

Balder Bryn Morsund

Analyse av effekten av å ha gått i barnehage på leseferdigheter i fjerdeklasse i Norge, Sverige, Nederland og Tyskland i PIRLS2001-undersøkelsen.

Bacheloroppgave i Samfunnsøkonomi

Veileder: Bjarne Strøm

Mai 2020

Sammendrag

Fra tidligere studier har man sett at barnehagedeltagelse kan ha en positiv sleeper effekt på leseforståelse, spesielt for elever med minoritetsspråklig bakgrunn. På bakgrunn av Progress in International Reading Literacy Study(PIRLS)2001 for Tyskland, Nederland, Norge og Sverige er det ikke grunnlag for å konkludere med det samme. Det var mulig å påvise en sammenheng for elever i Tyskland som aldri snakket testspråket hjemme, som fikk en signifikant økning i testscore. Det er likevel grunn til å tro at barnehagedeltagelse er positivt korrelert med leseferdigheter i fjerdeklasse, men ikke konklusivt i denne oppgaven.

Abstract

From earlier studies it has been seen that kindergarten attendance can have positive sleeper effect on literacy, especially for students from a minority language background. On the basis of the Progress in International Reading Literacy Study(PIRLS)2001 for Germany, Netherlands, Norway and Sweden it is not possible to conclude on the same. It was possible to show a correlation for students in Germany who never spoke the test language at home, who did get a significant boost in their test score. It is nevertheless reason to believe that kindergarten attendance is positively correlated with literacy for fourth graders, but not conclusive in this thesis.

Innhold

1	Problemstilling	4
1.1	Innledning	4
2	Teoretisk rammeverk	4
2.1	Regresjonsanalyse ved minste kvadraters metode	4
2.2	Ordinary least squares	5
2.3	Determinisjonskoeffisienten R^2	5
2.4	Multipel regresjons	6
2.5	Hypotesetesting	6
2.6	Skoleproduktfunksjonen	6
2.7	Tidligere studier	7
3	Datanalyse	7
3.1	PIRLS	7
3.2	Sammenligning av Norge, Sverige, Tyskland og Nederland	8
3.2.1	Barnehagedeltakelse	8
3.2.2	Andel fra vanskeligsstilte hjem	9
3.2.3	Andel født i et annet land	9
3.2.4	Inntekt og antall bøker	10
3.3	Deskriptiv statistikk	10
3.3.1	Avhengig variabel	10
3.3.2	Forklaringsvariabel	11
3.3.3	Kontrollvariable	13
3.4	Korrelasjon	14
3.4.1	Korrelasjonsforskjeller mellom landene	16
4	Empirisk strategi	16
4.1	Økonometrisk grunnmodell	16
4.2	Hypoteseformulering	17
5	Regresjonsanalyse	18
5.1	Direkte effekter	18
5.2	Krysseffekter	20

6 Hypotesetesting	21
6.1 Test av hovedhypotese	21
6.2 Test av alternativhypoteser	22
7 Diskusjon	22
8 Konklusjon	23
9 Bibliografi	23
10 Appendix	24

1 Problemstilling

Problemstillingen for oppgaven er å analysere effekten av å ha gått i barnehage på leseferdigheter i fjerdeklasse i Norge, Sverige, Nederland og Tyskland i PIRLS2001-undersøkelsen.

1.1 Innledning

I Tyskland, Nederland, Norge og Sverige går en stor andel av barna i barnehage. Hvordan påvirker dette deres leseferdigheter? Leseferdigheter er tidligere vist å gi mindre sjanse for å droppe ut (Hernandez, D.J., 2011 [5]). Hvis barnehagedeltagelse kan bidra til bedre leseferdigheter, kan det indirekte minske skolefravall.

Barnehager er i stor grad helt eller delvis finansiert av staten, og det er derfor nyttig å se på effekten av tiltaket. Det er også nyttig å se om tiltaket kan spisses. Kan det for eksempel påvises en større effekt av å gå i barnehage for folk med foreldre født i et annet land, eller for elever som ikke snakker testspråket hjemme?

Tyskland, Nederland, Norge og Sverige er tre land med høy andel av barn i barnehagealder i barnehage. Hvordan påvirker dette leseferdighetene til barna, kontra de som ikke har gått i barnehage? Er betydningen annerledes for barn av minoritetspråklige foreldre, eller barn fra lavinntektshusholdninger? Ønsker å benytte dataene fra Progress in international Reading Literacy Study, PIRLS for å undersøke effekten av barnehage på leseferdigheter i 4. klasse, og se om effekten er forskjellig i de fire nevnte landene, og undersøke hva dette i så fall kan skyldes.

2 Teoretisk rammeverk

2.1 Regresjonsanalyse ved minste kvadraters metode

For å se på effekten av barnehagedeltagelse vil regresjonsformen minste kvadraters metode benyttes. For to variabler, en avhengig variable Y og en forklaringsvariabel X antas forholdet linjært. X og Y er observasjoner, hvor hver observasjon er et reelt tall. Dette kan settes opp på formen

$$E(Y) = \alpha + \beta X \quad (1)$$

hvor $E(Y)$ er forventningsverdien til den avhengige variabelen, avhengig av X . α er der grafen til forventningsverdien krysser y -aksen og β er stigningstallet til grafen. $E(Y)$ er ikke nødvendigvis forventningsrett for Y og Y uttrykkes dermed på formen

$$Y = E(Y) + \epsilon \quad (2)$$

hvor ϵ er estimeringsfeilen som tar høyde for at $E(Y)$ ikke forklarer Y eksakt. Dette kan kombineres til

$$Y = \alpha + \beta X + \epsilon. \quad (3)$$

Gitt n antall observasjoner i X og Y kan vi skrive Y på formen

$$Y_i = \alpha + \beta X_i + \epsilon_i \quad (4)$$

hvor Y_i indikerer observasjon i av n . For valgte α og β vil det være en differanse mellom faktisk Y_i og estimert, \hat{Y}_i . Forholdet mellom disse er estimeringsfeilen for hver observasjon, på formen

$$\epsilon_i = Y_i - \hat{Y}_i \quad (5)$$

2.2 Ordinary least squares

Mer spesifikt vil metoden ordinary least squares, OLS, benyttes. Da er det fem forutsetninger som må oppfylles. X må være ikke-stokastisk. Forventningsverdien til estimeringsfeilen må være lik null, variansen til estimeringsfeilen må være lik forventningsverdien kvadrert og estimeringsfeilen må være normalfordelt, $\epsilon_i \sim N(0, \sigma^2)$.

Hvis disse forutsetningene er oppfylt, ønskes det å minimere estimeringsfeilen kvadrert, slik at

$$SSR = \sum \epsilon_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - a - bX_i)^2 \quad (6)$$

Hvor SSR er *square sum of residuals*, som ønskes minimert ved valg av a og b . Mer spesifikt ønskes

$$\frac{\partial SSR}{\partial a} = -2\sum (Y_i - a - b - X_i) = 0 \text{ og} \quad (7)$$

$$\frac{\partial SSR}{\partial b} = -2\sum (Y_i - a - b - X_i)X_i = 0, \quad (8)$$

et likningssett med to likninger og to ukjente, som først kan løses for a slik at

$$a = \frac{\sum Y_i}{n} + b \frac{\sum X_i}{n} \quad (9)$$

Bemerk at a er gitt av gjennomsnittsverdiene til Y og X , henholdsvis \bar{Y} og \bar{X} . Dermed kan likningssettet løses tilsvarende for b , med a innsatt. Som gir

$$b = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \quad (10)$$

Da er b en lineær estimator for β , altså at $E(b) = \beta$, med normalfordelt b . Fra forutsetningen om normalfordelt ϵ_i følger også normalfordelt Y_i .

2.3 Determinisjonskoeffisienten R^2

Determinisjonskoeffisienten er et mål på hvor godt modellen forklarer variasjonen i Y , og beregnes ut i fra de tre likningene

$$\begin{aligned}
SST &= \sum (Y_i - \bar{Y})^2 \\
SSE &= \sum (\hat{Y}_i - \bar{Y})^2 \\
SSR &= \sum \epsilon_i^2
\end{aligned}
\tag{11}$$

hvor SST er total variasjon, SSE er forklart variasjon og SSR er den uforklarte variasjonen slik at

$$R^2 = \frac{SSE}{SST} = \frac{SST - SSR}{SST} = 1 - \frac{SSR}{SST} \tag{12}$$

R^2 vil dermed ligge mellom 0 og 1, hvor 0 tilsvarer at modellen ikke forklarer noe av variasjonen i Y mens 1 tilsvarer at modellen forklarer all variasjon i Y .

2.4 Multippel regresjons

Antar k -antall variable og en modell på formen

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon \tag{13}$$

slik at hver enkelt observasjon uttrykkes

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \epsilon_i \tag{14}$$

For å nå få forventningsrette estimatorer må det antas at ingen av variablene X_{ij} er en kollineære med noen av de andre variablene. Ønsker fortsatt å minimere SSR , med k antal variabler så vi vår et likningssett med k -antall likninger på formen

$$\frac{\partial S}{\partial b_j} = 2 \sum X_{ji} (Y_i - b_1 - b_2 X_{2i} - \dots - b_k X_{ki}), \quad X_{1i} = 0 \tag{15}$$

2.5 Hypotesetesting

F-test vil benyttes for hypotesetesting, hvor hypotesen vil settes som restriksjon på modellen. Da vil en testobservatoren ha formen

$$TS = \frac{(SSR_u - SSR)/h}{SSR_u/(n-k)} \sim F_{h, n-k} \tag{16}$$

hvor SSR_u er uforklart variasjon uten restriksjoner, og SSR er uforklart variasjon med restriksjoner. h er antall variabler som blir testet og k er antall variabler totalt.

2.6 Skoleproduktfunksjonen

Skoleproduktfunksjonen er en produktfunksjon for utdanning. Hvor idéen er at skolerresultat kan beskrives av innsatsfaktoreri skolen, samt elevens bakgrunn og miljø. Idéen ble først presentert av

(Coleman, James S., Ernest Q. Campbell, 1966 [2]), og senere utviklet av blant andre (Hanushek, Eric A., 2007 [3]).

Ifølge skoleproduktfunksjonen så kan en elevs resultat beskrives av faktorer bestemt av skolemyndighetene, som klassestørrelse, lærerbakgrunn og tilgjengelige ressurser som PC. I tillegg kan resultatene beskrives av en elevs bakgrunn som for eksempel hvor de er født, om de snakker samme språk på skolen som hjemme, foreldrenes inntekt og akademiske bakgrunn og antall bøker hjemme. Skoleresultatene kan også tilskrives faktorer bestemt av medelevene, som for eksempel antall medelever fra vanskeligstilte hjem, og antall elever født i et annet land.

Resultatet av skoleproduktfunksjonen kan være antall år med utdanning fullført, lønn etter endt utdanning, eller i nyere tid standardiserte restresultater. Dette muliggjør sammeligning for å kunne gi statistisk signifikante resultater som veiledning for skolemyndighetene for å kunne allokere ressursene mest mulig effektivt.

Skoleproduktfunksjonen vil ha formen

$$R = f(S, F, P) \tag{17}$$

hvor R er resultat, og er en funksjon av skolefaktorer, S , familiefaktorer, F og peer-faktorer, P .

2.7 Tidligere studier

Tidligere studier har vist at barnehagedeltakelse kan ha en effekt på leseferdigheter senere i skoleløpet, men ikke entydig. Haslip, M., 2018 [4] har vist at barnehagedeltakelse gir klart bedre leseferdigheter i førsteklasse. (Magnuson, K. A., 2007 [6]) har vist en sleeper-effekt, hvor barnehagedeltakelse kan i utslag i bedre leseferdigheter senere i skoleløpet.

Barnehagedeltakelse kan være spesielt viktig for noen grupper. (Melby-Lervåg, M., Lervåg, A., 2011 [7]) har vist at spesielt barn med et annet morsmål kan få økte leseferdigheter som følge av å ha gått i barnehage. (Camilli, G., Vargas, S., Ryan, S., 2010 [1]) har vist over flere tiår at barn fra vanskeligstilte hjem kan få stort utslag i leseferdighet i senere år av barnehagedeltakelse.

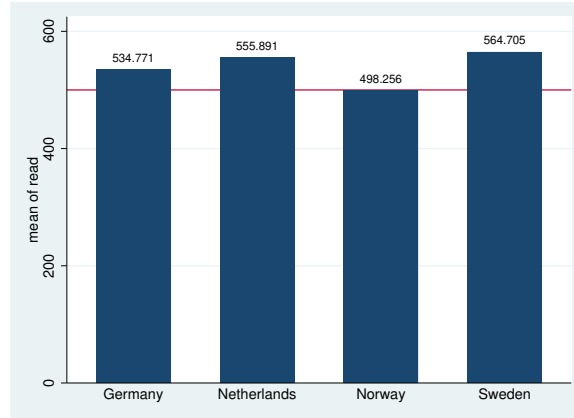
3 Datanalyse

3.1 PIRLS

Progress in International Reading Literacy Study (PIRLS), er en internasjonal leseundersøkelse arrangert av The International Association for the Evaluation of Educational Achievement (IEA). Den ble først arrangert i 1991. Tallene i denne undersøkelsen stammer fra PIRLS 2001 som omfattet 150 000 elever fra 35 forskjellige land.

PIRLS vektlegger forståelse, med både flervalgsspørsmål og åpne svaralternativer. Svarprosenten på elevene som gjennomførte PIRLS var på 92%.

Snittet for PIRLS er 500 poeng. De fire landene som skal undersøkes er Norge, Sverige, Tyskland og Nederland.



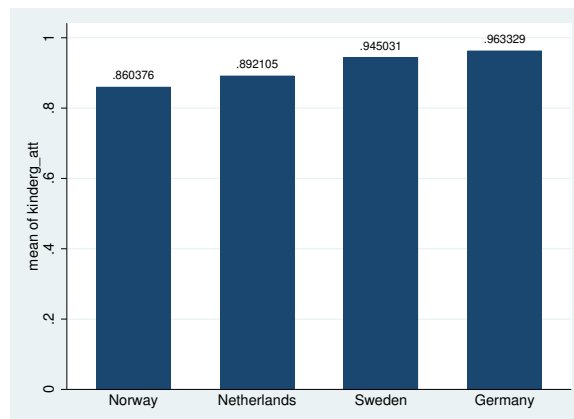
Figur 1: Gjennomsnittlig lesescore for de fire landene. Den røde linjen viser gjennomsnittet for alle landene som deltok i PIRLS 2001. Data fra PIRLS 2001.

Som man kan se av figur 1 ligger Norge så vidt under gjennomsnittet, mens Sverige har høyest score med 564.7, som også var det høyeste resultatet i PIRLS 2001. Også Tyskland og Nederland ligger godt over snittet.

3.2 Sammenligning av Norge, Sverige, Tyskland og Nederland

3.2.1 Barnehagedeltakelse

Alle de fire landene har høy barnehagedeltakelse, men det fortsatt store forskjeller mellom landene som man kan se av figur 2

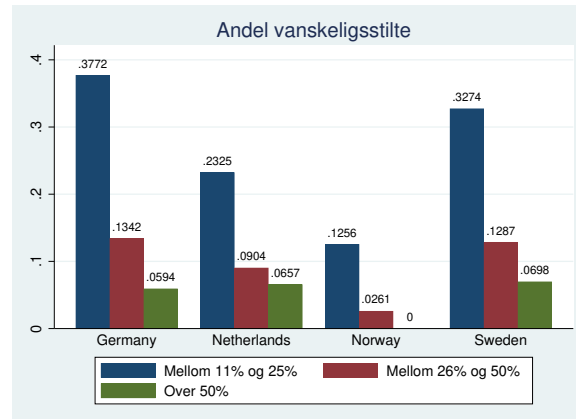


Figur 2: Prosentvis barnehagedeltakelse for de fire landene. Data fra PIRLS 2001

I Tyskland er det svært høy barnehagedeltakelse, hvor under under 4% ikke går i barnehage. Sverige har tilsvarende høy score, mens både Norge og Nederland har lavere barnehagedeltakelse med henholdsvis 86% og 92%.

3.2.2 Andel fra vanskeligsstilte hjem

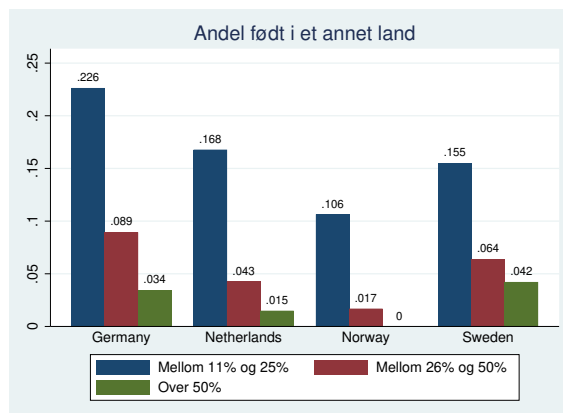
Som nevnt i det teoretiske rammeverket har (Camilli, G., Vargas, S., Ryan, S., 2010 [1]) vist at vanskeligstilte har bedre effekt av barnehage enn ikke-vanskeligstilte. Norge skiller seg fra de tre andre landene ved å ha lav andel fra vanskeligsstilte hjem, mens både Nederland, Tyskland og Sverige har sammeligbare tall, som man kan se av figur 3



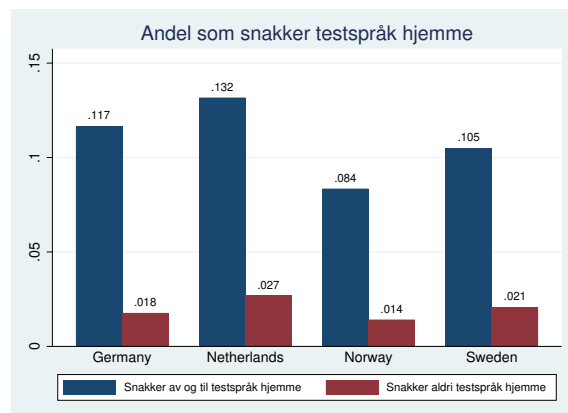
Figur 3: Andel fra vanskeligsstilte hjem. Data fra PIRLS 2001

3.2.3 Andel født i et annet land

(Melby-Lervåg, M., Lervåg, A., 2011 [7]) har vist at også morsmål er en viktig faktor å se på når man skal studere effekten av barnehagedeltakelse.



Figur 4: Andel fra født i et annet land. Data fra PIRLS 2001

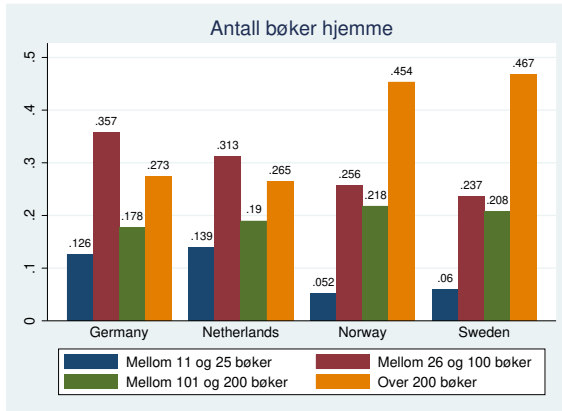


Figur 5: Andel som snakker morsmål hjemme. Data fra PIRLS 2001

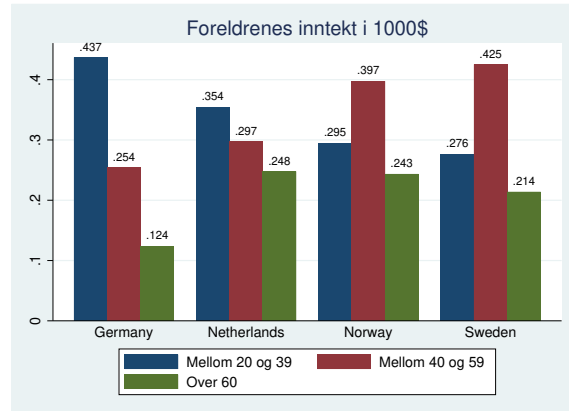
Hvor Norge igjen er annerledeslandet med en lavere andel født i et annet land, mens andelen som kun av og til og aldri snakker testspråket hjemme er relativt like for de fire landene.

3.2.4 Inntekt og antall bøker

Fra tidligere studier er det også vist at antall bøker hjemme og foreldrenes inntekt (sett inn referanse her) har stor innvirkning på leseferdighet, og vil dermed være viktige kontroll variable.



Figur 6: Antall bøker hjemme. Data fra PIRLS 2001



Figur 7: Foreldrenes inntekt. Data fra PIRLS 2001

Både Norge og Sverige er nær identiske på både inntekt og antall bøker hjemme, og en større middelklasse, inntekt mellom 40 000 og 59 000\$, mens Nederland har en større andel i klassen over 60 000\$. Tyskland skiller seg ut med noe lavere inntekt enn de andre landene, men skiller seg lite ut i forhold til Nederland på antall bøker hjemme.

3.3 Deskriptiv statistikk

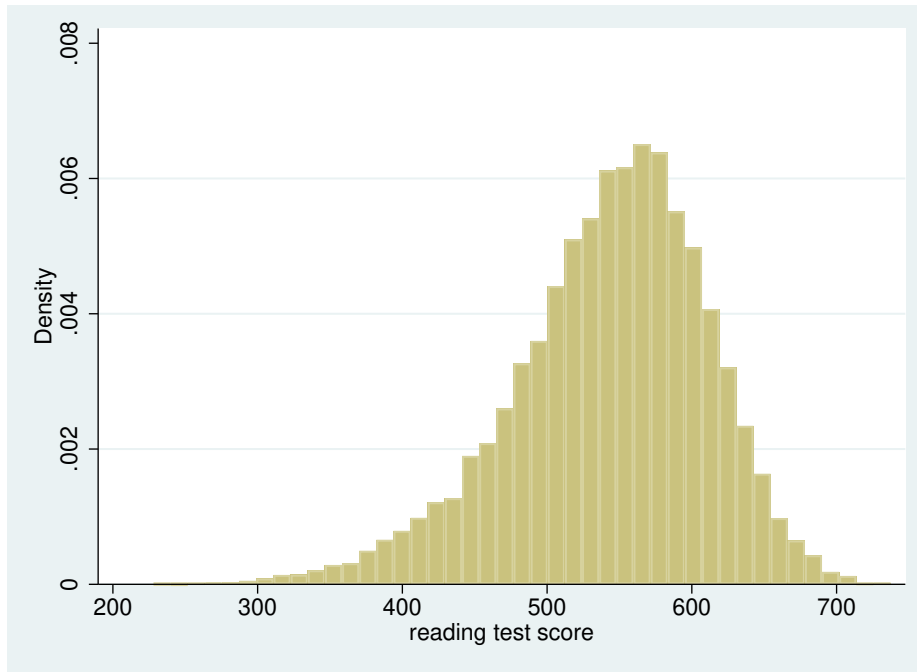
3.3.1 Avhengig variabel

Den avhengige variabelen for undersøkelsen er *read*, en kontinuerlig variabel definert som testscoren til hver enkelt elev på lesetesten til PIRLS.

Tabell 1: Deskriptiv statistikk for *read*

	count	mean	sd	min	max
read	22403	542.6288	67.9328	228.0606	737.3258

Read har et relativt standardavvik på 67.932, 8% av gjennomsnittsverdien. Det er stor forskjell mellom minimumsverdien, på 228, og maksimumsverdien på 737.



Figur 8: Histogram for *read*. Data fra PIRLS 2001

Histogrammet for *read* er tydelig forskjøvet mot høyre for gjennomsnittet for PIRLS 2001 totalt, grunnet de høye gjennomsnittlige testresultatene for Sverige, Tyskland og Nederland.

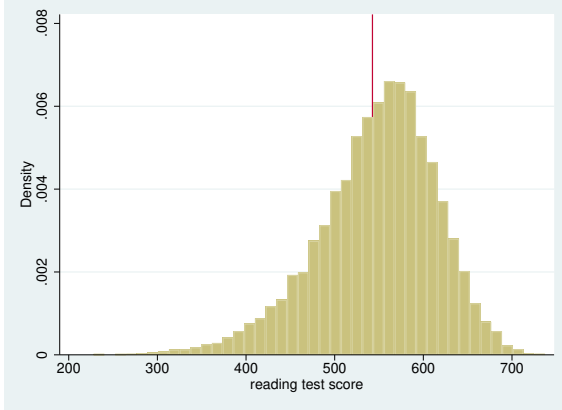
3.3.2 Forklaringsvariabel

Forklaringsvariabelen i undersøkelsen er *kinderg_att*, en dummyvariabel som tar verdien 1 om eleven har gått i barnehage og 0 ellers.

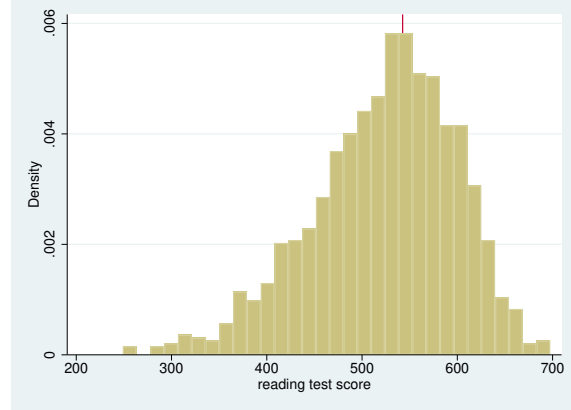
Tabell 2: Deskriptiv statistikk for *kinderg_att*

	count	mean	sd	min	max
<i>kinderg_att</i>	19009	.9300857	.2550092	0	1

kinderg_att har et høyt snitt, over hele tallgrunnlaget er det bare 7% som ikke går i barnehage. Av totalt antall observerte fra tabell 1, har 84.8% variabelen *kinderg_att*.

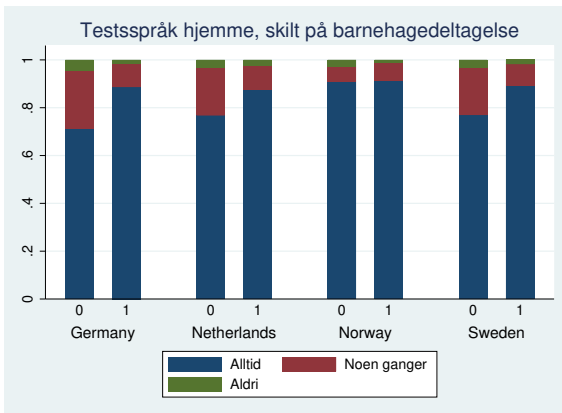


Figur 9: Histogram over *read* for barn som gått i barnehage. Data fra PIRLS 2001

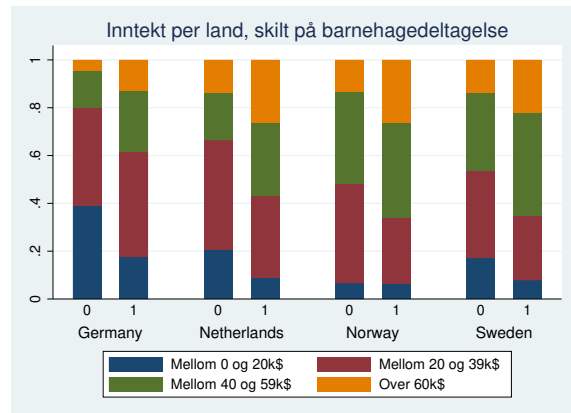


Figur 10: Histogram over *read* for barn som ikke gått i barnehage. Data fra PIRLS 2001

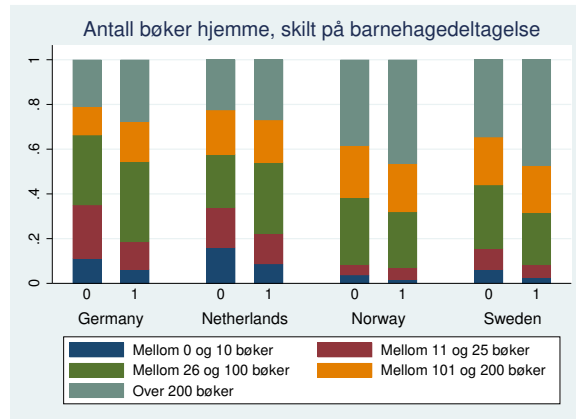
Den røde hjelpelinjen i figur 9 og 10 er gjennomsnittet av *read* fra tabell 2. En betraktelig større andel av elevene som ikke har gått i barnehage er under gjennomsnittet i tallmaterialet, kontra elevene som har gått i barnehage.



Figur 11: Andel som snakker testspråk hjemme. 0 hvis de ikke har gått i barnehage, 1 hvis de har gått i barnehage. Data fra PIRLS 2001



Figur 12: Foreldrenes inntekt. 0 hvis de ikke har gått i barnehage, 1 hvis de har gått i barnehage. Data fra PIRLS 2001



Figur 13: Antall bøker hjemme. 0 hvis de ikke har gått i barnehage, 1 hvis de har gått i barnehage. Data fra PIRLS 2001

Ser av figur 11, 12 og 13 en tydelig trend for samtlige land at karakteristikkene forbundet med høy teststcore er mer fremtredene for dem som går i barnehage. Altså at man med en bemiddlet bakgrunn har større sannsynlighet for å gå i barnehage.

3.3.3 Kontrollvariable

Kontrollvariablene er delt opp i tre kategorier, familiekarakteristika, medelevkarakteristika og skolekarakteristika.

Tabell 3: **Familiekarakteristika**(F):

Variabler	Beskrivelse
books_home11_25	tar verdien 1 om eleven har 11-25 bøker hjemme, 0 ellers
books_home26_100	tar verdien 1 om eleven har 26-100 bøker hjemme, 0 ellers
books_home101_200	tar verdien 1 om eleven har 101-200 bøker hjemme, 0 ellers
books_home200_plus	tar verdien 1 om eleven har over 200 bøker hjemme, 0 ellers
income20_39	tar verdien 1 om elevens foreldre tjener 20 000-39 000\$, 0 ellers
income40_59	tar verdien 1 om elevens foreldre tjener 40 000-59 000\$, 0 ellers
income60_plus	tar verdien 1 om elevens foreldre tjener over 60 000\$, 0 ellers
par_edu_higher	tar verdien 1 om elevens foreldre har høyere utdanning, 0 ellers
par_not_born	tar verdien 1 om elevens foreldre er født i et annet land enn testlandet, 0 ellers.
par_unemp	tar verdien 1 om begge elevens foreldre er uten arbeid, 0 ellers
speak_testlang_sometimes	tar verdien 1 om eleven prater testsspråket hjemme av og til, 0 ellers
speak_testlang_never	tar verdien 1 om eleven aldri prater testsspråket hjemme, 0 ellers

Familiekarakteristika beskriver det i elevens bakgrunn som kan ha innvirkning på testscoren.

Tabell 4: **Medelevkarakteristika**(P)

Variabler	Beskrivelse
pct_abroad11_25	tar verdien 1 om andelen elever på skolen født i et annet land er mellom 11 og 25 prosent, 0 ellers
pct_abroad26_50	tar verdien 1 om andelen elever på skolen født i et annet land er mellom 26 og 50 prosent, 0 ellers
pct_abroad50_plus	tar verdien 1 om andelen elever på skolen født i et annet land er over 50 prosent, 0 ellers
pct_disadv11_25	tar verdien 1 om andelen elever på skolen fra et vanskeligstilt hjem er mellom 11 og 25 prosent, 0 ellers
pct_disadv26_50	tar verdien 1 om andelen elever på skolen fra et vanskeligstilt hjem er mellom 26 og 50 prosent, 0 ellers
pct_disadv50_plus	tar verdien 1 om andelen elever på skolen fra et vanskeligstilt hjem er over 50 prosent, 0 ellers

Medelevkarakteristika beskriver medelevene til eleven på skolenivå, som kan ha innvirkning på elevens testscore.

Tabell 5: **Skolekarakteristika**(S):

Variabler	Beskrivelse
pc_class	tar verdien 1 om eleven har tilgang på PC i klasserommet, 0 ellers
teacher_cert	tar verdien 1 om læreren er sertifisert, 0 ellers
teacher_exp	beskriver antall år med erfaring læreren har
clsiz	beskriver antall elever i klassen

Skolekarakteristika beskriver skolefaktorene som kan ha innvirkning på elevens testscore. Deskriptiv statistikk for kontrollvariable er inkludert i tabell 15 i appendix.

3.4 Korrelasjon

Startpunkt for undersøkelsen er å se på korrelasjonen mellom avhengig variabel, forklaringsvariabel og kontrollvariable. Det gir et bilde av hvilke variable som er nødvendige for å ikke gi skjevhet i OLS-estimatoren, som dermed vil gi et feil bilde og potensielt være villedende.

Tabell 6: Korrelasjonsmatrise mellom *read*, *kinderg_att* og kontrollvariable

	read	kinderg_att
read	1	
kinderg_att	0.0943***	1
books_home11_25	-0.151***	-0.0295***
books_home26_100	-0.117***	0.00556
books_home101_200	0.0176*	-0.00329
books_home200_plus	0.251***	0.0352***
income20_39	-0.113***	-0.0354***
income40_59	0.0645***	0.0279***
income60_plus	0.187***	0.0500***
par_edu_higher	0.245***	0.0937***
par_not_born	-0.208***	-0.0863***
par_unemp	-0.0786***	-0.0261***
pc_class	0.0681***	-0.0243**
pct_abroad11_25	-0.00902	0.0171*
pct_abroad26_50	-0.0702***	-0.00296
pct_abroad50_plus	-0.110***	-0.0548***
pct_disadv11_25	0.0376***	0.0475***
pct_disadv26_50	-0.0299***	0.00317
pct_disadv50_plus	-0.122***	-0.0338***
speak_testlang_sometimes	-0.174***	-0.0605***
speak_testlang_never	-0.0947***	-0.0317***
teacher_cert	0.0184**	-0.0201**
teacher_exp	-0.00705	0.0141
clsize	0.166***	0.0251***

Signifikansnivåer: * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$
 Fullstending korrelasjonstabellen er vedlagt i appendix.

Bemerk først og fremst at *kinderg_att* er positivt korrelert med *read*, uten å ta høyde for bakenforliggende årsaker. Både *books_home200_plus*, *income60_plus* og *par_edu_higher* er positivt korrelert med *kinderg_att*, og alle disse variablene er positivt korrelert med *read*.

Tilsvarende er også *books_home11_25*, *income20_39*, *par_unemp* og flere av de andre kontrollvariablene negativt korrelert med *kinderg_att*, og negativt korrelert med *read*. Det gjør at det basert på korrelasjonsmatrisen ikke er mulig å si noe om den statistiske sammenhengen mellom *kinderg_att* og *read*, fordi det kan være karakteristika ved elever som gikk i barnehage som har større innvirkning på *read*, enn det faktum at de gikk i barnehage, og motsatt.

3.4.1 Korrelasjonsforksjeller mellom landene

Tabell 7: Korrelasjonsmatrise mellom *read*, *kinderg_att* og kontrollvariable for de fire landene

	Germany		Netherlands		Norway		Sweden	
	read	kinderg_att	read	kinderg_att	read	kinderg_att	read	kinderg_att
kinderg_att	0.0747***	1	0.1000***	1	0.0763***	1	0.0688***	1
books_home11_25	-0.202***	-0.0652***	-0.163***	-0.0401*	-0.139***	0.00635	-0.157***	-0.0367**
books_home26_100	-0.0945***	0.0181	-0.0640***	0.0540**	-0.179***	-0.0380*	-0.125***	-0.0254*
books_home101_200	0.0749***	0.0248*	0.0447*	-0.00654	-0.0122	-0.0158	-0.0210	-0.00305
books_home200_plus	0.306***	0.0285*	0.272***	0.0312	0.259***	0.0571**	0.249***	0.0585***
income20_39	-0.0922***	0.0104	-0.205***	-0.0722**	-0.120***	-0.106***	-0.110***	-0.0475***
income40_59	0.190***	0.0440***	0.0673*	0.0713**	-0.0110	0.0102	-0.00831	0.0473***
income60_plus	0.186***	0.0460***	0.231***	0.0866**	0.205***	0.103***	0.210***	0.0458***
par_edu_higher	0.231***	0.0536***	0.327***	0.0811***	0.294***	0.135***	0.248***	0.0590***
par_not_born	-0.293***	-0.140***	-0.260***	-0.222***	-0.176***	-0.0226	-0.218***	-0.0868***
par_unemp	-0.104***	-0.0696***	-0.0844***	-0.0670***	-0.0399*	0.0144	-0.0620***	-0.0119
pc_class	0.00399	0.00768	-0.0575***	-0.0185	-0.0245	0.0443*	0.0241*	-0.0170
pct_abroad11_25	-0.0184	-0.00602	-0.0946***	0.0242	0.00209	0.0101	0.00951	-0.00680
pct_abroad26_50	-0.120***	-0.0387**	-0.0496**	-0.0768***	-0.0174	-0.00536	-0.0987***	0.0167
pct_abroad50_plus	-0.137***	-0.0748***	-0.0874***	-0.136***	-	-	-0.207***	-0.0798***
pct_disadv11_25	0.0408**	0.0212	-0.0137	0.0477*	-0.00910	0.0489*	-0.0182	-0.00460
pct_disadv26_50	-0.166***	-0.0412**	-0.0342*	-0.0510*	-0.0114	-0.0142	0.0131	0.0174
pct_disadv50_plus	-0.180***	-0.0491***	-0.210***	-0.103***	-	-	-0.186***	-0.0373**
speak_testlang_sometimes	-0.270***	-0.0886***	-0.173***	-0.0970***	-0.144***	0.0186	-0.158***	-0.0864***
speak_testlang_never	-0.106***	-0.0413***	-0.0498**	-0.0196	-0.164***	-0.0568**	-0.121***	-0.0231
teacher_cert	0.0365**	0.00256	0.0264	-0.0223	-0.0289	0.0108	0.0318**	-0.00723
teacher_exp	-0.00471	-0.00443	0.0669***	0.0290	-0.0389*	-0.0590**	0.0685***	0.00211
clsiz	0.0909***	0.0190	0.180***	0.104***	0.0295	-0.0333	0.0444***	0.00494

Signifikansnivåer: * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Nederland har den sterkeste korrelasjonen mellom *read* og *kinderg_att* på 0.1, mens Sverige har den laveste korrelasjonskoeffisienten på 0.0668. Norge har sterkest korrelasjon mot de familiekaraktéristika som er kjent for å gi gode leseferdigheter og *kinderg_att*, *books_home200_plus*.

Bemerk at *kinderg_att* er positivt korrelert med *read* for alle fire landene, og at de samme trendene går igjen, med unntak av *teacher_exp* som er positivt korrelert mot *read* for Sverige, og Nederland, mot negativt korrelert for Tyskland og Norge.

4 Empirisk strategi

4.1 Økonometrisk grunnmodell

Ønsker å se på effekten av barnehagedeltagelse på leseferdigheter i 4. klasse. I tråd med (Hanushek, Eric A., 2007 [3]) presentert i det teoretiske rammeverket er det flere faktorer som må kontrolleres for å kunne se om barnehagedeltagelse har en statistisk signifikant effekt, eller om det er bakenforliggende årsaker som gjør at elever som har gått i barnehage har bedre leseferdigheter i fjerdeklasse. Kommer til å basere grunnmodellen på ligning 17, men anta uavhengig av medelevaraktéristika og skoleressursvariable.

$$R_i = f(F_i) = \beta_0 + \beta_1 K_i + \beta_{(1+j)} B_i^{(j)} + \beta_{(5+k)} I_i^{(k)} + \beta_9 H_i + \beta_{10} N_i + \beta_{11} U_i + \beta_{(11+l)} T_i^{(l)} + \varepsilon_i \quad (1)$$

R_i er *read* for elev i . K er *kinderg_att* for elev i . $B_i^{(j)}$ hvor $j = 1, 2, 3, 4$ er henholdsvis *books_home11_25*, *books_home26_100*, *books_home101_200* og *books_home200_plus* for elev i . $I_i^{(k)}$ hvor $k = 1, 2, 3$ er

henholdsvis *income20_39*, *income40_59* og *income60_pluss* for elev i . H_i er *par_edu_higher* for elev i . N_i er *par_not_born* for elev i . U_i er *par_unemp* for elev i . $T_i^{(l)}$, hvor $l = 1, 2$ er henholdsvis *speak_testlang_sometimes* og *speak_testlang_never* for elev i . ε_i er stokastisk additiv feil.

For å se effekten av skoleressursvariablene utvides modell 1 med skoleressursvariablene, nevnt i figur 5.

$$R_i = f(F_i, S_i) = \beta_0 + F_i + \beta_{14}D_i + \beta_{15}C_i + \beta_{16}E_i + \beta_{17}Z_i + \varepsilon_i \quad (2)$$

hvor D_i er *pc_class* for elev i . C_i er *teacher_cert* for elev i . E_i er *teacher_exp* for elev i og Z_i er *clsiz* for elev i .

Til slutte utvides modellen til å inkludere medelevkarakteristika med variablene fra figur 4.

$$R_i = f(F_i, S_i, P_i) = \beta_0 + F_i + S_i + \beta_{(17+m)}A_i^{(m)} + \beta_{(20+n)}V_i^{(n)} + \varepsilon_i \quad (3)$$

hvor $A_i^{(m)}$ hvor $m = 1, 2, 3$ er henholdsvis *pct_abroad11_25*, *pct_abroad26_50* og *pct_abroad50_plus*. $V_i^{(n)}$ hvor $n = 1, 2, 3$ er henholdsvis *pct_disadv11_25*, *pct_disadv26_50* og *pct_disadv50_plus*.

For å undersøke kryss effekter utvides modell 3 for å se på kryss effektene av å ha gått i barnehage med δ_1 som er krysskoeffisienten mellom N_i og K_i , δ_2 som er krysskoeffisienten mellom T_i^1 og K_i , og δ_3 som er krysskoeffisienten mellom T_i^2 og K_i .

4.2 Hypoteseformulering

På bakgrunn av problemstillingen, teorien og den økomometriske modellen formuleres en hovedhypotese og tre alternative hypoteser.

Hovedhypotesen er at barnehagedeltagelse bidrar positivt på leseferdigheter i fjerdeklasse i alle fire land med et signifikansnivå på 5%.

Første alternativhypotese er at for barn med foreldre født i et annet land enn testlandet bidrar barnehagedeltagelse bidrar positivt på leseferdigheter i fjerdeklasse i alle fire land med et signifikansnivå på 5%.

Andre alternativhypotese er at for barn som av og til snakker testpråket hjemme bidrar barnehagedeltagelse bidrar positivt på leseferdigheter i fjerdeklasse i alle fire land med et signifikansnivå på 5%.

Tredje alternativhypotese er at for barn som aldri snakker testpråket hjemme bidrar barnehagedeltagelse bidrar positivt på leseferdigheter i fjerdeklasse i alle fire land med et signifikansnivå på 5%.

5 Regresjonsanalyse

5.1 Direkte effekter

Tabell 8: Regresjon over hele tallmaterialet på modell (1), (2) og (3)

	(1)	(2)	(3)
	read	read	read
kinderg_att	16.99*** (2.238)	16.42*** (2.314)	14.45*** (2.502)
_cons	492.6*** (4.062)	458.0*** (5.274)	453.4*** (5.765)
N	13158	12007	10617
R^2	0.127	0.148	0.154

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Tabell 8 er en avkortet versjon av regresjonsresultatet fra modell (1), (2) og (3) på hele tallmaterialet. Fullstendig regresjonsresultater er i tabell 16 i appendix. Modell (3) er en bedre fit enn modell (1) og (2), som man kan se av R^2 -verdien. Det mest bemerkelsesverdige er koeffisienten til *kinderg_att* i alle tre modellene, som er svært høy, med høyt signifikansnivå. Dette er mest sannsynlig fordi, som man kan se av figur 1 og figur 2, så har Norge både betraktelig lavere testscore og lavere barnehagedeltakelse. En elev som går i barnehage er ofte ikke norsk, og har dermed høyere score på *read*. Nederland har også lav barnehagedeltakelse, i forhold til de andre, men også færrest observasjoner. Tyskland og Sverige har i tillegg til høy barnehagedeltakelse og høy gjennomsnittlig testscore, også flest observasjoner, som kan forklare den høye koeffisienten til *kinderg_att*.

Dermed er det mer nyttig å se på regresjonsanalysen land for land, gjort i tabell 9.

Tabell 9: Regresjon land for land, modell (1), (2) og (3)

	Germany (1)-(3)			Netherlands (1)-(3)			Norway (1)-(3)			Sweden (1)-(3)		
	read	read	read	read	read	read	read	read	read	read	read	read
kinderg_att	1.404 (0.793)	3.048 (0.589)	2.465 (0.686)	-0.608 (0.896)	-6.226 (0.214)	-6.469 (0.230)	7.892 (0.055)	7.567 (0.072)	6.478 (0.165)	1.771 (0.594)	0.819 (0.813)	0.508 (0.888)
books_home11_25	4.456 (0.372)	4.673 (0.384)	5.049 (0.388)	7.570 (0.221)	6.905 (0.283)	6.796 (0.336)	10.28 (0.384)	7.175 (0.548)	8.797 (0.488)	12.54* (0.034)	13.64* (0.031)	15.69* (0.015)
books_home26_100	18.90*** (0.000)	18.74*** (0.000)	15.97** (0.003)	15.00** (0.007)	15.08** (0.009)	11.38 (0.077)	20.94* (0.048)	20.62 (0.053)	22.00 (0.054)	24.84*** (0.000)	25.95*** (0.000)	28.18*** (0.000)
books_home101_200	32.26*** (0.000)	31.39*** (0.000)	27.93*** (0.000)	22.96*** (0.000)	24.99*** (0.000)	20.34** (0.003)	31.94** (0.003)	30.48** (0.005)	30.90** (0.007)	28.12*** (0.000)	28.78*** (0.000)	31.23*** (0.000)
books_home200_plus	45.54*** (0.000)	47.13*** (0.000)	45.88*** (0.000)	32.30*** (0.000)	33.20*** (0.000)	28.77*** (0.000)	44.43*** (0.000)	43.55*** (0.000)	46.09*** (0.000)	39.58*** (0.000)	40.39*** (0.000)	42.02*** (0.000)
income20_39	12.92*** (0.000)	10.06** (0.001)	9.216** (0.008)	-9.610 (0.079)	-6.188 (0.290)	-5.384 (0.407)	7.837 (0.240)	8.351 (0.222)	14.95* (0.044)	-5.417 (0.107)	-4.785 (0.171)	-6.525 (0.069)
income40_59	25.59*** (0.000)	22.70*** (0.000)	22.49*** (0.000)	0.457 (0.936)	2.548 (0.678)	-1.057 (0.877)	7.499 (0.262)	7.930 (0.246)	12.92 (0.081)	-3.670 (0.269)	-3.949 (0.253)	-5.620 (0.113)
income60_plus	26.74*** (0.000)	22.10*** (0.000)	18.05*** (0.000)	4.639 (0.439)	9.296 (0.149)	8.211 (0.249)	21.29** (0.003)	20.41** (0.005)	25.55** (0.001)	11.37** (0.002)	11.49** (0.002)	8.969* (0.020)
par_edu_higher	11.65*** (0.000)	12.04*** (0.000)	10.13*** (0.000)	18.79*** (0.000)	18.59*** (0.000)	19.28*** (0.000)	29.24*** (0.000)	29.42*** (0.000)	29.52*** (0.000)	19.11*** (0.000)	19.15*** (0.000)	18.64*** (0.000)
par_not_born	-10.21** (0.001)	-11.58*** (0.001)	-9.182* (0.012)	-21.06** (0.002)	-18.55* (0.011)	-16.37* (0.031)	-5.513 (0.460)	-2.885 (0.706)	-5.319 (0.517)	-3.048 (0.349)	-2.588 (0.449)	1.053 (0.771)
par_unemp	-1.448 (0.672)	-0.687 (0.850)	3.560 (0.368)	-7.199 (0.237)	-7.101 (0.123)	-10.67 (0.638)	-2.240 (0.773)	-2.317 (0.771)	-4.056 (0.638)	-6.312 (0.202)	-7.473 (0.149)	-1.255 (0.816)
speak_testlang_sometimes	-19.05*** (0.000)	-19.61*** (0.000)	-18.72*** (0.000)	-13.44** (0.003)	-10.84* (0.026)	-9.243 (0.073)	-19.94*** (0.001)	-21.18*** (0.000)	-20.61** (0.001)	-7.429* (0.018)	-7.888* (0.016)	-7.324* (0.031)
speak_testlang_never	-22.11** (0.003)	-23.64** (0.003)	-30.17*** (0.001)	-4.948 (0.598)	4.160 (0.675)	-0.799 (0.940)	-65.72*** (0.000)	-73.24*** (0.000)	-72.87*** (0.000)	-13.71* (0.042)	-17.68* (0.014)	-13.74 (0.060)
not_born	-29.69*** (0.000)	-31.70*** (0.000)	-30.41*** (0.000)	-8.371 (0.188)	-13.01 (0.067)	-14.31* (0.049)	-23.71*** (0.000)	-25.05*** (0.000)	-23.95*** (0.000)	-33.09*** (0.000)	-31.47*** (0.000)	-32.36*** (0.000)
pc_class	-2.285 (0.239)	-3.088 (0.156)			-16.22 (0.052)	-48.01*** (0.000)	-5.254 (0.198)	-4.825 (0.279)		8.140 (0.127)	7.375 (0.167)	
teacher_cert	8.800* (0.012)	4.560 (0.250)		22.36 (0.250)	25.36 (0.194)		-3.839 (0.645)	5.515 (0.581)		1.989 (0.539)	3.079 (0.366)	
teacher_exp	0.136 (0.105)	0.101 (0.279)		0.361** (0.005)	0.359* (0.015)		-0.161 (0.219)	-0.156 (0.275)		0.240*** (0.001)	0.211** (0.003)	
cls_size	0.631* (0.012)	0.489 (0.078)		0.783* (0.015)	0.973** (0.010)		0.172 (0.551)	-0.0780 (0.804)		0.191* (0.027)	0.291** (0.002)	
pct_abroad0_10		11.40 (0.086)		10.96 (0.569)				-2.729 (0.851)		20.28*** (0.001)		
pct_abroad11_25		14.05* (0.034)		7.114 (0.715)				9.124 (0.534)		10.87 (0.088)		
pct_abroad26_50		4.855 (0.497)		-1.112 (0.957)				0 (.)		10.80 (0.083)		
pct_disadv0_10		16.54** (0.010)		5.834 (0.517)				-1.407 (0.765)		12.78** (0.003)		
pct_disadv11_25		12.78* (0.044)		-0.403 (0.966)				0 (.)		8.064 (0.057)		
pct_disadv26_50		0.659 (0.919)		-10.20 (0.323)				-9.208 (0.452)		24.34*** (0.000)		
_cons	506.0*** (0.000)	484.3*** (0.000)	472.6*** (0.000)	546.3*** (0.000)	513.3*** (0.000)	528.1*** (0.000)	442.2*** (0.000)	450.2*** (0.000)	442.6*** (0.000)	530.7*** (0.000)	512.1*** (0.000)	480.3*** (0.000)
N	3444	3032	2543	1226	1066	915	2705	2588	2244	5623	5179	4789
R ²	0.269	0.278	0.285	0.194	0.217	0.229	0.151	0.154	0.161	0.144	0.144	0.159

p-values in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

Det gir et annet bilde, hvor for modell 2 og 3 er overraskende effekten av å gå i barnehage negativ på *read* i Nederland, hvor for modell 3 å gå i barnehage gir en reduksjon i testscore på 6.496 poeng. Det er klart høyest koeffisient i Norge, hvor barnehagedeltakelse gir en økning på 6.478 på modell 3, mens effekten er lav i Tyskland, i alle tre modeller. Sverige ligger midt mellom Norge og Tyskland. Resultatene i Tyskland er usikre, med høyt standardavvik, i forhold til utslag på 27% i modell 3.

Som forventet går det igjen i alle land at høyt antall bøker hjemme gir høye testresultater. Effekten av høy inntekt, 60k\$+ er størst i Tyskland og Norge hvor å ha foreldre med en inntekt på over 60k\$ i forhold til 0-19k\$ gir en økning på henholdsvis 18.05 og 25.55 poeng, mens den er lavere i Sverige og betrakelig lavere i Nederland.

Den negative effekten av å ikke snakke testspråket hjemme er svært høy i Norge, med -81.82 i Norge i modell 3, mens den i Nederland er svært mye lavere i absoluttverdi. Utslaget er også stort i Tyskland og Sverige, men godt under halvparten av effekten i Norge. Det kan bemerkes at den negative effekten på *read* av foreldre født i et annet land er størst i Nederland.

5.2 Krysseffekter

Neste steg er å se på krysseffektene i de forskjellige landene, basert på de tre alternativhypotesene.

Tabell 10: Regresjon med krysseffekt av *par_not_born* og *kinderg_att* på modell (3)

	Germany read	Netherlands read	Norway read	Sweden read
1.kinderg_att	1.758 (7.482)	-4.194 (5.642)	3.809 (4.789)	1.725 (3.996)
1.par_not_born	-12.91 (12.96)	-18.94 (15.24)	-32.46 (17.92)	-11.98 (9.701)
1.kinderg_att1.par_not_born	-1.967 (13.21)	0.227 (16.54)	24.69 (18.70)	7.984 (9.847)
<i>N</i>	2590	946	2273	4808
<i>R</i> ²	0.254	0.226	0.158	0.135

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Ser av tabell 10 at krysseffekten av *kinderg_att* og *par_not_born* kun er klart positiv i Norge og Sverige, hvor krysseffekten av å både ha gått i barnehage og ha foreldre født i et annet land gir økning i testscore på henholdsvis 24.69 og 7.984 poeng, men med høye standardavvik. Den er negativ i Tyskland, men også med høye standardavvik, dette blir diskutert i diskusjonen og relaterer trolig til dualitetet i krysseffekter.

Tabell 11: Regresjon med krysseffekt av *speak_testlang_sometimes* og *kinderg_att* på modell (3)

	Germany read	Netherlands read	Norway read	Sweden read
1.kinderg_att	1.820 (6.845)	-4.944 (5.701)	5.785 (4.776)	1.097 (3.947)
1.speak_testlang_sometimes	-20.40 (15.49)	-12.97 (14.38)	-16.23 (18.75)	-21.87* (10.15)
1.kinderg_att1.speak_testlang_sometimes	-3.705 (15.74)	5.751 (15.12)	-6.927 (19.50)	13.54 (10.40)
<i>N</i>	2590	946	2273	4808
<i>R</i> ²	0.254	0.226	0.157	0.135

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Noe overraskende er krysseffekten av *speak_testlang_sometimes* og *kinderg_att*, som for krysseffekten med *par_not_born* negativ både i Norge og Tyskland, mens den er positiv i Sverige og Nederland. Det er høye standardavvik, noe som gjør resultatene usikre.

Tabell 12: Regresjon med krysseffekt av *speak_testlang_never* og *kinderg_att* på modell (3)

	Germany read	Netherlands read	Norway read	Sweden read
1.kinderg_att	-0.486 (6.221)	-3.186 (5.381)	4.821 (4.700)	2.507 (3.713)
1.speak_testlang_never	-166.4** (53.98)	26.45 (28.32)	-95.61*** (23.77)	-36.75 (20.01)
1.kinderg_att1.speak_testlang_never	132.0* (54.70)	-36.41 (29.95)	22.05 (29.39)	17.32 (21.18)
<i>N</i>	2590	946	2273	4808
<i>R</i> ²	0.256	0.227	0.157	0.135

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Krysseffekten av *speak_testland_never* og *kinderg_att* mot *read* gir høyere utslag enn de tidligere hypotesene, dermed også høyere standardavvik, men for Tyskland er standardavviket relativt mindre sett opp mot hvor stor effekten er på *read*.

Som man kan se av tabell 10, 12 og 11 er den eneste krysseffekten som gir en p-verdi tilsvarende en statistisk signifikant effekt å gå i barnehage når man ikke snakker testspråket hjemme, i Tyskland, med et signifikansnivå mindre enn 5%. For de som aldri snakker testspråket hjemmet gir det å gå i barnehage en økning i testscore på 132 poeng, i Tyskland.

6 Hypotesetesting

6.1 Test av hovedhypotese

Tester hovedhypotesen hvorvidt barnehagedeltagelse gir bedre leseferdigheter i fjerdeklasse, på modell 3, siden den har determinisjonskoeffisient nærest 1. Skal altså sette opp den ensidige hypotesetesten

$$H_0 : \beta_1 = 0$$

$$H_A : \beta_1 \neq 0$$

Utfører dermed hypotesetesten ved hjelp av Stata for Tyskland, Nederland, Norge og Sverige. Resultatet er presentert i tabell 13

Tabell 13: Hypotesetest $H_0 : \beta_1 = 0$

Germany	Netherlands	Norway	Sweden
$F(1, 2518) = 0.16$	$F(1, 890) = 1.44$	$F(1, 2221) = 1.93$	$F(1, 4764) = 0.02$
Prob > $F = 0.6860$	Prob > $F = 0.2305$	Prob > $F = 0.1649$	Prob > $F = 0.8882$

For ingen av landene kan nullhypotesen forkastes med et signifikansnivå på 5%, og må dermed beholdes. Det bemerkes at i Sverige og Tyskland er resultatet svært lite signifikant, mens det har en høyere grad av signifikans i Norge og Nederland, men langt unna det statistiske signifikansnivået på 5 %.

6.2 Test av alternativhypoteser

Tester så alternativhypotesene. Første alternativhypotese at det er en kryseffekt mellom å gå i barnehage og å ha foreldre født i et annet land, og ander og tredje alternativhypotese om det er en kryseffekt mellom å snakke teststpråket henholdsvis av og til og aldri hjemme og barnehagedeltagelse.

Tabell 14: Resultat av test av alternativhypoteser

Hypotesetest	Germany	Netherlands	Norway	Sweden
$H_0 : \delta_1 = 0$	$F(1, 2565) = 0.02$	$F(1, 921) = 0.00$	$F(1, 2250) = 1.74$	$F(1, 4783) = 0.66$
$H_A : \delta_1 \neq 0$	Prob $> F = 0.8816$	Prob $> F = 0.9890$	Prob $> F = 0.1868$	Prob $> F = 0.4175$
$H_0 : \delta_2 = 0$	$F(1, 2565) = 0.06$	$F(1, 921) = 0.14$	$F(1, 2250) = 0.13$	$F(1, 4783) = 1.70$
$H_A : \delta_2 \neq 0$	Prob $> F = 0.8140$	Prob $> F = 0.7038$	Prob $> F = 0.7224$	Prob $> F = 0.1929$
$H_0 : \delta_3 = 0$	$F(1, 2565) = 5.82$	$F(1, 921) = 1.48$	$F(1, 2250) = 0.56$	$F(1, 4783) = 0.67$
$H_A : \delta_3 \neq 0$	Prob $> F = 0.0159$	Prob $> F = 0.2244$	Prob $> F = 0.4533$	Prob $> F = 0.4137$

Hvor δ_1 er krysskoeffisienten til N_i og K_i , hvor δ_2 er krysskoeffisienten til T_i^1 og K_i , og δ_3 er krysskoeffisienten til T_i^2 og K_i i modell 3.

Ingen av nullhypotesene kan forkastes, med unntak av $H_0 : \delta_3 = 0$ for Tyskland, som må forkastes til fordel for alternativhypotesen $H_A : \delta_3 \neq 0$, altså at barnehagedeltagelse gir en effekt for elever som aldri snakker testpråket hjemme i Tyskland med et signifikansnivå på 1.59%.

7 Diskusjon

Det finnes ikke grunnlag i disse dataene for å konkludere med at barnehagedeltagelse generelt gir bedre leseferdigheter i fjerdeklasse i Tyskland, Nederland, Norge eller Sverige. Selv om det ikke er bevist noe statistisk signifikant sammenheng på et 5% signifikansnivå, så har med unntak av i Nederland barnehagedeltagelse en positiv korrelasjon med leseferdighetene, som styrker at det bør forsøkes å gjøre mer rettede studier.

Som man kan se av R^2 , så er modellene heller ikke spesielt treffsikre, med høye standardavvik. Man kunne derfor prøvd å sette opp andre modeller, med flere variabler som ikke er tilgjengelige i datagrunnlaget. Man kunne trolig sett mer på språkferdigheter og mer konkret hvilke språk som snakkes. Det kan for eksempel være at å snakke et språk hjemme nærmere til testspråket gir et annet utslag enn et språk veldig forskjellig fra testspråket.

Kvaliteten på barnehage er ikke kjent i datagrunnlaget. Det er uvisst hvor mange år den enkelte elev gikk i barnehage, og for eksempel om det var et fokus på lesing og språk i barnehagen. Et videre mer rettet studie burde derfor rettes også mot å se på bakgrunn i selve barnehagedeltagelsen til eleven.

Som man kan se av p-verdiene i tabell 9, så er det større variasjon og usikkerhet i effekten av barnehagedeltagelse for Tyskland og Sverige, som begge har høy barnehagedeltagelse, mens for Norge og Nederland hvor barnehagedeltagelsen er lavere er det en lavere p-verdi. Dette kan peke på at når det er så stor generell barnehagedeltagelse som i Sverige og Tyskland, så er det andre bakenforliggende årsaker med det lille utvalget som ikke går i barnehage, som ikke plukkes opp av modellen.

Som sett i datanalysen i figur 11, 12 og 13 går bemidlede allerede, karkaterstikker forbundet med høy teststcore i større grad i barnehage. Dette skal være tatt høyde for i modellen, men det kan være

at det ikke et stort nok utvalg av vanskeligstilte i barnehage til å kunne få statistisk signifikante resultater, og at dermed datagrunnlaget er for lite.

Spesielt krysseffektene varierer stort fra land til land. Dette kan ha sammenheng med dualiteten i krysseffekter. Når koeffisienten er negativ for krysseffekt *par_not_born* og *kinderg_att* i Tyskland i tabell 10, så kan det både bety at hvis man har foreldre som er født i utlandet, får man negativ effekt av å gå i barnehage, men også at hvis man går i barnehage, så er det negativ effekt av å ha foreldre som ikke er født i utlandet. Dette ser man også trolig for krysseffektene i tabell 11 og 12. Dette gjør det vanskelig å fastslå hva som påvirker hva.

8 Konklusjon

Det finnes ikke grunnlag i disse dataene for å konkludere med at barnehagedeltagelse generelt gir bedre leseferdigheter i fjerdeklasse i Tyskland, Nederland, Norge eller Sverige. Det er påvist en statistisk sammenheng med et signifikansnivå på 5% av krysseffekten av å gå i barnehage og å aldri snakker testspråket hjemme på leseferdigheter i fjerdeklasse i Tyskland, som får en økning i testscore på 132 poeng.

Selv om det ikke er en statistisk signifikant sammenheng på et signifikansnivå på 5%, er det grunn til å tro at barnehagedeltagelse er positivt korrelert med leseferdigheter i fjerdeklasse. Dette gir grunnlag for et videre studie, som bør inkludere flere variabler knyttet direkte mot barnehagen, som antall år og kvalitet på barnhagen.

9 Bibliografi

Referanser

- [1] Camilli, G., Vargas, S., Ryan, S., 2010. «*Meta-analysis of the effects of early education interventions on cognitive and social development*». I: *The Teachers College Record*, 112(3) (2010), s. 579–620.
- [2] Coleman, James S., Ernest Q. Campbell, 1966. *Equality of Educational Opportunity*. Washington: U.S. Government Printing Office.
- [3] Hanushek, Eric A., 2007. *Education Production Functions*. Stanford University: U.S. Hoover Institution.
- [4] Haslip, M., 2018. «*The effects of public pre-kindergarten attendance on first grade literacy achievement: a district study*». I: *International Journal of Child Care and Education Policy volume 1* (2018).
- [5] Hernandez, D.J., 2011. «*Double jeopardy: How third-grade reading skills and poverty influence high school graduation*». I: (2011). URL: <https://www.fcd-us.org/assets/2016/04/DoubleJeopardyReport.pdf>.
- [6] Magnuson, K. A., 2007. «*Does prekindergarten improve school preparation and performance?*» I: *Economics of Education Review*, 26 (2007).
- [7] Melby-Lervåg, M., Lervåg, A., 2011. «*Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: a meta-analysis of the correlational evidence*». I: *Journal of Research in Reading*, 34(1) (2011), s. 114–135.

10 Appendix

Tabell 15: Deskriptiv statistikk for kontrollvariablene

	observasjoner	gjennomsnitt	sd
books_home11_25	18940	.0928194	.2901869
books_home26_100	18940	.292661	.4549961
books_home101_200	18940	.1966737	.397494
books_home200_plus	18940	.368849	.4825057
income20_39	16294	.3417823	.4743216
income40_59	16294	.3501903	.477044
income60_pluss	16294	.1911133	.3931901
par_edu_higher	17074	.5784233	.4938259
par_not_born	21272	.1196879	.3246038
par_unemp	17617	.0562525	.2304152
pc_class	20980	.8198761	.3842998
pct_abroad11_25	20676	.1732443	.3784675
pct_abroad26_50	20676	.0613755	.2400236
pct_abroad50_plus	20676	.0282453	.1656769
pct_disadv11_25	19603	.2961281	.4565599
pct_disadv26_50	19603	.1080447	.3104448
pct_disadv50_plus	19603	.0549916	.2279696
speak_testlang_sometimes	21757	.1104472	.3134536
speak_testlang_never	21757	.0198097	.1393491
teacher_cert	20647	.9373759	.2422915
teacher_exp	20645	18.25832	11.72774
clsize	20989	24.52756	6.63973

	(1)	(2)	(3)
	read	read	read
kinderg_att	16.99*** (2.238)	16.42*** (2.314)	14.45*** (2.502)
books_home11_25	3.244 (3.691)	3.163 (3.890)	4.238 (4.156)
books_home26_100	15.57*** (3.294)	15.90*** (3.461)	16.66*** (3.710)
books_home101_200	23.43*** (3.395)	23.36*** (3.561)	23.57*** (3.812)
books_home200_plus	35.53*** (3.346)	35.80*** (3.511)	36.65*** (3.760)
income20_39	4.263 (2.245)	3.551 (2.372)	4.221 (2.539)
income40_59	8.580*** (2.284)	6.158* (2.414)	6.990** (2.585)
income60_plus	18.23*** (2.494)	15.13*** (2.634)	16.11*** (2.825)
par_edu_higher	21.63*** (1.245)	21.95*** (1.302)	21.55*** (1.383)
par_not_born	-9.674*** (2.300)	-10.44*** (2.423)	-10.88*** (2.617)
par_unemp	-4.981 (2.801)	-4.845 (2.940)	-3.401 (3.142)
speak_testlang_sometimes	-16.06*** (2.237)	-15.55*** (2.337)	-14.56*** (2.478)
speak_testlang_never	-27.21*** (4.881)	-30.89*** (5.140)	-32.91*** (5.381)
pc_class		3.499* (1.649)	4.780** (1.840)
teacher_cert		-3.633 (2.531)	-1.533 (2.752)
teacher_exp		0.130** (0.0504)	0.0742 (0.0535)
clsiz		1.408*** (0.0811)	1.394*** (0.0854)
pct_abroad11_25			-3.358 (1.835)
pct_abroad26_50			-8.840** (3.160)
pct_abroad50_plus			-10.83* (4.975)
pct_disadv11_25			10.25*** (1.419)
pct_disadv26_50			14.55*** (2.479)
pct_disadv50_plus			9.015* (3.766)
_cons	492.6*** (4.062)	458.0*** (5.274)	453.4*** (5.765)
<i>N</i>	13158	12007	10617
<i>R</i> ²	0.127	0.148	0.154

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Tabell 16: Regresjonsanalyse over hele tallmaterialet for modell 1-3

