500$

Læsning:  $Score = ((0,8830 * \text{logit} - 0,5079) / 1,1002) * 100 + 500$  (elever med mgl. kode for køn).

Hvis den danske stikprøve er repræsentativ, så må man forvente, at beregningerne med de to sæt formler



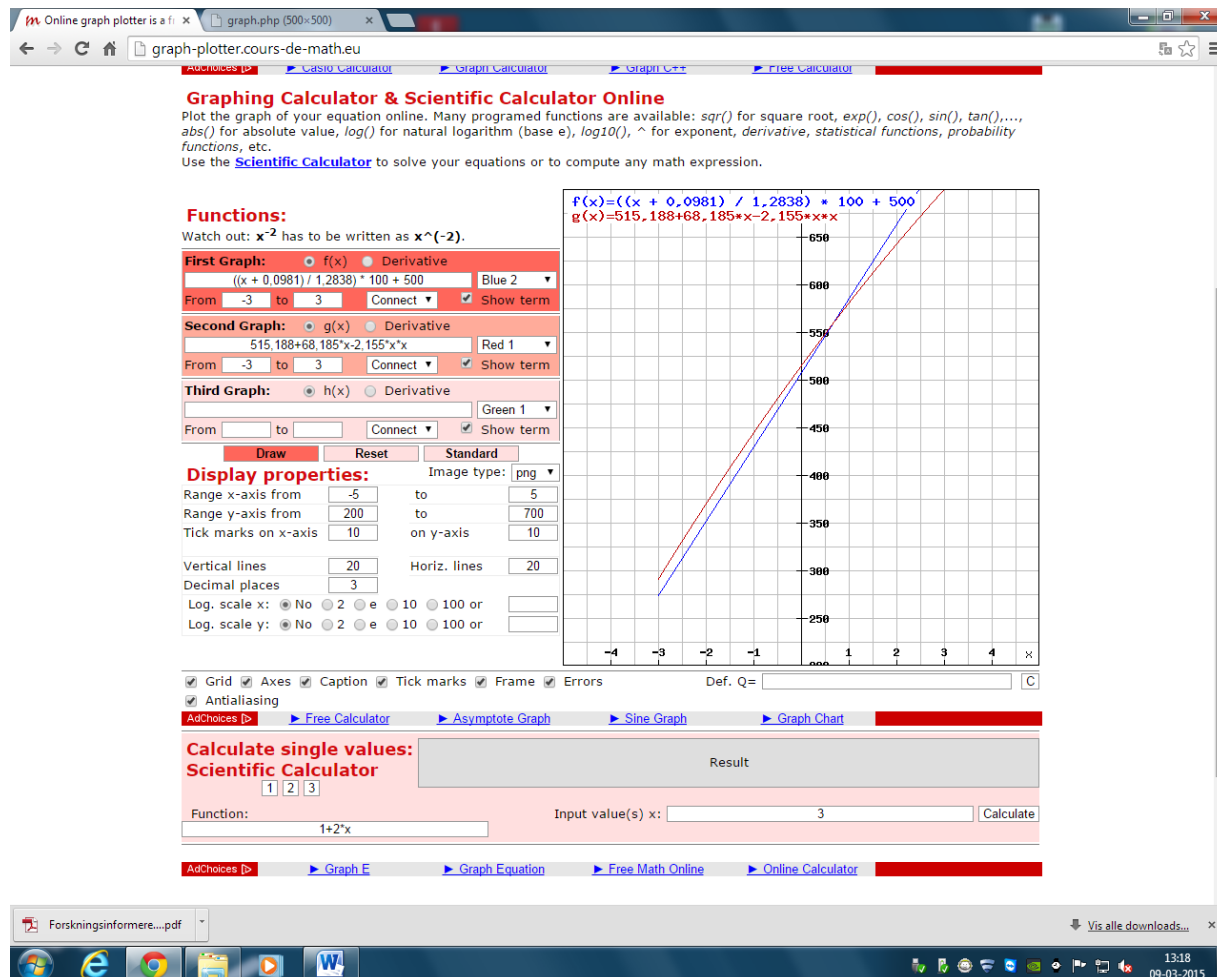
giver samme resultat. Altså må der for en repræsentativ stikprøve gælde:

$$\text{Matematik: } 515,188+68,185*\text{logit}-2,155*\text{logit}*\text{logit} = ((\text{logit}+0,0981)/1,2838)*100+500$$

$$\text{Læsning: } 471,374+79,663*\text{logit}-5,719*\text{logit}*\text{logit} = ((0,883*\text{logit}-0,4837)/1,1002)*100+500$$

Løser man de to udtryk med hensyn til logit (kaldet x i grafplotteren) fås følgende (grafisk) løsning:

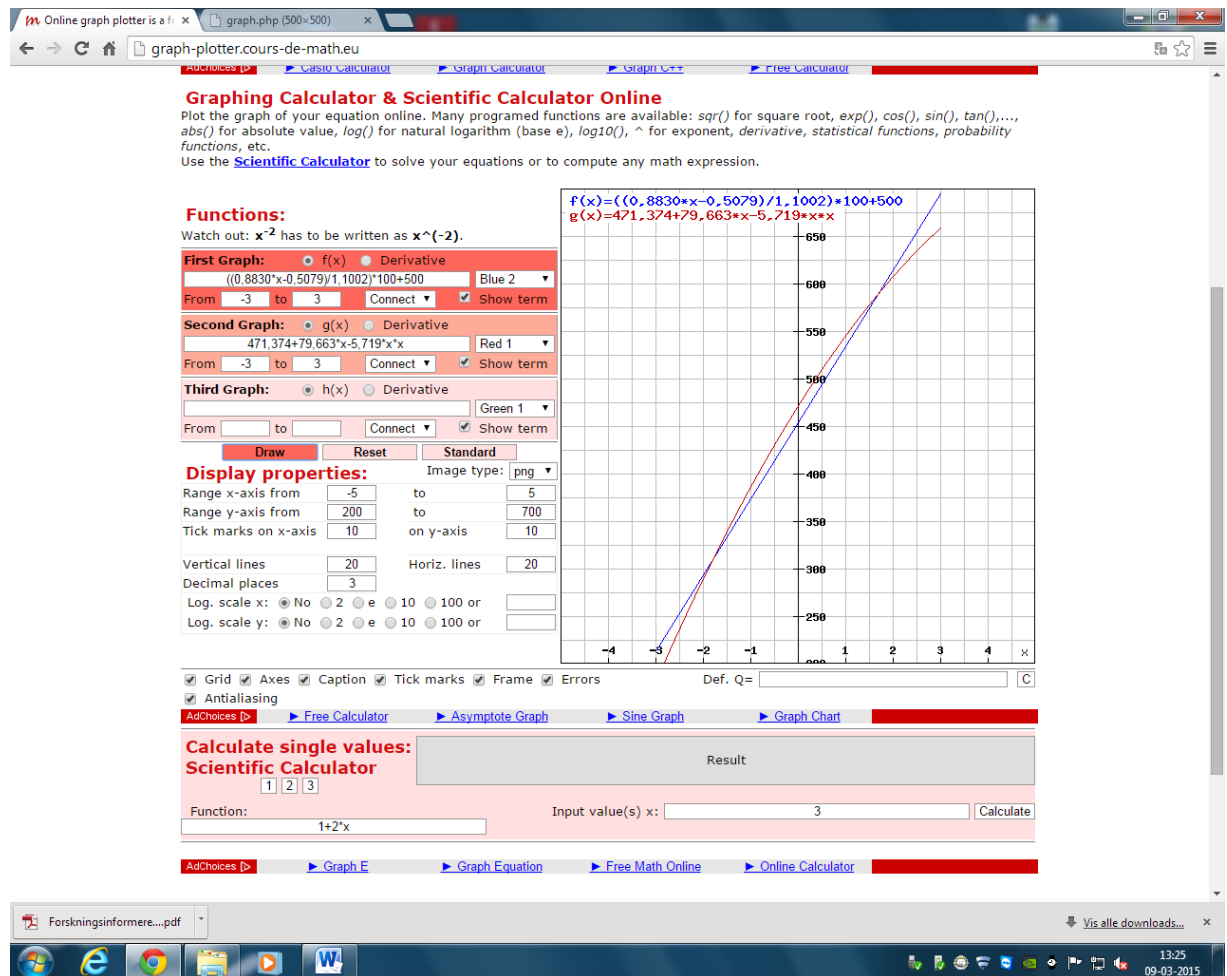
Matematik:



Logit = 0,7

Score = ca. 562 point.

Læsning:



Logit = 1,5.

Score = ca. 584 point.

(<http://graph-plotter.cours-de-math.eu/>)

Som det ses, er der to løsninger til hver ligning. Den "nederste" svarer formentligt til, at de danske krav ved fx FSA slet ikke matcher, det PISA måler.

- I de danske rapporter ser det ud til, at to-sprogethed er en indvandrerkategori. Det er det ikke i PISA – se side 307 i de internationale rapport. Skolelistens angivelse af andelen af indvandrere bygger på Danmarks Statistiks lister over indvandrere og efterkommere, men definitionen i Danmark og PISA er ikke den samme. I Danmark tager man fx forældrenes statsborgerskab med i definitionen, det gør man ikke i PISA. Stratificeringen af stikprøven er derfor ikke entydig i den forstand, at flere skoler formentlig ville blive placeret i et andet stratum med PISAs definition.

Niels Christoffersen

## Vægte i PISA

Beskrivelsen af vægtberegningerne i PISA fremgår af de internationale rapporter - kapitel 8. De vægte, der er angivet på side 161 i PISA 2012 resultatrapport er formentlig ikke rigtige. Vægtene er øjensynlig beregnet ved, at man har taget antal elever i populationen og divideret med antal elever som har gennemført testen.

Eksempel stratum 4:

Elever i populationen = 13400

Gennemført testen = 799

"Vægt" =  $13400/799 = 16,77$

Lad mig vise ved et eksempel, hvordan vægtene skal beregnes - stratum 4.

*Tabel 6.1 Fordeling af skoler og elever – i populationen og i stikprøven*

Stratum	Population		Stikprøve (estimeret antal baseret på institutionslisten fra Danmarks Statistik)	
	Antal elever	Antal skoler	Skoler	Elever
01 (+06) = Høj	4.255	131	113	2.803
02 = Mellem	14.190	257	100	2.000
03 = Lav	40.710	1.069	120	2.284
04 = Ingen	14.996	1.567	67	1.050
<b>I alt</b>	<b>74.151</b>	<b>3.024</b>	<b>355</b>	<b>8.137</b>

Side 155 i 2012 resultatrapporten.

Elevpopulationen (stratum 4) består af 14996 15 årige elever, der er indskrevet på 1567 skoler, altså i gennemsnit  $14996/1567 = \text{ca } 10$  elever pr skole.

**Elevvægtene beregnes som (skolens basis vægt)\*(skolens indre vægt)\*(5 justeringsfaktorer).**

### Skolens basisvægt er defineret som:

Udtræksinterval/skolens størrelse, hvis skolens størrelse er mindre end udtræksintervallet ellers er den lig med 1. Udtræksintervallet er skolernes samlede størrelse (14996) divideret med det antal skoler man ønsker udtrække fra stratum 4, altså 67.

I stratum 4 er udtræksintervallet altså  $14996/67=223$ . Skolernes gennemsnitlige størrelse er ca. 10

Basisvægten for en skole i stratum 4 er derfor i gennemsnit  $223/10= 22,3$ . Hver skole i stratum 4 repræsenterer altså omkring 22-23 skoler i stratummet.

### Skolens indre vægt

Skolens indre vægt er defineret som antal indskrevne 15 årige på testdagen divideret med det antal man vil udtrække (TCS). Hvis der er færre indskrevne 15 årige på skolen end det antal man ønsker at udtrække, så bliver alle udtrukket, og skolens indre vægt bliver lig med 1. Det fremgår af udtræksproceduren, at man ønsker at udtrække 28 (TCS=28) elever pr skole i stratum 4 og åbenbart også nogle flere. Da der i gennemsnit kun er ca. 10 femten årige pr skole i stratum 4 bliver alle udtrukket.

Så de gennemsnitlige elevvægte i stratum 4 er  $22,3*1*$ justeringer. De udtrukne elever i stratum 4 repræsenterer før justering altså hver omkring 22-23 elever.

I nedenstående skema er beregningerne lavet på både stikprøven i 2009 og 2012 og koblet sammen med skemaet for stratificering og udtræksproceduren. Vægtene er før justering. Skemaet følger udtræksproceduren som beskrevet i både 09 og 12. Betegnelserne "HØJ", "MELLEME", "LAV" og "INGEN" henviser til andelen af etniske elever.

**Oversigt over stratificering og vægte 2009 og 2012 – vægtene er før justering.**

<b>2009</b> stratificering	<b>Etniske</b>	<b>Danske</b>	Vægte 2009	<b>2012</b> stratificering	<b>Etniske</b>	<b>Danske</b>	Vægte 2012
Str. 1: INGEN	0	Alle	1,31	Str. 1: INGEN	0	Alle	1,16
Str. 2: LAV	Under 10 %	Mindst 90 %	2,91	Str. 2: LAV	Under 10 %	Mindst 90 %	2,57
Str. 3: MELLEEM	Mellem 10 % og 33 %	Mellem 67 % og 90 %	7,53	Str. 3: MELLEEM	Mellem 10 % og 33 %	Mellem 67 % og 90 %	8,91
Str. 4: HØJ	Mindst 33 %	Højst 67 %	14,10	Str. 4: HØJ	Mindst 33 %	Højst 67 %	23,39

Som det ses, får alle elever i skoler med etniske elever en højere vægt end elever i skoler, hvor der kun er danske elever. Og alle udtrukne elever på samme skole får tildelt den samme vægt. Om de fem justeringsfaktorer kan ændre på billedet vides ikke. Dog ved vi, at kun en elev og en skole har fået trimmet deres vægte i PISA 2012 – se side 161-162 i PISA 2012 resultatrapport.

Hans Bay skriver i en artikel i Nationaløkonomisk Tidsskrift 151 (2013): 247-258 følgende: Den sidste vægt (også kaldet final student vægt) bruges, når man skal kompensere for, at stikprøven kan være udtrukket skævt - jævnfør også den internationale rapport side 133 (t2ij is a final student weight trimming factor, used to reduce the weights of students with exceptionally large values for the product of all the preceding weight components). Trimming er brugt på en skole og en elev - det vil sige en skole, hvor der kun en elev er udtrukket eller har deltaget.

Niels Christoffersen

Ministeren svar på første henvendelse

Kære Niels Christoffersen

24-03-2015

Tak for din henvendelse af 9. marts 2015 om PISA-undersøgelsen.

Du spørger til overrepræsentationen af tosprogede elever i stikprøvedutrækket til PISA-undersøgelsen og baggrunden herfor.

I stikprøvedvælgelsen til den danske del af PISA-undersøgelsen har der i 2009 og 2012 været udtrukket en større andel af tosprogede elever, som ikke afspejler sammensætningen på landsplan.

I Danmark har vi valgt dette stikprøvedesign for at kunne analysere denne gruppe mere i dybden, end det ellers havde været muligt. Tidligere PISA-undersøgelser har vist, at tosprogede elever havde relativt dårligere resultater end etnisk danske elever. Undervisningsministeriet har derfor fundet det nødvendigt at teste et større antal tosprogede elever for at opnå et mere sikkert billede af tosprogede elevers kompetencer og dermed at kunne undersøge, om sådan en sammenhæng er statistisk sikker.

I de samlede analyser tages højde for overrepræsentationen af tosprogede elever ved at vægte stikprøven på en måde, der genskaber repræsentativiteten. Stikprøvedutræk og vægtning foretages af en ekspertgruppe i det internationale konsortium, der står bag PISA-undersøgelsen. Ekspertgruppen får adgang til de officielle lister med samtlige skoler og antallet af elever på 15-16 år på hver skole og designer herefter stikprøverne i alle deltagerlandene, herunder Danmark. Den internationale ekspertgruppe sikrer også, at vægtene genskaber repræsentativiteten.

Som du påpeger i din henvendelse, er der eksempler på, at metodebeskrivelserne i PISA-rapporten er formuleret upræcist og visse steder fejlagtigt. Et eksempel herpå er beskrivelsen af stikprøvens strata, hvor der på side 188 i den tekniske rapport for PISA 2009 er byttet om på beskrivelsen af de fire strata, som skolerne inddeles i. Skoler, som skulle have været defineret "Høj", betegnes her som "Ingen" og omvendt. Det har dog ingen betydning for, hvordan stikprøven er blevet trukket og analyseret, og det internationale konsortium fastslår, at de danske PISA-undersøgelser opfylder de faglige krav til repræsentativitet.

På trods af disse upræcise formuleringer i rapporten vil jeg understrege, at det fortsat er min opfattelse, at konsortiet bag undersøgelsen sikrer, at PISA-resultaterne er repræsentative for danske 15-16-årige

kompetencer. Det er naturligvis beklageligt, at der findes den slags uklarheder i rapporterne, men jeg ved, at medarbejdere i Undervisningsministeriet sammen med det danske PISA-konsortium arbejder på, at lignende fejl mindskes fremover.

Med venlig hilsen

Christine Antorini

Henvendelse nr 2.

Niels Christoffersen  
Rorupvej 9 Rorup  
4320 Lejre

Til undervisningsminister Christine Antorini

Angående svar på min henvendelse af 9. marts 2015 om PISA-undersøgelsen. Reference: 007.17V.981

Det er med nogen overraskelse, jeg kan konstatere, at ministeren er af den opfattelse, at de danske PISA resultater er repræsentative for de 15-16 årige indskrevet på en uddannelsesinstitution i den danske del af rigsfælleskabet.

Den del af svaret, der er overraskende lyder:

I de samlede analyser tages højde for overrepræsentationen af tosprogede elever ved at vægte stikprøven på en måde, der genskaber repræsentativiteten. Stikprøveudtræk og vægtning foretages af en ekspertgruppe i det internationale konsortium, der står bag PISA-undersøgelsen. Ekspertgruppen får adgang til de officielle lister med samtlige skoler og antallet af elever på 15-16 år på hver skole og designer herefter stikprøverne i alle deltagerlandene, herunder Danmark. Den internationale ekspertgruppe sikrer også, at vægtene genskaber repræsentativiteten.

Svaret ser dog ud til at være i overensstemmelse med det, der fremgår af de danske rapporter fx tabel 6.5 side 161 i 2012 rapporten, hvor man kan se, at vægtene beregnes som (sum af elever i populationen)/(elever der har gennemført testen).

Min overraskelse bygger på følgende:

**Citat fra svar 1:**

I de samlede analyser tages højde for overrepræsentationen af tosprogede elever ved at vægte stikprøven på en måde, der genskaber repræsentativiteten.

**Bemærkning:**

Det er vist rimeligt normalt, at man i surveyundersøgelser søger at forbedre repræsentativiteten via baggrundsoplysninger fra registre. Baggrundsoplysninger i PISA tilvejebringes imidlertid via deltagernes (forældre, skoleledere og elever) svar på spørgeskemaer og for elevernes vedkommende også svar på spørgsmålene i de faglige opgaver. Det virker derfor noget mærkværdigt, at man kan "genskabe repræsentativiteten" med mindre stikprøven er udtrukket repræsentativt - det er den danske stikprøve som bekendt ikke og derfor er de baggrundsoplysninger, man råder over heller ikke repræsentative. Repræsentativiteten sikres ved at udtrække en repræsentativ stikprøve.

**Citater fra svar 1:**

Stikprøveudtræk og vægtning foretages af en ekspertgruppe i det internationale konsortium, der står bag PISA-undersøgelsen.

**Bemærkning:**

Vægtningen foretages som angivet i kapitel 8 i de internationale rapporter. Der bruges 5 justeringsfaktorer, herunder en justeringsfaktor for skoler, hvor der kun deltager med elever fra klasser beregnet for 15 årige. Da Færøerne netop deltager som en dansk undergruppe med alle elever fra 9. klasse må denne justeringsfaktor være brugt i forhold til de

færøske skoler, men denne faktor ændrer ikke ved, at de færøske skoler ikke er en del af det danske skolesystem. Færøerne er nævnt i den internationale rapport side 221 - skoler i stratum 5.

Det nationale center er for øvrigt ansvarlig for blandt andet at

- forberede den ramme skolerne udtrækkes fra (prepared the school sampling frame and submitted this to the international contractor for the selection of the school sample)
- udvælge eleverne (selected the student sample from a list of eligible students provided by the School Coordinators).

Skolelisten bygger på lister fra Danmarks Statistik fra året før undersøgelsen. Udvælgelsen af eleverne foretages øjensynligt af det nationale center.

Listen over det nationale centers opgaver står på side 108 i den internationale rapport.

#### **Citat fra svar 1:**

Ekspertgruppen får adgang til de officielle lister med samtlige skoler og antallet af elever på 15-16 år på hver skole og designer herefter stikprøverne i alle deltagerlandene, herunder Danmark. Den internationale ekspertgruppe sikrer også, at vægtene genskaber repræsentativiteten.

#### **Bemærkning:**

Stratificeringen er et nationalt anliggende (PISA ser dog ud til at skulle overtales af stærke argumenter for at godkende ændringer i stratificeringen). I DNK har man både i 2009 og 2012 brugt en national option, hvor man oversampler skoler og elever, så man får en større andel af elever med anden etnisk baggrund end dansk i stikprøven end andelen i populationen viser. Da oversamlingen er en national option skal de oversamplede elever/skoler ikke indgå i den internationale sammenligning (se side 453 i den internationale rapport), men det gør de øjensynligt. Standarden for nationale optioner er dermed formentlig ikke blevet implementeret korrekt. Måske har PISA ikke opdaget oversamlingen og fjernet de oversamplede elever/skoler fra udtrækket, fordi der ikke er overensstemmelse mellem PISA's definition af indvandrere og den definition, der bruges i de officielle lister fra Danmarks Statistik.

PISA registrerer en elev som native/indfødt, når eleven angiver, at man er født i testlandet og en af forældrene også er født i testlande. Hvis man ser på tosprogede elever, så vil den stramme udlændinge politik, der førte de sidste 12-15 år, sikkert betyde, at de fleste elever med anden etnisk baggrund end dansk er indfødte set med PISA's definition.

Stratificeringen er derfor ikke entydig indenfor den ramme, hvor undersøgelsen foregår, fordi flere elever/skoler kan placeres i flere steder end der, hvor de rent faktisk bliver placeret med den danske definition. PISA har øjensynlig opdaget uoverensstemmelsen, idet der er skiftet navn fra "Minority Enrollment" i 2009 til "Immigrant Levels" i 2012. Men den danske definition/stratificering og udtræksprocedure er ens begge år.

#### **Citat fra svar 1:**

Det har dog ingen betydning for (forvirring omkring betegnelser HØJ ... INGEN), hvordan stikprøven er blevet trukket og analyseret, .....

#### **Bemærkning:**

Lad mig så se på det i lyset af følgende (mine fremhævelser):

The school sampling frame was defined in the *School Sampling Preparation Manual*2 as a **frame that would provide**



**complete coverage of the national defined target population without being contaminated by incorrect or duplicate entries or entries referring to elements that were not part of the defined target population.** It was expected that the school sampling frame would include any school that could have 15-year-old students, even those schools which might later be excluded, or deemed ineligible because they had no PISA-eligible students at the time of data collection. **The quality of the sampling frame directly affects the survey results through the schools' probabilities of selection and therefore their weights and the final survey estimates. NPMs were therefore advised to be diligent and thorough in constructing their school sampling frames (side 70 i den internationale rapport).**

### Fuldstændig dækning

Det fremgår af de internationale rapporter i 2009 og 2012, at Danmark ikke opnåede fuldstændig dækning. Den danske udtræksramme manglede nogle internationale skoler, som skulle have været inkluderet i rammen: Overall exclusions were greater than 5% (6.18%). Data were fully explained – there was a difficulty in defining the school population – some international schools were not included when they should have been - side 284 i den internationale rapport 2012.

### Færøerne

Det er uklart, om deltagelsen af de færøske skoler siden 2006 (jævnfør fx PISA 2006 undersøgelsen – en sammenfatning fremlagt for Uddannelsesudvalget (2. samling) UDU alm. del - Bilag 11 Offentligt) kan betegnes som contamination (**entries referring to elements that were not part of the defined target population**), men det er klart, at de optræder som stratum 5 i den internationale rapport 2012 (side 221) - "and all students in stratum 5 schools (Faroes)" - det er også klart, at det færøske skolesystem ikke er en del af det danske skolesystem. Men deres deltagelse som en del af det danske system må uvægerligt påvirke det danske resultat, da de færøske elever scorer markant lavere end selv de danske tosprogede elever.

### Oversampling af elever med anden etnisk baggrund end dansk.

Det er uklart, om oversamplingen af elever med anden etnisk baggrund end dansk (dansk definition) kan betegnes som en contamination (**incorrect or duplicate entries**) med det fører også til en oversampling af elever med svag socioøkonomisk baggrund jævnfør side 14 i delrapport 1 om kriteriebaseret af de nationale test (elever som både har deltaget i PISA 2012 og de nationale test i læsning/matematik). Socioøkonomisk baggrund har direkte indflydelse på resultatet via non-response mv.

I den tekniske rapport fra 2003 står således følgende - side 211:

In Denmark, students performed surprisingly poorly on booklet 9 when responding to both the science and the reading items. In contrast they performed quite well (relative to other booklets) on problem solving. In addition, it has been noted that the non-responding students (for each domain) have a lower value in the index of economic, social and cultural status (ESCS) than students who did respond to items on each domain. Given the positive correlation between ESCS and achievement, the lower values of ESCS for the students who were not assessed in a domain, and the lower than expected scores on booklet 9, it can be expected that the imputations for the non-assessed students will lead to a reduction in the mean scores in reading and science for Denmark.

Og så kan det i parentes bemærkes, at det også er derfor vægtene, der rapporteres i danske PISA 2012 tabel 6.5 ikke er rigtige (the students who were not assessed in a domain, and the lower than expected scores on booklet 9, it can be expected that the **imputations for the non-assessed students etc.**). Den måde at regne på der vist i tabel 6.5 (sum af elever i populationen)/(divideret med elever som har gennemført testen) forudsætter nok, at alle elever har løst alle opgaver, hvilket jo lang fra er tilfældet i PISA. Se fx diagrammet her <http://videnskab.dk/kultur-samfund/pisa-ranglisten-er-baseret-pa-computerskabte-data>

## Citat fra svar 1

... og det internationale konsortium fastslår, at de danske PISA-undersøgelser opfylder de faglige krav til repræsentativitet.

### Bemærkning

Det internationale konsortium fastslår, at de danske data er blevet inkluderet i databasen. Repræsentativiteten sikres ved at udtrække en repræsentativ stikprøve, hvilket sikres via udtræksrammen jævnfør side 447 i den internationale rapport og en korrekt implementering af nationale optioner:

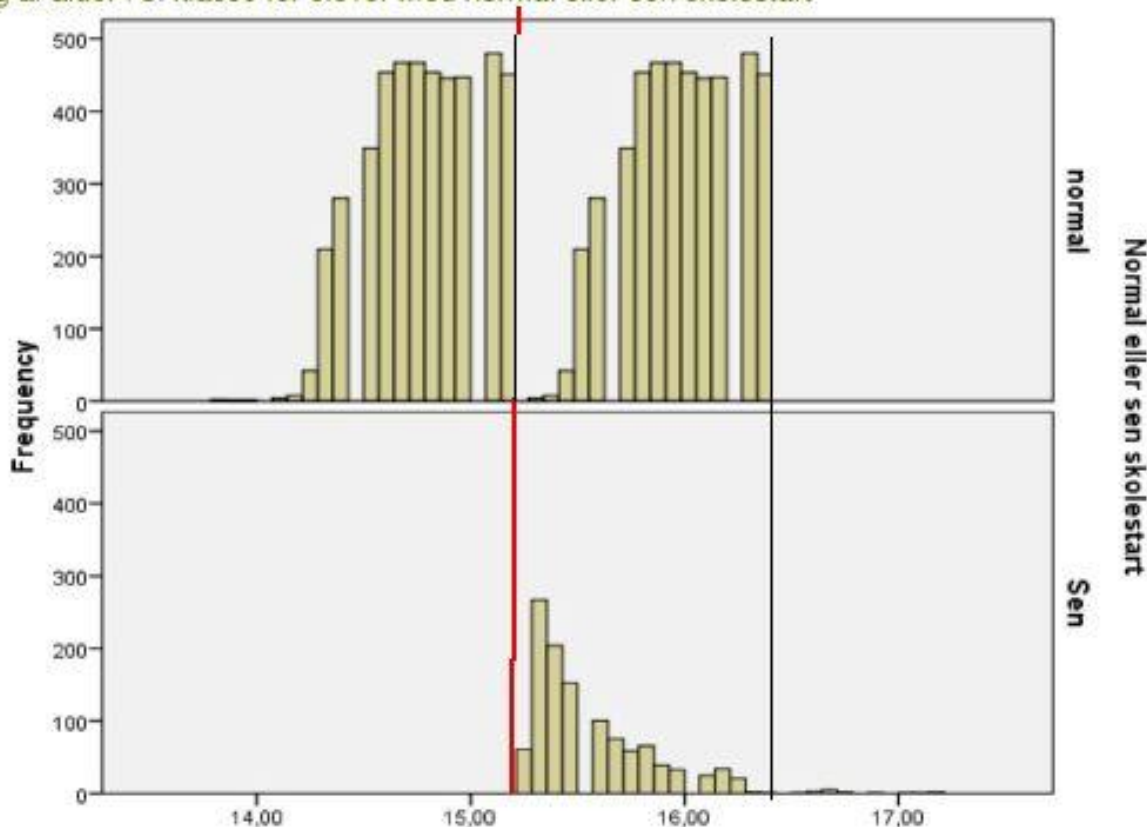
**Generalisability:** Data are collected from specific individuals, in a specific situation, and at a certain point in time. Individuals to be tested, test materials and tasks, etc. should be selected in a way that will ensure that the conclusions reached from a given set of data do not simply reflect the setting in which the data were collected but hold for a variety of settings and are valid in the target population at large. Thus, collecting data from a representative sample of the population, for example, will lead to results that accurately reflect the level of literacy of fifteen-year-old students in a country. (Min fremhævelse)

Det internationale konsortium sikrer ikke repræsentativiteten, de foretager en lang række kontroller, men de forudsætter øjensynlig, at de vedtagne standarder er kendt.

- **Timeliness:** Consistency, precision and generalisability of the data can be obtained in a variety of ways. However, the tight timelines and budgets in PISA, as well as the sheer number of participating countries, preclude the option of developing and monitoring local solutions to be harmonised at a later stage in the project. Therefore, the standards specify one clear-cut path along which data collection and data submission should progress.

Det fremgår af rapporterne om kriteriebaseret af de nationale test figur 4.10, at der har deltaget elever som aldersmæssigt ikke hører med i PISA.

Fordeling af alder i 8. klasse for elever med normal eller sen skolestart



I noten til figuren hedder det:

Alderen i 8. klasse er beregnet ud fra PISA registrering af alderen, idet der er trukket et år fra alderen blandt de elever, der er blevet PISA testet i 9. klasse.

I den øverste del af fig. har jeg "flyttet" (den røde streg) 9. klasse eleverne tilbage, hvor de hører hjemme, idet man har fundet aldersfordelingen i 8 klasse (2011) af de normale skolestartere ved trække et år fra deres alder i 9. klasse (2012). Så stikprøven indeholder elever, som ikke burde deltage. Det er formentlig sket, fordi man udtrukket flere elever end TCS angiver (national option), for eksempel hedder det, at der i stratum 02 blev udtrukket 20 elever (TCS=20) med dansk etnisk baggrund, og alle eleverne med anden etnisk baggrund end dansk blev tilføjet til udtrækket.

TCS er det antal elever (15 årige) man må udtrække.

I den tekniske rapport side 179 angives 104 elever som ineligble students, men der ser tydeligvis ud til at være flere, som faktisk har deltaget.

Sammenfattende:

1. Den danske udtræksramme manglede nogle internationale skoler som skulle have været inkluderet i rammen: Overall exclusions were greater than 5 % (6.18 %). Data were fully explained – there was a difficulty in defining the school population – some international schools were not included when they should have been - side 284.  
Så alle skoler med potentielle 15 årige deltog ikke i lodtrækningen, jævnfør også standard 1.5: Students are sampled using *agreed upon*, established and professionally recognised principles of scientific sampling and in a way that represents the full population of *PISA-Eligible students*.

2. Det kan vises, at den danske stikprøve indeholder elever, som ikke er PISA elever, de kan være smuttet med, fordi man tilføjer alle tosprogede elever udover TCS til udtrækket og så glemte, at der var tale om en national option. Den internationale rapport angiver 104 elever, som værende ineligibile. Men det ser dog ud til, at der er flere.
3. Den danske stikprøve indeholder skoler (stratum 5), som er beliggende på Færøerne. Færøerne klarer sig langt dårligere end selv tosprogede danske elever. Det påvirker selvfølgelig resultatet i nedadgående retning og det kan man ikke vægte sig fra.
4. Den manglende overensstemmelse mellem definitionerne på indvandrere (PISA/Danmarks Statistik) gør, at man vil få to forskellige stratificeringer og dermed vægte. Det endelige resultat afhænger derfor af hvilken definition man bruger. **The quality of the sampling frame directly affects the survey results through the schools' probabilities of selection and therefore their weights and the final survey estimates.**
5. I realiteten har man bare oversamlet elever med svag socioøkonomisk status (se side 14 delrapport 1 om kriteriebaseret af de nationale test).
6. Det ser faktisk ud til, at de oversamlede skoler/elever (national option) deltager i PISA undersøgelsens internationale del, hvilket åbenbart strider mod standarden for implementering af nationale optioner.

#### Spørgsmål

Er ministeren stadig af den overbevisning, at de danske PISA resultater er repræsentative for de 15 årige under uddannelse i den danske del af rigsfælleskabet?

Med venlig hilsen  
Niels Christoffersen  
Rorupvej 9  
4320 Lejre

Kopi til Børne- og Undervisningsudvalget

Ministerens svar på anden henvendelse

Kære Niels Christoffersen

30-04-2015

Tak for din mail fra den 10. april 2015 og dine opfølgende spørgsmål og interesse i PISA.

Da der er tale om meget tekniske spørgsmål, har jeg bedt Danmarks Statistik, som indgår det danske PISA-konsortium, om at kommentere din henvendelse. Danmarks Statistik er eksperter i repræsentative stikprøver, og de gennemfører i 2015 dataindsamlingen på de danske skoler i forbindelse med PISA 2015.

Danmarks Statistik anfører følgende:

”Når man stratificerer en undersøgelse med henblik på at opnå viden om fx tosprogede elever, defineres stratificeringen på forhånd ud fra registre og skolernes indrapportering. Delpopulationen er således fastlagt, før stikprøven udtrækkes. Udvælgelsen sker ved hjælp af det stikprøveprogram, som det internationale PISA-konsortium kræver, at landene bruger. Da det internationale konsortium kender listerne, kan de kontrollere, at udvælgelsen er sket korrekt.

Du har ret i, at vægten er dannet ud fra princippet om, at summen af elever i delpopulationen divideres med summen af gennemførte test i samme delpopulation. Vægtene afhænger af, hvilken delpopulation man tilhører, når stikprøven trækkes. Opregningen laves derudover skole for skole. Hvis der på en skole fx er valgt hver anden tosprogede elev og fjerde dansksprogede elev, har de tosprogede elever på denne skole den halve vægt i forhold til dansksprogede elever, fordi de har den dobbelte udvalgschance. Det er måden, man sikrer repræsentativitet på.

Det mest sikre resultat opnås, hvis man bruger alle observationer i en repræsentativ undersøgelse. Hvis man i eksemplet ovenfor tilfældigt smed halvdelen af de tosprogede testpersoners besvarelser væk, ville de resterende besvarelser for tosprogede og dansksprogede elever få den samme vægt, fordi de har den samme udvalgschance. Dette resultat ville være lidt anderledes, fordi det er baseret på færre observationer, men det ville også være repræsentativt. Sikkerheden ville dog være lidt mindre, fordi man ikke har brugt hele datagrundlaget. Repræsentativitet og sikkerhed er således ikke det samme. Repræsentativitet kræver, at man kender hele populationen på forhånd, trækker tilfældigt i hver delgruppe, og med vægte korrigerer for udvalgschancen. Repræsentativitet er en egenskab, der har med designet, udvælgelsen og opregningen at gøre. Det har ikke noget at gøre med, hvor mange man spørger. En repræsentativ stikprøve bliver mere og mere sikker, jo flere man spørger. Det er derfor, at det internationale PISA-konsortium for Danmark – og for de andre lande med ekstra stikprøver – har valgt ikke at smide den ekstra stikprøve væk. Det ville også se mærkeligt ud, hvis der for Danmark var to lidt forskellige tal: Et med den lille stikprøve i den internationale rapportering, og et lidt andet i den danske rapport, hvor man analyserer forskellen mellem tosprogede og dansksprogede elevers resultater. At de ikke kan være helt ens skyldes, at det er to forskellige repræsentative mål for det samme, med hver deres statistiske sikkerhed.

Eleverne får forskellige delmængder af det samlede opgavesæt. Det fastlægges forud for hver test på skolen og sker på en balanceret måde i det stikprøveprogram, som alle landene skal bruge. Nogle får fx delmængde 1 og 2, andre 2 og 3, og andre igen 3 og 1. Man kan korrigere for, at man ikke har svaret på alle

opgaverne på mange måder, men den bedste er at imputere den delmængde, eleverne ikke har svaret på ud fra den delmængde, de har svaret på. Det er også denne løsning det internationale konsortium har valgt.

I alle lande er der skoler eller elever, der ikke ønsker at deltage, samt specialskoler og internationale skoler, der ikke deltager. Alle landene skal forud for, at stikprøven trækkes, sende lister med samtlige skoler i landet til det internationale PISA-konsortium. Det internationale konsortium har uafhængige kontrollanter i alle landene, der styres og betales af det internationale konsortium. På baggrund af de officielle lister over samtlige skoler estimeres, hvor mange skoler og elever der har deltaget, og deres andel af elever. Dette sker på samme måde i Danmark som i alle andre lande”.

Angående dine betragtninger om Færøerne, skal det anføres, at de færøske data ikke indgår i det danske datasæt.

Jeg henholder mig til Danmarks Statistiks besvarelse, og det er fortsat min vurdering, at de danske PISA-resultater er repræsentative for 15-16-årige under uddannelse i Danmark.

Hvis du har yderligere spørgsmål til den praktiske gennemførelse af stikprøve mv., vil jeg anbefale, at du kontakter Danmarks Statistik. Du kan henvende dig til kontorchef Peter Linde i DST Survey ([pli@dst.dk](mailto:pli@dst.dk)).

Med venlig hilsen

Christine Antorini