

*Notat for Ungdoms- og Uddannelsesforvaltningen
i Københavns Kommune*

**Elevsammensætning i Københavns Kommunes folkeskoler
belyst ved PISA-København-data**

Af Beatrice S. Rangvid, akf, amternes og kommunernes forskningsinstitut

1. Sammenfatning

Dette notat anvender talmaterialet fra PISA-København-undersøgelsen, som i 2004 er blevet gennemført i Københavns Kommune som en særligt tilpasset undersøgelse i det internationale PISA-koncept. Formålet med notatet er at se på den overordnede sammenhæng mellem elevsammensætning i skolen og de dansk- og tosprogede elevers test-scores i læsning, matematik og naturvidenskab. Undersøgelsen er baseret på deskriptive analyser, samt regressionsanalyser, hvor der tages højde for elevernes forskelle i baggrundskaraktistika.

Helt konkret besvares følgende spørgsmål:

Klarer tosprogede elever, der går i skoler med lav tosprogsandel, sig bedre end elever på skoler med høj andel?

Klarer elever med en svag social baggrund sig bedre i skoler med i gennemsnit høj social baggrund end i skoler med en gennemsnitlig lav social baggrund?

Resultaterne peger på, at:

- ⇒ De deskriptive analyser viser, at der er en signifikant negativ sammenhæng mellem andel tosprogede elever på skolen og både de dansk- og tosprogede elevers testscores i læsning, matematik og naturvidenskab. Der er dog ikke forskel i test-scores for elever, hvad enten de går i skoler med mange tosprogede (omtrent 50-85%) eller i såkaldte højkonzentrationsskoler (mere end 85% tosprogede). Når man i regressionsanalyser tager højde for forskelle i elevernes baggrund, modificeres resultaterne i retning af, at der op til en tosprogsandel på omkring 50% på skolen ikke er forskelle i elevernes testscores, dvs. at der ikke er (en statistisk signifikant) forskel på gennemsnitlige testscores for elever, der går i skoler med 10% tosprogede og for elever, der går i en skole med 30% tosprogede. For lave tosprogs-konzentrationer på skolerne synes der efter hensyntagen til social baggrund således ikke at være sammenhæng mellem andel tosprogede elever på skolen og elevernes testresultater. Det samme gælder for høje tosprogsprocenter. Men når man sammenligner elever på skoler med lav tosprogsandel med elever

på skoler med mange tosprogede, så scorer både dansk- og tosprogede elever højere på skoler med lav tosprogsprocent.

- ⇒ Resultaterne for elevsammensætning, når den måles med social baggrund ligner meget mønsteret ovenfor, men resultaterne er gennemgående lidt mindre sikre (statistisk signifikante). Der kan ikke skelnes empirisk mellem, hvilken af de to faktorer (tosprogethed eller social baggrund), der er den afgørende i sammenhængen mellem skolens elevsammensætning og testscores.
- ⇒ Sammenhængen mellem elevsammensætning og elevernes testscores er lige kraftig for dansk- og tosprogede elever. Sammenhængen mellem elevsammensætning og elevernes testscores er stærkest og mest sikker for læsefærdighederne og lidt mindre sikker og kraftig for naturvidenskab og (især) matematik.

2. Indledning

Københavns Kommune har i 2004 gennemført en særligt tilpasset undersøgelse i det internationale PISA-koncept. Det blev besluttet, at testen skulle omfatte alle 9.-klasses-elever – uanset alder. Undersøgelsen indgår i et større kvalitetsudviklingsprojekt af folkeskolerne i København. I den forbindelse har Ungdoms- og Uddannelsesforvaltningen i Københavns Kommune finansieret dette notat vedrørende sammenhængen mellem elevsammensætningen i skolen og elevernes PISA-testscores.

2.1 Formål

Helt konkret belyser analysen følgende spørgsmål:

- ⇒ Klarer tosprogede elever, der går i skoler med lav tosprogsandel, sig bedre end elever på skoler med høj andel?
- ⇒ Klarer elever med en svag social baggrund sig bedre i skoler med en gennemsnitlig høj social baggrund end i skoler med en gennemsnitlig lav social baggrund?

2.2 Datagrundlag

Datagrundlaget for analysen er først og fremmest den københavnske PISA-undersøgelse, som omfatter i alt 2352 elever fordelt på 83 folke- og privatskoler i Københavns Kommune^{1 2}. Populationen for analyserne er dog noget mindre, da den begrænses til folkeskolerne. Ud af de 1667 folkeskoleelever i datamaterialet er der 476 tosprogede elever fra tredjelande³ (uden for Vesteuropa, USA, Canada, Japan, Australien og New Zealand), svarende til 28% af folkeskoleeleverne i analysedatasættet (jf. også boks 1)⁴.

PISA-testscores er blevet normeret, så det internationale gennemsnit af alle OECD-landene er 500 testpoint og spredningen omkring gennemsnittet (målt ved standardafvigelsen) er 100 point. For en mere detaljeret beskrivelse af testdesignet henvises til Egelund (2005a).

2.3 Metode

Den anvendte *metode* er først og fremmest deskriptive analyser, dog suppleret med (multiple) regressionsanalyser for at tage højde for elevernes forskellige sociale baggrund. Regressionsanalyser er ét skridt i retning af at afdække selvstændige effekter af elevsammensætningen på elevernes testcores.

2.4 Baggrund for notatet

Elevsammensætningen på skolen menes at have en selvstændig betydning for elevernes boglige færdigheder. Der er tre nordiske undersøgelser, der peger på, at eleverne opnår bedre færdigheder, når de andre elever på skolen kommer fra ressourcerstærke hjem, end hvis der er en overvægt af elever fra mindre velstillede hjem (Skolverket 2003, Rangvid 2003, Arbejderbevægelsens Erhvervsråd 2003). Ønsket til denne undersøgelse var at se nærmere på sammenhængen mellem elevsammensætningen og elevernes testede færdigheder i PISA-København-datamaterialet. Emnet ønskes belyst ud fra to dimensioner: tosprogethed og socioøkonomisk baggrund. Analyserne inkluderer alene folkeskoleelever.

Boks 1: Definition af tosprogethed

Tosprogede elever er i denne undersøgelse defineret som elever, der i Københavns Kommunes administrative elevregister er opført som tosprogede (dvs. at deres forældre ved barnets skolestart har oplyst, at det sprog, der hovedsageligt bliver talt i hjemmet, er et andet sprog end dansk), og som har oprindelse i et land uden for den vestlige kulturkreds (dvs. uden for Vesteuropa, USA, Canada, Japan, Australien og New Zealand). Der er således tale om tosprogede elever fra *ikke*-vestlige lande.

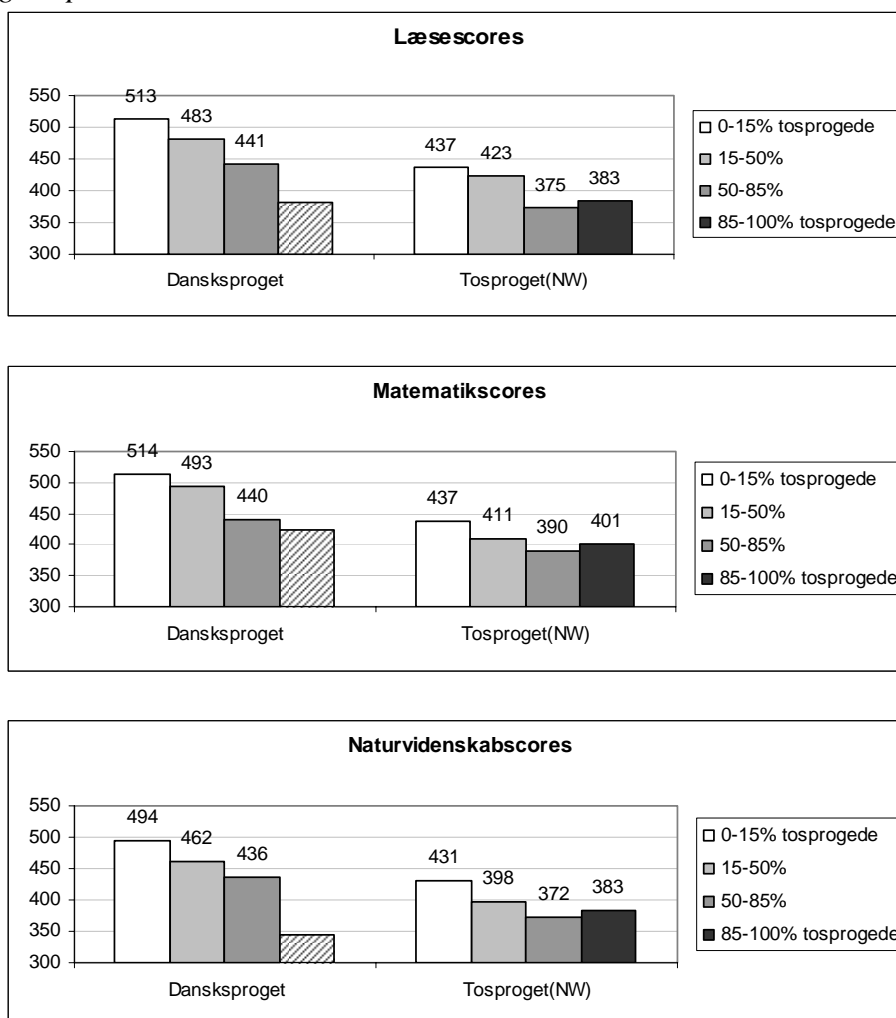
3.1 Elevsammensætning målt ved andel tosprogede elever

For at se på, hvordan eleverne i skoler med forskellig høj tosprogsprocent klarer sig, er skolerne opdelt efter, hvor stor en andel af 9.-klasses-eleverne, der er tosproget. Målet for elevsammensætningen er baseret på PISA-København-datamaterialet. I PISA-København er der ikke oplysninger for alle 9.-klasses-elever, men kun for de elever, der rent faktisk har deltaget i PISA-testen. Som vist i Rangvid (2005a), er der på nogle sko-

ler et stort frafald ved PISA-testen, det vil sige, at der er en del 9.-klasser-elever, der ikke er med i PISA-datasættet. Fra Københavns elevregister haves dog informationer om alle elever fra efteråret 2003 for folkeskolerne⁵. Disse data er i Bilag A brugt til at se på, om anvendelsen af PISA-data, til at beregne andelen af tosprogede elever, er en acceptabel tilnærmelse. Resultaterne fra denne analyse viser, at sammenfaldet af de to mål er rimeligt pænt. I den videre analyse anvendes derfor det på PISA-data beregnede mål for andel tosprogede elever.

I hovedanalysen tages udgangspunkt i andel tosprogede blandt alle 9.-klasser-elever på skolen, mens der i Bilag B er medtaget en analyse, som er baseret på andelen af tosprogede på klasseniveau. Det viser sig, at konklusionerne er uændrede, hvad enten elevsammensætning måles på skole- eller klasseniveau.

Figur 3.1 : Gennemsnitlige testscores for dansk- og tosprogede elever efter andel tosprogede på skolen



Note: Resultaterne for dansksprogede elever i skoler med mere end 85% tosprogede elever er upålidelige (på grund af meget få observationer i denne kategori), og resultaterne er derfor særligt markeret.

3.1.1 Gennemsnitlige færdigheder og andel tosprogede

Figur 3.1 (og tabel A3.1) viser de gennemsnitlige færdigheder i læsning, matematik og naturvidenskab for dansksprogede og tosprogede folkeskoleelever. Eleverne er opdelt efter, hvor stor tosprogsprocenten er i deres skoler. Der er fire kategorier: elever, der går i skoler med (i) op til 15% tosprogede elever, (ii) mellem 15 og 50%, (iii) mellem 50 og 85%, og (iv) mere end 85%. Der er kun ganske få dansksprogede elever, der går i en skole med en tosprogsprocent over 85, hvorfor resultaterne er for usikre til at drage konklusioner for den elevgruppe (resultater for denne gruppe er vist i figurerne og tabellerne, men er særligt markeret).

Resultaterne i figur 3.1 viser, at de gennemsnitlige testscores falder gennem de første tre kategorier for både dansksprogede og tosprogede elever. For de tosprogede elever sker der dog derefter ikke et yderligere fald i testscore mellem 3. og 4. kategori (dvs. forskellen i de gennemsnitlige testscores mellem 3. og 4. kategori er ikke statistisk signifikante).

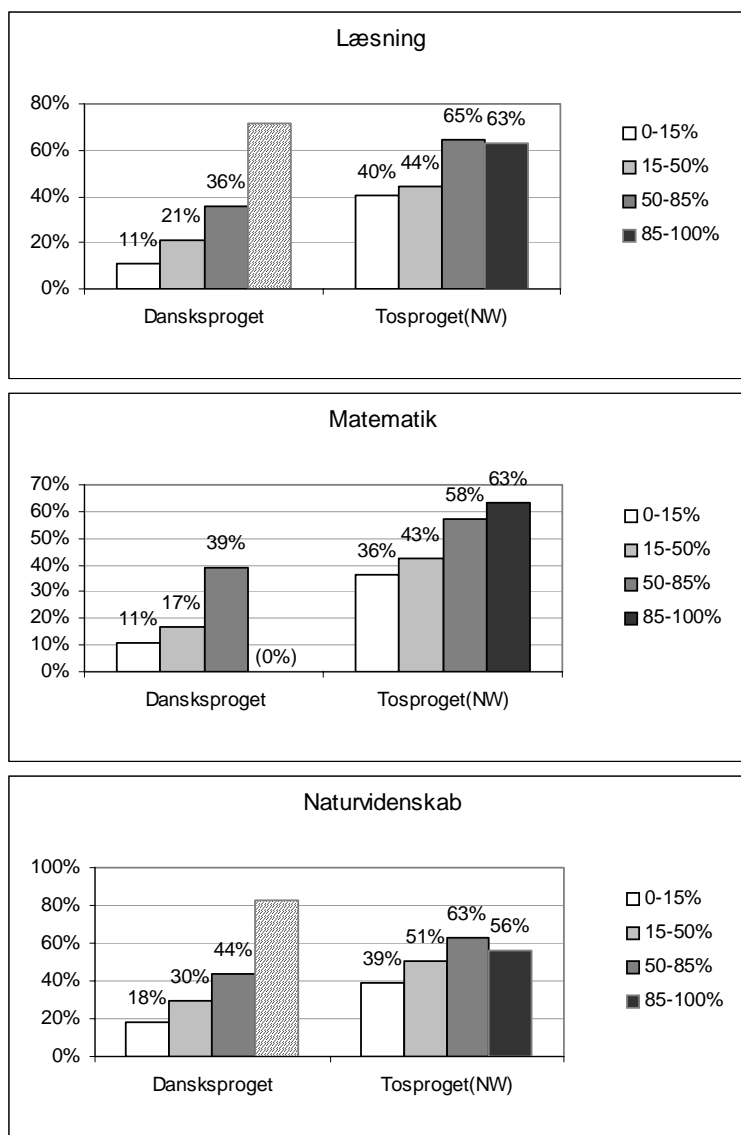
Det bemærkes, at de gennemsnitlige resultater for tosprogede elever er ikke lavere for højkonzentrationsskoler (med flere end 85% tosprogede) end i skoler med en tosprogsprocent mellem 50 og 85. For at se, om resultaterne er følsomme over for måden, kategorierne er defineret på, beregnes gennemsnittene for lidt anderledes inddelinger af kategorierne. Når man ændrer kategori grænserne til 0-25%, 25-50%, 50-75%, 75-100%, er konklusionerne uændrede (resultaterne er ikke vist her).

3.1.2 Elever med særligt lave testscores

Figur 3.1 ovenfor viste elevernes *gennemsnitlige* færdigheder efter tosprogsprocenten i deres skole. Der er dog særdeles stor interesse omkring de svage elever i folkeskolerne. Derfor viser figur 3.2 (og tabel A3.1) andelen af elever med svage læsefærdigheder (på PISA-niveau 1 eller derunder) for de fire kategorier af skoler. Figuren skal læses således: 11% af de dansksprogede elever, der går i skoler med lav tosprogsprocent (0-15% tosprogede) har svage læsefærdigheder, mens 40% af de tosprogede elever, der går i sådanne skoler, er blandt de svage læsere. Af de tosprogede elever, der går i skoler med et flertal af tosprogede elever, har flere end 60% svage læsefærdigheder.

Alt i alt peger resultaterne i figur 3.2 på, at der er flere elever med særligt lave scores (både dansk- og tosprogede) i både læsning, matematik og naturvidenskab i skoler med højere tosprogsprocent – dog synes andelen af svage læsere ikke at stige yderligere fra den næsthøjeste til den højeste kategori efter tosprogsprocenten, da forskellen i de gennemsnitlige testscores mellem 3. og 4. kategori ikke er statistisk signifikant (for tosprogede elever)⁶.

Figur 3.2: Andel elever med svage færdigheder for dansk- og tosprogede elever efter andel tosprogede på skolen



Note: Resultaterne for dansksprogede elever i skoler med mere end 85% tosprogede elever er upålidelige (grundet meget få observationer i denne kategori), og resultaterne er derfor særligt markeret.

3.1.3 Samvariation mellem andel tosprogede elever på skolen og elevernes testcores belyst ved regressionsanalyse

Regressionsanalysen kan give svar på følgende forhold, som er vanskelige at belyse ved den deskriptive analyse i de foregående afsnit:

- Er der en selvstændig sammenhæng mellem elevsammensætningen på skolen og elevernes læsefærdigheder, eller fremkommer det ovenfor observerede mønster

udelukkende, fordi ressourcestærke elever i højere grad går på skoler med en mere favorabel elevsammensætning?

- Er denne sammenhæng stærkere for tosprogede end for dansksprogede elever?
- Er sammenhængen den samme for høje og lave tosprogskoncentrationer, eller er den forskellig? For eksempel så vi før, at der ikke synes at være et yderligere fald for tosprogede i skoler med meget høje tosprogsprocenter.
- Er det elevernes tosprogethed eller deres sociale baggrund der betyder noget for deres færdigheder? Eller kan vi slet ikke sige noget om det med statistiske metoder (belyses i afsnit 3.3)?

Multipel regressionsanalyse er en statistisk metode, der sikrer, at der groft sagt sammenlignes elever, der ligner hinanden med hensyn til en række karakteristika (fx forældreuddannelse, bolig, orientering i hjemmet, m.m.). Man kan herved, ud fra den statistiske model, forudsige forskellen i læsefærdigheder for to elever, der ligner hinanden, men som går i skoler med forskellig tosprogsprocent. Med andre ord sikres ved denne metode, at der kun sammenlignes elever, der ligner hinanden med hensyn til de i regressionsanalysen inkluderede karakteristika (se også boks 1).

Der er i regressionerne taget højde for følgende baggrundskarakteristika⁷:

- Elevens køn
- Antal søskende
- Om eleven bor sammen med begge forældre
- Forældrenes højeste uddannelse
- Forældrenes højeste stillingsplacering
- Om moderen og faderen har fuldtidsarbejde
- Om eleven er »dansker«, indvandrer eller efterkommer⁸
- Indikatorer for såkaldt kulturel kommunikation⁹, for social kommunikation¹⁰, for uddannelsesressourcer i hjemmet¹¹, for kulturelle besiddelser¹², og for antallet af bøger i hjemmet.

Resultaterne beregnes for dansksprogede og tosprogede adskilt, da det kan tænkes, at effekten ikke er ens for elevgrupperne.

Hovedresultaterne fra regressionsanalyserne er vist i tabel 3.1. Følgende hovedresultater kan trækkes ud af tabellen:

- Der er en stærk og statistisk sikker sammenhæng (i negativ retning) mellem tosprogsprocenten og testscores for dansksprogede og tosprogede elever (sammenhængen for tosprogede i matematik er dog kun marginalt sikker (på 10%-niveau)).
- Der er ikke en statistisk sikker forskel mellem koefficientstørrelserne for dansk- og tosprogede elever, dvs. sammenhængen er lige stærk for dansk- og tosprogede elever¹³.

Tabel 3.1: Regressionsresultater – Estimeret (lineær) sammenhæng mellem tosprogsprocent og danske og tosprogede elevers testscores efter korrektion for social baggrund

Tosprogsprocent (Andel tosprogede elever på skolen)		Læsescores			
		Koeff.	Std.fejl	Sign.niveau	R-sq
Dansksprogede elever	Obs=1191	-59.44	16.18	0.001	0.330
Tosprogede elever	Obs=476	-76.05	20.61	0.001	0.210
		Matematikscores			
		Koeff.	Std.fejl	Sign.niveau	R-sq
Dansksprogede elever	Obs=656	-41.84	18.41	0.027	0.294
Tosprogede elever	Obs=279	-34.34	20.35	0.097	0.210
		Naturvidenskabscorer			
		Koeff.	Std.fejl	Sign.niveau	R-sq
Dansksprogede elever	Obs=668	-54.02	19.63	0.008	0.287
Tosprogede elever	Obs=257	-52.12	25.25	0.044	0.298

Note: Koefficienter, der er signifikante på 5- hhv. 10%-niveau er angivet med hhv. fed og kursiv skrift.

Desuden er der i regressionsanalysen forsøgt at se på, om sammenhængen mellem en højere tosprogsprocent og faldende testcores er forskellig for forskellige niveauer af tosprogsprocenten (tabel 3.2). Resultaterne må dog mere ses som et fingerpeg end som håndfaste konklusioner, fordi mange resultater er statistisk usikre. Tendensen bekræfter dog den mere simple analyse i figur 3.2 ovenfor: at en højere tosprogsprocent samvarierer med lavere testcores for dansksprogede elever¹⁴, mens der ikke sker et yderligere fald i testcores for de tosprogede elever for skoler med meget høje tosprogsprocenter. Når man som her tager højde for elevernes baggrund, er der dog ikke – som var tilfældet i den deskriptive analyse i afsnit 3.1.1 – længere signifikant forskel på de gennemsnitlige testcores i skoler med op til 15% tosprogede elever sammenlignet med elever på skoler med 15-50% tosprogede.

Der er desuden foretaget analyser, hvor andel tosprogede måles på klasseniveau frem for på skoleniveau (se Bilag B). Resultaterne viser, at det ikke ændrer konklusionen, hvad enten elevsammensætningen måles på klasse- eller på elevniveau¹⁵.

Tabel 3.2: Regressionsresultater – Estimeret sammenhæng mellem tosprogsprocent og danske og tosprogede elevers testscores efter korrektion for social baggrund (ikke-lineær sammenhæng: 4 kategorier for tosprogsprocent)

Tosprogsprocent (Andel tosprogede elever)		Læsescores				
		Koeff.	Std.fejl	Sign.niveau	R-sq	
Dansksprogede elever	Obs=1191	0-15%	Sammenligningskategori			0.33
		15-50%	-11.29	7.830	0.155	
		50-85%	-32.58	9.923	0.002	
		85-100%	-43.47	16.470	0.011	
Tosprogede elever	Obs=476	0-15%	Sammenligningskategori			0.22
		15-50%	-6.19	17.320	0.720	
		50-85%	-53.07	18.500	0.006	
		85-100%	-47.66	21.750	0.033	
		Matematikscores				
		Koeff.	Std.fejl	Sign.niveau	R-sq	
Dansksprogede elever	Obs=656	0-15%	Sammenligningskategori			0.30
		15-50%	-2.05	6.960	0.77	
		50-85%	-26.87	8.650	0.003	
		85-100%	-52.78	37.160	0.160	
Tosprogede elever	Obs=279	0-15%	Sammenligningskategori			0.21
		15-50%	-14.76	22.840	0.521	
		50-85%	-31.27	23.010	0.18	
		85-100%	-26.66	25.365	0.298	
		Naturvidenskabscores				
		Koeff.	Std.fejl	Sign.niveau	R-sq	
Dansksprogede elever	Obs=668	0-15%	Sammenligningskategori			0.29
		15-50%	-10.77	8.650	0.22	
		50-85%	-28.91	9.660	0.004	
		85-100%	-53.30	33.690	0.119	
Tosprogede elever	Obs=257	0-15%	Sammenligningskategori			0.30
		15-50%	-25.54	20.150	0.21	
		50-85%	-47.16	18.318	0.013	
		85-100%	-39.53	25.320	0.124	

Note: Koefficienter, der er signifikante på 5%-niveau er angivet med fed skrift.

3.2 Elevsammensætning målt ved socioøkonomisk baggrund

Tosprogethed er dog kun én dimension af elevernes baggrund. Familiernes socioøkonomiske status er en anden. Selv om det ville være meget nyttigt at kunne afgøre, om det er tosprogethed i sig selv, eller dens samvariation med en lav social baggrund, der er af betydning for eksistensen af korrelationen mellem elevsammensætningen i skolen og elevernes færdigheder, så er det dog i praksis meget vanskeligt med statistiske metoder at skelne mellem disse to faktorer (se afsnit 3.3 herom). Det skyldes, at samvariationen mellem det at være tosproget og en lav social baggrund er ret stor. For eksempel har 54% af de tosprogede børns mødre højst en grundskoleuddannelse. Blandt danske elever er tallet kun 18%. For fædrene er der en lignende forskel.

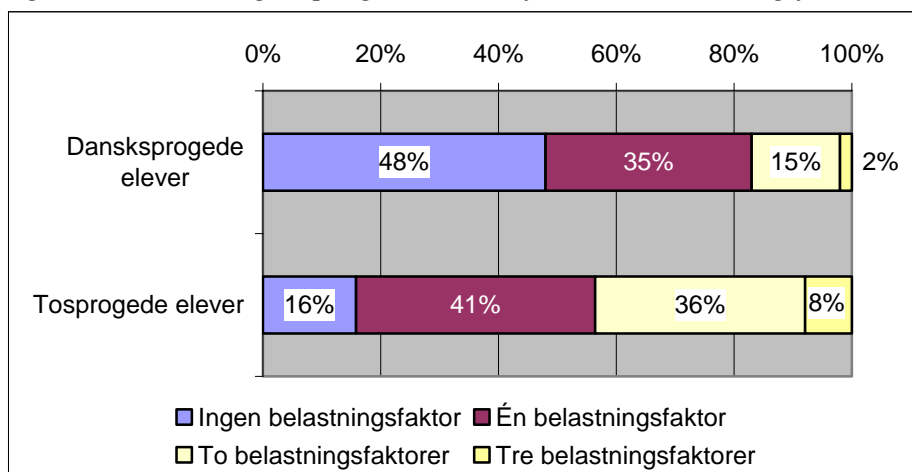
For at foretage en sådan analyse er der brug for et éndimensionalt mål for social baggrund. Som beskrevet i boks 2 er der mange måder at »koge« de forskellige dimensioner af social baggrund ned i en éndimensionel indikator. Her er det som i notatet af Arbejderbevægelsens Erhvervsråd (2003) valgt at benytte summen af tre mulige såkaldte belastningsfaktorer (ingen af forældrene har mere end en grundskoleuddannelse, barnet bor ikke sammen med mor og far, mindst én af forældrene er ikke i arbejde). Elever med to eller tre belastningsfaktorer siges at have en »svag« baggrund, elever med én belastningsfaktor er i mellemgruppen, mens elever uden belastningsfaktor har en »stærk« social baggrund. Den valgte metode udmærker sig ved, at den er ret ligefrem og intuitivt forståelig. Det har til gengæld måske den svaghed, at den er lidt »grov i masken«. Derfor er de samme analyser foretaget med et alternativt mål for socioøkonomisk baggrund (dannet ved principalkomponentanalyse), og resultaterne derfra er vist i bilag.

Boks 2: Definition af socioøkonomisk baggrund

Socioøkonomisk baggrund består af mange dimensioner, fx forældreuddannelse, -beskæftigelse, -indkomst og familiemønstret. Der er i empiriske undersøgelser brugt forskellige måder at opsummere disse dimensioner på. For at kunne opdele elever i forskellige sociale grupper, er det i dette notat valgt at bruge (næsten) den samme metode, som er blevet brugt i Arbejderbevægelsens Erhvervsråds (2003) undersøgelse (for hovedanalyserne)¹⁶. Der defineres her tre såkaldte »belastningsfaktorer« – og jo flere belastningsfaktorer, desto svagere siges elevens sociale baggrund at være: (i) ingen af forældrene har mere end én grundskoleuddannelse, (ii) barnet bor ikke sammen med mor og far, og (iii) mindst én af forældrene er ikke i arbejde (dvs. er ledig eller står uden for arbejdsmarkedet). For at se, om resultaterne, opnået ved denne metode, er robuste, er hovedanalysen gentaget med en alternativ metode (principalkomponentanalyse), som beskrevet i afsnit 3.3. Desuden er den alternative metode nyttig på grund af rekvirentens ønske om at kunne se på de 15% elever med hhv. højeste og laveste forældrebaggrund i kapitel 3. Det kræver en kontinuert variabel, hvor man kan »skære igennem« præcist ved et bestemt procenttal (her: 15%).

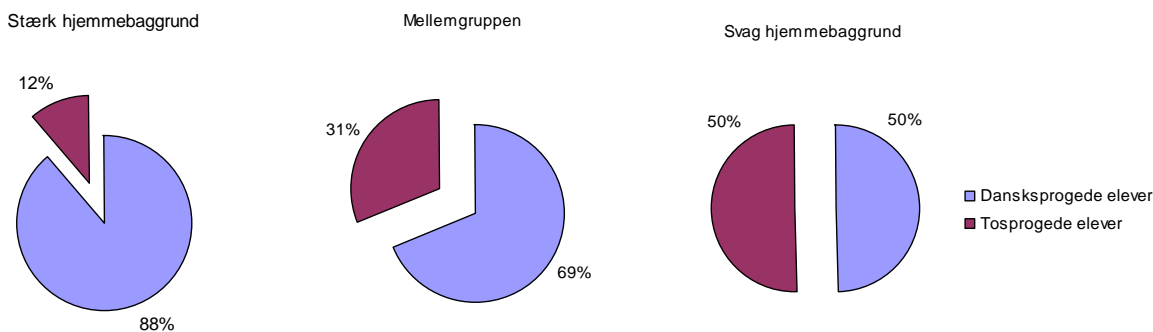
Ovenfor blev det antydnet, at der er en sammenhæng (som dog på ingen måde er deterministisk) mellem at være tosproget og at have en svag social baggrund. Figur 3.3 viser, hvor stor en andel af de hhv. dansk- og tosprogede elever, der befinder sig i hver af de fire kategorier. Næsten halvdelen af de dansksprogede elever har ingen belastningsfaktorer og er dermed i kategorien »stærk social baggrund«, mod kun 16% af de tosprogede elever. Omvendt har kun 17% af de dansksprogede elever en svag baggrund (to eller tre belastningsfaktorer), mod 44% af de tosprogede elever.

Figur 3.3: Dansk- og tosprogede elever efter antal belastningsfaktorer



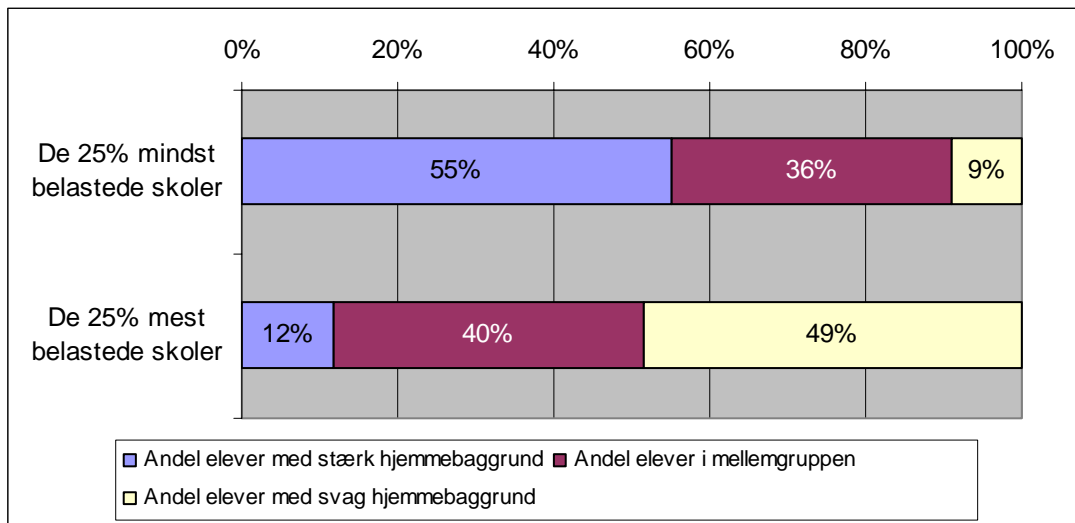
Ser man på, hvordan de tre kategorier (stærk/svag baggrund, samt mellemgruppen) er sammensat (figur 3.4), efter om eleverne er dansk- eller tosprogede, ses, at de tosprogede elever er underrepræsenteret i gruppen med stærk social baggrund (hvoraf de udgør 12%), jævnt repræsenteret i mellemgruppen (31%)¹⁷, mens de udgør hele 50% af gruppen med svag social baggrund.

Figur 3.4: Sammensætning af de tre kategorier for social baggrund



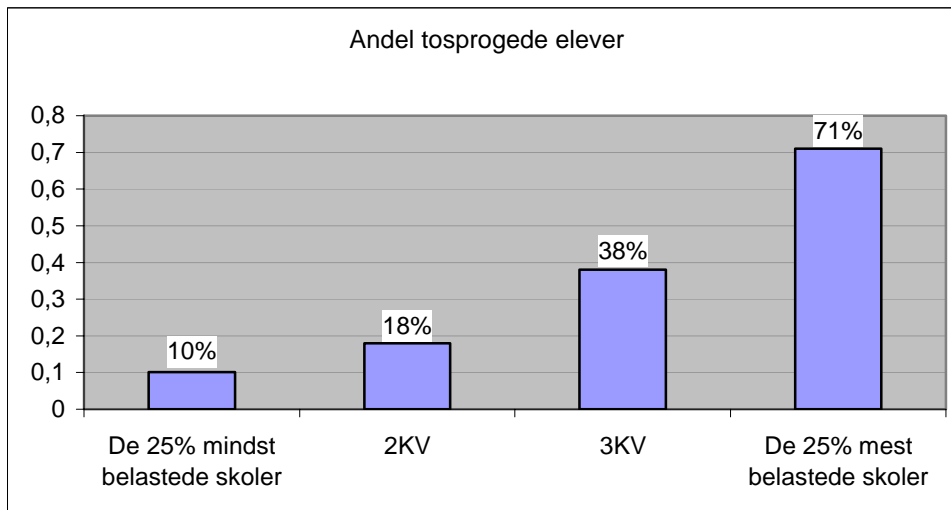
Skolerne deles nu ind i fire kvartiler efter, hvor stor en andel børn med svag baggrund (dvs. med to eller tre belastningsfaktorer), der går på skolen (på 9. klassetrin). I første kvartil (dvs. de 25% af skolerne med den laveste andel af elever med svag baggrund) har skolerne 3-14% elever med svag baggrund. I anden, tredje og fjerde kvartil har skolerne en andel på 16-26%, 26-39% og 40-73% elever med svag baggrund.

Figur 3.5: Elevsammensætning efter social baggrund på de 25% hhv. mest og mindst belastede skoler



Figur 3.5 viser den gennemsnitlige elevsammensætning på de 25% mindst belastede og de 25% mest belastede skoler. Hvor andelen af elever med stærk baggrund (dvs. uden en enkelt af ovenstående belastningsfaktorer) er i gennemsnit 55% i de 25% mindst belastede skoler, så er tallet helt nede på 12% i de mest belastede skoler. Omvendt er der blandt eleverne i de mindst belastede skoler kun 9% med svag baggrund, hvor næsten halvdelen af alle elever i de mest belastede skoler har en svag baggrund.

Figur 3.6: Andel tosprogede elever på de mest og mindst socialt belastede skoler



Når vi ser på sammenhængen mellem andelen af tosprogede elever og andelen af elever med svag social baggrund, viser der sig en tydelig sammenhæng (figur 3.6). I de 25%

mindst belastede skoler er der i gennemsnit 10% tosprogede 9.-klasser-elever, mens 71% af eleverne i de 25% mest belastede skoler er tosprogede.

Gennemsnitlige færdigheder

Figur 3.7 (og tabel A.3.3) viser de gennemsnitlige testscores for elever med hhv. stærk, mellem og svag social baggrund efter, hvor belastet skolerne er med hensyn til andelen af elever med svag social baggrund. Figuren læses således (fx panelet øverst til venstre): elever med stærk social baggrund (dvs. ingen belastningsfaktor), som går i en af de 25% mindst belastede skoler, scorer i gennemsnit 518 PISA-point i læsetesten, mens de, der går i en af de 25% mest belastede skoler, kun scorer 415 point.

To hovedpunkter opsummerer resultaterne fra Figur 3.7:

- Uanset om eleverne selv har en stærk, mellem eller svag social baggrund, så har elever, der går på mindre belastede skoler (dvs. skoler med få elever med svag social baggrund) i gennemsnit højere scores end elever på mere belastede skoler (for matematik og naturvidenskab er sammenhængen dog ikke helt så tydelig for gruppen af elever med svag familiebaggrund).
- Resultaterne for naturfag antyder desuden, at eleverne i de mest belastede skoler har omtrent lige lave testscores, uanset om de selv har en stærk eller svag social baggrund (ligger på mellem 392 og 395 punkter).

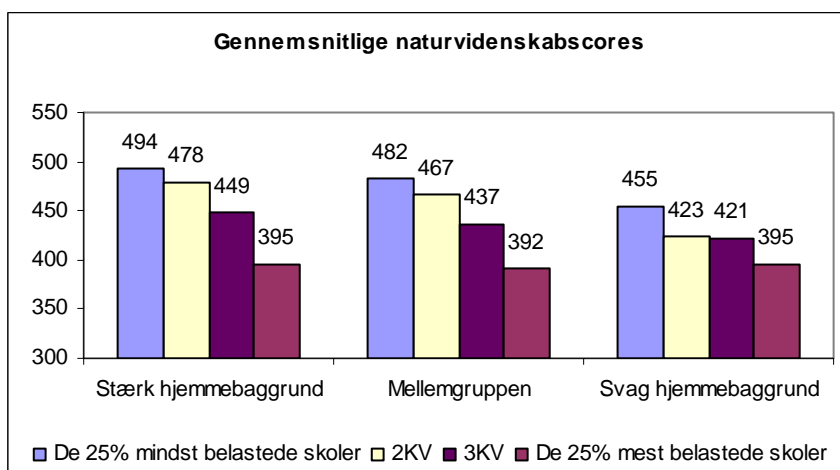
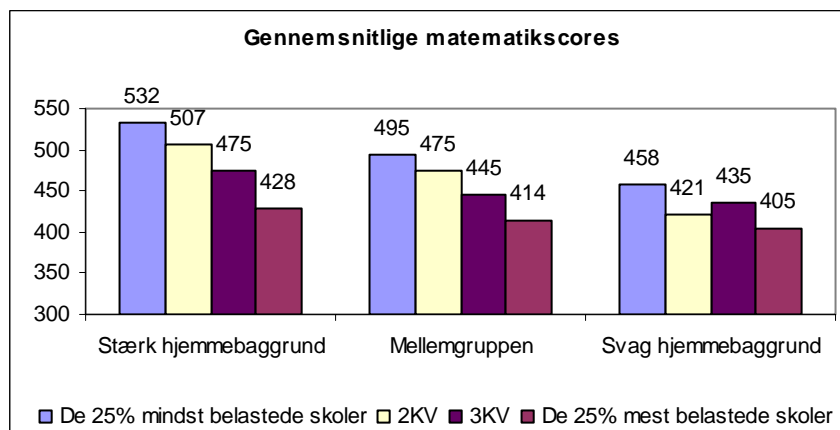
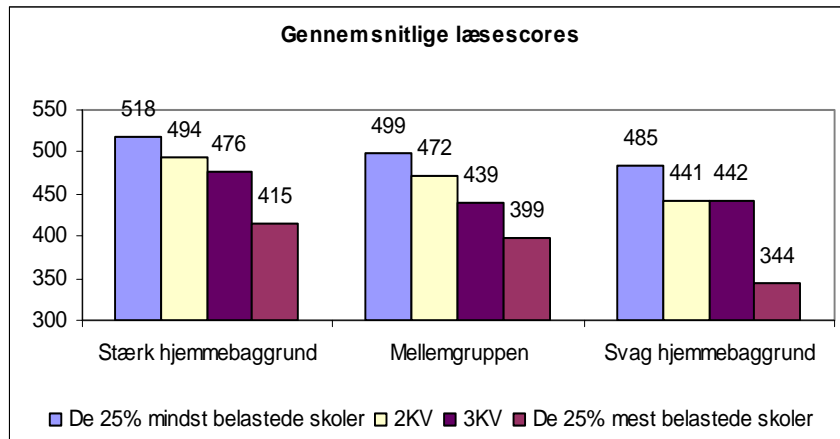
Figur A3.1 (og tabel A3.4) i appendiks viser de tilsvarende analyser beregnet ved det alternative mål for social baggrund dannet ved principalkomponentanalyse (PCA). I store træk tegner resultaterne herfra et lignende billede som ovenfor.

Analysen her er dog ret grovkornet, fordi der kun er fire niveauer for elevsammensætningen i skolen, og kun tre kategorier for elevens egen sociale baggrund. I det følgende afsnit benyttes derfor regressionsanalyse, som tillader en mere forfinet analyse, for at tjekke ovenstående resultater.

3.2.1 Regressioner for elevsammensætning efter socioøkonomisk baggrund

Resultaterne fra regressionsanalyserne, hvor elevsammensætningen måles efter, hvor stor en andel af eleverne, der har en svag social baggrund (dvs. 2 eller 3 belastningsfaktorer), ses i tabel 3.4¹⁸. Resultaterne for *alle elever* (med korrektion for social baggrund) viser, at når man ser på alle elever på én gang, så er der en statistisk sikker sammenhæng mellem andel elever med svag social baggrund og elevernes læse- og naturvidenskabs-testscores (det sidste dog kun på 10%-niveau), mens sammenhængen ikke er statistisk sikker for matematik.

Figur 3.7: Gennemsnitlige testscores for elever med stærk, mellem, og svag social baggrund efter elevsammensætningen i skolen



Holder man disse resultater op mod, hvad der findes med den alternative SES-specifikation for socioøkonomisk baggrund, så ses i tabel 3.5, at mønsteret også her er, at sammenhængen er svagest for matematik (som kun er lige signifikant på 5%-niveau)¹⁹. Den statistiske sikkerhed er dog gennemgående højere, når man anvender det alternative mål for socioøkonomisk baggrund.

Tabel 3.4: Regressionsresultater – Estimeret (lineær) sammenhæng mellem andel elever med svag social baggrund på skolen og elevernes testscores efter korrektion for egen sociale baggrund

Socioøkonomisk baggrund (Andel med svag baggrund)

			Læsescores			
			Koeff.	Std.fejl	Sign.niveau	R-sq
Alle elever	Uden korrektion	Obs=1667	-234.68	36.10	0.000	0.104
	Med korrektion	Obs=1667	-58.82	24.78	0.021	0.380
Stærk hjemmebaggrund (Ingen belastningsfaktor)	Uden korrektion	Obs=642	-182.12	54.78	0.002	0.044
	Med korrektion	Obs=642	-28.67	33.67	0.398	0.373
Mellemgruppen (1 belastningsfaktor)	Uden korrektion	Obs=606	-230.04	43.52	0.000	0.096
	Med korrektion	Obs=606	-44.23	30.31	0.150	0.404
Svag hjemmebaggrund (2/3 belastningsfaktorer)	Uden korrektion	Obs=398	-161.39	40.31	0.000	0.057
	Med korrektion	Obs=398	-82.42	42.86	0.059	0.364
			Matematikscores			
			Koeff.	Std.fejl	Sign.niveau	R-sq
Alle elever	Uden korrektion	Obs=935	-228.42	34.43	0.000	0.105
	Med korrektion	Obs=935	-32.93	26.87	0.225	0.368
Stærk hjemmebaggrund (Ingen belastningsfaktor)	Uden korrektion	Obs=371	-236.91	58.26	0.000	0.078
	Med korrektion	Obs=371	-61.93	54.35	0.260	0.400
Mellemgruppen (1 belastningsfaktor)	Uden korrektion	Obs=342	-181.42	46.66	0.000	0.067
	Med korrektion	Obs=342	-6.27	41.66	0.880	0.380
Svag hjemmebaggrund (2/3 belastningsfaktorer)	Uden korrektion	Obs=212	-69.95	47.18	0.144	0.013
	Med korrektion	Obs=212	9.39	52.42	0.858	0.349
			Naturvidenskabsscores			
			Koeff.	Std.fejl	Sign.niveau	R-sq
Alle elever	Uden korrektion		-206.64	33.29	0.000	0.078
	Med korrektion		-43.56	24.09	0.076	0.320
Stærk hjemmebaggrund (Ingen belastningsfaktor)	Uden korrektion	Obs=353	-191.41	57.38	0.002	0.049
	Med korrektion	Obs=353	-65.35	40.42	0.110	0.401
Mellemgruppen (1 belastningsfaktor)	Uden korrektion	Obs=341	-184.29	46.47	0.000	0.058
	Med korrektion	Obs=341	-5.68	37.01	0.879	0.358
Svag hjemmebaggrund (2/3 belastningsfaktorer)	Uden korrektion	Obs=220	-127.69	39.4	0.002	0.036
	Med korrektion	Obs=220	12.94	45.86	0.280	0.779

Note: Koefficienter, der er signifikante på 5- hhv. 10%-niveau er angivet med hhv. fed og kursiv skrift.

Tabel 3.5: Regressionsresultater – Estimeret sammenhæng mellem elevernes gennemsnitlige sociale baggrund på skolen og elevernes testscores efter egen social baggrund (sekstiler)

			Læsescores			
			Koeff.	Std.fejl	Sign.niveau	R-sq
Alle elever	Uden controls	Obs=1667	60.59	5.68	0.000	0.190
	Med controls	Obs=1667	23.92	4.69	0.000	0.393
Laveste SES (1. sekstil)	Uden controls	Obs=279	34.90	7.99	0.000	0.088
	Med controls	Obs=279	33.23	8.15	0.000	0.365
2. sekstil	Uden controls	Obs=278	36.72	8.89	0.000	0.072
	Med controls	Obs=278	13.00	10.41	0.217	0.360
3. sekstil	Uden controls	Obs=277	44.30	11.62	0.000	0.093
	Med controls	Obs=277	21.51	11.12	0.058	0.390
4. sekstil	Uden controls	Obs=288	61.43	13.20	0.000	0.123
	Med controls	Obs=288	19.40	9.94	0.056	0.479
5. sekstil	Uden controls	Obs=271	51.85	11.92	0.000	0.090
	Med controls	Obs=271	20.46	12.91	0.119	0.470
Højeste SES (6. sekstil)	Uden controls	Obs=274	31.12	14.85	0.043	0.023
	Med controls	Obs=274	11.79	15.24	0.444	0.299
			Matematikscores			
			Koeff.	Std.fejl	Sign.niveau	R-sq
Alle elever	Uden controls	Obs=935	52.90	5.90	0.000	0.157
	Med controls	Obs=935	11.49	5.68	0.048	0.371
Laveste SES (1. sekstil)	Uden controls	Obs=171	16.95	7.28	0.024	0.028
	Med controls	Obs=171	6.37	8.69	0.467	0.385
2. sekstil	Uden controls	Obs=145	23.71	12.22	0.058	0.028
	Med controls	Obs=145	-3.07	12.39	0.804	0.467
3. sekstil	Uden controls	Obs=151	39.96	13.36	0.004	0.082
	Med controls	Obs=151	17.89	15.75	0.261	0.436
4. sekstil	Uden controls	Obs=161	55.47	14.24	0.000	0.104
	Med controls	Obs=161	11.41	16.62	0.496	0.526
5. sekstil	Uden controls	Obs=155	53.54	13.55	0.000	0.089
	Med controls	Obs=155	23.20	22.41	0.306	0.378
Højeste SES (6. sekstil)	Uden controls	Obs=152	32.84	16.68	0.056	0.034
	Med controls	Obs=152	4.67	20.89	0.820	0.295
			Naturvidenskabscorer			
			Koeff.	Std.fejl	Sign.niveau	R-sq
Alle elever	Uden controls	Obs=925	52.37	5.84	0.000	0.138
	Med controls	Obs=925	18.24	5.10	0.001	0.331
Laveste SES (1. sekstil)	Uden controls	Obs=143	30.58	8.53	0.001	0.070
	Med controls	Obs=143	19.42	12.97	0.141	0.336
2. sekstil	Uden controls	Obs=169	28.88	7.97	0.001	0.052
	Med controls	Obs=169	4.90	8.98	0.587	0.442
3. sekstil	Uden controls	Obs=150	42.83	17.28	0.017	0.066
	Med controls	Obs=150	18.77	16.72	0.267	0.537
4. sekstil	Uden controls	Obs=156	69.88	12.41	0.000	0.158
	Med controls	Obs=156	30.19	14.71	0.046	0.494
5. sekstil	Uden controls	Obs=156	41.12	14.99	0.010	0.054
	Med controls	Obs=156	5.71	20.79	0.785	0.409
Højeste SES (6. sekstil)	Uden controls	Obs=151	18.28	14.04	0.201	0.007
	Med controls	Obs=151	10.79	13.64	0.434	0.383

Note: Koefficienter, der er signifikante på 5- hhv. 10%-niveau er angivet med hhv. fed og kursiv skrift.

Også her vil man være interesseret i at se på, om elevsammensætningen har mere betydning for nogle elevgrupper end for andre. Det vanskeliggøres dog til en vis grad af, at mange resultater bliver insignifikante, når man deler eleverne op i mindre grupper. Derfor er der tillige lavet analyser med alle elever, hvor man ser på krydseffekter (med interaktionsled mellem elevsammensætning og elevens egen baggrund) i stedet for at dele datasættet op. Resultaterne fra regressionerne for de enkelte elevgrupper er vist i tabellerne 3.4 og 3.5. Resultaterne fra analyserne med krydseffekter er vist i tabel A.3.5.

Som nævnt bliver resultaterne for de enkelte elevgrupperne (opdelt efter egen social baggrund) kun meget sjældent statistisk signifikant, da usikkerheden i analyserne er stor (dvs. store standardfejl på estimerne). Det eneste (marginalt) signifikante resultat, der opnås fra regressionerne på de enkelte elevgrupper i tabel 3.4 er, at der for elever med svag social baggrund er en sammenhæng mellem andel elever med svag baggrund på skolen og elevernes læsetestscores: en højere andel af svage elever på skolen samvarierer for disse elever med lavere læsetestscores. Dette resultat bekræftes af analyserne på den alternative baggrundsspecifikation (tabel 3.5): For de ca. 15% svageste elever (laveste/1. sekstil) er der en positiv sammenhæng mellem gennemsnits-SES for eleverne på skolen og deres egen læsetestscores. For hverken matematik eller naturvidenskab er der dog signifikante delgrupperresultater (undtagen for naturfag i 4. kvartil) heller ikke i den alternative specifikation. Dette til trods for, at der (som nævnt) er en signifikant sammenhæng for alle elever under ét. Der er også lavet analyser på hele samplet med interaktionsled for at undgå problemet med de få observationer i regressionen. Resultaterne er dog ikke systematisk forskellige, idet de fleste interaktionsled er insignifikante (jf. tabel A3.5 i bilag).

Alt i alt har det vist sig vanskeligt at finde ud af, om sammenhængen mellem andel svage elever på skolen og elevernes testcores er forskellig for forskellige elevgrupper. Det skyldes sandsynligvis, at der er ret få observationer, når der deles op, kombineret med, at effekten af andel svage elever på skolen er forholdsvis svag/usikker, når der kontrolleres for egen sociale baggrund og tosprogethed. For alle elever under ét gælder dog, at der er en sikker sammenhæng mellem elevsammensætningen og læsetestcores og testcores i naturvidenskab, mens det er lidt mere tvivlsomt, om der er en sammenhæng med matematikscores.

Resultaterne fra regressionsanalyserne viser, at en højere andel tosprogede eller en elevsammensætning med flere elever fra svage hjem samvarierer med lavere testcores hos eleverne. Men skyldes det, at eleverne er tosprogede eller blot, at de tosprogede elever i gennemsnit har en svagere social baggrund? Det søges belyst i næste afsnit.

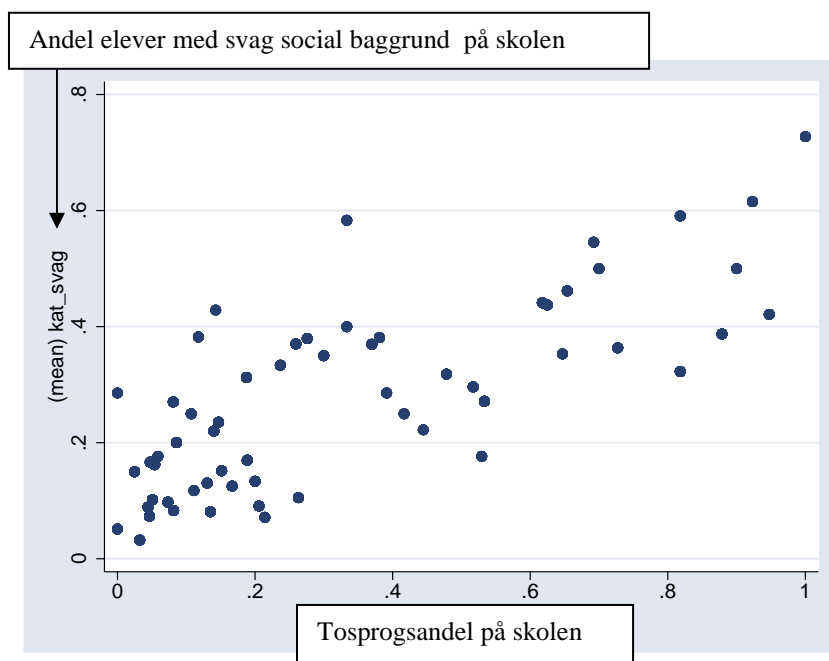
3.3 Tosprogethed eller socioøkonomisk baggrund?

Det sidste forhold, der ses på ved hjælp af regressionsanalyse er, om det er elevsammensætningen efter tosprogethed i sig selv eller sammensætningen efter social baggrund, der betyder noget for læsefærdigheder.

Figur 3.8 viser sammenhængen mellem tosprogsandelen og andel elever med svag social baggrund på hver enkelt folkeskole i PISA-København-datamaterialet. Hvis alle

punkter lå på en lige linje, så var der en 100% sammenhæng mellem tosprogsandelen og andelen af elever med svag social baggrund. Det ses af figuren, at selv om der ikke er en 100% sammenhæng, så er de to mål dog stærkt samvarierende (med en korrelationskoefficient på 77%). Man kunne derfor være bekymret for, at man ikke med statistiske metoder ville kunne afgøre, om en af de to forhold, er en mere afgørende faktor end den anden.

Figur 3.8: Samvariation mellem andel elever med svag social baggrund og tosprogsandel på skolen



Resultaterne fra en regressionsanalyse, der medtager begge mål for elevsammensætningen på én gang (dvs. både tosprogsandelen og andel elever med svag social baggrund på skolen), viser klare tegn på, at der er betydelige problemer med at skelne mellem de to påvirkninger (resultaterne ikke vist her). Man kan således ikke sige, om den ene eller den anden faktor er stærkest, fordi de to forhold (en høj tosprogsandel og en stor andel elever med svag social baggrund på skolen) som regel forekommer sammen.

3.4 Diskussion

De deskriptive analyser har vist, at der er en signifikant negativ sammenhæng mellem andel tosprogede elever på skolen og både de dansk- og tosprogede elevers testscores i læsning, matematik og naturvidenskab. Der er dog ikke forskel i testscores for elever, hvad enten de går i skoler med mange tosprogede (omtrent 50-85%) eller i såkaldte højkoncentrationsskoler (mere end 85% tosprogede). Når man i regressionsanalyser tager højde for forskelle i elevernes baggrund, modificeres resultaterne i retning af, at der op

til en tosprogsandel på omkring 50% på skolen ikke er forskelle i elevernes testscores, dvs. at der ikke er (en statistisk signifikant) forskel på gennemsnitlige testscores for elever, der går i skoler med 10% tosprogede og for elever, der går i en skole med 30% tosprogede. For lave tosprogsprocentager på skolerne synes der efter hensyntagen til social baggrund således ingen sammenhæng at være mellem andelen tosprogede elever på skolen og elevernes testresultater. Det samme gælder for høje tosprogsprocenter. Men når man sammenligner elever på skoler med lav tosprogsandel med elever på skoler med mange tosprogede, så scorerer både dansk- og tosprogede elever højere på skoler med lav tosprogsprocent.

Resultaterne for elevsammensætning, når den måles med social baggrund, ligner meget mønsteret ovenfor, men resultaterne er gennemgående lidt mindre sikre (statistisk signifikante). Der kan ikke skelnes empirisk mellem, hvilken af de to faktorer (tosprogethed eller social baggrund), der er den afgørende i sammenhængen mellem skolens elevsammensætning og testscores.

Sammenhængen mellem elevsammensætning og elevernes testscores er lige kraftig for dansk- og tosprogede elever. Sammenhængen mellem elevsammensætning og elevernes testscores er stærkest og mest sikker for læsefærdighederne og lidt mindre sikker og kraftig for naturvidenskab og (især) matematik.

Resultaterne i nærværende undersøgelse afviger fra de i Egelund (2005b) rapporterede resultater. Der har vist sig at være opstået en fejl i databehandlingen i den tidligere rapport, hvorved resultatet i Egelund (2005b) er modsat det, som dette notat viser.

Nyere amerikansk forskning støtter disse resultater. Angrist & Lang (2004) har set på effekterne af et desegregeringsprojekt (Metropolitan Council for Educational Opportunity (Metco) i Boston), hvor elever fra ressourcetsvage hjem fra den indre by (hovedsageligt »black Americans«) bliver kørt ud til skoler i mere velstående forstæder. Dette medfører, at procenten af minoritetselever i modtageskolerne er steget fra 7,5% til 12,5% i gennemsnit. Analysen ser udelukkende på effekten på de elever, der var i modtageskolerne i forvejen, men kan ikke identificere effekten på de elever, der bliver kørt ud fra indre bydistrikter. Analysens resultater tyder ikke på, at denne stigning i andelen af minoritetselever har en (negativ) effekt på de (hvide) elevers testscores i modtageskolerne²⁰. Forholdene omkring segregering i USA er dog noget forskellige fra danske forhold, hvorfor resultaterne ikke umiddelbart kan overføres.

Det er vigtigt at bemærke, at undersøgelserne i notatet ikke afdækker årsagssammenhæng mellem elevsammensætningen og elevernes testscores, men udelukkende leverer en overordnet analyse af statistiske sammenhænge. Det skyldes først og fremmest, at der på trods af ret gode muligheder for at tage højde for forskelle i elevernes sociale baggrund, vil være usikkerhed, om øvrige ikke-observerede faktorer spiller en rolle, hvilket kan »farve« resultaterne. Resultaterne kan derfor ikke anvendes til at afgøre, hvorvidt elevsammensætningen i skolen har en *effekt* på elevernes færdigheder (jf. også boks 3).

For en effektvurdering af elevsammensætningen stilles der særligt store krav til det tallemateriale som undersøgelsen baseres på:

- En anden type datagrundlag/metoder end der har været til rådighed for dette notat, for eksempel (i) eksperimenter (hvor man fx fordeler eleverne tilfældigt på skoler med forskellig tosprogsprocent), eller (ii) paneldata, hvor elevernes testes gentagne gange (fx i starten og slutningen af skoleåret), således at man kan se på den enkelte elevs udvikling over en bestemt tidshorisont.
- Talmaterialet bør indeholde oplysninger om en række forhold vedrørende læringsmiljøet, som forekommer relevante i forbindelse med en vurdering af effekten af elevsammensætning: fx lærernes uddannelsesbaggrund i forhold til at kunne håndtere etnisk kompleksitet fagligt, især ud fra, hvor nyt/gammelt dette forhold er i det pågældende lokalsamfund og ud fra skolens fag- og undervisningsplaner. Et andet forhold, der kan gøre sig gældende, er på hvilken måde den sproglige og etniske heterogenitet indgår i den lokale sammenhæng. Er der tale om en jævn udvikling eller tilstræbes en heterogenitet hurtigt som led i en bredere social integrationspolitik, fx som et middel mod ghettoisering og segregation?

Boks 3: Rå korrelationer, rensede korrelationer og effekter/årsagssammenhænge

Rå korrelationer er simple sammenhænge mellem to størrelser, fx tosprogsandelen i skolen og elevens testscores. Rå korrelationer afspejler ikke blot, i hvilket omfang der måtte være en direkte sammenhæng mellem tosprogsandel og testscores, men kan også afspejle indirekte sammenhænge, fx at elever, der går i skoler med højere tosprogsandel, typisk kommer fra mindre ressourcerstærke hjem, hvilket i sig selv kan have en negativ indflydelse på disse elevers testscores.

Ved såkaldte *rensedede korrelationer* er der ved regressionsanalyse taget højde for de forskelle i elevernes karakteristika, som forskeren har data for. De rensede korrelationer kan derfor siges at være fri af forskelle i målte elevkarakteristika.

Effekter/årsagssammenhænge. Der vil dog altid være karakteristika, der ikke er data for – og som kan skævvride resultaterne, da man ikke eksplicit kan tage højde for disse forhold. For på troværdig vis at kunne identificere årsagssammenhæng, stilles der særligt store krav til talmaterialet, og der kræves en anden type datagrundlag, end der har været til rådighed for dette notat. Eksempler på sådanne datasæt/metoder er (i) eksperimenter (hvor man fx fordeler eleverne tilfældigt på skoler med forskellig tosprogsprocent), eller (ii) paneldata, hvor elevernes testes gentagne gange (fx i starten og slutningen af skoleåret), således at man kan se på den enkelte elevs udvikling over en bestemt tidshorisont.

Litteratur:

Angrist, J. D. & K. Lang (2004): Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program. The American Economic Review. Vol. 94, No. 5, December 2004.

Arbejderbevægelsens Erhvervsråd (2003): Det delte Danmark. Notat om social opdeling af skoler. Kan downloades fra: http://www.ugebreveta4.dk/smmedia/niels-nov-03.pdf?mb_GUID=45514A79-6F69-41C6-A3C9-27E639735CBF.pdf

Egelund, N. (2005a): OECD-programmet PISA. Kap. 2 i: PISA-København. Akf forlaget.

Egelund, N. (2005b): Sammenhænge med skoledemografiske forhold. Kap. 5 i: PISA-København. Akf forlaget.

Rangvid, B.S. (2003): Educational Peer Effects: Quantile Regression Evidence from Denmark with PISA2000 Data. Working Paper No. 03-2, Handelshøjskolen Århus

Rangvid, B.S. (2005a): Datamaterialet i PISA-København. Kap. 3 i: PISA-København. Akf forlaget.

Rangvid, B.S. (2005b): Sources of Immigrants' Underachievement – Results from PISA-Copenhagen. akf working paper 9:2005.

Skolverket (2003): Läsförståelse hos elever med utländsk bakgrund. Skolverket.

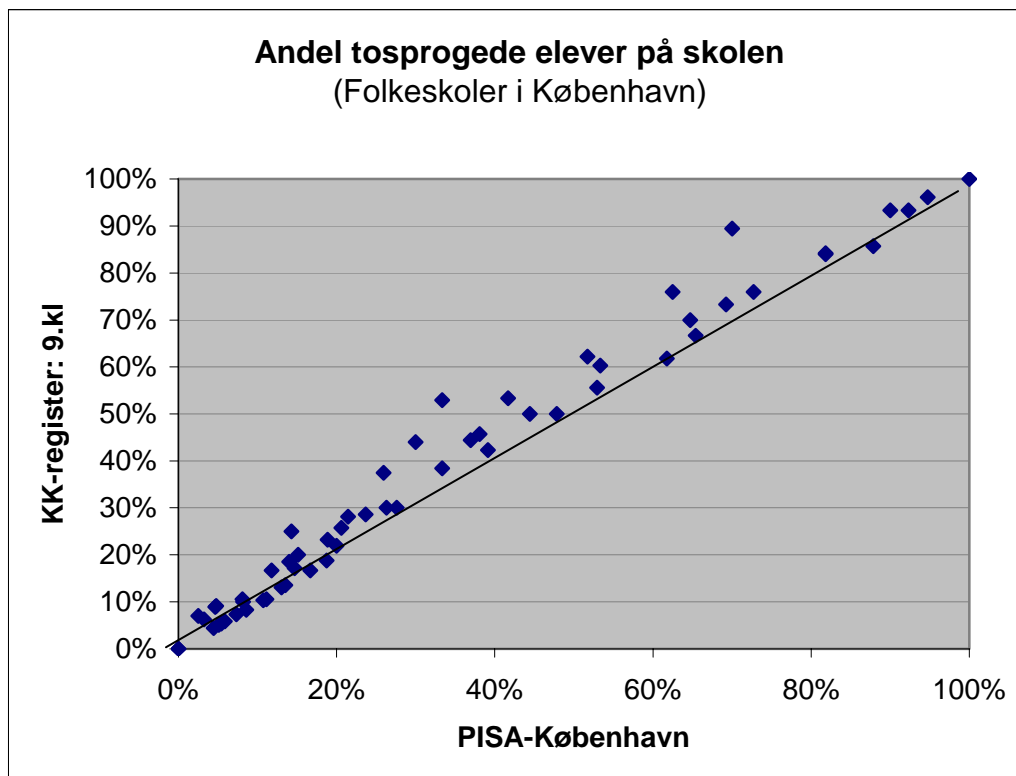
Bilag A

Validering af datagrundlag for beregning af andel tosprogede

I PISA-København er der ikke oplysninger for alle 9.-klasses-elever, men kun for de elever, der rent faktisk har deltaget i PISA-testen. Som vist i Rangvid (2005b), er der på nogle skoler et stort frafald ved PISA-testen, det vil sige at der er mange 9.-klasses-elever, vi mangler at have med i PISA-datasættet. Hvis frafaldet er skævt, dvs. hvis det er relativt flere dansksprogede eller tosprogede elever, der ikke har deltaget i testen, så vil den med PISA-data beregnede andel tosprogede ikke være lig den *faktiske* andel tosprogede, som er den, der er relevant for denne analyse. Fra Københavns Kommunes elevregister haves dog informationer om alle elever fra efteråret 2003 for folkeskolerne. Disse data kan derfor bruges til at se på, om anvendelsen af PISA-data til at beregne andelen af tosprogede elever på klassetrinnet er acceptabelt.

I figur B.1 er andelen af tosprogede på skolens 9. klassetrin beregnet med (i) PISA-København-data, og med (ii) data fra elevregistret, og så indtegnet mod hinanden. Hvis målene var fuldstændig identiske, ville alle punkter ligge på den viste skrå linje. Det ses, at selv om alle punkter ikke ligger præcist på linjen, så forekommer sammenfaldet af de to mål rimeligt pænt²¹. Det skønnes derfor, at frafaldet i PISA-København ikke bør give anledning til bekymring i forhold til beregningen af den faktiske andel tosprogede i klassen, og i den videre analyse anvendes derfor målet for andel tosprogede elever beregnet med PISA-data.

Figur B.1: Validering af mål for elevsammensætning fra PISA med registerdata



Bilag B

Klassesammensætning og testscores

Afsnit 3.1 er baseret på elevsammensætningen blandt alle 9.-klasses-elever i PISA-København. Via tilknytningen af PISA-Københavns data med Københavns Kommunes elevregister har vi dog yderligere information om klassekoden for hver enkelt elev. Det er derfor muligt også at se på elevsammensætningen på *klasseniveau*. For at give et indtryk af, hvor stor forskel der er på klassesammensætningen i skolernes 9. klassetrin, viser tabel A.3.2 (vedlagt som bilag) for hver enkelt skole den (samlede) gennemsnitlige sammensætning på 9. klassetrin (som analysen i afsnit 3.1 er baseret på), samt den gennemsnitlige elevsammensætning i skolens enkelte 9.-klasser. Som det ses, er der på enkelte skoler stor forskel i andelen af tosprogede elever i 9.-klasserne. Det er derfor relevant at revurdere hovedresultaterne fra afsnittet ovenfor med klassesdata. Det ses dog også, at der i 20 ud af 59 skoler kun er én 9. klasse, hvorfor variabelen for elevsammensætning på disse skoler vil være ens på klasse- og klassetrinsniveau.

Figur B.2 viser de samme analyser som figur 3.2, men nu også for tilsvarende analyser for klassesammensætningen. Som figur B.2 viser, synes det overordnet ikke at have den store betydning, om man bruger skole- eller klassevariabelen som mål for elevsammensætningen: på trods af nogle afvigelser – som hovedsageligt forekommer i kategorier, hvor der er få observationer – er mønsteret stort set det samme.

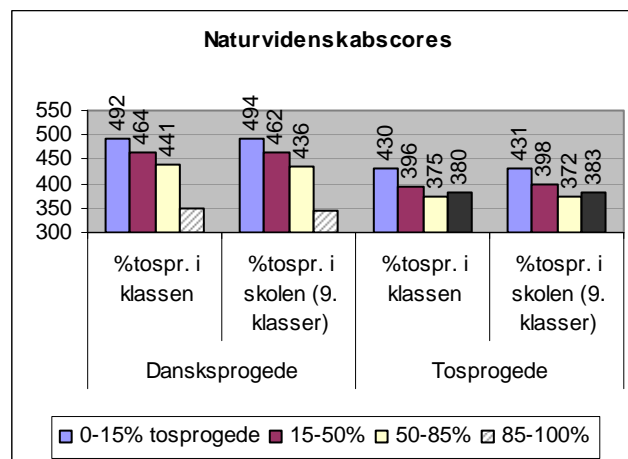
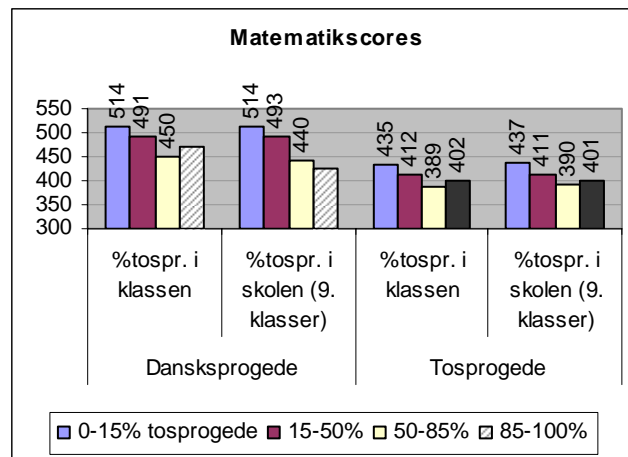
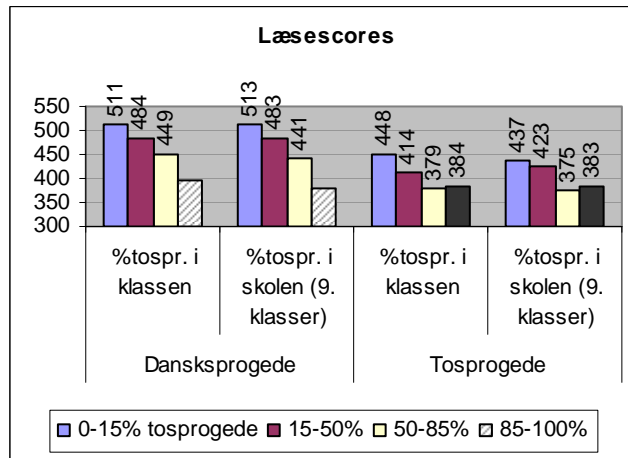
I tabel B.1 er resultaterne fra tabel 3.2 gentaget og tilføjet resultater for elevsammensætningen målt på klasseniveau. Der er to sæt resultater på klasseniveau: først med alle elever, og dernæst udelades elever i klasser med kun 10 eller færre elever. Herved sikres en vis pålidelighed for målet af tosprogsprocenten i klassen.

Resultaterne fra regressionsanalyser for klassesammensætningen viser, at:

- elevsammensætningen, både på skole- og klasseniveau, samvarierer (statistisk signifikant på mindst 10%-niveau) med både dansk- og tosprogede elevers test scores på alle tre testområder. Resultatet for matematik, som kun er statistisk signifikant på 10%-niveau for tosprogede elever, når andelen af tosprogede elever måles på skoleniveau, bliver mere sikkert bestemt (på 5%-niveau), når der måles på klasseniveau (og når man begrænser analysen til klasser med flere end 10 elever). Til gengæld går resultatet for dansksprogede elever i matematik fra signifikant til kun marginalt signifikant (på 10%-niveau), og signifikansniveauet for resultatet for tosprogede i naturvidenskab ligger nu lige over 5%-niveauet. Men på trods af disse små forskydninger i signifikansniveauerne, er alle estimater dog signifikante på mindst 10%-niveau, hvad enten de estimeres på skole- eller klasseniveau.
- resultaterne fra en pooled regression (dvs. med både et- og tosprogede elever) med en indikatorvariabel for tosprogethed viser, at der heller ikke er en statistisk sikker forskel mellem koefficienterne til andelen af tosprogede elever, når elevsammensætningen måles på klasseniveau.

Samlet set vil det sige, at uanset, om elevsammensætningen måles på klasse- eller på elevniveau, er konklusionen grundlæggende den samme: at en højere tosprogsprocent er korreleret med lavere test scores i alle tre fag for både danske og tosprogede elever.

Figur B.2: Gennemsnitlige testscores beregnet for tosprogsprocent på klasse- og skole-niveau (9.-klasser)



Tabel B.1: Regressionsresultater for elevsammensætningen målt på (i) skoleniveau og (ii) klasseniveau

Klasseniveau

Tosprogsprocent (Andel tosprogede elever)		Læsescores			
		Koeff.	Std.fejl	Sign.niveau	R-sq
<i>På skoleniveau (9. klasser)</i>					
Dansksprogede elever	Obs=1191	-59.44	16.18	0.001	0.330
Tosprogede elever	Obs=476	-76.05	20.61	0.001	0.210
<i>På klasseniveau</i>					
Dansksprogede elever	Obs=1191	-48.15	14.39	0.001	0.326
Tosprogede elever	Obs=476	-69.70	20.09	0.001	0.203
<i>På klasseniveau (flere end 10 elever i klassen)</i>					
Dansksprogede elever	Obs=1138	-46.91	14.91	0.003	0.332
Tosprogede elever	Obs=448	-75.82	21.80	0.001	0.212
		Matematikscores			
		Koeff.	Std.fejl	Sign.niveau	R-sq
<i>På skoleniveau (9. klasser)</i>					
Dansksprogede elever	Obs=656	-41.84	18.41	0.027	0.294
Tosprogede elever	Obs=279	-34.34	20.35	0.097	0.21
<i>På klasseniveau</i>					
Dansksprogede elever	Obs=656	-32.44	17	0.062	0.29
Tosprogede elever	Obs=279	-32.32	20.27	0.12	0.21
<i>På klasseniveau (flere end 10 elever i klassen)</i>					
Dansksprogede elever	Obs=626	-32.74	17.49	0.067	0.298
Tosprogede elever	Obs=262	-42.59	19.99	0.038	0.23
		Naturvidenskabscorer			
		Koeff.	Std.fejl	Sign.niveau	R-sq
<i>På skoleniveau (9. klasser)</i>					
Dansksprogede elever	Obs=668	-54.02	19.63	0.008	0.287
Tosprogede elever	Obs=257	-52.12	25.25	0.044	0.298
<i>På klasseniveau</i>					
Dansksprogede elever	Obs=668	-46.16	18.82	0.017	0.285
Tosprogede elever	Obs=257	-46.96	25.78	0.074	0.293
<i>På klasseniveau (flere end 10 elever i klassen)</i>					
Dansksprogede elever	Obs=641	-41.54	19.03	0.034	0.288
Tosprogede elever	Obs=241	-50.3	26.76	0.066	0.28

Note: Koefficienter, der er signifikante på 5- hhv. 10%-niveau er angivet med hhv. fed og kursiv skrift.

Tabel A3.1: Gennemsnitlige testcores for dansk- og tosprogede elever efter andel tosprogede på skolen

			Kat 1	Kat 2	Kat 3	Kat 4
			0-15%	15-50%	50-85%	85-100%
Dansksproget	Læsning	Gns.	513	483	441	381
		% uden fkt. læsekompetence	0.112	0.208	0.355	0.714
		Antal	696	384	102	9
	Matematik	Gns.	514	493	440	424
		% under "niveau 2"	0.111	0.171	0.390	0.000
		Antal	395	205	55	1
	Naturvidenskab	Gns.	494	462	436	345
		% under "niveau 2"	0.181	0.300	0.441	0.833
		Antal	381	214	65	8
Tosproget(NW)	Læsning	Gns.	437	423	375	383
		% uden fkt. læsekompetence	0.404	0.440	0.648	0.633
		Antal	57	141	185	93
	Matematik	Gns.	437	411	390	401
		% under "niveau 2"	0.364	0.425	0.575	0.630
		Antal	33	87	106	53
	Naturvidenskab	Gns.	431	398	372	383
		% under "niveau 2"	0.387	0.507	0.630	0.558
		Antal	31	75	99	52

Tabel A3.2: Andel tosprogede elever på 9. klassetrin, samt i de enkelte 9. klasser

Elevsammensætning (andel tosprogede fra ikke-vestlige lande)

Skole	Alle 9. classes PISA-elever	Enkelte 9. klasser	Antal klasser på 9. trin	Skole	Alle 9. classes PISA-elever	Enkelte 9. klasser	Antal klasser på 9. trin
1	5%	0-10%	3	31	53%	53%	1
2	13%	13%	1	32	5%	0-15%	2
3	37%	27-50%	3	33	19%	13-29%	3
4	8%	5-11%	2	34	63%	63%	1
5	11%	0-21%	2	35	20%	10-25%	2
6	55%	33-67%	4	36	82%	82%	1
7	24%	(24-) 24%	2	37	95%	95%	1
8	44%	44%	1	38	6%	6%	1
9	30%	13-42%	2	39	11%	11%	1
10	26%	19-36%	2	40	4%	0-10%	2
11	14%	14%	1	41	21%	15-27%	2
12	69%	69%	1	42	88%	(88-) 88%	2
13	28%	20-36%	2	43	26%	26%	1
14	3%	0-7%	3	44	9%	0-17%	2
15	3%	0-7%	2	45	73%	73%	2
16	14%	6-22%	3	46	19%	18-20%	2
17	15%	10-21%	2	47	65%	50-83%	2
18	12%	11-13%	2	48	0%	0%	2
19	52%	47-57%	2	49	8%	0-17%	2
20	39%	38-40%	2	50	15%	14-16%	2
21	90%	90%	1	51	5%	0-10%	2
22	21%	11-31%	2	52	5%	0-11%	2
23	14%	5-22%	2	53	0%	0%	1
24	17%	7-30%	2	54	92%	92%	1
25	33%	33%	1	55	82%	76-88%	2
26	48%	45-50%	2	56	65%	65%	1
27	42%	42%	1	57	88%	88%	1
28	100%	100%	1	58	62%	47-76%	2
29	33%	33%	1	59	7%	5-10%	2
30	38%	33-44%	2				

Tabel A3.3: Gennemsnitlige læsescores for elever med stærk, mellem og svag social baggrund efter, hvor belastet skolerne er med hensyn til andelen af elever med svag hjemmebaggrund (kvartiler, hvor 1. kvartil er de 25% af skoler med færrest svage elever)

Læsning	Elevsammensætning (SES) på skolen			
	<i>1. kvartil</i>	<i>2. kvartil</i>	<i>3. kvartil</i>	<i>4. kvartil</i>
Stærk social baggrund	518	494	476	415
<i>Antal elever</i>	283	216	116	28
Mellemgruppen	499	472	439	399
<i>Antal elever</i>	186	184	141	96
Svag social baggrund	485	441	442	344
<i>Antal elever</i>	46	103	135	118
Matematik				
Stærk social baggrund	532	507	475	428
<i>Antal elever</i>	167	127	63	14
Mellemgruppen	495	475	445	414
<i>Antal elever</i>	92	108	87	56
Svag social baggrund	458	421	435	405
<i>Antal elever</i>	27	54	73	61
Naturvidenskab				
Stærk social baggrund	494	478	449	395
<i>Antal elever</i>	150	118	66	20
Mellemgruppen	482	467	437	392
<i>Antal elever</i>	108	101	79	53
Svag social baggrund	455	423	421	395
<i>Antal elever</i>	26	54	74	66

Tabel A3.4: Gennemsnitlige testscores for elever med høj, mellemhøj og lav familie-SES efter elevsammensætningen i skolen (resultater for definition af hjemmebaggrund ved hjælp af PCA-analyse)

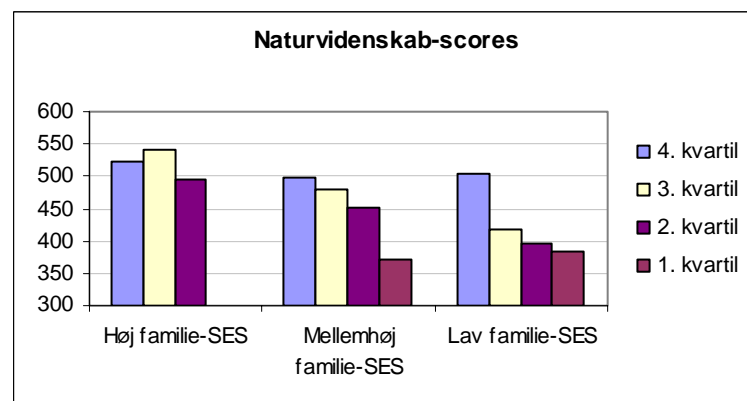
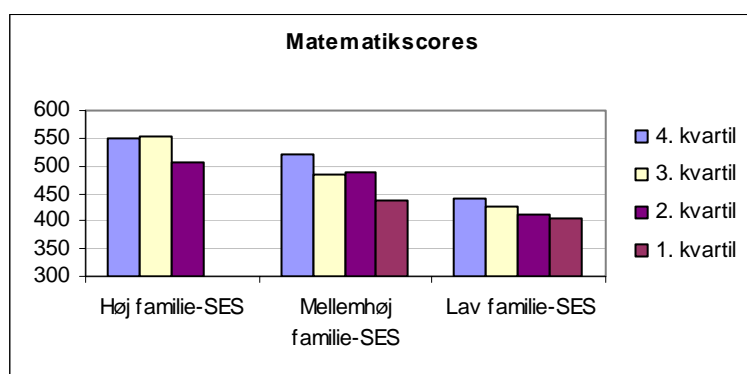
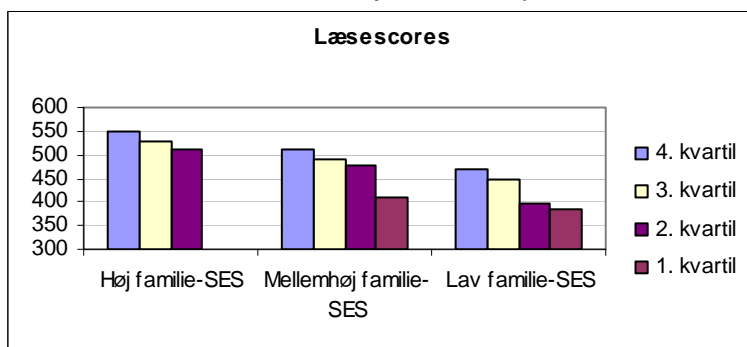
		Lav skole-SES ----- Høj skole-SES			
Læsning		1. kvartil	2. kvartil	3. kvartil	4. kvartil
Lav familie-SES	Gns.	383	396	447	468
(15% med laveste SES)	Antal	114	70	40	18
Mellemhøj familie-SES	Gns.	408	477	492	510
(30% med mellemhøj SES)	Antal	37	109	138	156
Høj familie-SES	Gns.	0	512	530	550
(15% med højeste SES)	Antal	0	35	42	109
Matematik		1. kvartil	2. kvartil	3. kvartil	4. kvartil
Lav familie-SES	Gns.	406	412	425	440
(15% med laveste SES)	Antal	68	40	26	11
Mellemhøj familie-SES	Gns.	438	487	485	520
(30% med mellemhøj SES)	Antal	22	53	81	85
Høj familie-SES	Gns.	.	507	552	548
(15% med højeste SES)	Antal	0	20	26	59
Naturvidenskab		1. kvartil	2. kvartil	3. kvartil	4. kvartil
Lav familie-SES	Gns.	383	395	417	504
(15% med laveste SES)	Antal	64	36	21	11
Mellemhøj familie-SES	Gns.	371	451	479	497
(30% med mellemhøj SES)	Antal	20	67	68	92
Høj familie-SES	Gns.	.	496	541	524
(15% med højeste SES)	Antal	0	20	23	58

Tabel A3.5: Resultater fra regressioner med krydseffekter for sammenhængen mellem elevernes gennemsnitlige sociale baggrund på skolen og elevernes testcores efter egen sociale baggrund

SEES-mål	Obs	Elevsammensætning			Interaktionsled med kat1			Interaktionsled med kat2			R-sq
		Koeff.	Std.fejl	Sign.niveau	Koeff.	Std.fejl	Sign.niveau	Koeff.	Std.fejl	Sign.niveau	
Andel elever med svag social baggrund	1667	-62.56	28.45	0.032	-8.09	23.50	0.732	21.15	34.81	0.55	0.375
	935	-50.27	39.44	0.208	6.79	35.51	0.849	53.32	48.75	0.28	0.365
	914	-76.77	34.15	0.028	46.82	33.49	0.168	49.79	42.80	0.25	0.320
Gennemsnits-SES på skolen	1667	25.31	4.70	0.000	2.20	2.12	0.305				0.39
	935	15.65	5.86	0.010	6.19	2.470	0.015				0.38
	914	20.25	5.48	0.000	3.52	3.020	0.248				0.33

Note: Når kat1=1 betyder det, at eleven har én (social) belastningsfaktor. Når kat2=1, har eleven 2 eller flere belastningsfaktorer.

Figur A3.1: Gennemsnitlige testscores for elever med høj, mellemhøj og lav familie-SES efter elevsammensætningen i skolen (resultater for definition af hjemmebaggrund ved hjælp af PCA-analyse)



Note: Der er ikke nogen søjler for 1. kvartil for elever med høj familie-SES, fordi der ikke er elever i denne kategori.

Noter

- 1 Alle 59 folkeskoler med 9. klassetrin deltog i analysen. Desuden valgte 24 ud af 39 privatskoler med 9. klassetrin at deltage i den københavnske PISA-undersøgelse. Analyserne i dette notat er dog udelukkende baseret på folkeskoleelever. På grund af den ringe datakvalitet for privatskoleeleverne ville der af datamæssige årsager være problemer med at udvide analysen til også at omfatte privatskoler. Det skyldes både, at Danmarks Statistik ikke har kunnet knytte cpr-numre til en større gruppe af privatskoleelever fra PISA-København, og derfor kan der hellere ikke tilkobles oplysninger fra Københavns Kommunes elevregister. Desuden er der i elevregistret kun oplysninger for privatskoleelever, der bor i Københavns Kommune. Det vil sige, at den relevante definition af tosprogede ikke kan dannes for en stor del af privatskoleeleverne. Et yderligere problem er, at ikke alle privatskoler med 9.-klasser har deltaget i PISA-K. Deltagelse var frivillig for privatskolerne, og derfor er det ikke sikkert, at man har et repræsentativt udsnit af privatskolerne med i undersøgelsen, hvilket vil kunne skævvride resultaterne.
- 2 Til analysen er der tillige anvendt information fra Københavns Kommunes administrative elevregister, hvorfra der haves oplysninger, om barnet er tosproget, som begrebet anvendes i Ungdoms- og Uddannelsesforvaltningen. Det har været rekvirentens ønske at kunne koble denne oplysning på PISA-København-data, således at definitionen af tosprogede elever kan baseres herpå. Det viste sig imidlertid, at der i datagrundlaget fra PISA-København-undersøgelsen ikke findes cpr-numre, men blot elevens navn og fødselsdag. Danmarks Statistik har ved hjælp af disse oplysninger kunnet tilknytte cpr-numrene for 96% af folkeskoleeleverne i PISA-København, mens de øvrige 4% ikke har kunnet indgå i denne analyse. Ud af de 2352 elever i PISA-datasættet, er der i alt 1746 folkeskoleelever (74%). 79 elever må udgå af analyserne på grund af manglende oplysninger i Københavns Kommunes Elevadministrationsregister og/eller cpr-match, hvilket som udgangspunkt giver en samlestørrelse på i alt 1667 dansk- og tosprogede elever.
- 3 Oprindelse i et tredjeland defineres her som enten (i) at elevens oprindelsesland er et tredjeland, eller (ii) mindst én af forældrene er født i et tredjeland.
- 4 Ud af de 1667 folkeskoleelever i datasættet er 523 (31%) tosproget. Af disse kommer 476 (28%) fra tredjeverdenslande.
- 5 Som det ses, er trækket fra det administrative elevregister fra efteråret 2003, mens PISA-testen blev foretaget i foråret 2004. Det vil sige, at hvis nogle elever har flyttet skole i den periode, vil data fra registret ikke være helt dækkende for elevbestanden på skolerne i forår 2004.
- 6 Som nævnt før er tallene for dansksprogede børn i skoler med en tosprogsprocent over 85% upålidelige grundet meget få danske elever i datasættet i disse skoler.
- 7 Manglende værdier for enkelte variabler håndteres ved, at der er inkluderet indikatorvariabler for hver baggrundskarakteristika. Standardfejlene er korrigeret for within-school clustering af eleverne.
- 8 Ifølge Danmarks Statistiks definition af hhv. danskere, indvandrere og efterkommere.
- 9 Kulturel kommunikation dækker over svar på spørgsmålene: Hvor tit sker det, at dine forældre: (i) diskuterer politiske eller sociale emner med dig, (ii) diskuterer bøger, film eller fjernsynsprogrammer med dig, (iii) lytter til klassisk musik sammen med dig?
- 10 Social kommunikation dækker over svar på spørgsmålene: Hvor tit sker det, at dine forældre: (i) diskuterer, hvordan det går dig i skolen, (ii) sidder og spiser et hovedmåltid sammen med dig, (iii) bruger tid til at tale med dig?
- 11 Uddannelsesressourcer i hjemmet dækker over spørgsmål, om eleverne har følgende i hjemmet: en ordbog, et stille sted til at læse/studere, et skrivebord til at læse/studere ved, lærebøger og lommeregnerne.
- 12 Kulturelle besiddelser dækker over, om eleven har følgende i hjemmet: klassisk litteratur, digtsamlinger og kunstværker.
- 13 Dette er testet i en pooled estimation for dansk- og tosprogede elever med interaktionsled mellem egen sprogbaggrund og andel tosprogede i skolen.

-
- 14 Igen: resultaterne for den 4. kategori er estimeret med stor usikkerhed for dansksprogede elever, men heller ikke for dansksprogede elever er der statistisk signifikant forskel mellem 3. og 4. kategori.
 - 15 Desuden vil der med den type data, som vi har for PISA-København (nemlig, hvor der er samlet hele klasser i skolerne – i princippet (og i modsætning til de internationale PISA-data) – være mulighed for at lave en såkaldt »school fixed-effects«-estimation, som ville mindske skævheden i estimerne mht. uobserverede faktorer, som påvirker forældrenes skolevalg, og med hensyn til skolekarakteristika. I praksis er datasættet dog for lille (der er ikke variation nok i classesammensætningen på skolen i datagrundlaget) for at få pålidelige/brugbare resultater. I stedet for har jeg prøvet at inkludere et begrænset sæt af skolekarakteristika (navnlig de karakteristika, som i Rangvid (2005) er blevet vist at være ulige fordelt mellem skoler, som danske hhv. tosprogede elever er indskrevet i). Resultaterne fra denne analyse viser, at estimerne for elevsammensætning er stort set uændrede.
 - 16 De éndimensionale socioøkonomiske variabler er alene brugt til inddeling i elevgrupper, ikke som forklarende variabler i regressionsanalyser. I regressionsanalyserne er inkluderet det fulde sæt af baggrundsfaktorer. Jf. afsnit 3.1.3 for en oversigt over de inkluderede baggrundsfaktorer.
 - 17 I forhold til en samlet andel tosprogede folkeskoleeleverne på 28%.
 - 18 Der er brugt det samme sæt af baggrundsvariabler som i afsnit 3.1.4, dog er der her også inkluderet en indikator for tosprogethed, som ikke var nødvendig i afsnit 3.1.4, hvor dansk- og tosprogede blev analyseret hver for sig.
 - 19 Bemærk, at fortegnene her er omvendt sammenlignet med målet »andel svage elever«: dvs. at et positivt fortegn betyder, at en højere gennemsnitlig socioøkonomisk baggrund samvarierer med højere testcores.
 - 20 Der findes svage tegn på en negativ effekt på minoritets elever (der var i modtageskolerne i forvejen), men Angrist & Lang vurderer, at disse effekter er små og kortvarige.
 - 21 Andelen af tosprogede på 9. klassetrin undervurderes dog for en del skoler noget ved brug af PISA-København-data, hvilket kunne tyde på, at ikke-deltagelse i PISA-testen er hyppigere blandt tosprogede end dansksprogede elever.