

Nødtvedt, Vetle Christian  
Røstbakken, Åsmund  
Urke, Tor Inge

# En empirisk undersøkelse av leseferdigheter for elever som undervises av samme lærer i fire år

Bacheloroppgave i Samfunnsøkonomi

Veileder: Bjarne Strøm

Mai 2020



Nødtvedt, Vetle Christian  
Røstbakken, Åsmund  
Urke, Tor Inge

# **En empirisk undersøkelse av leseferdigheter for elever som undervises av samme lærer i fire år**

Bacheloroppgave i Samfunnsøkonomi  
Veileder: Bjarne Strøm  
Mai 2020

Norges teknisk-naturvitenskapelige universitet  
Fakultet for økonomi  
Institutt for samfunnsøkonomi



**NTNU**

Kunnskap for en bedre verden



En empirisk undersøkelse av leseferdigheter for elever som undervises av samme lærer i fire år. Skrevet av Vetle Christian Nødtvedt, Åsmund Røstbakken og Tor Inge Urke.

## Innholdsfortegnelse

<b>1. Innledning</b>	<b>3</b>
1.1. Problemstilling	3
<b>2. Teoretisk rammeverk og oversikt over tidligere studier</b>	<b>4</b>
2.1. Innledning	4
2.2. Teoretisk rammeverk	4
2.3. Tidligere studier	4
2.3.1. Lærerstabilitet	4
2.3.2. Kontaktlærer	5
2.4. Økt elev-lærer-forhold	5
2.5. Valgte land	6
<b>3. Datamaterialet</b>	<b>8</b>
3.1. Innledning	8
3.2. Om datamaterialet	8
3.2.1. Definisjoner	8
3.2.1.1. Avhengig variabel	8
3.2.1.2. Kontrollvariabler	8
3.2.1.3. Dummyvariabler	9
3.3. Deskriptiv statistikk	10
3.3.1. Deskriptiv statistikk for variablene	11
3.3.2. Kryssplott	12
3.3.3. Kovarians, korrelasjon og korrelasjonsmatrise	13
3.4. Oppsummering	14
<b>4. Økonometrisk modell</b>	<b>15</b>
4.1. Innledning	15
4.2. Valg av funksjonsform	15
4.2.1. Lineær grunnmodell:	15
4.3. Empirisk strategi	15

4.3.1. Estimeringsmetode	16
4.3.2. Modeller	16
4.3.3. Forskjeller mellom modellvariantene	18
4.4. Analyseopplegg	18
<b>5. Empiriske resultater</b>	<b>19</b>
5.1. Innledning	19
5.2. Empiriske hovedresultater	20
5.2.1. Fordeling 1 (samlet)	20
5.2.2. Fordeling 2 (EU)	23
5.2.3. Fordeling 3 (Ikke EU)	25
5.2.4. Sammenligning av EU land og ikke EU land	26
5.3. Oppsummering	28
<b>6. Avslutning</b>	<b>28</b>
6.1. Generell oppsummering	28
6.2. Konklusjon	29
6.3. Styrker og svakheter	29
6.4. Videre forskning	30
<b>7. Litteraturliste</b>	<b>31</b>
<b>8. Vedlegg</b>	<b>33</b>
8.1. STATA Do-file	33

# 1. Innledning

Skolen er et lærested hvor en tilegner seg kunnskap og ferdigheter. Barn skal gjennom et skolesystem som har en målsetting om å gi barnet en best mulig utdanning. I Norge måler vi dette blant annet etter elevprestasjoner. Tallene har i flere år vært middelmådige sammenlignet med andre nasjoner. Målet er å bedre kvaliteten på skoleprestasjonene (Briseid, 2019).

I flere nasjoner måles elevenes prestasjoner i lys av nasjonale prøver som PISA, TIMSS og PIRLS, der skolens kvalitet kan tolkes ut i fra prøvenes resultater. «Elevprestasjon» kan være et begrep med stor appell-verdi, for hvilket land vil ikke ha et skolesystem med kvalitet, der elevene presterer? Ved bruk av nasjonale prøver kan en med utgangspunkt i flere variabler gå i dybden på elevenes prestasjoner og skolesystemets kvaliteter. Hva står egentlig bak elevenes prestasjoner? Kan elever med samme lærer over tid bedre sine prestasjoner? Eller har ikke lærerens tidsperiode i klassen betydning for elevenes læring? (Briseid, 2019).

I denne oppgaven skal vi undersøke om det forekommer forskjeller på elevprestasjoner mellom elever som har hatt samme lærer i fire år, mot elever som har byttet lærer en eller flere ganger i løpet av disse fire årene. Dette gjelder fra første til fjerde trinn. Ved bruk av kvantitativ metode skal vi empirisk undersøke datasettet fra den internasjonale undersøkelsen fra PIRLS (Progress in International Reading Literacy Study), fra 2001. Den internasjonale undersøkelsen PIRLS er en lese-undersøkelse som omhandler elevenes leseferdigheter ved fjerde trinn. Utvalget ved denne undersøkelsen består av 150 000 elever fra 35 ulike nasjoner.

Vi tar utgangspunkt i nasjonene Tyrkia, Ungarn, Tyskland, Norge og Russland. Datamaterialet fra disse landene viser en jevn fordeling i andel elever som har hatt samme lærer i fire år, og de som ikke har hatt det. Nasjonene deles deretter inn i to grupper, de som er medlem av EU, og de som ikke er det. Norge er medlem av EØS-avtalen, men i denne oppgaven så tar vi med Norge under EU-landene. Effekten av å ha samme lærer i fire år skal vi kontrollere opp mot flere kontrollvariabler. Oppgaven skal belyse følgende problemstilling:

## 1.1. Problemstilling

*«Hva er det potensielle utbyttet av å undervises av samme lærer i fire år eller mer? Vil elever som bytter lærer oftere ha svakere elevprestasjoner enn de som beholder samme lærer over fire år? Og hvilke forskjeller er det mellom utvalgte EU-land og utvalgte ikke EU-land?»*

## **2. Teoretisk rammeverk og oversikt over tidligere studier**

### **2.1. Innledning**

I dette kapitlet skal vi ta for oss deler av teorien om elevprestasjoner i skolen. Først skal vi presentere tidligere funn fra internasjonale undersøkelser (PIRLS) og andre forskningsinstitutt. Funnene fra forskningsresultatene er lagt inn i en oversikt. Tilslutt vil vi belyse forskjeller som finnes i de forskjellige landenes skolesystem.

### **2.2. Teoretisk rammeverk**

Er det forskjeller i leseferdigheter mellom elever som har hatt samme lærer i fire år sammenlignet med de som ikke har hatt det? Tidligere forskning vi tar utgangspunkt i har kun sammenlignet prestasjonene til elever som har hatt samme lærer i to år og ikke mer.

Forskning på elevprestasjoner gjøres årlig av mange institusjoner verden over. Land praktiserer ulike måter å gjennomføre skolegang på. Et av målene med skolegang er å oppnå en høyest mulig elevprestasjon. I Norge gjennomfører Utdanningsdirektoratet jevnlig endringer for å få mest mulig utbytte av ressursene de har forbehold skolen. I Norge har endringer i lærernes utdanningslengde blitt forandret. Vil elevene yte bedre med å ha en lærer med mer utdanning enn tidligere?

Det er mange ulike faktorer som viser store forskjeller i elevprestasjonene. Foreldrenes utdanningsnivå er en faktor som gang på gang slår kraftig ut i barnas elevprestasjoner i skolen (Wiborg, Arnesen, Grøgaard, Støren & Opheim, 2011, s. 8). I Norge har SSB (Statistisk sentralbyrå), publisert en oversikt fra august 2019, som viser store forskjeller i gjennomsnittlige grunnskolepoeng i fagene engelsk skriftlig, norsk hovedmål og matematikk. Elever som har foreldre med høyere utdanning fra universitet og lignende, gir en gjennomsnittlig høyere elevprestasjon hos barn i skolen (Wiborg et al., 2011, s. 8).

### **2.3. Tidligere studier**

En doktorgrad fra 2002 skrevet av Vada S. Bogart fra East Tennessee State University viser til at det er en akademisk fordel å ha samme lærer i to år. Denne forskningen ble gjennomført fra starten av tredje klasse til slutten av fjerde klasse, hvor det ble tatt tester både før og etter prosjektet (Bogart, 2002, s. 2). Forskningen viser at elever som ble satt i et looping program, samme lærer i to år, scoret bedre på lesehastighet enn elever som byttet lærer mellom tredje- og fjerde klasse.

#### **2.3.1. Lærerstabilitet**

I en rapport av PIRLS (Progress in International Reading Literacy Study) konkluderes det med at norske elever i skoler med svakt lesegjennomsnitt har mindre lærerstabilitet (flere lærerbytter) enn elever med høyt lesegjennomsnitt (Solheim & Tønnessen, 2003).



### 2.3.2. Kontaktlærer

Kontaktlærer (også kalt klasseforstander) følger vanligvis samme klasse gjennom flere trinn på enten barneskolen, ungdomsskolen eller videregående. En kontaktlærer blir ansett som klassens hovedlærer, som har de fleste timene med den samme klassen. I Norge er norskfaget et vanlig ansvar for en kontaktlærer på barneskolen, hvor lesing er en viktig del av faget.

### 2.4. Økt elev-lærer-forhold

I en amerikansk undersøkelse fra 2018 presenteres empirisk data på økte prestasjoner ved et økt elev-lærer-forhold (samme lærer i to år) ved barneskolen. Små men tydelige fordeler viser størst effekter for minoriteter. "Dårlige" lærere har størst prestasjonseffekt av et økt elev-lærer-forhold over tid (Hill & Jones, 2018). Under ramser vi opp potensielle fordeler og ulemper ved å beholde samme lærer over flere år.

#### Fordeler:

- Forholdet mellom elev og lærer styrkes over tid.
- Bedre tilpasset læring til hver enkelt elev.
  - Over tid vil en som er lærer over flere år i større grad kunne tilpasse undervisningen til å møte de behovene som passer eleven best både individuelt og kollektivt.
- Mer tid brukes på å lære, mindre på å finne læringsnivået eller å bli kjent
  - Læreren kjenner allerede eleven og har i større grad kontroll på hvor langt elevene har (spesielt i starten av et skoleår) kommet i fagene, og hvor dyktige de faktisk er.
- Reduserte atferdsproblemer
  - Atferdsnormer er mer vanlig i stressende perioder som ved starten eller slutten av et skoleår, hvor sosiale aspekter kan ta mye tid. Å starte et nytt år med en kjent lærer kan skape en trygg og forutsigbar atmosfære.

#### Ulemper:

- Personlige konflikter
  - Det kan oppstå personlige konflikter mellom lærer, elev eller elevens foreldre. Konfliktene kan fortsette over flere år og kan skape store problemer for læringsmiljøet både for elev og lærer.
- Mangel på variasjon
  - Nye tilpasninger og varierte læringsmiljøer kan være en fordel for enkelte elever. Spesielt med tanke på videre utdanning i fremtiden, hvor lærere og læringsmiljø vil variere.
- Mindre tilpasningsdyktige
  - Når elever som er vant med samme lærer over flere år får en ny lærer, vil ikke-tilpasningsdyktige elever få større vansker med å tilpasse seg enn elever som er vant til at det ofte skjer endringer.

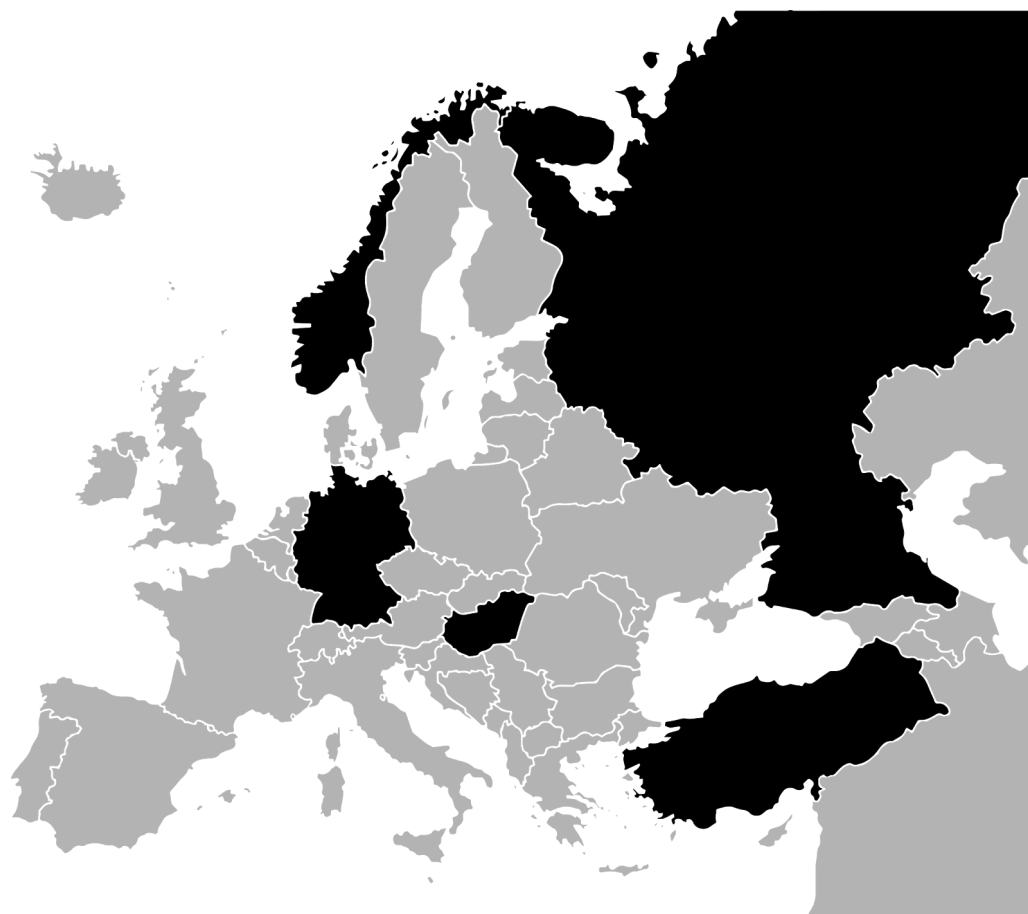
## 2.5. Valgte land

Vi har valgt å ta utgangspunkt i landene Tyrkia (Turkey), Ungarn (Hungary), Tyskland (Germany), Norge (Norway) og Russland (Russia). Alle disse landene har en jevn andel av elever som enten har hatt samme lærer i fire år eller ikke har det. Slik ser oversikten ut:

Navn	ID	Samme lærer 4 år	Ikke samme lærer 4 år	Ingen data
Turkey	792	56.39%	42.99%	0.62%
Hungary	348	54.09%	44.15%	1.77%
Germany	276	51.81%	43.26%	4.93%
Norway	578	39.17%	59.76%	1.07%
Russia	643	38.77%	61.23%	0.00%

Tabell 2.1: Fordeling lærer fire år

Det er lite trolig at disse landene har en enhetlig skolepolitikk for eller imot at samme lærer følger en skoleklasse i fire år eller mer. Dette kan i de forskjellige landene være bestemt av geografisk lokasjon, individuelt på de forskjellige skolene eller bare ved ren tilfeldighet.



Figur 2.2: Geografisk lokasjon til de fem landene vi har valgt

**Tyrkia**

Skolesystemet i Tyrkia er organisert med åtte års obligatorisk grunnskole fra barna er seks år. Første fremmedspråk er engelsk. Tyrkia mangler ressurser til å dekke hele landets skolebehov, det er blant annet mangel på skolebygg og andre ressurser. For å begynne på høyere utdanning i Tyrkia må elever bestå en statlig eksamen (Welle-Strand, 2009).

**Ungarn**

Skolesystemet er obligatorisk for alle fra alderen seks til 16 år. Barneskolen er på åtte år. Russisk var obligatorisk språk, men både engelsk og tysk er blitt de dominerende fremmedspråkene de siste årene. For de aller minste barna (under tre år), er det stor mangel på barnehageplasser, men det er obligatorisk barnehage for femåringer (Welle-Strand, 2013).

**Tyskland**

Året de fyller seks år begynner barna på barneskole. Det er obligatorisk skole i minst ni år. I 2009 gikk fem prosent av landets elever på private skoler. Elevene blir i all hovedsak undervist av en klasselærer de første to årene. Deretter blir elevene i økende grad undervist av faglærere (Pedersen, 2018).

**Norge**

Grunnskolen som gjelder barne- og ungdomsskole er obligatorisk og gratis for barn fra 6-årsalderen til 16 år. Barnehagene skal være for alle barn, uavhengig av foreldrenes økonomi. Norge har som mål at alle barn skal ha barnehageplass (Thune, Reisegg, Askheim, 2019).

**Russland**

Elevene i Russland har ni-årig obligatorisk skolegang. De har mulighet til å begynne på barneskole i seks- eller sju-års alderen. Velger de førstnevnte, må de ha fire år med barneskole, velger de å begynne som sju-åringer er det kun tre år med barneskole. Deretter begynner ungdomstrinnet (Papirleksikonet Store norske leksikon, 2013).

**2.5.1. Data**

	Antall elever	Gjennomsnittsalder	Minimum	Maximum	Gjennomsnittlig Read
Tyrkia	5125	10.2	6.9	14.9	449.0
Ungarn	4759	10.7	7.6	14.4	538.8
Tyskland	7633	10.5	8.5	13.3	534.8
Norge	3459	10.0	6.7	13.1	498.3
Russland	4093	10.3	8.1	14.2	530.0

Tabell 2.3: Data fra utvalgte land

## 3. Datamaterialet

### 3.1. Innledning

I dette kapitlet sammenligner vi deskriptiv statistikk fra undersøkelsen vi har tilgang på. Vi starter med å presentere datamaterialet. Deretter avklarer og definerer vi variablene som anvendes, før vi ser nærmere på gjennomsnittsverdi, standardavvik og kovarians.

### 3.2. Om datamaterialet

Datasettet som anvendes i vår studie er PIRLS (Progress in International Reading Literacy Study) fra 2001. Det er en internasjonal leserundersøkelse gjennomført av fjerdeklassinger i over 35 land, hvor vi har valgt fem land av interesse for vår undersøkelse. Datainnsamlingen er gjennomført i år 2001 i løpet av månedene april, mai og juni. Det var over 150 000 elever som deltok i undersøkelsen. I tillegg til leserundersøkelsen fikk elevene sammen med foreldre et spørreskjema om diverse elev-karakteristika (alder, kjønn, innvandrerstatus, antall bøker i hjemmet, foreldrenes utdanning, inntekt etc). Lærere og ledere på skolene fikk også et spørreskjema om skole-karakteristika (f.eks. antall elever på skolen, prosentandelen av elevene på skolen som ikke er født i landet, etc).

#### 3.2.1. Definisjoner

Her forklarer vi hva de forskjellige variablene betyr.

##### 3.2.1.1. Avhengig variabel

*Read*, leseferdighet, er den avhengige variabelen. Vi ønsker å finne ut hvilken effekt de forskjellige kontrollvariablene har på den avhengige variabelen.

- $read_i$  vårt mål på leseferdighet

##### 3.2.1.2. Kontrollvariabler

- $sameteacher4\_plus_i$  vårt mål på samme lærer i fire år eller mer (dummy)
- $clsiz_i$  vårt mål på klassestrørrelse
- $par\_edu_i$  vårt mål på foreldrenes utdanningsnivå (dummy)
- $girl_i$  vårt mål på elevens kjønn (dummy)
- $books\_home_i$  vårt mål på antall bøker elven har tilgang på hjemme (dummy)
- $kinderg\_att_i$  vårt mål på barnehage deltagelse (dummy)
- $teacher\_exp_i$  vårt mål på lærerens erfaring i antall år
- $teacher\_fem_i$  vårt mål på lærerens kjønn (dummy)

## 3.2.1.3. Dummyvariabler

Av variablene vi har valgt, så er seks av dem dummyvariabler. Dummyvariablene *sameteacher4\_plus*, *girl*, *kinderg\_att* og *teacher\_fem* fulgte med datasettet som vi tar utgangspunkt i. Variablene *par\_edu* og *books\_home* er kategorivariabler som vi har omkodet til dummyvariabler. Dette har vi valgt å gjøre fordi det er enklere å arbeide med dummyvariabler som alle har verdien 1 eller 0. Vi har valgt å kalle disse nye dummyvariablene for *highpar\_edu* og *highbooks\_home*. Oversikt over hva verdiene på de seks forskjellige dummyvariablene betyr:

sameteacher4_plus	Undervist av samme lærer i 4 år	1	
	Ikke undervist av samme lærer i 4 år	0	
highpar_edu	University degree	1	1
	Post secondary (not undeversity)	2	
	Upper secondary	3	0
	Lower secondary	4	
	Not completed lower secondary	5	
girl	Jente	1	
	Gutt	0	
highbooks_home	0 til 10 bøker	1	0
	11 til 25 bøker	2	
	26 til 100 bøker	3	
	101 til 200 bøker	4	1
	Mer enn 200 bøker	5	
kinderg_att	Gikk i barnehage	1	
	Gikk ikke i barnehage	0	
teacher_fem	Kvinne	1	
	Mann	0	

### 3.3. Deskriptiv statistikk

Vi viser her frem samlet deskriptiv statistikk for de utvalgte landene. Dette gir en enkel forståelse av variablene vi har valgt. Senere vil vi få bruk for disse tallene når vi begynner med regresjonsanalyse.

*Minimum* er den laveste observasjonen og *maximum* er den høyeste observasjonen.

Først ser vi på *gjennomsnitt*, den viser summen av observasjonene dividert på antall observasjoner:

$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{1}{n} * \sum_{i=1}^n X_i$	n - antall observasjoner $\bar{X}$ - gjennomsnitt
---	--

*Utvalgsvarians* sier oss noe om variasjonen i observasjonene. Vi kan benytte oss av to varianser, *utvalgsvariansen* og *populasjonsvariansen*. Ved *populasjonsvarians* deler en kun på n. Ved *utvalgsvarians* deler en den kvadrerte summen av differansen mellom observasjoner og snitt med n-1, som vi gjør for å ikke underestimere variansen.

$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$	n-antall observasjoner $\bar{X}$ - gjennomsnitt $s^2$ - utvalgsvarians
--	--

For å finne *standardavviket* tar vi roten av *variansen*. Dvs at *utvalgsstandardavviket* er roten av *utvalgsvariansen*.

$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$	n - antall observasjoner $\bar{X}$ - gjennomsnitt s - utvalgsstandardavvik
---	--

### 3.3.1. Deskriptiv statistikk for variablene

Variabel	Obs	Mean	Std. Dev.	Min	Max	Var
<b>Read</b>	25 069	512.1988	77.9376	199.94	740.39	6074.27
<b>Sameteacher4_plus</b>	24 540	0.5037	0.4999	0	1	0.2499
<b>Clsize</b>	24 155	26.3393	9.5109	2	69	90.4572
<b>highpar_edu</b>	20 820	0.5075	0.4999	0	1	0.2499
<b>Girl</b>	24 844	0.4891	0.4999	0	1	0.2499
<b>highbooks_home</b>	23 275	0.4290	0.4949	0	1	0.2449
<b>Kinderg_att</b>	23 299	0,7760	0.4169	0	1	0.1738
<b>teacher_exp</b>	24 052	18.9990	10.8928	1	55	118.6553
<b>teacher_fem</b>	24 270	0.8206	0.3837	0	1	0.1472

Tabell 3.1: Deskriptiv statistikk for våre utvalgte variabler

#### **Read**

Read er en variabel som måler leseferdighetene hos elevene. Vi ser at den laveste scoren er på 199.94 og den høyeste er 740.39. Gjennomsnitt er på 512.1988, standardavviket er på 77.9376 og variansen er på 6074.26.

#### **Sameteacher4\_plus**

Sameteacher4\_plus er en dummyvariabel som viser om elevene har hatt samme lærer i fire år eller ikke. Gjennomsnittet er på 0.5037 som betyr at det er en jevn fordeling i nasjonene vi har valgt. Standardavviket er på 0.4999 og variansen på 0.2499 Det er gjort 24.540 observasjoner.

#### **Clsize**

Clsize er en variabel som måler hvor mange elever det er i hver klasse. Vi ser at den minste klassen som har deltatt i undersøkelsen er på to elever, og den klassen med flest er på 69 elever. Gjennomsnittet er på 26.33, standardavviket på 9.51 og variansen på 90.4572.

#### **Highpar\_edu**

Dette er en tidligere kategorivariabel som er gjort om til en dummyvariabel. Highpar\_edu er en dummyvariabel som viser foreldrenes utdanningsnivå. Gjennomsnittet er på 0.4215, standardavviket på 0.4938 og variansen på 0.2499.

#### **Girl**

Denne dummyvariablen forteller oss om eleven er gutt eller jente. Det er gjort 24 844 observasjoner. Totalt i undersøkelsen har vi 25 069 elever som har deltatt, og 24 844 har svart om de er gutt eller jente. Her ser vi et eksempel på mangel statistikk. Gjennomsnittet er på 0.4891, det er en ganske jevn fordeling av gutter og jenter i dette datasettet. Standardavviket er på 0.4999 og variansen på 0.2499.

### Highbooks\_home

Dette er en tidligere kategorivariabel som er gjort om til en dummyvariabel. Den sier noe om hvor mange bøker elevene har i husholdningen. Gjennomsnittet er på 0.4290, standardavviket på 0.4949 og variansen på 0.2449.

### Kinderg\_att

Dummyvariabelen `kinderg_att` gir oss en oversikt over hvorvidt eleven har deltatt i barnehage før skolestart eller ikke. Gjennomsnittet er på 0.8860 som betyr at flertallet har gått i barnehage. Standardavviket er på 0.4169 og variansen på 0.1738.

### Teacher\_exp

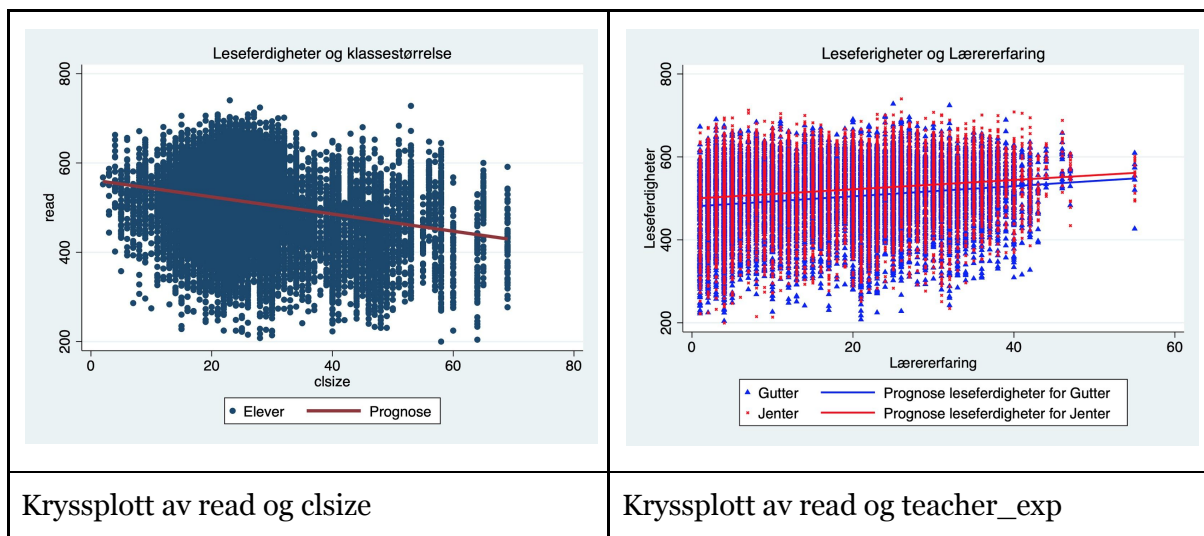
Denne variabelen forklarer hvor mange års erfaring læreren har. Gjennomsnittet er på 18.9990 år, standardavviket på 10.8928 og variansen på 118.6553.

### Teacher\_fem

Denne dummyvariabelen forteller oss hvilket kjønn læreren har. Gjennomsnittet er på 0.8206 som betyr at det er flertall av kvinnelige lærere i dette datasettet. Standardavviket er på 0.3837 og variansen på 0.1472.

## 3.3.2. Kryssplott

Under har vi to diagrammer for kryssplott av to variabler. Vi har valgt et kryssplott av variablene `read` og `clsiz`, og et av variablene `read` og `teacher_exp`. Vi har valgt disse kontrollvariablene grunnet tidligere forskning viser til at det er store sammenhenger mellom leseferdigheter og klassestørrelser, samt leseferdigheter og lærerens erfaringer (Norgård & Harsvik, 2017).





### 3.3.3. Kovarians, korrelasjon og korrelasjonsmatrise

#### Kovarians

*Kovarians* viser sammenhengen mellom tallene vi har i observasjonene. Den viser oss et estimat for den lineære avhengigheten mellom to variabler. For eksempel mellom leseferdigheter og samme lærer i over fire år.

$v_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$	n-antall observasjoner $\bar{X}$ - gjennomsnitt X $\bar{Y}$ - gjennomsnitt Y $v_{xy}$ - kovarians mellom X og Y
--	--

#### Korrelasjon

*Korrelasjon* er en statistisk sammenheng mellom to variabler. *Kovariansen* relativt til produktet av standardavvikene til de to variablene. *Korrelasjonen* vil alltid være et tall mellom (-1 og 1). Der 1 er perfekt korrelert, mens -1 er perfekt ukorrelert (Thomas 2005, s. 194).

$R = \frac{v_{xy}}{s_x s_y} = \frac{\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}} * \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$	
n-antall observasjoner $\bar{X}$ - gjennomsnitt X $\bar{Y}$ - gjennomsnitt Y	$v_{xy}$ - kovarians mellom X og Y R - korrelasjonen mellom X og Y $s_x, s_y$ - standardavviket til X og Y

## Korrelasjonsmatrise

Variabler	Sameteacher4_plus	Clsize	Highpar_edu	girl	Highbooks_home	Kinderg_att	teacher_exp
Clsize	-0.0503						
Highpar_edu	-0.0193	-0.0569					
Girl	-0.0016	0.0021	-0.0072				
Highbooks_home	-0.0380	-0.1976	0.2770	-0.0046			
Kinderg_att	0.0040	-0.3979	0.1833	0.0007	0.2968		
Teacher_exp	0.0381	-0.0783	0.1264	0.0085	0.0650	0.1521	
Teacher_fem	0.0233	-0.2667	0.0888	0.0078	0.1745	0.3061	0.0335

Tabell 3.2: Korrelasjonsmatrise for utvalgte variabler

Ut fra tallene i korrelasjonsmatrisen får vi en enkel ide om hvordan variablene korrelerer med hverandre. Siden flere av variablene vi har med er dummyvariabler, vil ikke matrisen gi noen stor mening. De eneste variablene som er med som gir mening, er korrelasjonen mellom variablene *teacher\_exp* og *clsize* som er  $-0.0783$ . Det er lav toveis kausalitet mellom våre variabler, derfor er det usikkert hvorvidt det finnes ukjente spuriøse sammenhenger.

### Kausalitet

Beskriver forholdet mellom årsaken til utfallet og virkningen. At to utfall korrelerer eller samvarierer betyr ikke nødvendigvis at det foreligger kausalitet. Toveis kausalitet betyr at to utfall påvirker hverandre. Eksempelvis, når vil X påvirke Y, vil Y påvirke X (Thomas 2005, s. 259).

### Spuriøs sammenheng

Vil si at det er en tredje faktor (faktor Z), som ligger bak (skjult) og er den egentlig årsaken til X og Y (Thomas 2005, s. 258).

## 3.4. Oppsummering

Vi har i dette kapittelet forklart datasettet vi anvender i denne oppgaven. Samtlige variabler som senere skal være med i regresjonsanalysen har vi definert. Vi har sett på deskriptiv statistikk for variablene, samt laget to krysslott-diagrammer. Til slutt så vi på hvordan variablene korrelerer med hverandre, og fant ut at det ikke var noen sterk sammenheng mellom noen av variablene i matrisen.

## 4. Økonometrisk modell

### 4.1. Innledning

I dette kapitlet skal vi presentere funksjonsform, empirisk strategi og komme med en kort forklaring på minste kvadraters metode (forkortet OLS på engelsk).

### 4.2. Valg av funksjonsform

Utgangspunkt. La elevprestasjonene måles ved testscore,  $T$

(1)  $T = f(S, C, F)$  er skole-produktfunksjonen.

Hvorav:

- $T$  = Test Score read
- $S$  = samme lærer fire år (dummy)
- $C$  = klassestørrelse
- $F$  = Skolefaktorer

#### 4.2.1. Lineær grunnmodell:

Velger en lineær approksimasjon av vårt utgangspunkt (1)

(2)  $read_i = \beta_1 + \beta_2 sameteacher\_4plus_i + \beta_3 clsize_i + \varepsilon_i$

Hvorav:

- $read_i$                       vårt mål på leseferdighet
- $sameteacher\_4plus_i$       vårt mål samme lærer fire år
- $clsize_i$                       vårt mål på klassestrørrelse

Med tilhørende koeffisientvektor  $\beta$  og stokastisk restledd  $\varepsilon$  som representerer alle andre variabler som påvirker  $read_i$ .

### 4.3. Empirisk strategi

Det kreves at vi kontrollerer for andre forhold enn klassestørrelse i denne multiple regresjonsmodellen, for at vi skal få troverdige anslag av effekten som "samme lærer i fire år" har på testscore "read".

Vi har fem andre variabler som kan representere  $F$  i den teoretiske modellen.

- Foreldrenes utdanning (dummy)
- Kjønn (dummy)
- Antall bøker eleven har tilgang til hjemme (dummy)
- Barnehage deltagelse (dummy)
- Lærerens erfaring
- Lærerens kjønn (dummy)

### 4.3.1. Estimeringsmetode

Vi er først og fremst interessert i estimering av parameteren  $\beta_2$  i den lineære grunnmodellen, og vi bruker Minste kvadraters metode (eller OLS) som estimeringsmetode.

Vi antar at:

$X$  ( $sameteacher\_4plus_i$  og  $clsize_i$ ) består av én variabel, slik at vår multiplere regresjonsmodell består av to forklaringsvariabler.

Vi forutsetter at:

Betinget forventning til stokastisk restledd  $\varepsilon$  er lik null. Dette er fordi  $X$  er ukorrelet med forklaringsvariabelen i vår lineære grunnmodell (2).

OLS-estimatorene til  $\beta_1$  og  $\beta_2$  er forventningsrette, variansminimale og oppfyller standard-forutsetningene.

Forventer at det stokastisk restledd  $\varepsilon$  er normalfordelt slik at vi kan benytte t- og f-tester når vi tester hypoteser om parametrene i vår lineære grunnmodell (2).

### 4.3.2. Modeller

Under er (antall) modeller, (A)-(E):

$$(A) \text{ read}_i = \beta_1 + \beta_2 \text{sameteacher4\_plus}_i + \varepsilon_i$$

$$(B) \text{ read}_i = \beta_1 + \beta_2 \text{sameteacher4\_plus}_i + \beta_3 \text{clsize}_i + \beta_4 \text{highpar\_edu}_i + \varepsilon_i$$

$$(C) \text{ read}_i = \beta_1 + \beta_2 \text{sameteacher\_4plus}_i + \beta_3 \text{clsize}_i + \beta_4 \text{highpar\_edu}_i + \beta_5 \text{girl}_i +$$

$$\beta_6 \text{highbooks\_home}_i + \beta_7 \text{kinder}_i + \beta_8 \text{teacher\_exp}_i + \beta_9 \text{teacher\_fem}_i$$

Hvorav:

- $clsize_i$  vårt mål på klassestrørrelse
- $highpar\_edu_i$  vårt mål på foreldrenes utdanning (dummy)
- $girl_i$  vårt mål på elevens kjønn (dummy)
- $highbooks\_home_i$  vårt mål på antall bøker eleven har tilgang til hjemme (dummy)
- $kinder\_att_i$  vårt mål på barnehagedeltagelse (dummy)
- $teacher\_exp_i$  vårt mål på lærerens erfaring
- $teacher\_fem_i$  vårt mål på lærerens kjønn (dummy)

Vi gjennomfører en empirisk analyse for alle modeller, fordelt på tre forskjellige måter:

<b>Fordeling 1 (Samlet)</b>	<b>Fordeling 2 (EU)</b>	<b>Fordeling 3 (ikke EU)</b>
Tyskland Ungarn Norge Russland Tyrkia	Tyskland Ungarn Norge* (EØS)	Russland Tyrkia

Tabell 4.1: Oversikt over våre tre forskjellige fordelinger

\* Vi tar Norge med i EU-fordelingen på grunn av tette bånd med EU gjennom EØS-avtalen.

### 4.3.3. Forskjeller mellom modellvariantene

- (A) Grunnmodell. I denne modellen undersøker vi kun leseferdighet mellom elever med og uten samme lærer i fire år eller mer. Det vil si at vi i denne modellen estimerer kun forskjellene mellom de to.
- (B) Kontrollerer for klassestørrelse og foreldrene utdanningsnivå
- (C) Kontrollerer for klassestørrelse, foreldrenes utdanningsnivå, elevenes kjønn, antall bøker hjemme, barnehage deltagelse, lærerens erfaring og lærerens kjønn. Dette er for å korrigere for mulig utelatte variabler-problemer p.g.a multikollinearitet.

## 4.4. Analyseopplegg

### Minste kvadraters metode

Minste kvadraters metode (MKM/OLS) er en estimeringsmetode som går ut på å finne en løsning der spriket mellom observasjonene og løsningen er minst mulig. Vi vil finne en sammenheng mellom en avhengig variabel og en eller flere forklaringsvariabler. Vi bruker denne metoden for å estimere en fremtidig verdi for en variabel. Den bestemmer en lineær graf som passer alle punktene for lesehastigheten. Minste kvadraters metode gir oss linja som minimerer summen av det kvadrerte avviket (Thomas 2005, s. 266 - 270).

Forutsetninger for minste kvadraters metode estimatoren:

1. Restleddet har forventning lik
2. Alle forklaringsvariablene er ukorrelerte med restleddet
3. Restleddene er ukorrelerte med hverandre målt over tid
4. Restleddet har konstant varians
5. Restleddet er normalfordelt

### Multikollinearitet

Graden av lineær sammenheng i modellens forklaringsvariabel. Først beskrevet av den norske økonomen Ragnar Frisch i 1934. Perfekt multikollinearitet forekommer når korrelasjonen er 1 eller -1. Ved høy korrelasjon mellom de uavhengige variablene, vil det være vanskelig å isolere effekten av hver enkelt variabel, noe som igjen kan forårsake høyst usikre estimater. (Thomas 2005, s 402-415)

## 5. Empiriske resultater

### 5.1. Innledning

I dette kapitlet presenteres resultatene fra den empiriske analysen. Regresjonsanalysen av modell A, B og C, tidligere definert. Samt forskjeller mellom EU-land (Tyskland, Ungarn og Norge) og ikke EU-land (Tyrkia og Russland). Analysen er utført ved hjelp av den statistiske programvaren STATA. Vi vil undersøke hvorvidt resultatene fra regresjonsanalysene er statistisk signifikant, og tolkningen av dem.

#### Signifikansnivå $\alpha$

Signifikansnivå  $\alpha$  (alfa) forteller hva som skal til for å avvise en sann nullhypotese, altså en Type I-feil. En Type II-feil får du ved å akseptere en usann nullhypotese, og har notasjon  $\beta$  (beta).  $\alpha$  og  $\beta$  har et inverst forhold, altså dersom  $\alpha$  synker vil  $\beta$  øke. (Thomas 2005, s.129)

#### T-test

Er en hvilken som helst statistisk hypotesetest der teststatistikken følger studentens t-distribusjon under nullhypotesen. Den brukes for å sammenligne begge utvalgenes gjennomsnittsverdier, og om det er signifikant forskjell mellom disse. Testen brukes også for å sjekke om stigningstallet til en regresjonslinje er signifikant forskjellig fra null. (Thomas 2005, s. 157 - 160).

#### R-squared

R-squared (eller  $R^2$ ) er determinasjonskoeffisienten (coefficient of determination). Den beskriver modellens sin andel av variasjon i den avhengige variabelen som kan forklares ut fra de uavhengige variablene, der  $0 < R^2 < 1$ . Det vil si at jo nærmere  $R^2$  er lik 1, jo bedre er forklaringskraften til modellen, og jo nærmere  $R^2$  er lik 0 så er det lavere forklaringskraft til modellen. (Thomas 2005, s 273).  $R^2$  er en viktig størrelse som forklarer hvor godt modellene passer til datasettet. Jo flere variabler som inngår i modellen, desto høyere forklaringskraft vil modellen ha. Hadde vi hatt mange flere variabler med i undersøkelsen kunne det vært mere relevant å bruke adjusted  $R^2$ , siden det tar hensyn til mange variabler. Dette er ikke tilfellet i denne undersøkelsen og vi forholder oss til den originale  $R^2$ .

## 5.2. Empiriske hovedresultater

### 5.2.1. Fordeling 1 (samlet)

	(A)		(B)		(C)	
VARIABLES	read		read		read	
sameteacher4_plus	-0.783	(0.997)	-2.800***	(0.989)	-0.147	(0.974)
clsiz			-1.928***	(0.0517)	-0.398***	(0.0557)
highpar_edu					26.02***	(1.020)
girl					15.99***	(0.971)
highbooks_home					37.85***	(1.268)
kinderg_att					33.77***	(1.365)
teacher_exp					0.493***	(0.0465)
teacher_fem					18.08***	(1.368)
Constant	512.0***	(0.707)	563.8***	(1.558)	426.4***	(2.666)
Observations	24,540		23,800		19,209	
R-squared	0.000		0.055		0.248	

Standardavvik i parentes, \*\*\*  $\alpha < 0.01$  (1%), \*\*  $\alpha < 0.05$  (5%), \*  $\alpha < 0.1$  (10%)

#### 5.2.1.1. Modell A

- $\beta_2$  - Å beholde samme lærer i fire år fører til en lavere leseferdighet på 0.783

#### 5.2.1.2. Modell B

- $\beta_2$  - Å beholde samme lærer i fire år fører til en lavere leseferdighet på 2.800
- $\beta_3$  - 1 standardavviks høyere klassestørrelse minker leseferdigheter med 1.928

#### 5.2.1.3. Modell C

- $\beta_2$  - Å beholde samme lærer i fire år fører til en lavere leseferdighet på 0.147
- $\beta_3$  - 1 standardavviks høyere klassestørrelse minker leseferdigheter med 0.398
- $\beta_4$  - Elever med høyt utdannede foreldre har en økt leseferdighet på 26.02
- $\beta_5$  - At eleven er jente, vil gi en økning i leseferdigheter på 15.99
- $\beta_6$  - Det å ha over 100 bøker i hjemmet, vil øke leseferdigheter med 37.85
- $\beta_7$  - At elever har gått i barnehage, vil gi en økning i leseferdigheter på 33.77

- $\beta_8$  - Å ha en lærer med erfaring, vil gi en økning i leseferdigheter på 0.493 per år erfaring frem til den minker
- $\beta_9$  - Å ha kvinnelig lærer, vil gi en økning i leseferdighet på 18.08

Tabellen viser estimerte forskjeller på leseferdigheter fra elever som har hatt samme lærer i fire år. Vi ser her at det er forskjeller i modellene A, B og C. Dette kommer av at det er ulike sett av kontrollvariabler i de forskjellige modellene. Beta-koeffisienten foran *sameteacher4\_plus* sier at hvis denne variabelen er lik 1 så minker leseferdighetene med 0.783 i modell A. Modell A er en bivariat regresjonsanalyse siden den bare inneholder to variabler, *read* og *sameteacher4\_plus*. Når vi øker variabler i de resterende modellene kalles det for en multivariat regresjonsanalyse. Den estimerte forskjellen i *sameteacher4\_plus* varierer fra -2.800 til -0.128. Leseferdighetene på samme lærer i fire år varierer mest i modell B. I denne modellen inngår kontrollvariabelen *clsiz*, klassestørrelse. Størrelsen på klassen har en negativ effekt på leseferdigheter når klassen har hatt samme lærer i fire år. Det vil si at hvis det er en ekstra person i klassen så minker leseferdighetene med 1.928 i modell B. I modell C har vi den minste forskjellen på *sameteacher4\_plus*. Her har de nye kontrollvariablene *highpar\_edu*, *girl*, *highbooks\_home*, *kinderg\_att*, *teacher\_exp* og *teacher\_fem* en positiv effekt på leseferdighetene.

I tabellen ser vi at bak flere av koeffisientene så er det merket med \*, \*\* eller \*\*\*. Det indikerer om variabelen er signifikant på et 1%-, 5%-, og 10% signifikansnivå. Vi tar videre utgangspunkt i fem prosent signifikansnivå. T-tester på variabelen *sameteacher4\_plus* på modell A ikke er statistisk signifikant med et signifikansnivå på fem prosent. I modell B er den statistisk signifikant, men det er den ikke i modell C. Alle de resterende kontrollvariablene i modell C er signifikante.

Vi ser fra denne tabellen at det å ha samme lærer i fire år har en minimal negativ effekt på leseferdigheter fra samtlige modeller. Ut i fra dette datasettet er det en fordel å ikke ha samme lærer i fire år, men siden variablene ikke er signifikante kan resultatet være en tilfeldighet.

Konstanten tilsvarer det vi har betegnet som  $\beta_1$  i likning 2. Modell A har 0.000 i forklaringskraft, det vil si at det finnes ingen lineær sammenheng mellom variablene *read* og *sameteacher4\_plus*. *Sameteacher4\_plus* forklarer 0.00 prosent av variansen i modell A. Modell C har høyest forklaringskraft på 0.250. Dette kommer av at denne modellen har flest variabler.

### Heteroskedastisitet

I modell C tester vi for heteroskedastisitet. Det betyr at variansen til restleddet ikke er konstant, men at den varierer med størrelsen på variablene. Dette kan føre til at estimert standardavvik og varians for variablene blir feil, som igjen kan føre til t-testen får feil konklusjon. (Thomas 2005, s.479 - 482).

Ettersom vi først og fremst har dummyvariabler, vil heteroskedastisitet være et trekk ved slike variabler, siden:  $VAR(X) = p(1 - p) = p - p^2$

Som er en kvadratisk funksjon av:  $p = P(X = A)$



Kan du forvente lignende avvik i to grupper bare for tilfeller når:  $p_a = 1 - p_b$

$$VAR(X_a) = VAR(X_b)$$

$$p_a(1 - p_a) = p_b(1 - p_b) = [1 - p_a](1 - [1 - p_a]) = (1 - p_a)p_a$$

(Wilhelm, 2015)

Derfor gjør vi ikke noe med den klare antydningen til heteroskedastisitet som testen viser.

## 5.2.2. Fordeling 2 (EU)

	(A)		(B)		(C)	
VARIABLES	read		read		read	
sameteacher4_plus	-4.799***	(1.118)	-2.788**	(1.145)	-0.363	(1.231)
clsize			1.782***	(0.121)	1.544***	(0.124)
highpar_edu					28.14***	(1.243)
girl					15.57***	(1.218)
highbooks_home					38.15***	(1.871)
kinderg_att					33.32***	(2.809)
teacher_exp					0.195***	(0.0567)
teacher_fem					8.035***	(1.930)
Constant	530.0***	(0.801)	488.6***	(2.924)	399.9***	(4.627)
Observations	15,354		14,657		11,061	
R-squared	0.001		0.016		0.136	

Standardavvik i parentes, \*\*\*  $\alpha < 0.01$  (1%), \*\*  $\alpha < 0.05$  (5%), \*  $\alpha < 0.1$  (10%)

## 5.2.2.1. Modell A

- $\beta_2$  - Å beholde samme lærer i fire år fører til en lavere leseferdighet på 4.799

## 5.2.2.2. Modell B

- $\beta_2$  - Å beholde samme lærer i fire år fører til en lavere leseferdighet på 2.788
- $\beta_3$  - 1 standardavviks høyere klassestørrelse øker leseferdigheter med 1.782

## 5.2.2.3. Modell C

- $\beta_2$  - Å beholde samme lærer i fire år fører til en lavere leseferdighet på 0.363
- $\beta_3$  - 1 standardavviks høyere klassestørrelse øker leseferdigheter med 1.544
- $\beta_4$  - Elever med høyt utdannede foreldre har en økt leseferdighet på 28.14
- $\beta_5$  - At eleven er jente, vil gi en økning i leseferdigheter på 15.57
- $\beta_6$  - Å ha over 100 bøker i hjemmet, vil øke leseferdigheter med 38.15
- $\beta_7$  - At elever har gått i barnehage, vil gi en økning i leseferdigheter på 33.32
- $\beta_8$  - Å ha en lærer med erfaring, vil gi en økning i leseferdigheter på 0.195 per år erfaring frem til den minker

- $\beta_9$  - Å ha kvinnelig lærer, vil gi en økning i leseferdighet på 8.035

Denne tabellen viser estimerte forskjeller på leseferdigheter fra elever som har hatt samme lærer i fire år innad i utvalgte EU land. Den estimerte forskjellen varierer fra -4.799 til -0.363 i variabelen *sameteacher4\_plus*. I motsetning til datasettet for samlede land der det er størst variasjon i modell B, så er det størst variasjon i modell A for EU landene. I denne modellen inngår kun den avhengige variabelen *read* og interessevariabelen *sameteacher4\_plus*. I modell C har vi den minste forskjellen på *sameteacher4\_plus*. Her har samtlige variabler bortsett fra *sameteacher4\_plus* en positiv effekt på leseferdighetene.

Ved å gjennomføre en t-test finner vi ut at variabelen *sameteacher4\_plus* er signifikant på et 5% signifikansnivå i modell A og B, men ikke i modell C. Samtlige resterende kontrollvariabler er signifikante.

Likt som tabellen for alle nasjonene, så er det minimale negative forskjeller på variabelen *sameteacher4\_plus* for samtlige modeller. Det vil si at det å ha samme lærer i fire år vil ha en liten negativ effekt på leseferdighetene til elevene i EU landene.

Konstanten tilsvarer det vi har betegnet som  $\beta_1$  i likning 2. Forklaringskraften  $R^2$  er i modell A 0.001, det vil si den er minimalt større enn i modell A for samlede land. Denne modellen forklarer dermed datasettet til en veldig liten grad. Forklaringskraften i modell B er lavere enn for samlede land. Modell C har den høyeste forklaringskraften med 0.136. Som nevnt tidligere så vil modeller med flere variabler ha høyere forklaringskraft.

### 5.2.3. Fordeling 3 (Ikke EU)

	(A)	(B)	(C)
VARIABLES	read	read	read
sameteacher4_plus	3.151* (1.761)	0.994 (1.701)	1.926 (1.555)
clsize		-1.874*** (0.0694)	-0.647*** (0.0685)
highpar_edu			26.42*** (1.807)
girl			15.28*** (1.544)
highbooks_home			29.47*** (1.847)
kinderg_att			21.17*** (1.839)
teacher_exp			1.009*** (0.0784)
teacher_fem			24.98*** (1.954)
Constant	483.5*** (1.230)	544.8*** (2.548)	424.9*** (3.513)
Observations	9,186	9,143	8,148
R-squared	0.000	0.074	0.285

Standardavvik i parentes, \*\*\*  $\alpha < 0.01$  (1%), \*\*  $\alpha < 0.05$  (5%), \*  $\alpha < 0.1$  (10%)

#### 5.2.3.1. Modell A

- $\beta_2$  - Å beholde samme lærer i fire år fører til en økt leseferdighet på 3.157

#### 5.2.3.2. Modell B

- $\beta_2$  - Å beholde samme lærer i fire år fører til en økt leseferdighet på 0.994
- $\beta_3$  - 1 standardavviks høyere klassestørrelse minker leseferdigheter med 1.874

#### 5.2.3.3. Modell C

- $\beta_2$  - Å beholde samme lærer i fire år, fører til en høyere leseferdighet på 1.926
- $\beta_3$  - 1 standardavviks høyere klassestørrelse minker leseferdigheter med -0.647
- $\beta_4$  - Elever med høyt utdannede foreldre, har en økt leseferdighet på 26.42
- $\beta_5$  - At eleven er jente, vil gi en økning i leseferdigheter på 15.57
- $\beta_6$  - Å ha over 100 bøker i hjemmet, vil øke leseferdigheter med 38.15
- $\beta_7$  - At elever har gått i barnehage, vil gi en økning i leseferdigheter på 33.32
- $\beta_8$  - Å ha en lærer med erfaring, vil gi en økning i leseferdigheter på 0.195 per år erfaring frem til den minker

- $\beta_9$  - Å ha kvinnelig lærer, vil gi en økning i leseferdighet på 8.035

Denne tabellen viser estimerte forskjeller på leseferdigheter fra elever som har hatt samme lærer i fire år innad i utvalgte land som ikke er medlem av EU. Tabellen viser estimerte forskjeller på leseferdigheter fra elever som har hatt samme lærer i fire år. Den estimerte forskjellen varierer fra 3.151 til 0.994 i variabelen *sameteacher4\_plus*. I motsetning til tabellene for samlede land og EU land så har vi her positive tall. Dette forteller oss at modellene vi bruker i dette datasettet gir en positiv effekt av å ha samme lærer i fire år på leseferdigheter. Leseferdighetene på elever som har hatt samme lærer i fire år varierer mest i modell A. I denne modellen inngår kun den avhengige variabelen *read* og interessevariabelen *sameteacher4\_plus*. I modell B har vi den minste forskjellen på *sameteacher4\_plus*. I modell C så har *sameteacher4\_plus* en positiv effekt med 1.926 på leseferdigheter.

Ved å gjennomføre en t-test ser vi at variabelen *sameteacher4\_plus* ikke er signifikant i modell A, modell B og modell C på et fem prosent signifikansnivå. Det betyr at resultatene fra tabellen kan være tilfeldig.

Likt som tabellene for alle nasjonene og EU-landene, så er forskjellene minimale for variabelen *sameteacher\_4plus*. Forskjellen fra de tidligere tabellene er at den minimale forskjellen her er positiv for alle modellene.

Konstanten tilsvare det vi har betegnet som  $\beta_1$  i likning 2. Forklaringskraften  $R^2$  er i modell A 0.000, likt som for samlede nasjoner. det vil si den er minimalt mindre enn i modell A for samlede land. Denne modellen forklarer dermed datasettet til ingen grad. Forklaringskraften i modell B er større enn i for samlede land og EU-landene, men forklaringskraften er fortsatt lav for samtlige modeller. Modell C i ikke-EU-landene har den høyeste forklaringskraften av samtlige modeller i alle tabellene med 0.285. Som nevnt tidligere vil modeller med flere variabler ha høyere forklaringskraft.

#### 5.2.4. Sammenligning av EU land og ikke EU land

I problemstillingen har vi blant annet ønsket å sammenligne utvalgte EU-nasjoner med utvalgte ikke-EU-nasjoner. Vi ville se på forskjeller i leseferdigheter der elever har hatt samme lærer i fire år i EU-land og i ikke EU-land. Vi ser at i samtlige av de tre modellene vi har fra regresjonsanalysen av EU-land at det er svakt negativt å ha samme lærer i fire år. Fra regresjonsanalysen av ikke EU-land at det er en positiv effekt i alle modellene å ha samme lærer i fire år. Hvis vi sammenligner modell C fra EU-land og ikke-EU-land, så er det differansen på det å ha samme lærer i fire år på 2.289 i fordel ikke EU-land. Disse forskjellene på variabelen *sameteacher4\_plus* er så minimale at det kanskje ikke har så mye for seg. Denne variabelen er som nevnt tidligere ikke statistisk signifikant, som vil si at resultatene kan være tilfeldige.

### 5.3. Oppsummering

I dette kapitlet har vi undersøkt om det å beholde samme lærer påvirker leseferdighetene til elever i undersøkelsen PIRLS gjennomført i 2001.

De tre forskjellige modellene (A, B, C) og de tre fordelingen (samlet, EU, ikke EU) gir alle et tydelig grunnlag for å si at å ha samme lærer i fire år eller mer ikke har innvirkning på elevenes leseferdigheter.

I tillegg til dette ser vi ut ifra analysen av modell C, at foreldres utdanning, elevenes kjønn, antall bøker hjemme, barnehage og lærerens kjønn, har betydning for leseferdighet.

## 6. Avslutning

### 6.1. Generell oppsummering

I denne bacheloroppgaven har vi gjennomført en analyse av leseferdigheter for elever på fjerde trinn, som undervises av samme lærer i fire år eller mer, i fem forskjellige land.

Det er ikke blitt gjennomført mye forskning på dette området, og den forskningen som er utført, varierer i gjennomførelse, datamateriale og tidsramme. Mye av den tidligere forskningen tyder på at et økt elev-lærer-forhold har ulike fordeler. Disse fordelene varierer veldig og er spekulative.

Det er grunn til å tro at det er veldig naturlig for mange lærere å føle at de blir flinkere til å undervise elever dersom læreren får anledning til å bli godt kjent med elevene.

I kapittel 1 definerte vi vår problemstilling: «*Hva er det potensielle utbyttet av å undervises av samme lærer i fire år eller mer?*» med to tilleggsspørsmål «*Vil elever som bytter lærer oftere ha svakere elevprestasjoner enn de som beholder samme lærer over mer enn fire år?* Og hvilke forskjeller er det mellom utvalgte EU-land og utvalgte ikke-EU-land?»

I kapittel 2 utforsker vi relevant forskning på området og ser på skolesystemet i de utvalgte landene. Landene ble valgt ut fordi de hadde en jevnere fordeling av elever som underviser av samme lærer i fire år og ikke, enn de andre landene i vårt datasett.

I kapittel 3 presenterte vi vårt datamateriale fra PIRLS 2001, og definerte variablene vi bruker i våre økonometriske modeller.

I kapittel 4 definerte vi tre modeller, A, B og C. Modell A isolerer problemstillingen vår. Modell B inkluderer variabelen klassestørrelse, mens modell C inkluderer sju andre variabler for å motvirke multikollinearitet.

I kapittel 5 la vi frem empiriske resultater av regresjonsanalysen for de tre forskjellige fordelingene vi opererer med. Vi gjennomførte t-tester for alle variablene og testet for heteroskedastisitet.

## 6.2. Konklusjon

I denne oppgaven har vi gjennomført en analyse basert på data fra PIRLS undersøkelsen i 2001. Vi har benyttet data fra fem av 35 land: Tyrkia, Ungarn, Tyskland, Norge og Russland.

Våre resultater viser at det ikke er forskjell i leseferdigheter blant elever som har og ikke har hatt samme lærer i fire år. Forskjellene mellom leseferdigheter er ikke signifikante, og ved vår beste modell (modell C), er forskjellen innenfor standardavviket i alle fordelingene.

Modell C har høyest forklaringskraft i alle tabellene, som betyr at denne modellen gir størst forklaringsgrad. Variabelen *sameteacher4\_plus* er ikke signifikante i noen av de modellene, og vi kan konkludere med at resultatet kan være tilfeldig. Dette vil si at vi verken kan si det er en fordel eller en ulempe å ha samme lærer i fire år.

Vi har ikke funnet tidligere forskning som har brukt tilsvarende data. Forskning med annet datamateriale har sett tendenser til at økt elev-lærer-forhold på grunn av at samme lærer underviser elever i to år eller flere, har en signifikant positiv virkning.

Det er lite eller ingen forskjell mellom de tre forskjellige modellene med tanke på problemstillingen. Vi kan tydelig se ut fra modell C at variabler som potensielt har stor betydning på leseferdighet er om eleven har bøker i hjemmet, har gått i barnehage, har høyt utdannede foreldre og er jente.

Ut ifra analysen vi gjennomførte i kapittel 5 har vi ikke noe grunnlag for å konkludere om det er en fordel eller ulempe å ha samme lærer i fire år. Mest sannsynlig har dette lite eller ingen virkning.

## 6.3. Styrker og svakheter

Det er flere styrker og svakheter ved datamaterialet vi har brukt og vår analyse. Dataene våre er 19 år gamle, og selv om mye utdanning gjennomføres på samme måte som i 2001, har det i løpet av disse årene foregått en digital modernisering av skolen. Det kan ha ført til at undersøkelsen i sin helhet ikke lenger er like relevant for dagens undervisningssituasjon. Det vi har undersøkt kan tenkes å være mindre påvirket enn andre faktorer, men det er ikke sikkert.

I empiriske undersøkelser er det et mål å finne kausale effekter. Det er svært mange faktorer som potensielt kan påvirke leseferdigheten til elever. Disse faktorene kan være både vanskelig å identifisere, kvantifisere og måle.

PIRLS undersøkelsen fra 2001 er et øyeblikksbilde av situasjonen. Vi kan ikke si noe om utviklingen i leseferdigheten til elevene over lengre tid. Vi vet ikke om elever som har hatt samme lærer i fire år eller mer påvirkes negativt eller positivt av det i videre skolegang, eller hva som var utgangspunktet.

Det er ikke sikkert at de variablene vi er benyttet, er blitt målt på likt grunnlag og lik måte. Forskjeller i både språk, kultur og politikk kan ha hatt en stor innvirkning på testresultatene. Selv om forskerne som har gjennomført PIRLS har gjort mange gode grep for å forsikre seg om at prøven er mest mulig lik i alle landene, er det høyst usikkert om de har lyktes 100 prosent. Russland er for eksempel det eneste land som ikke har blanke svar i

dummyvariabelen *Sameteacher4\_pluss*. Akkurat hvorfor de ikke har noen mangler i denne variabelen er vanskelig å vite, men det gir grunnlag for å være skeptisk.

En klar styrke ved analysen er at den inneholder mange observasjoner. Dette bidrar til å øke forklaringskraften for Modell C betraktelig.

#### **6.4. Videre forskning**

I denne analysen har vi bare sett på elever med samme lærer i fire år eller mer. Det mulig det finnes en positiv korrelasjon på leseferdigheter, dersom elever har samme lærer mellom ett og fire år. Det kunne vært interessant å sett mer på denne virkningen. For å kunne bedre kartlegge virkningen av å ha samme lærer over flere år burde en gjennomføre en undersøkelse både før og etter. Ved å se forandringer i leseferdighet over tid vil en kunne få mer enn bare et øyeblikksbilde av leseferdigheten til elevene.



## 7. Litteraturliste

Bogart, V. S. (2002). *The Effects of Looping on the Academic Achievement of Elementary School Students* (Doktoravhandling, East Tennessee State University) Hentet: 10.04.2020, fra: <https://dc.etsu.edu/cgi/viewcontent.cgi?article=1864&context=etd>

Briseid, L. G. (2019). Det nasjonale kvalitetsvurderingssystemet – en trussel mot skolen dannelsoppdrag. Hentet: 22.03.2020, fra: <https://utdanningsforskning.no/artikler/det-nasjonale-kvalitetsvurderingssystemet--en-trussel-mot-skolens-dannelsoppdrag/>

Hill, A. J., & Jones, D. B. (2018). A teacher who knows me: The academic benefits of repeat student-teacher matches. *Economics of Education Review*. Hentet: 01.05.2020, fra: <https://www.sciencedirect.com/science/article/pii/S0272775717306635>

Norgård, J. D., & Harsvik, T. (2017). Klassestørrelse og læringsutbytte - hva viser forskningen? Hentet: 22.04.2020, fra: <https://www.utdanningsforbundet.no/var-politikk/kunnskapsgrunnlag/publikasjoner/2011/klassestorrelse-og-laringsutbytte--hva-viser-forskningen/>

Papirleksikonet Store norske leksikon, (2013). *Skole og utdanning i Russland*. Hentet: 26.03.2020, fra: [https://snl.no/Skole\\_og\\_utdanning\\_i\\_Russland](https://snl.no/Skole_og_utdanning_i_Russland)

Pedersen, J. W. (2018). *Skole og utdanning i Tyskland*. Hentet: 26.03.2020, fra: [https://snl.no/Skole\\_og\\_utdanning\\_i\\_Tyskland](https://snl.no/Skole_og_utdanning_i_Tyskland)

Thomas, L. (2005): *Using Statistics in Economics*. Berkshire: McGraw Hill

Thune, T. Reisegg Ø. Askheim S. (2019). *Skole og utdanning i Norge*. Hentet: 19.03.2020, fra: [https://snl.no/Skole\\_og\\_utdanning\\_i\\_Norge](https://snl.no/Skole_og_utdanning_i_Norge)

Solheim, S. G. & Tønnessen, F. E. (2003). PIRLS Hvorfor leser klasser så forskjellig? Hentet: 27.04.2020, fra: [https://lesesenteret.uis.no/getfile.php/13108401/Lesesenteret/PIRLS\\_Hvorfor%20leser%20klasser%20s%C3%A5%20forskjellog.pdf](https://lesesenteret.uis.no/getfile.php/13108401/Lesesenteret/PIRLS_Hvorfor%20leser%20klasser%20s%C3%A5%20forskjellog.pdf)

Welle-Strand, A. (2009). *Skole og utdanning i Tyrkia*. Hentet: 23.03.2020, fra: [https://snl.no/Skole\\_og\\_utdanning\\_i\\_Tyrkia](https://snl.no/Skole_og_utdanning_i_Tyrkia)

Welle-Strand, A. (2013). *Skole og utdanning i Ungarn*. Hentet: 23.03.2020, fra: [https://snl.no/Skole\\_og\\_utdanning\\_i\\_Ungarn](https://snl.no/Skole_og_utdanning_i_Ungarn)

Wiborg, Ø., Arnesen, C. Å., Grøgaard, J. B., Støren, L. A. & Opheim, V. (2011). Elevers prestasjonsutvikling – hvor mye betyr skolen og familien? Hentet: 11.05.2020, fra: <https://www.udir.no/globalassets/filer/tall-og-forskning/forskningsrapporter/elevers-prestasjonsutvikling--hvor-mye-betyr-skolen-og-familien.pdf>

Wilhelm, Jochen. (2015). Re: Which test of Heteroscedasticity is best when we have binary dummy dependent variable?. Hentet: 09.05.2020, fra:

[https://www.researchgate.net/post/Which\\_test\\_of\\_Heteroscedasticity\\_is\\_best\\_when\\_we\\_have\\_binary\\_dummy\\_dependent\\_variable/555d9a435e9d97aa4f8b457d/citation/download](https://www.researchgate.net/post/Which_test_of_Heteroscedasticity_is_best_when_we_have_binary_dummy_dependent_variable/555d9a435e9d97aa4f8b457d/citation/download)

## 8. Vedlegg

### 8.1. STATA Do-file

Kopi av do-fil brukt for å gjennomføre analysen i STATA.

\*\*\*Fordeling 1 (Samlet)\*\*\*

\*Åpne datasett

```
Use "data2901-Germany.dta"
append using "data2901-Hungary"
append using "data2901-Norway"
append using "data2901-Russia"
append using "data2901-Turkey"
```

\*Lager kryssplott for leseferdigheter og lærererfaring

```
twoway (scatter read teacher_exp if girl==0, mcolor(blue)) (lfit read teacher_exp if girl==0) (scatter
read teacher_exp if girl==1, mcolor(red)) (lfit read teacher_exp if girl==1)
```

\*Lager kryssplott for leseferdigheter og clsize

```
twoway (scatter read clsize) (lfit read clsize)
```

\*Lager dummyvariabel for antall bøker

```
gen highbooks_home=1
replace highbooks_home=. if books_home==.
replace highbooks_home=. if books_home==.
replace highbooks_home=0 if books_home==1
replace highbooks_home=0 if books_home==2
```

\* Modell A

```
reg read sameteacher4_plus
test sameteacher4_plus = 0
est store m_A
```

\* Modell B

```
reg read sameteacher4_plus clsize
test sameteacher4_plus = 0
test clsize = 0
est store m_B
```

estat hetttest

\* Modell C

```
reg read sameteacher4_plus clsize highpar_edu girl highbooks_home kinderg_att teacher_exp
teacher_fem
test sameteacher4_plus = 0
test clsize = 0
test highpar_edu = 0
test girl = 0
test highbooks_home = 0
test kinderg_att = 0
test teacher_exp = 0
test teacher_fem = 0
est store m_C
```

\* Tabell

```
outreg2 [m_A m_B m_C] using modell_ABC_1.doc, replace
```

shellout using "modell\_ABC\_1.doc"

\*\*\*Fordeling 2 (EU)\*\*\*

\*Åpne datasett

Use "data2901-Germany.dta"

append using "data2901-Hungary"

append using "data2901-Norway"

\*Lager dummyvariabel for foreldres utdanning

gen highpar\_edu=1

replace highpar\_edu=. if par\_edu==.

replace highpar\_edu=0 if par\_edu==3

replace highpar\_edu=0 if par\_edu==4

replace highpar\_edu=0 if par\_edu==5

\*Lager dummyvariabel for antall bøker

gen highbooks\_home=1

replace highbooks\_home=. if books\_home==.

replace highbooks\_home=0 if books\_home==1

replace highbooks\_home=0 if books\_home==2

\* Modell A

reg read sameteacher4\_plus

test sameteacher4\_plus = 0

est store m\_A

\* Modell B

reg read sameteacher4\_plus clsiz

test sameteacher4\_plus = 0

test clsiz = 0

est store m\_B

estat hetttest

\* Modell C

reg read sameteacher4\_plus clsiz highpar\_edu girl highbooks\_home kinderg\_att teacher\_exp

teacher\_fem

test sameteacher4\_plus = 0

test clsiz = 0

test highpar\_edu = 0

test girl = 0

test highbooks\_home = 0

test kinderg\_att = 0

test teacher\_exp = 0

test teacher\_fem = 0

est store m\_C

\* Tabell

outreg2 [m\_A m\_B m\_C] using modell\_ABC\_eu.doc, replace

shellout using "modell\_ABC\_eu.doc"

\*\*\*Fordeling 3 (Ikke EU)\*\*\*

\*Åpne datasett

Use "data2901-Russia.dta"

append using "data2901-Turkey"

\*Lager dummyvariabel for foreldres utdanning

gen highpar\_edu=1

replace highpar\_edu=. if par\_edu==.

replace highpar\_edu=0 if par\_edu==3

replace highpar\_edu=0 if par\_edu==4

replace highpar\_edu=0 if par\_edu==5

\*Lager dummyvariabel for antall bøker

gen highbooks\_home=1

replace highbooks\_home=. if books\_home==.

replace highbooks\_home=0 if books\_home==1

replace highbooks\_home=0 if books\_home==2

\* Modell A

reg read sameteacher4\_plus

test sameteacher4\_plus = 0

est store m\_A

\* Modell B

reg read sameteacher4\_plus clsiz

test sameteacher4\_plus = 0

test clsiz = 0

est store m\_B

estat hetttest

\* Modell C

reg read sameteacher4\_plus clsiz highpar\_edu girl highbooks\_home kinderg\_att teacher\_exp  
teacher\_fem

test sameteacher4\_plus = 0

test clsiz = 0

test highpar\_edu = 0

test girl = 0

test highbooks\_home = 0

test kinderg\_att = 0

test teacher\_exp = 0

test teacher\_fem = 0

est store m\_C

\* Tabell

outreg2 [m\_A m\_B m\_C] using modell\_ABC\_nEU.doc, replace

shellout using "modell\_ABC\_nEU.doc"

