

Ingvild Eliassen
Eira Flatmo Enodd

En empirisk undersøkelse av om lærernes erfaring og utdanning påvirker elevprestasjonene i matematikk

Masteroppgave i Samfunnsøkonomi
Veileder: Hans Bonesrønning
Trondheim, mai 2018

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi



Forord

Denne oppgaven markerer slutten på vårt femårige masterstudium i samfunnsøkonomi ved Norges teknisk-naturvitenskapelige universitet. Prosessen har vært krevende, innsiktsfull og veldig lærerik. Med både glede og vemod leverer vi fra oss oppgaven og forbereder oss på å forlate studietiden på Dragvoll.

Vi ønsker å rette en stor takk til våre studievenner, spesielt Sigurd Melsom, for alle relevante og ikke like relevante samtaler i løpet av studietiden. Årene på Dragvoll hadde ikke blitt de samme uten dere.

Vi viser Daniel Flatmo Bakken en dypfølt takk for eksemplarisk hjelp i korrekturlesingen av oppgaven.

Avslutningsvis ønsker vi å takke vår veileder Hans Bonesrønning for alltid å ha vært tilgjengelig og bistått oss med god veiledning. Besøkene på ditt kontor har hjulpet oss å løse mange av de frustrerende hindrene vi har møtt på veien.

Masteroppgaven er et resultat av et felles arbeid mellom Ingvild Eliassen og Eira Flatmo Enodd. Tolkninger og synspunkter i denne oppgaven er våre egne, og eventuelle mangler kan kun klandres oss selv. Det har vært en stor glede å gjennomføre arbeidet sammen.

Ingvild Eliassen og Eira Flatmo Enodd

Trondheim, 31. mai 2018

Sammendrag

Denne oppgavens formål er å undersøke om det finnes observerbare lærerkarakteristikk som er av betydning for elevprestasjon i matematikk. Analysen tar utgangspunkt i et tverrsnittsmateriale for skoleåret 2016/17 i et prosjekt som omhandler smågruppeundervisning. De observerbare lærerkarakteristikkene er hentet fra en spørreundersøkelse for lærerne i prosjektet.

For å undersøke problemstillingen benyttes minste kvadraters metode til å estimere sammenhengen mellom avhengig variabel, posttestresultat, lærererfaring og antall studiepoeng i matematikk.

Vi finner at erfaring til smågruppelærere korrelerer positivt med elevprestasjon, og at effekten av mer erfaring er sterkere jo mindre erfaring lærerne initialt har. Tilsvarende konklusjon kan ikke trekkes på statistisk grunnlag for storgruppelærerne.

Den separate effekten av antall studiepoeng for lærerne viser seg å være av ikke-signifikant betydning for elevprestasjon. Dog finner vi en negativ, signifikant effekt av lærernes samlede antall studiepoeng på elevprestasjon.

Sett i sammenheng med at oppgaven kun baserer seg på data fra startfasen av 1+1 prosjektet, kan resultatene kun gi indikasjoner på effekten av inkluderte lærerkarakteristikk for elevprestasjon. Muligheten for at resultatene endres ved prosjektslutt er sannsynlig når et mer komplett datamateriale kommer på plass. Da kan man inkludere eventuelle nye mål for lærerkarakteristikk og ha mulighet til å kontrollere for elevene i prosjektet.

Abstract

This dissertation gives an evaluation of observable teacher characteristics and to what extent they affect student achievement in mathematics. A cross section data set from the school year 2016/17 is used in the analysis. The data set is sourced from an ongoing school project in Norway, *I+I prosjektet*, where the observed teacher characteristics are from a survey for the teachers in the project.

To evaluate our hypothesis, we use ordinary least squares to estimate the relationship between dependent variable, post-test scores, teacher experience and credits in mathematics from college.

We find that the small-group teacher's experience correlates with student achievement and that the effect is stronger the less initial experience. We do not find the same, on statistical basis, for the big-group teachers.

The separate effect of the teacher's credits in mathematics from college is nonsignificant and does not affect student achievement. We find in contrast a negative significant effect of the interaction term.

This thesis is based on data from the start of the project, and the results can only give an indication of the effects of observable teacher characteristics. The potential for including new measurements of the characteristics, and a possibility to control for student characteristics may give different results at the end of the project.

Innhold

1. INNLEDNING	1
2. OM 1+1 PROSJEKTET OG LITTERATURGJENNOMGANG	5
2.1 BESKRIVELSE AV 1+1 PROSJEKTET	5
2.2 TIDLIGERE LITTERATUR	7
2.2.1 BETYDNING AV LÆRERERFARING	7
2.2.2 BETYDNING AV UTDANNING	7
2.2.3 BETYDNING AV LÆRERKVALITET	8
2.2.4 OPPSUMMERING	9
3. TEORI	11
3.1 HUMANKAPITAL	11
3.2 PRODUKTFUNKSJON FOR UTDANNING	11
3.3 TILBUD OG ETTERSØRSEL ETTER LÆRERE	13
4. DATAMATERIALET	15
4.1 DATA	15
4.2 TESTRESULTATER	15
4.2.1 MINSTE KVADRATERS METODE MODELL	17
4.2.2 VALUE ADDED SPESIFIKASJON	18
4.3 FORKLARINGSVARIABLER	18
4.3.1 LÆRERKARAKTERISTIKKER	19
4.3.2 PRETEST	22
4.3.3 VARIASJON	23
5. ØKONOMETRISK RAMMEVERK	25
5.1 MODELLSPESIFIKASJON	27
5.2 ØKONOMETRISKE UTFORDRINGER	28
5.2.1 SELEKSJONSSKJEVHET	28
5.2.2 UTELATT VARIABLEL PROBLEM	30
5.2.3 MÅLEFEIL	31
5.2.4 ENHETSSPESIFIKKE VARIABLER OG SKOLEFASTE EFFEKTER	32
5.2.5 MANGLENDE OBSERVASJONER	34
5.2.6 EFFISIENS	35
5.2.7 MULTIKOLLINEARITET	36
6. ANALYSE OG RESULTATER	37
6.1 MINSTE KVADRATERS METODE	37

6.2 ALTERNATIVE MÅL PÅ LÆRERKOMPETANSE	43
6.2.1 MATEMATIKK PÅ VIDEREGÅENDE SKOLE	44
6.2.2 ALTERNATIVT MÅL PÅ STUDIEPOENG	46
6.3 KORT OPPSUMMERING	47
7. SENSITIVITETSANALYSE	49
7.1 VALUE ADDED SPESIFIKASJON	49
7.2 TRINNENE SEPARAT	53
7.3 UTVIDELSE	55
7.3.1 ANTALL UKER I SMÅGRUPPEN	56
7.3.2 INKLUDERING AV EKSTREMVERDIER PÅ STUDIEPOENG	57
7.3.3 INKLUDERING AV INTERAKSJONSLEDD FOR ERFARING	59
8. OPPSUMMERING OG DISKUSJON	61
8.1 STYRKER, SVAKHETER OG VIDERE FORSKNING	63
9. KONKLUSJON	65
REFERANSER	67
A. APPENDIKS	I
A.1 DESKRIPTIV STATISTIKK	I
A.1.1 ANDRE TRINN	I
A.1.2 TREDJE TRINN	II
A.2 TETTHETSFUNKSJONER	III
A.3 RESULTATER TRINNVIS	IV
A.3.1 ANDRE TRINN	IV
A.3.2 TREDJE TRINN	IV
A.4 KORRELASJONSMATRISER	V
A.4.1 FOR TRINNENE SAMLET	V
A.4.2 ANDRE TRINN	V
A.4.3 TREDJE TRINN	V
A.4.4 VALUE ADDED SPESIFIKASJON	V
A.5 FIGUR VA	VI

Figurer og tabeller

Figur 1: Tetthetsfunksjon for standardisert pretest.....	16
Figur 2: Tetthetsfunksjon for standardisert posttest	16
Figur 3: Tetthetsfunksjon for erfaring til smågruppelærer	20
Figur 4: Tetthetsfunksjon for erfaring til storgruppelærer	20
Figur 5: Tetthetsfunksjon for smågruppelærers studiepoeng.....	22
Figur 6: Tetthetsfunksjon for storgruppelærers studiepoeng	22
Figur A.5: Grafisk framstilling- høy VA lærer.....	vi
Tabell 1: Deskriptiv statistikk 2. og 3.trinn	17
Tabell 2: Deskriptiv statistikk lærerkarakteristikk	19
Tabell 3: Avhengig variabel er posttest standardisert etter trinn	38
Tabell 4: Utvidelse - Betydning av antall år matematikk på VGS.....	44
Tabell 5: Robusthetstest - Alternativt mål på studiepoeng	46
Tabell 6: Avhengig variabel er forskjell i standardisert testscore for 2. og 3. trinn samlet.....	51
Tabell 7: Utvidelse - Betydning av antall uker i smågruppen.....	56
Tabell 8: Robusthetstest - Inkludering av ekstremverdier for studiepoeng i matematikk fra høyere utdanning	57
Tabell 9: Robusthetstest - Interaksjonsledd mellom erfaring som samarbeidsindikator	59
Tabell A.1.1: Deskriptiv statistikk 2. trinn.....	i
Tabell A.1.2: Deskriptiv statistikk 3.trinn.....	ii
Tabell A.3.1: MKM resultater andre trinn.....	iv
Tabell A.3.2: MKM resultater tredje trinn.....	iv
Tabell A.4.1 - A.4.4: Korrelasjonsmatriser.....	v

1. Innledning

I Norge utgjør utgiftene til utdanningssektoren anslagsvis 5 prosent av bruttonasjonalprodukt (BNP). For 2013 utgjorde dette en totalsum på 150 milliarder kroner, et beløp tilsvarende 113 934 kroner per elev (Statistisk Sentralbyrå, 2016). Det investeres mye kapital i utdanning, og derfor er kunnskap om forvaltning av skoleressurser sentral for å generere avkastning i skolesektoren. Utfordringen til politikerne som bevilger pengene er at det per i dag ikke finnes et fasitsvar på hvordan skolesektoren bør utformes for å sikre best mulig læringsutbytte for elevene. PISA-undersøkelsen i 2012 viser at økt ressursbruk alene ikke har vært tilstrekkelig for å øke elevprestasjoner i Norge. Elever i norsk grunnskole presterer ikke godt nok sammenlignet med andre OECD-land, til tross for at Norge er et av de OECD-landene som bevilger mest penger til utdanning (Regjeringen, 2013).

Fra empirisk utdanningslitteratur finnes det ikke et tydelig svar for hvordan skoleressurser påvirker elevprestasjon. I de analysene som finner en sammenheng, er tolkningen av denne tvetydig. Det vanlige i studier som betrakter ressursbruk i skolen, er at påvirkningen en lærer har på elevprestasjon ikke tas hensyn til. Dette gjelder spesielt for studier som analyserer effekten av å redusere klassestørrelse. For å kunne identifisere en kausaleffekt¹ av ressursbruk i skolesektoren, vil det være nødvendig å ta hensyn til lærerkarakteristikker. På den måten kan man kartlegge i hvilken grad ressursene i skolen og læreren som gjennomfører undervisningen er av betydning for elevprestasjon.

Et godt eksempel er Mueller (2013) som finner i sin analyse av lærererfaring med datamateriale fra STAR-prosjektet², at det kun er lærere med erfaring som er i stand til å oppnå økt gjennomsnittlig elevprestasjon ved redusert klassestørrelse. Et annet eksempel er Hammond (2000), som viser til at lærere som har brukt mest tid til egen utdanning før start i yrket vil ha større suksess i interaksjonen med elever. Det ser dermed ut til at erfaring og utdanning som mål på kvaliteten til en lærer er avgjørende for å finne signifikante effekter av smågruppeundervisning.

¹ En *ceteris paribus* («alt annet likt»)-endring i en variabel som har en effekt på en annen variabel.

² Prosjektet STAR er et randomisert eksperiment på klassestørrelseeffekt fra Tennessee i USA fra perioden 1999-2003.

Det kan også være essensielt hvordan undervisningen ved en skole gjennomføres. Vanligvis vil ikke elevprestasjon påvirkes av en klassestørrelsesreduksjon hvis det ikke skjer noe med hvordan undervisningsopplegget struktureres i tillegg. Dobbie og Fryer (2013) finner signifikant effekt av smågruppeundervisning. Webster, Blatchford og Russell (2011) finner tilsvarende resultat når nødvendige retningslinjer følges av lærer og lærerassistent. For øvrig er det ingen av disse studiene som rapporterer kausale effekter av smågruppeundervisning. Dette gir samlet motivasjon til å undersøke følgende hypotese:

Avhenger effekten av smågruppeundervisning i matematikk av lærernes erfaring og utdanning?

I denne oppgaven utnytter vi data om smågruppeundervisning i matematikk fra det norske 1+1 prosjektet³. Vi vil benytte elev- og lærerdata for skoleåret 2016/2017 for å se på hvilke faktorer som er av betydning for hvordan elevene presterer. Formålet med oppgaven er å undersøke om eventuelle positive effekter av smågruppeundervisning avhenger av små- og storgruppelærerens utdanning og erfaring, dette gjøres ved at vi utnytter variasjonen i læringsutbytte mellom behandlingsskolene i prosjektet.

Hanushek (2011) rapporterer at lærere har stor innflytelse på elevprestasjon. Imidlertid finner Rivkin, Hanushek og Kain (2005) at effekt på elevprestasjon av uobserverbare lærerkarakteristikk er av betydning. For å finne forventningsrette prognoser for lærers betydning på elevprestasjon, dokumenterer Chetty (2014) at value added modeller som kontrollerer for elevs tidligere testresultater gir forventningsrette prediksjoner.

Utgangspunktet for analysen er en minste kvadraters metode (MKM) modell samt en value added spesifisering (VA). Vi utfører ulike korrelasjonsanalyser for å undersøke om læringsutbyttet korrelerer med observerbare karakteristikk ved lærerne. For øvrig er vi tydelige på at våre resultater ikke direkte kan tolkes som kausale, som følge av uobserverbar seleksjon. Elevprestasjoner vil i tillegg avhenge av uobserverbare heterogene elevkarakteristikk, som vi heller ikke er i stand til å kontrollere for. Analysen vår er derfor ment som et bidrag i kartleggingen av hvilke faktorer som driver effekten av smågruppeundervisning.

³ 1+1 prosjektet beskrives i kapittel 2.1.

Oppgaven vil være strukturert i ni kapitler. I kapittel to vil vi først presentere 1+1 prosjektet og deretter tidligere litteratur på forskningsområdet. Det vil gis en oversikt over aktuell teori i kapittel tre, mens datamaterialet presenteres i kapittel fire. Kapittel fem gir en diskusjon av det økonometriske rammeverket, vår modellspesifikasjon og utfordringene knyttet til den. Resultater fra den empiriske analysen presenteres i kapittel seks, og i kapittel sju vil vi gjennomføre sensitivitetsanalyser. I kapittel åtte oppsummerer og drøfter vi resultatene, før vi til slutt konkluderer i kapittel ni.

2. Om 1+1 prosjektet og litteraturgjennomgang

I dette kapittelet redegjør vi for 1+1 prosjektet og presenterer tidligere litteratur som ser på lærere som ressurs i utdanningssektoren. Først presenteres funn for betydningen av lærererfaring, deretter for utdanning. Vi redegjør så for betydningen av lærerkvalitet før vi avslutter med en kort oppsummering.

2.1 Beskrivelse av 1+1 prosjektet

1+1 prosjektet, heretter 1+1, er et pågående smågruppeundervisningsprosjekt i matematikk for elever i norsk grunnskole. Prosjektet ledes av NIFU (Nordisk institutt for studier av innovasjon, forskning og utdanning) og gjennomføres sammen med Senter for økonomisk forskning ved NTNU og institutt for samfunnsforskning. 1+1 er finansiert og vedtatt av Norges forskningsråd.

Bakgrunnen for prosjektet bygger på resultater fra forskning (e.g. Hanushek, 2002, 2008, 2012) hvor resultatene viser til at elevers prestasjon i matematikk er av betydning for både individuell framtidig inntekt og veksten i økonomien som helhet. Formålet med 1+1 er å teste om det finnes en effekt på elevers matematikkferdigheter ved allokering av én ekstra lærer i faget. Ekstralæreren skal oppfylle gitte mattekvalifikasjoner og gi skreddersydd undervisning i matematikk i en begrenset periode, for homogene elevundergrupper. Denne type undervisning gis til alle typer elever, slik at stigmaet ved å bli plassert i smågrupper unngås.

Lengden på prosjektet er fire år og startet i 2016. Totalt ti kommuner spredt over hele landet deltar: Asker, Bærum, Bodø, Drammen, Sandefjord, Sarpsborg, Stavanger, Trondheim, Tromsø og Ålesund. Deltakelseskommunene ligger i tettbefolkede områder for å sikre tilstrekkelig tilgang til kvalifiserte matematikklærere.

Prosjektet er utformet som et tilfeldig kontrollert forsøk, RCT, og omfatter 160 skoler. Disse skolene er likt fordelt mellom en behandling- og kontrollgruppe, hvor behandlingsskolene mottar ekstra ressurser i form av ett ekstra lærerårsverk. Det essensielle for analysen av 1+1 er at fordelingen til de to gruppene er helt tilfeldig, slik at forskjell i læringsressurser er det eneste som skiller de. Det gjør at man kan sammenligne og undersøke en potensiell effekt på elevers læringsutbytte.

Ved behandlingsskolene har det blitt bevilget midler til å ansette ett ekstra årsverk. Det har videre vært opp til hver enkelt skole hvor mange lærere de har ansatt til smågruppeundervisning innenfor årsverket. Basert på dette vil det være av betydning i

hvilken grad effekter av behandlingen avhenger av lærernes kvalitet. Når smågruppeundervisningen i matematikk foregår vil de øvrige elevene i hovedklassen oppleve redusert klassestørrelse og undervises av den vanlige mattelæreren.

Definisjonen for smågruppeundervisning i 1+1 er en gruppe av maksimum seks elever som møtes tre til fire ganger i løpet av en uke til relativt korte økter, i en periode på minimum fire uker. Alle elevene skal i løpet av skoleåret ha smågruppeundervisning i to perioder.

Retningslinjene i 1+1 instruerer til høy grad av informasjonsflyt⁴ mellom de to lærerne i matematikk, for å sikre at alle elever undervises i samme tema, opp mot felles læreplan.

Kontaktlæreren er den som fordeler elevene til smågruppene i hver periode, mens læreren i smågruppen tilpasser matematikkundervisningen til elevene. Skoleledelsen vil være instruert til å utarbeide formative vurderinger for å kartlegge elevene som til enhver tid er i smågruppen og progresjonen de har.

Det gjennomføres tester i matematikk både ved skoleårets start og slutt for å følge elevenes utvikling. Prøvene er korte og utformet med hensikt i å gi minst mulig belastning for skolene i prosjektet. Jevnlig registrering av gjennomføringen, årlige spørreundersøkelser for skoleledere og matematikklærere samt intervjuer og skolebesøk er også en viktig del av dokumenteringen av 1+1.

⁴ Parallell til de nødvendige retningslinjene presentert i Webster, Blatchford og Russell (2011) som ble nevnt i innledningen.

2.2 Tidligere litteratur

Dette delkapittelet presenterer tidligere litteratur fra utdanningsøkonomi.

Hanushek (2008) finner at avkastning fra utdanning, målt ved individuell elevprestasjon, dirkete kan knyttes innsatsfaktorer i skolen. Funnet gjelder faktorer som skole- og lærerkarakteristikk. Øvrige faktorer, eksempelvis familie og venner, som ligger utenfor de formelle rammene for utdanning viser seg å være av langt mindre betydning. At nevnte komponenter er av betydning er altså bevist, men hvordan de kan benyttes effektivt er utfordrende å besvare som følge av at faktorene ikke er direkte målbare.

I det videre presenteres litteratur som omhandler læreres påvirkning på elevprestasjon.

2.2.1 Betydning av lærererfaring

Fra studier som forsker på effekten av læreres betydning for elevprestasjon rapporteres det at lærere tidlig i karrieren er mindre effektive enn sine mer erfarne kollegaer (Harris & Sass, 2011, Rockoff et al., 2011).

Harris og Sass (2011) finner at de største avkastningene av lærererfaring for grunnskolelærere genereres i løpet av de fem første årene i karrieren. Det rapporteres at avkastningen av erfaring etter disse årene fortsatt vil være positiv, men effekten er ikke signifikant.

Muller (2013) finner med utgangspunkt i STAR at det bare er erfaringen til seniorlærere som har betydning for elevprestasjon. Effekten gjelder både for lavt og høyt presterende elever, men er mindre framtrædende hos de lavt presterende. For øvrig er det kun i smågruppeundervisning seniorlærerne genererer større læringsutbytte enn nyutdannede.

Kraft og Papay (2014) dokumenterer at gjennomsnittlig avkastning av lærererfaring i stor grad varierer mellom individuelle lærere, og varierer på tvers av lærergrupper ved ulike skoler. I sin analyse rapporterer de videre at lærere som arbeider i støttende miljø i større grad er i stand til å øke elevprestasjon enn lærere som ikke gjør det.

2.2.2 Betydning av utdanning

I forskning på betydningen av lærerutdanning rapporteres det ingen konsensus for elevprestasjon. Hammond (2000) løsriver seg fra en fundamental kritikk av betydningen av lærerutdanning, og setter søkelyset på at det finnes en rekke bevis for at forberedelse til yrket har effekt. Lærere som har brukt mer tid før inntreden i yrket har i større grad suksess og er mer selvsikker i interaksjon med studenter. Jacob og Lefgren og likeså Garet et al. (som sitert

i Harris & Sass, 2011) benytter eksperimentdesign og finner signifikante endringer i undervisningspraksis av videreutdanning av lærere. For øvrig er ikke effekten overførbart til elevprestasjon.

Harris og Sass (2011) gir som Hammond et overblikk over forskning om lærerutdanning og lærertrening, men fokuser på nyere tids rapporteringer. Det henvises til indikasjoner på at forholdet mellom elevprestasjon og lærerutdanning enten er negativt eller ikke-signifikant. I egen analyse finner Harris og Sass (2011) ingen sammenheng mellom læreres humankapital tilegnet *før* start i læreryrket og opptakseksamen til universitet med produktivitet i klasserommet. Det rapporteres heller ikke et konsistent forhold mellom kursing av lærere i jobb og lærerproduktivitet.

Enzi (2017) forsker på en tysk masseundersøkelse som tar for seg determinanter for å tre inn i læreryrket. Med fokus på lærereffektivitet baserer analysen seg på resultat fra videregående skole og svar på to nasjonale undersøkelser for å måle lærerkvalitet. Fra sin value added analyse finner Enzi at lærere som befinner seg i øvre kvartil av analyserte evnemål knyttes til høyere grad av effektivitet i matematikkundervisning. Enzi presenterer funn om at høyt evnenivå hos lærere kan assosieres med større grad av effektivitet i klasserommet.

I utdanningslitteraturen eksisterer det omfattende forskning på virkningen av klassestørrelse på elevprestasjon. Temaet er likevel av stor forskningsinteresse, fordi det mangler grunnlag for å fatte en generell konklusjon på effekten av klassestørrelse. Det amerikanske prosjektet STAR fra 1980-tallet er det eneste eksisterende randomiserte forsøket på effekt av klassestørrelse. Konklusjonen for betydningen av en størrelsesreduksjon er vanskelig å kartlegge, som følge av at effekten avhenger av elevenes alder og lærernes erfaring.

Indikasjoner fra tidligere forskning om at lærernes utdanning ikke er av betydning, motiverer til å undersøke om konklusjonen kan trekkes inn i norsk grunnskole.

2.2.3 Betydning av lærerkvalitet

I dette underkapittelet presenteres forskning som tegner et bilde av at intervensjoner i skolen med bruk av læringsassistenter ikke øker elevprestasjon, dersom ikke bestemte retningslinjer følges.

Storskalaprojektet Deployment and Impact of Support Staff (DISS) finner at elever som eksponeres for mye undervisning og oppfølging av lærerassistenter har signifikant lavere

prestasjonsutvikling enn elever som i større grad eksponeres for en kvalifisert lærer i undervisningen.

Som en tilslutning til funnene fra DISS utførte Webster, Blatchford og Russell (2013) prosjektet Effective Deployment of Teaching Assistants (EDTA). EDTAs formål er utvikling og evaluering av alternative strategier til lærerassistenter grad av forberedelse til og gjennomføring av undervisning. Det eksisterer dog et visst potensiale hos lærerassistentene som ikke blir utnyttet, hvor Alborz et al. og Slavin et al. (som sitert i Webster et.al, 2013) viser til at assistentene kan ha en positiv påvirkning når de har fått opplæring og er forberedt etter læreplanen.

Dobbie og Fryer (2013) finner positiv effekt av å inkludere læringsassistenter i USA. Konklusjonen er imidlertid todelt. Den første delen av konklusjonen bygger på at velfungerende charterskoler⁵ karakteriseres ved hyppig bruk av smågruppeundervisning. Det andre funnet bygger på et eksperiment (Fryer, 2014) for smågruppeundervisning i matematikk, hvor det vises til positive effekter for lavt presterende offentlige skoler.

Fra Danmark rapporterer Andersen et. al (2015) en positiv, gjennomsnittlig effekt på leseferdigheter for 13-åringene når det benyttes lærerassistenter. Det rapporteres derimot ingen effekt på matematikkferdigheter. Den ikke-signifikante effekten i matematikk forklares med forskjeller i optimal læringsteknologi mellom fagene. Det pekes med andre ord på viktigheten av lærerkvalitet, som følge av at undervisning i matte krever mer presisjon i undervisningen.

Det er hensiktsmessig å skille mellom begrepene smågruppeundervisning og redusert klassestørrelse. Redusert klasse viser til tilfeller hvor antallet elever i én klasse reduseres permanent. Undervisning i smågrupper reduserer klassestørrelsen kun for den perioden smågruppeundervisningen pågår, samtidig som at elevene ute i smågrupper opplever i den pågående periode en redusert klassestørrelse.

2.2.4 Oppsummering

I den eksisterende litteraturen som betrakter sammenhengen mellom ulike lærerkaraktistikker og elevprestasjon er det vanskelig å se et mønster, fordi resultatene er forskjellige og effektene ofte er små. Årsakene til at det er slik, kan være mange. En naturlig forklaring vil ligge i at det benyttes ulike tester og prøver som mål på elevprestasjon. Det vil derfor være urimelig å forvente at studiene skal komme fram til samme konklusjon.

⁵ Amerikansk statlig finansiert, privat skole, hvor elevene ikke er pålagt å betale.

Fra litteraturen er det klare indikasjoner på at lærererfaring er av størst betydning tidlig i karrieren, men for utdanningsnivå er det liten enighet mellom forskerne å skimte.

Betydningen av observerbare kjennetegn viser seg å være vanskelig å kartlegge. Å måle betydning av lærerutdanning på elevprestasjon vil gi ulike resultater for ulike lærere og ulike elever. Det finnes heller ikke en tydelig ramme for hvordan utdanningsnivå skal måles, og tilnærmingen vil blant annet variere ut fra forskerpreferanser, analysetype og tilgjengelig informasjon om elever og lærere.

En annen forklaring på ulike funn i ulike studier kan komme av hvordan forskerne har valgt å kontrollere for uobserverbare karakteristikk ved individene. Samtidig er metodene brukt for å undersøke sammenhengen mellom lærerkvalitet, undervisningsform og elevprestasjon ofte relativt enkle. Få politiske tiltak og randomiserte forsøk har gjort det utfordrende å finne en kausal effekt av lærerkarakteristikk på elevprestasjon nettopp på grunn av manglende mulighet til å kontrollere for uobserverbare karakteristikk hos lærere, elever, foreldre og nærmiljø. Det randomiserte 1+1 kan derfor trolig kunne bidra i litteraturjakten på kausaleffekter av lærerkarakteristikk på elevprestasjon.

3. Teori

I dette kapittelet vil vi ta for oss deler av teorien som danner grunnlaget for ressursbruk innen utdanningssektoren. Først presenteres begrepet humankapital før vi beveger oss over til produktfunksjonen for utdanning (PFU) som danner grunnlaget for analysen vi gjennomfører i kapittel 6 og 7. Deretter diskuteres markedet for etterspørsel og tilbud av lærere, der vi også redegjør for utfordringer knyttet til seleksjon av markedsaktørene.

3.1 Humankapital

Humankapital benyttes som økonomisk mål på et individs totale nivå av kunnskap, evner og kvalifikasjoner (Mincer, 1985). Begrepet danner grunnlag for en forståelse om hvorfor utdanning og tilegnelse av ny kunnskap er av betydning for samfunnet. Mincer betrakter humankapital som en viktig faktor i samfunnets økonomiske vekstprosess og den betraktes både som årsak til økonomisk vekst og som et resultat av den.

Økt teknologibruk i arbeidsmarkedet har vært med på å øke etterspørselen etter kvalifiserte arbeidere (Mincer, 1995). For å møte den økte etterspørselen etter utdannede arbeidere kreves mer kunnskap om hvordan utdanning kan effektiviseres og optimaliseres tidlig i skoleløpet. For å sikre økonomisk vekst i fremtiden kreves tilsvarende vekst i nivået av humankapital. Det vil derfor være av interesse å kartlegge om hvorvidt spesifikke lærerkarakteristikk har betydning for elevens utbytte av utdanning, og videre for nivået på individuell humankapital. Humankapitalteori gir forskere et empirisk rammeverk som til en viss grad gjør det mulig å måle de økonomiske kostandene og gevinstene til individ og samfunn som ligger i utdanning⁶.

3.2 Produktfunksjon for utdanning

Produktfunksjonen for utdanning (PFU) er en enkel produksjonsmodell som ligger til grunn i de fleste eksisterende analyser innen utdanningsøkonomi. Den typiske produktfunksjonen relaterer et mål for elevprestasjon opp mot en rekke innsatsfaktorer, som lærerkvalitet og familieegenskaper (Hanushek, 2008).

For å presentere teorien som danner rammen for PFU benyttes Todd og Wolpin (2003). Artikkelen gir en gjennomgang av tilnærminger til modellering av en funksjon for kognitive

⁶ En samling av artikler om humankapital presenteres i Scott R. Sweetland (1996): Human Capital theory; Foundations of a field of inquiry.

ferdigheter. Det antas at barns oppnåelse, målt ved testscore, er en kumulativ prosess av kunnskapstilegnelse.

Utgangspunktet for analysen er følgende generelle produktfunksjon:

$$T_{ija} = T_a [F_{ij}(a), S_{ij}(a), \mu_{ij0}, \epsilon_{ija}] \quad (1)$$

Der T_{ija} representerer et prestasjonsmål for elev i , i husholdning j med alder a . $F_{ij}(a)$ er vektor av innsatsfaktorer fra foreldre opp til alder a og $S_{ij}(a)$ er tilsvarende vektor for skoler. μ_{ij0} måler elevenes medfødte evnenivå. Restleddet ϵ_{ija} fanger opp uobserverbare faktorer som påvirker prestasjon. For vår oppgave er $S_{ij}(a)$ den sentrale innsatsfaktoren vi prøver å kartlegge betydningen av, og den representerer de lærerkaraktistikkene vi inkluderer i likning (4) i kapittel 5.1.

Produktfunksjonen betraktes ofte som den ideelle funksjonen innen utdanningsøkonomi, men hvordan den framkommer er en underliggende og kompleks prosess. I standard produksjonsteori benytter bedrifter produktfunksjonen for å maksimere profitt og minimere kostnader til gitte innsatsfaktorer og priser. Enhver innsatsfaktor som ikke bidrar til økt produksjon, til gitte kostnader, ekskluderes fra funksjonen og produksjonen til bedriftene.

Bedrifter kan benytte nøyaktige mål på inkluderte innsatsfaktorer i sin produktfunksjon⁷. På sin side er PFU mer utfordrende å estimere som følge av at inkluderte innsatsfaktorer, som kognitive ferdigheter, ikke direkte kan måles.

Det er langt mer utfordrende å kontrollere for insentiver hos lærere og skoleledelse som kan ha betydning for hvordan inkluderte innsatsfaktorer i skolesektoren påvirker elevprestasjon. Ved inkludering av målbare faktorer for elev- og lærerkaraktistikker i PFU kan man til en viss grad kontrollere for adferdsutfordringer som oppstår når funksjonen estimeres. Likevel vil en slik modell kreve en enorm mengde med data, og få datasett oppfyller kravet om komplett historisk informasjon for alle relevante innsatsfaktorer. Det benyttes derfor i

⁷ Utdanning inngår i markedet som benytter seg av kapital – *skolebygg* og arbeidskraft – *lærere* til produksjonsfunksjonen. Innsatsfaktorene vil variere mellom hvilke marked som betraktes og inkluderer i tillegg energi og råvarer.

utdanningsøkonomi alternative spesifikasjoner for å estimere den generelle PFU som følge av manglende informasjon. Vår tilnærming presenteres ved likning (4) i kapittel 5.1.

Value added analyse er den mest kjente og betrodde metoden for å estimere PFU. Metoden måler lærernes bidrag (value added) til elevprestasjon og kontrollerer samtidig for andre faktorer⁸ som påvirker prestasjon for ett gitt skoleår. Dette gjøres ved en sammenligning av nåværende og tidligere testscore. Den grunnleggende strategien i VA-modellering er å isolere bidraget som hver lærer gir et gitt år, slik at det kan sammenlignes med bidraget fra andre lærere (Kinsler, 2012). Med andre ord vil VA ved anvendelse av tidligere testresultat kunne redusere problemene knyttet til den kumulative prosessen.

Som nevnt i kapittel 1 finner Chetty (2014) at value added gir forventningsrette prognoser på lærerkvalitet. Ved å kontrollere for lærerkaraktistikkene erfaring, alder og studiepoeng i matematikk vil vi kunne redegjøre for om det er noen av disse som gir gode prediksjoner på lærerkvalitet. Gjennom å analysere value added kan vi få en større forståelse av påvirkningen fra observerbare lærerkaraktistikker⁹.

I kapittel 5.2 tar vi for oss potensielle utfordringer knyttet til estimering av PFU.

3.3 Tilbud og etterspørsel etter lærere

I standard arbeidsmarkedsteori styres hovedsakelig tilbudet og etterspørselen etter arbeidskraft av lønnen i markedet. Dersom tilbudet av arbeidskraft er for høyt, må lønnen reduseres for at etterspørselen skal øke, slik at markedet kan klareres til likevekt (Riis og Moe, 2012). For lønnsfastsettelse i skolesektoren er det andre faktorer som er avgjørende, der lønnen forhandles fram nasjonalt og framkommer av KS Hovedtariffavtalen. Lønnen fastsettes slik at lik ansiennitet og utdanning kvalifiserer til lik lønn uavhengig av alder, kjønn og geografi. På denne måten har ikke skolene mulighet til å bruke lønn som et virkemiddel for å tiltrekke seg de høyest kvalifiserte lærerne i markedet.

En typisk utfordring i norske småkommuner er at tilbudet av kvalifiserte lærere er for lavt. Norske skoler er fastbundet til sentrale tariffes, og småkommuner kan ikke benytte seg av et høyere lønnsnivå for å tiltrekke seg lærere som oppfyller nasjonale krav. Resultatet er at skolene ansetter vikarer uten formell kompetanse i midlertidige stillinger. Vikarene ansettes i perioder på inntil ett år, fram til skoleårets slutt. En slik situasjon vil naturlig være utfordrende

⁸ Faktorer som individuelle evner, familiemiljø, tidligere læring og medelever.

⁹ En generell grafisk framstilling av gjennomsnittlig elevprestasjon ved inntreden av en høy VA lærer presenteres ved figur A.1 i appendiks.

for både elever og faste ansatte, men det gjør det også vanskelig å analysere påvirkningen en lærer har for elever over tid. 1+1 prosjektet omgår problemet med manglende tilbud av kvalifiserte lærere ved å kun inkludere kommuner der etterspørselen etter tilleggslærere blir møtt.

Tilbudet av lærere i markedet dannes av en selektert individgruppe. Dette er mennesker som søker seg til lærerutdanning etter fullført videregående skole eller har en annen form for pedagogisk utdanning i ryggsekken. Gruppen er selektert både fordi tiltredelse i yrket avhenger av kvalifikasjonskrav og fordi etterspørselen etter lærere drives av formelle kvalifikasjonskrav.

Det er skoleledelsen som ansetter lærere til det tildelte årsverket. At skoleledelsen ønsker de dyktigste lærerne som finnes på tilbudssiden er en rimelig antakelse. Derimot vil det være ulike oppfatninger blant lærerne om jobben som smågruppelærer, og det kan tenkes at lærere med særegen interesse for matematikk vil finne disse jobbene attraktive. Hvis skoleledelsen avviser alle lærerne som de definerer som mindre effektive, og det kun er lærere med matematikkinteresse som søker seg til jobbene, vil vi få et seleksjonsproblem som er av betydning for estimeringen.

En mer detaljert presentasjon av seleksjonsproblematikk gis i kapittel 5.2.1.

4. Datamaterialet

I dette kapittelet presenterer vi datamaterialet som danner grunnlaget for vår empiriske analyse. Vi benytter et datamateriale bestående av to ulike datasett som er hentet fra 1+1 prosjektet. Informasjonen vi har er for elever i 2. og 3. trinn, hvor det ene datasettet gir informasjon om hver enkelt elevs resultat på pre- og posttest, koblet opp mot elevenes respektive smågruppelærere. Det andre bidrar med variabler fra en spørreundersøkelse lærerne i prosjektet har besvart. Vi har data for to tester utført på to ulike tidspunkt, men fordi testene gjennomføres innenfor samme skoleår jobber vi med et tverrsnittsmateriale for skoleåret 2016/2017.

4.1 Data

Datasettet vi benytter i analysen gir informasjon om hver enkelt elevs testscore, observerbare lærerkarakteristikker og et sett av kontroller. Totalt har vi data for 18479 elever og 888 lærere. I oppgaven ligger fokuset på behandlingsskolene og elevene som eksponeres for en tilleggslærer i matematikk. Når vi ekskluderer alle kontrollskolene har vi observasjoner for 9682 elever. I hovedanalysen i kapittel 6 benyttes 2183 observasjoner når vi inkluderer interesse- og kontrollvariabler. Dette tilsvarer omtrent 100 lærere og vi kan lene oss på sentralgrenseteoremet om normalfordelt utvalg¹⁰.

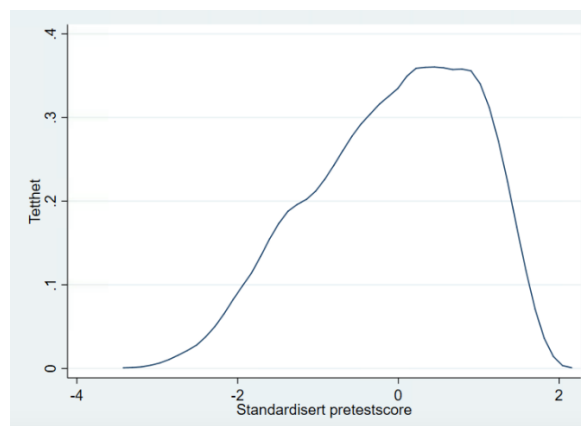
Videre har vi variabler som er aggregerte på ulike nivåer. De inkluderte lærerkarakteristikkene for smågruppelærerne er aggregert på klassenivå, som gjør at vi har mulighet til å koble hver enkelt smågruppelærer til sine respektive klasser og dermed elever. Storgruppelæreren aggregeres på skolenivå og vi har kun mulighet å koble lærerne til skolen de jobber på. Det gjør at vi ikke har mulighet til å kontrollere for storgruppelæreren på tilsvarende måte som smågruppelæreren.

4.2 Testresultater

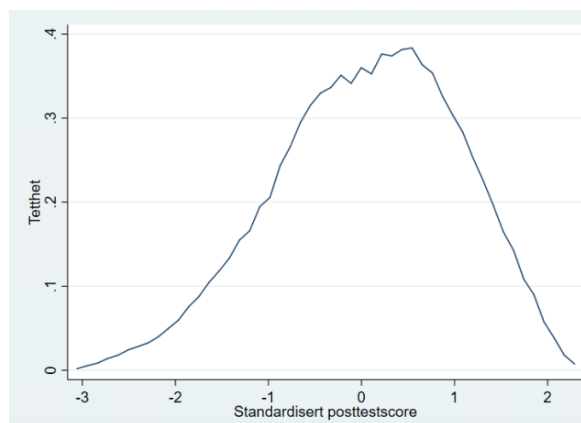
I hovedanalysen benytter vi en modellspesifikasjon som er bygd opp med posttest som avhengig variabel. I sensitivitetsanalysen benyttes en alternativ spesifikasjon med forskjell i posttest og pretest som avhengig variabel.

¹⁰ Nøkkresultat fra sannsynlighetsteori. Impliserer at summen av antall uavhengige variabler, og svakt avhengige variabler, følger en standard normalt fordeling når utvalgsstørrelsen vokser og variablene er standardisert ved eget standardavvik.

Pretest gjennomføres av elever i 2. og 3. trinn ved alle grunnskoler i prosjektet høsten 2016, og gjennomføres før smågruppeundervisningen i matematikk settes i gang. Posttest gjennomføres av de samme elevene ved skoleårets slutt, våren 2017, etter ett skoleår med smågruppeundervisning. Vi har informasjon om testresultat for 2. og 3. trinn, men fordi testene er av ulik måleskala for trinnene, kreves en standardisering av testene fordi i analysen betraktes trinnene samlet. Pretest og posttest er heller ikke direkte sammenlignbare fordi vanskelighetsnivå og maksimal score er forskjellig. Dermed standardiseres begge testene på trinnene.



Figur 1: Tetthetsfunksjon for standardisert pretest



Figur 2: Tetthetsfunksjon for standardisert posttest

Figur 1 og 2 viser tetthet og fordeling av elevenes testscore på de to testene. Grafene viser to tilnærmet standardnormalfordelte variabler med gjennomsnitt lik null og varians lik 1. Vi observerer en lengre venstrehale, særlig for pretest. Dette gir en indikasjon på at det er flere elever som presterer under gjennomsnittet på pretest, mens fordelingen har jevnet seg ut når

posttest har blitt gjennomført. Deskriptiv statistikk gir dermed ytterligere motivasjon for å undersøke vår problemstilling¹¹.

Tabell 4.1: Deskriptiv statistikk 2. og 3.trinn

Variabel	Gjennomsnitt	Standardavvik	Min	Max	N
Posttest(standardisert)					
Hele utvalget	≈0	≈1	-2.903754	2.321815	13 296
Behandling	0.0577864	0.978709	-2.903754	2.1359	6 231
Kontroll	-0.0509649	1.015613	-2.903754	2.321815	7 065
Pretest(standardisert)					
Hele utvalget	≈0	≈1	-3.268351	2.001751	16 396
Behandling	-0.0568641	1.005508	-3.268351	2.001751	6 933
Kontroll	0.041661	0.9938295	-3.268351	1.908091	9 463
VA(standardisert)					
Hele utvalget	-0.0269563	0.833912	-4.106485	3.86512	12 795
Behandling	0.0877808	0.8213043	-3.848887	3.709488	6 060
Kontroll	-0.130194	0.8317346	-4.106485	3.86512	6 735

4.2.1 Minste kvadraters metode modell

Hovedmodellen er en minste kvadraters metode (MKM) modell med posttest som avhengig variabel. I denne modellformuleringen søker vi etter observerbare faktorer som korrelerer med elevprestasjon målt ved individuelt resultat på posttest for trinnene samlet. Ved MKM tillates det at resultatet fra pretest forklarer noe av variasjonen i resultat på posttest, slik at koeffisienten foran pretest kan være ulik én.

Fra tabell 1 rapporteres i snitt et bedre resultat på pretest for skolene i kontrollgruppen. I behandlingsskolene er resultat på posttest ti prosent standardavvik høyere enn i kontrollskolene i prosjektet¹². For trinnene separat rapporteres også høyere gjennomsnittlig testscore i behandlingsklassene enn i kontrollklassene¹³.

¹¹ Standardavviket til en variabel gir et mål på spredningen mellom alle utvalgsgjennomsnittene og er enklere å tolke enn variansen siden den er i samme dimensjon som forklaringsvariabelen. Standardavviket til et estimat gir et mål på hvor mye predikert verdi avviker fra observert. Jo større standardavvik i estimatene, jo større usikkerhet om riktig predikert verdi.

¹² Den deskriptive statistikken gir indikasjoner på at det finnes en effekt av 1+1 prosjektet.

¹³ Tabell A1.1 og A.1.2 i appendiks.

Lavest mulig posttestscore er null for begge trinnene, og det rapporteres at 0.2% av elevene har oppnådd dette resultatet. Separat for trinnene er maksimal testscore ulik, hvor 2.trinn kan oppnå maksimalt 24 poeng og 3.trinn maksimalt 28. Trinnsvis vil variabelen derfor følge ulike utvalgsfordelinger.

4.2.2 Value added spesifikasjon

I kapittel 7 gjennomføres en sensitivitetsanalyse med en value added spesifikasjon der endring i elevprestasjon inngår som avhengig variabel. Hensikten med denne spesifikasjonen er å måle lærerbidraget, målt ved endring i testscore ved skoleårets slutt. Spesifikasjonen brukes mye i den empiriske litteraturen og derfor er det av interesse å undersøke hvordan effekten av lærernes erfaring og utdanning slår ut i spesifikasjonen.

Det gjennomsnittlige VA-målet varierer fra positivt til negativt lærerbidrag i henholdsvis behandlings- og kontrollskolene. For klassene i behandling rapporterer VA-målet en gjennomsnittlig endring i testscore lik 0.07646 med et standardavvik lik 0.8479091. Et så høyt standardavvik relativt til gjennomsnittet viser at det er tilstrekkelig med variasjon i avhengig variabel til å estimere forskjeller i påvirkningen fra ulike lærere.

4.3 Forklaringsvariabler

I vår empiriske analyse benytter vi forklaringsvariablene presentert nedenfor for å forklare utviklingen i elevprestasjoner. Deskriptiv statistikk for lærerne er hentet fra spørreundersøkelsen de har besvart. For enkelte av de rapporterte verdiene finner vi små avvik, men ved undersøkelse av begge datasettene finner vi at de manglede observasjonene ikke bærer preg av systematikk og våre resultater vil ikke påvirkes av avvikene. I kapittel 7 utføres sensitivitetsanalyser med flere kontroller for elevprestasjon og den alternative modellformuleringen. Dette vil innebære inkludering av andre forklaringsvariabler som potensielt kan tenkes å drive observerte effekter av smågruppeundervisning i 1+1.

Tabell 4.2: Deskriptiv statistikk lærerkarakteristikk

Variabel	Gjennomsnitt	Standardavvik	Min	Max	N
Erfaring					
Hele utvalget	13.10143	9.5466	0	43	547
Små	12.09813	9.549385	0	36	107
Stor	13.78155	9.92144	0	40	206
Alder					
Hele utvalget	41.44314	10.96878	22	67	510
Små	40.08824	11.40532	24	66	102
Stor	41.74869	11.42019	24	67	191
Poeng					
Hele utvalget	41.22479	22.78388	10	135	476
Små	55.14433	26.2071	10	135	97
Stor	37.65556	20.58239	10	120	180

4.3.1 Lærerkarakteristikk

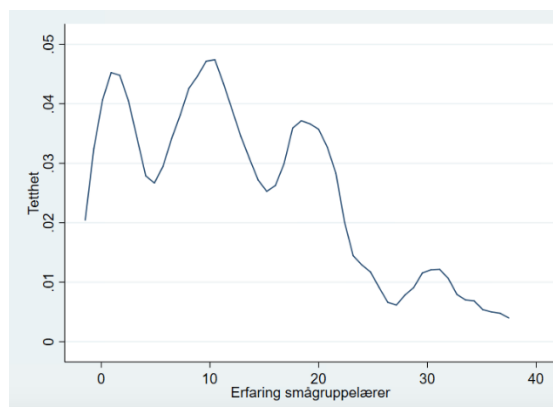
Utgangspunktet for analysen er observerbare karakteristikk ved lærerne i 1+1 prosjektet. Elevene ved behandlingsskolene eksponeres for både små- og storgruppelærer i matematikkundervisning og vi benytter derfor variabler for begge lærerne i modellen. Erfaring og studiepoeng i matematikk inkluderes for begge lærerne.

Erfaring

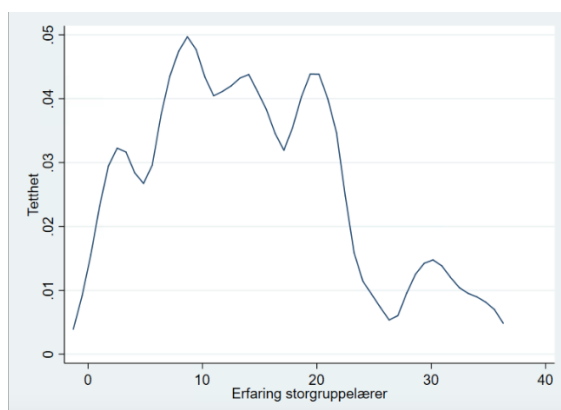
Erfaringsvariabelen varierer fra null til 43 år i yrket og defineres ved antall år med erfaring. Som følge av at elevene eksponeres for en kunnskapsproduksjon av to lærere inkluderes erfaringen til begge i estimeringen.

Gjennomsnittlig erfaring for lærerne i smågruppen er tolv år med et standardavvik på 9.55 hvor intervallet går fra null til 36 år. Storgruppelærerne har gjennomsnittlig erfaring lik 13.8 år med et standardavvik på 9.55 og erfaringsintervallet strekker seg fra null til 40.

Erfaring i snitt for de to lærerne er tilnærmet lik, og det finnes ikke tegn til systematikk i fordelingen av lærertypene i prosjektet. Dette betyr at det ikke finnes spesielle kjennetegn som er typisk for hverken små- og storgruppelærer, og det observeres ikke at smågruppelærerne har mer eller mindre erfaring enn lærerne i storgruppen.



Figur 3: Tetthetsfunksjon for erfaring til smågruppelærer



Figur 4: Tetthetsfunksjon for erfaring til storgruppelærer

Figur 3 og 4 illustrerer tetthetsfunksjonen for små- og storgruppelærers erfaring. Tettheten er for begge lærertypene størst i området med ti år erfaring.

Studiepoeng i matematikk fra høyere utdanning

Vi inkluderer to kontinuerlige variabler som fanger opp effekten av studiepoeng i matematikk for små- og storgruppelærerne. Disse er generert slik at null- og ekstremverdier utelates. I praksis er dette gjennomført ved at lærere som befinner seg utenfor intervallet -1 standardavvik til $+2$ standardavvik er droppet fra analysen, for å hindre at enkeltobservasjoner driver resultatene. Etter denne ekskluderingen er spredningen i variablene fra ti til 135 studiepoeng.

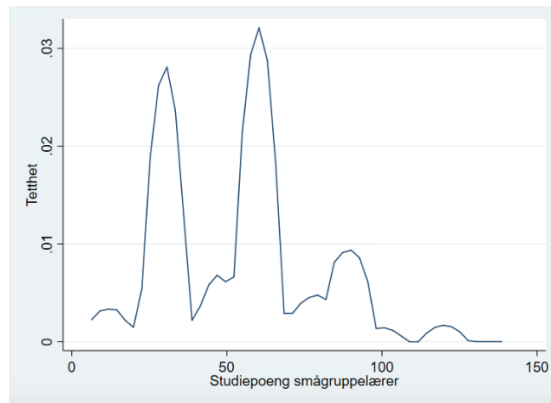
Årsaken for å droppe svært lave verdier for studiepoeng fra datasettet kommer av kravet om minimum 30 studiepoeng på lærerstudiet (I. O. Jakobsen, personlig kommunikasjon, 24.april 2018). Likevel kan vi ikke begrense utvalget for mye siden PPU¹⁴ ikke har et slikt krav og vi har lærere i datamaterialet som har fullført sin utdanning før innføring av matematikkkravet.

¹⁴ PPU er praktisk-pedagogisk utdanning. Ett årig utdanning som bygger på en bachelor- eller mastergrad. Gir kvalifikasjoner til å undervise i videregående skole og grunnskole 5.- 10. trinn (NTNU, 2018).

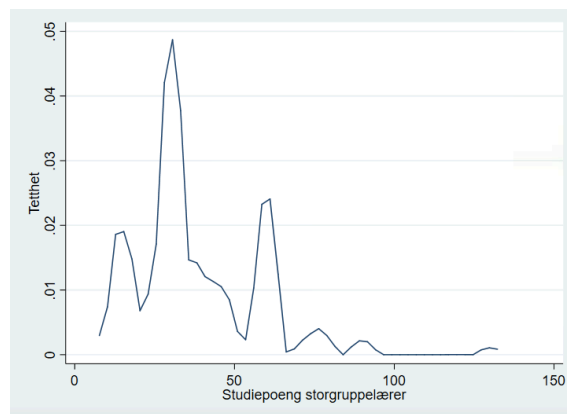
Tilsvarende utelates lærere med svært mange studiepoeng i matematikk, da det er lite trolig at de ønsker å undervise i matematikk på barneskolen. Derfor kan det tenkes at disse rapporteringene er feilrapportert.

Som et forsøk på å måle graden av samarbeid mellom lærerne inkluderes et interaksjonsledd mellom studiepoengene. Med en slik samarbeidsindikator har vi mulighet til å teste en tilleggshypotese om at *like barn leker best*, der en lærer med mye studiepoeng hypotetisk vil kunne samarbeide bedre med en annen lærer hvis denne læreren også har mange poeng. Når vi betrakter en økning i variablene vil det være snakk om en poengsøkning, slik at det ikke er hensiktsmessig å se på en enhetsøkning, men heller en økning på 5, 7.5, 10 eller 15.

I datasettet er det 100 lærere som underviser i smågruppene. Kun to av disse lærerne er uten formell lærerutdanning, 64 har fullført allmennlærerutdanning og ti har universitet- eller høyskoleutdanning med PPU. Storgruppelærerne er sterkere representert i datasettet med 203 observasjoner, hvor tre av lærerne ikke har en formell utdanning. 154 har formell utdanning, 139 har allmennlærerutdanning og 15 har universitet- eller høyskoleutdanning med PPU. De resterende små- og storgruppelærerne har rapportert *annet*. Utdanningsnivået for de to ulike lærertypene bærer i snitt ikke preg av systematikk, det indikerer at lærerne som søker seg til jobbene som smågruppelærer er relativt like som storgruppelærerne når utdanningsnivået betraktes.



Figur 5: Tetthetsfunksjon for smågruppelærers studiepoeng



Figur 6: Tetthetsfunksjon for storgruppelærers studiepoeng

Figur 5 og 6 viser tetthetsfunksjonene for de to lærernes studiepoeng. Vi observerer for begge lærertypene et toppunkt på omtrent 30 studiepoeng og et toppunkt på omtrent 60 poeng.

Alder

For alle lærerne i datasettet er rapportert gjennomsnittsalder 41.4 år med et standardavvik lik 10.97. I behandlingsskolene er gjennomsnittsalderen 40 år med et standardavvik på 11.4. De yngste lærerne i behandlingsskolene er 24 år og de eldste er 66 år. Vi har altså informasjon om lærere helt fra de er nyutdannede til de snart er ferdig i yrket. Variabelen inkluderes som en kontroll, slik at vi redegjør for alderen til lærerne når vi estimerer betydningen av lærerkarakteristikker. Aldersvariabelen er kontinuerlig og inkluderes kun for smågruppelærere.

Som følge av at mer erfaring genereres med økt alder, finner vi sterk positiv korrelasjon mellom erfaring og alder og variablene samvarierer sterkt.

4.3.2 Pretest

Når vi utfører minste kvadraters metode vil pretestvariabelen inngå som forklaringsvariabel. Gjennomsnittsscoren på pretesten er 26.6 og 24.8 med et standardavvik på 8.32077 og

8.83412 for 2. og 3.trinn i behandling. Høyst mulig score på denne testen er 41 poeng for 2.trinn og 42 for 3.trinn. Maksimal score på posttest er lavere for begge trinn, og med null som laveste mulig score.

4.3.3 Variasjon

Interesse- og kontrollvariablene vi anvender i analysen er rapportert med tilstrekkelig variasjon. Dette ser vi fra tilhørende standardavvik for hver variabel, som er høye relativt til øvrige verdier. Vi kan derfor på grunnlag av utvalget, gitt at utvalget er stort nok¹⁵ og forutsetning I.- IV. presentert i kapittel 5. holder, ta konklusjoner for populasjonen som helhet.

¹⁵ minimumskrav for utvalgsstørrelse, $n \geq 20$ (Thomas, 2005).

5. Økonometrisk rammeverk

Å skulle finne en økonometrisk sammenheng mellom elevprestasjon og smågruppeundervisning vil som i de fleste økonometriske analyser by på metodologiske utfordringer. Dette kapitlet presenterer metoden vi benytter for å håndtere de økonometriske utfordringene vi står overfor. Teorien er basert på Wooldridge (2015).

Et tverrsnittsdatasett inneholder observasjoner fra én enhet, for eksempel et individ eller et land, for ett bestemt tidspunkt. Vi har altså kun variasjon over én dimensjon, enhet, som benevnes med fotskrift i . For vår modell vil *elev* representere variasjon over enheter, slik at gitt vår struktur på data har vi mulighet til å undersøke variasjon i individuell elevprestasjon.

Ved å benytte minste kvadraters metode for å estimere sammenhengen mellom elevprestasjon og andre variable får vi benyttet all informasjon som finnes i data. Dette er en klar fordel, men det fører med seg noen ulemper også. MKM som estimeringsmetode minimerer summen av kvadrerte avvik mellom de inkluderte variabelenes estimerte og observerte verdi.

For å presentere forutsetningene ved MKM, samt utlede vår modell, benytter vi følgende enkle formulering:

$$Posttest_{igsk} = \partial Pretest_{igsk} + X_{lgsk}\beta + X_{sk}\gamma + u_{igsk} \quad (2)$$

Hvor $Posttest_{igsk}$ er elevprestasjon for elev i , ved skole s i kommune k , X_{lgsk} angir matrise av inkluderte forklaringsvariabler for smågruppelærer med tilhørende koeffisientvektor β . X_{sk} angir tilsvarende for storgruppelærer med tilhørende koeffisientvektor γ . u_{igsk} er det stokastiske restleddet. Restleddet er elevspesifikt og fanger opp alle uobserverbare variabler som påvirker elevprestasjon, $Posttest_{igsk}$, som er faste for hver elev. Dette kan for eksempel være forskjell i klasseromsfasiliteter mellom skoler.

For at MKM skal gi konsistente¹⁶ og forventningsrette¹⁷ estimat har vi at følgende forutsetninger må være oppfylt (Wooldridge, 2015, s. 74-82):

- I. Populasjonsmodellen er lineær i sine parametere.
- II. Utvalget er tilfeldig trukket fra populasjonen¹⁸.
- III. Ingen perfekt multikollinearitet. Ingen av utvalgets uavhengige variable er konstante og det finnes heller ikke noe eksakt lineært forhold mellom dem¹⁹.
- IV. Forklaringsvariablene er strengt eksogene. Det eksisterer ingen korrelasjon mellom restleddet og de uavhengige variablene, restleddet har null i forventningsverdi gitt enhver verdi på forklaringsvariablene:

$$E(u_{igsk} | X_{lgsk}, X_{sk}, Pretest_{igsk}) = 0 \quad (3.1)$$

For at MKM-estimatene videre skal være BLUE (best linear unbiased estimators), og dermed effisiente, må følgende forutsetning også oppfylles:

- V. Stokastisk restledd med konstant varians for enhver verdi på forklaringsvariablene, og er på den måten homoskedastisk. Restleddet er ikke seriekorrelert:

$$E(u_{igsk}u_{jhzc} | X_{lgsk}, X_{sk}, Pretest_{igsk}) = \begin{cases} \sigma_u^2 & \text{for } i = j, g = h, s = z \text{ og } k = c \\ 0 & \text{ellers} \end{cases} \quad (3.2)$$

Brudd på forutsetning V. gir ikke et problem med inkonsistente eller forventningsskjev estimat, som følge av at V. kun har betydning for om de estimerte koeffisientene er effisiente eller ikke. Et brudd på forutsetning I.-IV. vil derimot ha som konsekvens at MKM gir misledende prediksjoner av populasjonsmodellen og alternative estimeringsmetoder vil kunne

¹⁶ En estimator er konsistent dersom de empiriske momentene konvergerer mot sanne populasjonsverdier når utvalgsstørrelsen øker. $\text{plim}\hat{\beta} = \beta + \frac{\text{cov}(X, u_{igsk})}{\text{var}(X)} = \beta$

¹⁷En estimator er forventningsrett hvis gjennomsnittet ved gjentatte utvalg er lik sann verdi. $E(\hat{\beta}) = \beta$

¹⁸ Vårt utvalg er selektert, men randomisert blant deltakelsesskolene i prosjektet. Derfor vil, isolert sett, våre koeffisienter estimeres konsistente og forventningsrette, selv om forutsetningen ikke holder.

¹⁹ Krav om ikke-konstante forklaringsvariabler er en nødvendig betingelse for at vi skal kunne observere en effekt av klassestørrelse på elevprestasjon.

egne seg bedre. Videre er det avgjørende om forutsetning IV. er oppfylt, og i kapittel 5.2 drøftes ulike utfordringer knyttet til forutsetningen. Forutsetning I. og III. antar vi er oppfylt.

5.1 Modellspesifikasjon

I dette delkapittelet presenterer vi spesifikasjonen vi estimerer i hovedanalysen og deretter den alternative value added spesifikasjonen. Ved å inkludere ulike lærerkarakteristikker er målet å redegjøre for om noen av de observerbare karakteristikkene korrelerer med elevprestasjon. Vi tar utgangspunkt i et tverrsnittsdata som følger 2183 elever i skoleåret 2016/2017 og lærerne knyttet opp mot elevene.

I hovedanalysen benyttes følgende spesifikasjon:

$$\begin{aligned} Posttest_{igsk} = & \alpha_s + \beta_1 pretest_{igsk} + \beta_2 erfarsm\ddot{a}_{lgsk} + \beta_3 erfarsm\ddot{a}_{lgsk}^2 \\ & + \beta_4 aldersm\ddot{a}_{lgsk} + \beta_5 poengsm\ddot{a} + \beta_6 erfstor_{sk} \\ & + \beta_7 erfstor_{sk}^2 + \beta_8 poengstor_{sk} + \beta_9 interstudie_{sk} + \varepsilon_{igsk} \end{aligned} \quad (4)$$

Her er $Posttest_{igsk}$ standardisert testscore på posttest for elev i , klasse g , ved skole s i kommune k . $pretest_{igsk}$ er standardisert testscore på pretest for elev i , klasse g , ved skole s i kommune k .

$erfarsm\ddot{a}_{lgsk}$ er erfaring for småklasselærere og rapporterer erfaring til lærer l i klasse g ved skole s i kommune k . $erfarsm\ddot{a}_{lgsk}^2$ angir kvadrert erfaring for den samme læreren.

$aldersm\ddot{a}_{lgsk}$ noterer alder til smågruppelærer l , i klasse g , ved skole s i kommune k , og $poengsm\ddot{a}_{lgsk}$ er studiepoeng i matematikk fra høyere utdanning for den samme læreren.

Variablene $erfstor_{sk}$ og $erfstor_{sk}^2$ angir erfaring til storgruppelærer ved skole s i kommune k , mens $poengstor_{sk}$ er antall studiepoeng for storgruppelæreren. I datamaterialet er alle observasjoner for storgruppelærerne aggregert på skolenivå. Dermed kan ikke hver enkelt elev eller klasse knyttes direkte opp mot storgruppelæreren.

$interstudie_{sk}$ er interaksjonsledd mellom små- og storgruppelærers samlede antall studiepoeng og α_s er modellens skolefaste effekter og følgelig modellens konstantledd. Variabelen vil variere mellom skolene i behandling og fanger opp forskjeller som ligger mellom dem.

Vi inkluderer et kvadratisk erfaringsledd for begge lærerne for å kontrollere for at erfaringsfunksjonen ikke er lineær. Alder inkluderes som kontroll og studiepoeng i matematikk for begge lærere inngår for å måle effekt av utdanningsnivå. Variablene kontrollerer også for at begge lærerne bidrar i elevenes kunnskapsproduksjon. Interaksjonsleddet inngår i tillegg for å forsøke å måle lærersamarbeid presentert ved tilleggshypotesen i kapittel 4 og variabelen benyttes som en proxy for samarbeid.

I sensitivitetsanalysen benyttes:

$$\begin{aligned}
 VA_{igsk} = & \alpha_s + \mu_1 erfarsm\ddot{a}_{igsk} + \mu_2 erfarsm\ddot{a}_{igsk}^2 + \mu_3 aldersm\ddot{a}_{igsk} \\
 & + \mu_4 poengsm\ddot{a}_{igsk} + \mu_5 erfstor_{sk} + \mu_6 erfstor_{sk}^2 \\
 & + \mu_7 poengstor_{sk} + \mu_8 interstudie_{sk} + u_{igsk}
 \end{aligned} \tag{5}$$

Der VA_{igsk} er value added målet for elev i , klasse g , ved skole s i kommune k . Den avhengige variabelen er gitt ved elev is forskjell mellom standardisert post- og pretestresultat. Inkluderte forklaringsvariabler i (5) er tilsvarende som i likning (4).

5.2 Økonometriske utfordringer

Empiriske undersøkelser vil til enhver tid by på metodiske utfordringer. I dette kapittelet redegjør vi for utfordringer som er sentrale for vår oppgave. Vi presenterer først hovedutfordringen som er seleksjon, deretter fokuserer vi på utelatte variabler og målefeil. Skolefaste effekter belyses så grundig, før manglende observasjoner, effisiens og multikollinearitet presenteres.

5.2.1 Seleksjonsskjevhhet

Ved et brudd på forutsetning IV. om strengt eksogene forklaringsvariabler vil MKM gi forventningsskjeve estimatorer. I et slik tilfellet vil det i modellen være en eller flere variabler som er utelatt og korrelerer med inkluderte forklaringsvariabler. Graden og retningen på skjevheten til estimatorene vil avhenge av effekten til den utelatte variabelen og fortegnet på korrelasjonen mellom den inkluderte og utelatte variabelen.

Det er ikke utenkelig at lærere av relativt lik karakter søker seg til jobbene som smågruppelærer ved behandlingsskolene, fordi dette er jobber som trolig vil betraktes som attraktive. Vi har omfattende informasjon fra lærerspørreundersøkelsen og vi kan kontrollere for om smågruppelærerne har homogene karakteristikk. Ved undersøkelser av datasettet for

inkluderte variabler finner vi at variasjonen for de to lærertypene i prosjektet er stor, slik at fordelingen av små- og storgruppelærere er tilfeldig. Vi kan da undersøke om det finnes kjennetegn ved lærere, på et generelt nivå, som er av betydning for hvordan elever presterer. Likevel er det viktig å være klar over at vi ikke kan kontrollere for alle typer seleksjon, der vi for eksempel ikke er i stand til å kontrollere for seleksjon til læreryrket.

Vi har ikke mulighet til å kontrollere for denne typen yrkesseleksjon av lærere, slik at restleddet i populasjonsmodellen vil korrelere med de inkluderte forklaringsvariablene for læreren. Årsaken er at variabler som trolig fanger opp uobserverbare individuelle faktorer som er avgjørende for å velge læreryrket, høyst sannsynlig korrelerer med både lærernes erfaring og studiepoeng i matematikk. Å utelate variabler som fanges opp av restleddet gjør at vi ikke kan trekke kausale sammenhenger i analysen, slik at resultatene våre kun kan tolkes som korrelasjoner og ikke årsakssammenhenger.

Seleksjon av elever til klasser som mottar smågruppeundervisning utgjør ikke en bekymring for vår analyse, som følge av at prosjektet er randomisert på skolenivå. Når fordelingen skjer på skolenivå og ikke individnivå, vil en skole være randomisert til enten behandling eller ikke behandling. Dermed unngår vi problemer med selvseleksjon til behandlingsklassene fra foreldre og elever. En elev vil fordeles tilfeldig til en behandling- eller kontrollklasse ut fra hvilken gruppe elevens skole ble fordelt til ved prosjektstart. Elevgruppene i prosjektet er derfor homogene på tvers av skolene, slik at vi kan sammenlikne klassene direkte med hverandre.

En annen potensiell kilde til seleksjon er fra skoleledelse og rektorer. De kan stille seg motvillig til å delta i eksperimentet dersom elevene ved sine skoler mottar behandling ulikt. En skoleledelse kan selektivt velge hvilke elever eller lærere som skal påvirkes av en potensiell politikkendring. Eksempelvis kan skoleledelsen plassere lavtpresterende elever i små klasser eller lærere av høy kvalitet til de lavtpresterende elevene. Konsekvensen er at effekten av smågruppeundervisningen feilestimeres og potensielle feilaktige indikasjoner på at tiltaket er effektivt kan oppstå. Denne typen seleksjon vil ikke gi empiriske utfordringer da randomisering av skolene er kontrollert for ved prosjektstart.

Utfordringen ligger altså i seleksjon til yrket og vi har ikke datamateriale til å kontrollere for dette. Oppgaven vår gir en redegjørelse av mulige korrelasjoner, og ikke kausale effekter, for elevprestasjoner. De øvrige seleksjonsproblemene gir ut fra diskusjonen over ikke estimeringsutfordringer.

5.2.2 Utelatt variabel problem

Å utelate en relevant forklaringsvariabel som korrelerer med en inkludert forklaringsvariabel, gjør at man med høy sannsynlighet vil oppnå forventningsskjev estimat. Modellen vil være underspesifisert, slik at restleddet og inkludert forklaringsvariabel vil fange opp effekten av den utelatte variabelen. Resultatet er at inkludert forklaringsvariabel korrelerer med restleddet og forutsetning IV. er ikke oppfylt.

I vår modell for elevprestasjon, målt ved testscore, er det naturlig å tro at det eksisterer uobserverbar heterogenitet i form av utelatte variable både på individnivå, skolenivå og kommunenivå.

På klasse- og skolenivå er vi i stand til å inkludere kontrollvariabler for svært mange faktorer som potensielt kan ha en påvirkning på elevprestasjoner i matematikk. En slik kontroll vil for eksempel være lærerkvalitet, og fordi vi har omfattende data for lærerkarakteristikker, vil ikke problemet med utelatte variable ligge på dette nivået.

Vi benytter variasjon innad i skolene for å betrakte effekten på elevprestasjon, slik at utelatte variable på skolenivå ikke vil drive resultatet.

Når det kommer til individnivå vil det være utfordrende å inkludere alle kontrollvariabler som har betydning for elevprestasjon. Vanskeligheten ligger i spesifiseringen av en produktfunksjon som inkluderer kognitive ferdigheter. Dette er ferdigheter som er medfødte og ikke målbare på linje med for eksempel lærerkvalitet. Som følge av at det er vanskelig å finne et godt mål på kognitive ferdigheter er det sannsynlig at dette kan gi en utfordring i form av et utelatt variabel problem. Som eksempel er det slik at hvis elever med høyt evnenivå tenderer å havne i samme klasse, vil det være en mulighet for at vi finner et forhold mellom høyt presterende elever og høyt prestasjonsnivå som trolig ikke har noe med inkluderte lærerkarakteristikker å gjøre.

Til tross for at vi har mulighet for å kontrollere for observerbare lærerkarakteristikker, kan det ikke utelukkes at det finnes andre variabler vi ikke er i stand til å kontrollere for. Dette vil i hovedsak omfatte mål for hvordan særtrekk ved elever og skoler har betydning for gjennomføringen av smågruppeundervisningen. Potensielt kan smågruppens størrelse, sammensetning og tidsbruk være av stor betydning for elevprestasjon.

5.2.3 Målefeil

Problemer med målefeil oppstår dersom det eksisterer systematiske målefeil i vårt datamateriale, der sann og observert verdi fraviker hverandre. Vi vil her presentere problemer knyttet til målefeil, og konsekvensene som potensielt oppstår ved slike feil.

Til tross for at det er lite som tyder på at vårt materiale inneholder systematiske målefeil, er det verdt å bite seg merke i at det alltid vil eksistere en grad av målefeil i innhentet data. Sannsynligheten for tilfeldig målefeil i avhengig variabel i vårt data er til stede. Det vil eksistere tilfeller hvor det har skjedd en feil i retting av prøvene eller ved registrering av prøveresultat. Gitt at disse tilfellene er tilfeldige og ikke systematiske, vil det dog ikke by på særlige problemer.

Likevel kan vi ikke se bort fra at de øvrige forklaringsvariablene våre inneholder målefeil siden disse er innhentet fra en spørreundersøkelse. Variablene vi benytter for smågruppelæreren vil trolig ikke ha systematiske feil siden det er lite trolig at noen lærere feilrapporterer egen erfaring og antall studiepoeng i matematikk fra høyere utdanning.

Informasjonen vi har om storgruppelærerne vil derimot inneholde målefeil som følge av at lærerne er aggregert på skolenivå. Aggregeringsnivået gjør at vi ikke har mulighet til å koble storgruppelærerne til sine respektive klasser og elever. Ved at observasjonene for disse er et gjennomsnitt av alle storgruppelærerne på en skole vil hver observasjon inneholde systematiske målefeil. Konsekvensen er at vi ikke observerer de faktiske verdiene for erfaring og studiepoeng slik at estimatoren for variablene er forventningsskjev. Retningen på skjevheten går mot null med økt *noise to signal ratio*²⁰. Med andre ord vil MKM systematisk underestimere absoluttverdien til koeffisientene for storgruppelærerne.

Basert på humankapitalteori kan de standardiserte testresultatene i vår modell brukes som et mål på kognitive ferdigheter. Dersom kognitive ferdigheter fanger opp all variasjon i

$$^{20} \text{plim}(\widehat{\beta}_{stor}) = \beta_{stor} \left(\frac{1}{1 + \sigma_e^2 / \sigma_{X_{sk}}^*} \right)$$

Der σ_e^2 representerer variansen til målefeilen og $\sigma_{X_{sk}}^*$ variansen til uobserverbar variabel. Økt σ_e^2 vil redusere brøken og vi underestimerer β_{stor} . Eventuelt ved økt varians i X_{sk}^* relativt til variansen vil målefeilen i MKM bli mindre.

individuell humankapital, vil testresultatene være et direkte mål på humankapital. Da er det være uten betydning å inkludere andre forklaringsvariabler for humankapital, fordi alle relevante ferdigheter fanges opp av testresultat. At kognitive ferdigheter fanger opp all variasjon i humankapital er lite trolig. I stedet vil testscore inngå som et mål på humankapital som vil inneholde målefeil, μ :

$$H_{igsk} = \text{testresultat}_{igsk} + \mu \quad (6)$$

Å estimere en likning der testresultat inngår som eneste mål på humankapital gir problem i form av målefeil. Det vil eksistere uobserverbare faktorer, deriblant evner, som ikke tas hensyn til slik at målet for individuell gevinst av utdanning bærer preg av feil.

Problemet overføres til produktfunksjonen for utdanning, dersom vi i vår modell benytter *pretest* som et strengt mål på humankapital. For øvrig vil ikke dette gi problematikk for tolkningen av resultatene våre fordi vi ikke benytter testene direkte for å måle humankapital. Vi inkluderer tidligere testresultat som en forklaringsvariabel når vi undersøker om en potensiell endring i testresultat kan knyttes mot lærerkarakteristikker.

Et annet økonometrisk problem med estimering av PFU, jfr. diskusjon i kap.3.2, er dersom det ikke tas hensyn til at funksjonen skiller seg fra standard produktfunksjon som maksimerer profitt gitt et sett av produksjonsmuligheter. I skolesektoren er ikke dette tilfellet, og faktisk effekt av å redusere klassestørrelse avhenger av skolenes insentiv. Da vil utfordringen være at effekten varierer mellom skoler med ulike insentiv, og ytterligere vil det være utfordrende for politikere å identifisere en kausaleffekt av et slikt tiltak. Heller ikke dette vil utgjøre problem for analysen som følge av randomiseringen av prosjektet er på skolenivå.

5.2.4 Enhetsspesifikke variabler og skolefaste effekter

Siden prosjektet omfatter 160 skoler fordelt over ti kommuner ønsker vi å kontrollere for mulige forskjeller som er av betydning for elevprestasjon både mellom skoler og kommuner. For å kontrollere for slike forskjeller inkluderer vi skole – og kommunespesifikke variabler i analysen.

I datamaterialet vårt har vi mer enn én enhetsdimensjon. De nivåene vi ser på er klasse, skole og kommune. Fordelen ved å ha flere dimensjoner er at vi kan benytte faste effekter, som betyr én dimensjon holdes konstant mens den andre enhetene betraktes. Det vil være liten variasjon mellom klassene innen en skole, derimot vil skolene innen kommunen variere i

større grad. Skolene i en kommune kan være plassert i svært ulike områder, således kan elev- og lærersammensetningen i skolen variere mye. Skoleadministrasjonen²¹ er selektert til hver enkelt skole, slik at det er en risiko for at det eksisterer forskjeller i graden av samarbeid, trivsel og faglig utvikling mellom skolene.

Ved å benytte skolefaste effekter har vi mulighet til å kontrollere for slike ulikheter. Praktisk gjennomføres dette ved at vi benytter skoledummier i regresjonen. Dette gir hver enkelt skole i kommunen sitt særegne konstantledd og vi får koeffisientverdier som rapporterer gjennomsnittlig marginal effekt for skolene.

Under antakelse om at skolene i prosjektet kun har tilhørighet til én kommune, vil den kommunefaste effekten være konstant for hver skole og dermed bli utelatt fra analysen. Forøvrig gir ikke dette et problem i form av utelatt variabel skjevhet fordi den skolefaste effekten automatisk tar hensyn til den kommunefaste. Vi er ikke i stand til å direkte estimere de kommunefaste effektene, men dette byr ikke på utfordringer i tolkningen av modellen fordi vi fortsatt har variasjonen vi trenger i estimeringen. Faktorer som kontrolleres for ved inkluderingen kan være forskjeller i kommunestruktur, geografi, bevilgning av øvrige ressurser, størrelse og skoledrift.

Hvis vi ikke benytter faste effekter risikerer vi å få forventningsskjevne prediksjoner på effekten av inkluderte lærerkarakteristikker, ettersom vi innen en kommune kan ha skoler som karakteriseres som lavt og høyt presterende. Eksempelvis vil vi overestimere effekten av inkluderte variabler dersom en skole har flere lærere som er over gjennomsnittlig effektiv. Skjevheten kommer som en respons på at det eksisterer en utelatt skolekonstant variabel.

Vi benytter en enkel modell med faste effekter for å illustrere:

$$Posttest_{igsk} = \partial Pretest_{igsk} + X_{lgsk}\beta + X_{sk}\gamma + \alpha_s + \varepsilon_{igsk} \quad (8)$$

α_s fanger opp uobserverbar skoleheterogenitet som er faste for klassene og elevene i skolen. Det idiosynkratiske restleddet ε_{igk} representerer uobserverbare faktorer som endres over alle dimensjoner. En nødvendig forutsetning er at det skolespesifikke restleddet er ukorrelet med inkluderte forklaringsvariabler. Restleddet er sammensatt, og kan skrives: $u_{igsk} = \alpha_s + \varepsilon_{igsk}$

²¹ Rektor og administrasjon ved én skole.

I minste kvadraters metode antar vi at u_{igs_k} er ukorrelert med $Pretest_{igs_k}$, X_{lgs_k} og X_{sk} . Tross i antakelsen gjort om ε_{igs_k} vil MKM være forventningsskjev og inkonsistent hvis α_s er korrelert med variablene. Vi pålegger dermed en kritisk restriksjon om at det skolespesifikke restleddet vårt er ukorrelert med forklaringsvariablene i modellen. Hvis vi derimot hadde hatt en tidsdimensjon kunne vi ha brutt restriksjonen pålagt det skolespesifikke restleddet og enten transformert eller differensiert det bort fra modellen.

5.2.5 Manglende observasjoner

Konsekvensen av at vi arbeider med et omfattende datamateriale er at det vil naturlig mangle en betydelig andel observasjoner. I standard regresjonsanalyse kan vi ikke benytte en observasjon dersom det mangler data for avhengig variabel eller for en av forklaringsvariablene til observasjonen. Gitt at manglende observasjoner er rapportert riktig, vil *Stata* ignorere observasjoner som mangler informasjon.

I modellen vår betrakter vi hvordan lærernes erfaring og utdanning påvirker elevprestasjonene i matematikk. Sett bort fra pretest inkluderer vi ingen kontroller for elevene, størrelsen på små- og storklassen eller elevsammensetningen i gruppene. Konsekvensen av å ikke inkludere tilstrekkelig med kontrollvariabler blir som diskutert i kapittel 5.2.2 at estimatorene kan bli forventningsskjeve.

Vår problemstilling fokuserer på behandlingsskolene, slik at observasjonene for kontrollskolene utelates. For enkelte av våre inkluderte variabler mangler vi observasjoner slik at disse ikke inngår i analysen og rapporterte standardavvik øker. Dette vil ikke medføre problemer for analysene våre fordi sentralgrenseteoremet er gyldig og utvalget er representativt for populasjonen.

Hawthorne effekt defineres som en kunstig effekt som oppstår dersom individer som observeres endrer sin adferd som følge av det. Alle aktører i et eksperiment vet at de er en del av et eksperiment (Hoxby, 2000).

I eksperiment der skoler og lærere skjønner at en politikkendring ikke vil generere ønsket effekt, kan de ha insentiver til å manipulere prosjektet. Skoleadministrasjon kan for eksempel ved en klassestørrelsesreduksjon plassere de svakeste elevene i små klasser eller fordele de best kvalifiserte lærerne til disse elevene. Eventuelt kan lærerne i eksperimentperioden øke egen innsats.

Tilstedeværelse av Hawthorne effekt i 1+1 prosjektet kan til en viss grad kontrolleres for. Ved at utgangspunktet for analysen er fordelingen av elevene til smågruppene kan vi sikre fravær av manipulasjon fra skoleadministrasjonen. Videre vil vi ikke kunne kontrollere for atferdsendringer hos skoleledelse og lærere, årsaken er at adferd ikke er målbart og det kan ikke forventes at atferdsendringen rapporteres inn. Dermed kan det ikke utelukkes at lærerne i behandlingsskolene har endret adferd og innsats fordi de deltar i eksperiment.

5.2.6 Effisiens

Målet når man benytter MKM som estimeringsmetode er å oppnå effisiente estimatorer. Dette betyr at vi jakter de estimerte verdiene som gir et mest mulig presist anslag på parameterne i populasjonsmodellen. Graden av denne presisjonen måles ved variansen til estimatorene, og en estimator er effisient dersom det ikke finnes andre estimatorer for parameteren som har lavere varians.

Forutsetning V. krever at alle observasjoner i modellen har lik restleddsvarians, altså at restleddene i modellen er homoskedastiske. Dersom denne ikke er oppfylt er restleddene heteroskedastiske og varierer mellom observasjonene:

$$\text{Var}(u_{igsk} | X_{igsk}, X_{sk}, \text{Pretest}_{igsk}) = \sigma_i^2 \quad (7)$$

MKM estimatoren i en slik situasjon vil fortsatt være konsistent og forventningsrett, gitt at forutsetningen om forventet restleddsverdi lik null (IV.) er oppfylt. Forskjellen er derimot at variansen til estimerte verdier vil påvirkes når det eksisterer heteroskedastisitet i restleddet, og standard t- og F-fordeling er ikke lenger gyldig. Følgelig vil man møte på utfordringer med tester hvor disse fordelingene benyttes.

Restleddet i modellen vår er elevspesifikt slik at den tilhørende matrisen ikke vil ha standard struktur, og restleddet er seriekorrelert innen klassene på grunn av aggregeringsnivået. Ved seriekorrelasjon vil ikke standard formel for varians og standardavvik være gyldig, slik at seriekorrelasjon også gir utslag i estimat som ikke er effisiente.

For å ta høyde for variasjon og seriekorrelasjon i restleddet vil vi anvende robuste standardavvik og faste effekter, hvor det tillates for heteroskedastisitet og tilfeldig seriekorrelasjon mellom elevene.

5.2.7 Multikollinearitet

For at MKM skal gi konsistente og forventningsrette estimat, må III. om at det ikke eksisterer en perfekt lineær sammenheng mellom inkluderte forklaringsvariabler være oppfylt. Derfor kan man ha forklaringsvariabler som er høyt, men ikke perfekt, korrelerte uten at det bryter med MKM-forutsetningene.

Fra tabell A.5 rapporteres en korrelasjon lik 0.839 mellom erfaring og alder. En så sterk korrelasjon kan gjøre det vanskelig å skille effekten disse separat har på posttest, men fordi det kun er usikkerheten til disse to som påvirkes når to så sterkt korrelerte variable inkluderes i samme regresjon, inkluderes begge. Dette vil ikke utgjøre et problem for inkludering av flere kontroller i regresjonen så lenge alder og erfaring er uavhengige av de andre variablene.

Tabell A.5 bekrefter dette for blant annet pretest og standardavviket til estimatet øker ikke nevneverdig av inkluderingen i noen av modellene presentert i kapittel 6.1.

6. Analyse og resultater

I dette kapittelet presenteres resultatene fra vår empiriske analyse. Som følge av at vi benytter informasjon om smågruppelærerne er alle regresjoner basert på testscore for elever i behandlingsklassene. Regresjonene er gjennomført i dataprogrammet *Stata* og vi anvender robuste standardavvik i alle estimeringene.

I analysen gjennomfører vi mange regresjoner og ser på elevprestasjon både for trinnene samlet og hver for seg, slik at vi ikke presenterer alle resultatene i dette kapittelet. Vi vil grundig kommentere funnene for trinnene samlet i hovedmodellen. I tillegg presenterer vi hovedmodellen med to alternative mål lærerkompetanse. De separate regresjonene og value added spesifikasjonen blir gitt litt mindre fokus og resultatene inkluderes i appendiks og kapittel 7.

Vi vil som kjent ikke være i stand til å kontrollere for all variasjon hos lærerne som har betydning for elevprestasjon, slik at resultatene må tolkes med varsomhet. Videre i oppgaven benyttes begrepene posttest og pretest for de standardiserte testene.

6.1 Minste kvadraters metode

Vi presenterer og kommenterer i dette delkapittelet resultatene fra minste kvadraters metode for 2. og 3. trinn samlet. Hovedresultatene er presentert i tabell 3. Modell (7) viser resultater fra estimering av likning (4) presentert i kapittel 5.1.

Modell (1) danner utgangspunktet med posttest relatert til pretest og smågruppelærers erfaring. Erfaringsleddet inkluderes som en kvadrert funksjon for å kontrollere for at erfaringsfunksjonen ikke er lineær. Erfaringsleddet inngår positivt, mens det kvadrerte leddet er negativt slik at erfaring inngår som en positiv, avtakende funksjon. Forøvrig er funksjonen ikke-signifikant.

I modellene i kolonne (2) og (3) inkluderes kontroll for smågruppelærers alder og deretter for antall studiepoeng i matematikk. Modellene i kolonne (4) og (5) inkluderer erfaringsvariable for storgruppelærer og for antall studiepoeng i matematikk. Interaksjonsleddet mellom lærernes studiepoeng inkluderes i modell (6). Modell (7) er identisk med (6), men kontrollerer for skolefaste effekter når alle kontroller er pålagt. Modellen i kolonne (7) er vår prefererte modellspesifikasjon, og vi tar utgangspunkt i denne for å forsøke og besvare hypotesen om at lærers erfaring og utdanning kan forklare noe av variasjonen i læringsutbyttet mellom behandlingsskolene.

Tabell 6.1: Avhengig variabel er posttest standardisert etter trinn

VARIABLER	(1)	(2)	(3)	(4)	(5)	(6)	(7)
std_pretest	0.638*** (0.0128)	0.642*** (0.0130)	0.638*** (0.0140)	0.655*** (0.0154)	0.646*** (0.0167)	0.646*** (0.0167)	0.658*** (0.0166)
erfarsmå	0.000525 (0.00392)	0.0201*** (0.00474)	0.0216*** (0.00530)	0.0292*** (0.00612)	0.0358*** (0.00728)	0.0364*** (0.00747)	0.0457*** (0.0113)
erfarsmå ²	-0.000185 (0.000126)	-0.000315** (0.000127)	-0.000387** (0.000169)	-0.000620*** (0.000188)	-0.000628*** (0.000208)	-0.000643*** (0.000211)	-0.000674** (0.000285)
aldersmå		-0.0155*** (0.00223)	-0.0150*** (0.00226)	-0.0161*** (0.00240)	-0.0202*** (0.00311)	-0.0204*** (0.00315)	-0.0278*** (0.00510)
poengsmå			0.000620 (0.000585)	0.00192*** (0.000641)	0.000371 (0.000689)	0.000673 (0.00134)	0.00120 (0.00182)
erfstor				-0.0167*** (0.00615)	-0.0223*** (0.00649)	-0.0227*** (0.00659)	0.0238*** (0.00908)
erfstor ²				0.000437*** (0.000168)	0.000444** (0.000177)	0.000457** (0.000181)	-0.00104*** (0.000260)
poengstor					0.00246*** (0.000764)	0.00293 (0.00198)	0.000734 (0.00256)
interstudie						-7.22e-06 (2.80e-05)	-6.82e-05* (3.75e-05)
Konstant	0.173*** (0.0250)	0.596*** (0.0668)	0.535*** (0.0827)	0.612*** (0.105)	0.751*** (0.131)	0.737*** (0.144)	0.504*** (0.193)
Skolefaste effekter	Nei	Nei	Nei	Nei	Nei	Nei	Ja
Observasjoner	3,819	3,634	3,261	2,632	2,183	2,183	2,183
R ²	0.422	0.431	0.424	0.443	0.442	0.442	0.478

Robust standardavvik i parentes

***Signifikant på 1% signifikansnivå (p<0.01)

**Signifikant på 5% signifikansnivå (p<0.05)

*Signifikant på 10% signifikansnivå (p<0.1)

Erfaring

I modellen fanges erfaring opp ved en lineær variabel som gir et mål på antall år i yrket og deretter inkluderes variabelen på kvadrert form. Smågruppelærers erfaringsfunksjon inngår som en statistisk signifikant determinant for elevprestasjon i alle spesifikasjonene, sett bort i fra modell (1). Tilsvarende gjelder for storgruppelærers erfaring, men verdien på estimatene er lavere. Dette kan tyde på at lærerne som underviser i smågruppen har tettere interaksjon og sterkere innflytelse på elevene, slik at erfaring trolig er av større betydning for undervisning i smågruppene. Potensielt kan resultatet også drives av at målet for storgruppelærers erfaring er aggregert på skolenivå, inneholder målefeil og er basert på mindre variasjon.

Effekten av å inkludere et kvadratisk erfaringsledd inngår som forventet negativt signifikant for begge lærerne. Resultatet er i tråd med tidligere forskning som peker på at effekten på elevprestasjon av lærererfaring er en kvadratisk funksjon, med høyt stigningstallet i starten av karrieren som med tiden vil avta.

For å redegjøre for når effekten av erfaring til smågruppelærer har nådd sitt toppunkt, deriveres likning (4) (tabell 3. modell (7)) med hensyn på erfaring:

$$\frac{\partial Posttest_{lgsk}}{\partial erfarsm\grave{a}_{lgsk}} = 0.0457 - 2 * 0.000674erfarsm\grave{a}_{lgsk} = 0 \quad (9)$$

$$0.0457 - 0.001348erfarsm\grave{a}_{lgsk} = 0$$

$$erfarsm\grave{a}_{lgsk} = 33.9$$

Etter 33.9 år vil effekten av smågruppelærers erfaring på elevprestasjon være lik null.

For storgruppelærer finner vi:

$$\frac{\partial Posttest_{lgsk}}{\partial erfarstor_{lgsk}} = 0.0238 - 2 * 0.00104erfarstor_{lgsk} = 0 \quad (10)$$

$$0.0238 - 0.00208erfarstor_{lgsk} = 0$$

$$erfarstor_{lgsk} = 11.4$$

Storgruppelærers erfaring har ingen positiv effekt på elevprestasjon etter 11.4 år. Betydningen for elevene i behandling er en positiv predikert økning i posttest når små- og storgruppelæreren har under henholdsvis 33 og 11 år med erfaring.

Gjennomsnittlig erfaring for smågruppelærer er 12.1 år, mens for storgruppelærer er gjennomsnittet lik 13.8 år. Resultatet gir at den gjennomsnittlige smågruppelæreren befinner seg i området under toppunktet av erfaringsfunksjonen, mens storgruppelæreren i snitt har mer erfaring enn toppunktet.

Den samlede effekten predikeres til 0.5736, som tilsvarer en 57.4% standardavviks økning i predikert posttestresultat. Effekten av samlet erfaring vil være lik null når $erfarsm\hat{a}=0$ eller $=67.8$ og $erfstor=0$ eller $=22.88$. Undersøkelser av effekten til leddene separat gir ingen informasjon siden erfaring inntreer som en funksjon og ikke kan betraktes som to separate ledd.

Toppunktet i erfaringsfunksjonen til de to lærerne inntreffer til svært ulike tidspunkt. Den marginale effekten av ett år ekstra erfaring vil også påvirke elevprestasjon i ulikt. Når vi betrakter to lærere med gjennomsnittlig erfaring lik fem år, finner vi at en marginal endring for smågruppelærers erfaring øker predikert posttestscore med 0.04233. Tilsvarende endring hos storgruppelærer gir en økning lik 0.0186, og vi ser at en økning i erfaring gir en forskjell i testscore lik 6.1% av et standardavvik. Resultatet kan komme av at undervisningssituasjonen de to lærerne står overfor er svært forskjellige.

Smågruppelæreren underviser en liten gruppe på fire til seks elever der elevsammensettingen endrer seg flere ganger i løpet av skoleåret. Den samlede variasjonen av elevtyper i en liten gruppe er liten sammenliknet med en storgruppe. En lærer som underviser i en smågruppe har mulighet til å benytte et mer dynamisk, pedagogisk opplegg for å tilpasse undervisning til hver elev i hver gruppe.

Storgruppelæreren møter til enhver tid en større elevgruppe bestående av mange ulike typer elever og har ikke samme mulighet til å skreddersy undervisningen opp til hver enkelt elevs behov. Utbyttet av mer erfaring har muligens en naturlig begrensning i storgruppen, fordi der smågruppelæreren kan tilpasse undervisningsopplegget til elevene, vil nok storgruppelæreren være nødt til å gjøre dette på mer generelt grunnlag for at alle individ skal oppnå læringsutbytte.

Det er naturlig å anta at den totale elevsammensettingen av en storklasse er relativ lik fra år til år. Da er det også naturlig å tro at dette har sammenheng med at effekten av storgruppelærers erfaring avtar raskere. Likevel er dette kun spekulasjoner fordi variablene inneholder målefeil.

Koeffisienten foran storgruppelærers erfaring inngår negativt fram til inkludering av skolefaste effekter i (7). Dette indikerer at når vi ikke kontrollerer for de initiale nivåforskjellene mellom skolene i behandling, feilestimeres effekten nedover. En potensiell årsak kan være at vi har noen skoler med lavt presterende elever i datasettet der gjennomsnittlig storgruppelærer har mye erfaring. Den sterke signifikante korrelasjonen mellom pretest og posttest indikerer altså at resultat på posttest også vil være lavt. Når vi da kontrollerer for ulikhetene mellom skolene vil estimatene korrigeres for nivåforskjeller, og følgelig vil estimeringen av modellen skje på et mer nøyaktig grunnlag og gi mer forventningsrette estimater.

Erfaringen for smågruppelæreren er klart robust for inkludering av *poengsmå* og *interstudie*, endringene påfaller når vi inkluderer *aldersmå* og *poengstor*. Inkludering av *aldersmå* gir signifikant effekt fra erfaringsfunksjonen fordi den er en utelatt variabel som korrelerer med funksjonen. *Poengstor* og funksjonen er negativt korrelerte, dermed øker effekten av erfaring når *poengstor* legges til i modellen og koeffisienten inngår positivt.

Studiepoeng i matematikk fra høyere utdanning

Matematikkompetansen til lærerne i prosjektet måles ved antall studiepoeng i faget og fanges opp i modellen av tre ulike variabler.

Effekten av økt antall studiepoeng til smågruppelærer avhenger av hvor mange studiepoeng storgruppelærer har:

$$\frac{dPosttest}{dstupoe} = \beta_5 + \beta_9 poengstor \quad (11)$$

β_9 inngår negativt og jo flere studiepoeng storgruppelærer har, jo sterkere vil effekten fra dette leddet bli. Tolkningen blir tilsvarende for storgruppelærer.

Inkludering av kontinuerlig studiepoengvariabel til små- og storgruppelærer er av positiv, ikke-signifikant betydning for elevprestasjon. Inkludering av interaksjonsleddet gjør at separate effekt av antall studiepoeng må ses i sammenheng med leddet og tolkes med forsiktighet.

Produktet for begge lærernes samlede antall studiepoeng, *interstudie*, inngår negativt signifikant til ti prosent i modellen. Resultatet gir indikasjoner på en at økning i den ene lærerens studiepoeng, gitt nivået til den andre, vil ha negativ betydning for elevprestasjon. Dersom begge lærerne har det høyeste rapporterte antall studiepoeng, vil den negative effekten bli lik 110% av et standardavvik. En mulig tolkning av en slik sterk negativ effekt kan være at lærere med mange studiepoeng i matematikk utgjør en selektert gruppe med særskilt interesse for fagområdet. Dermed kan det spekuleres i om effekten vi observerer er et resultat av at dette er lærere som er mer opptatt av matematikk enn i undervisning og læringsmiljø i klasserommet.

Smågruppelærerens studiepoeng inntreer kun signifikant i modell (4), og får ved inkludering av *poengstor* en substansiell reduksjon i sin koeffisientverdi. Verdien er tilnærmet lik i modell (7), men med klart større standardavvik. Storgruppelærerens studiepoeng inntreer signifikant i modell (5). Estimaten får høyere standardavvik ved inkludering av interaksjonsleddet som følgelig endrer signifikansnivå²². Ingen av variablene for studiepoeng er robuste når skolefaste effekter pålegges.

Alder

I hovedmodellen inngår også smågruppelærers alder som en uavhengig variabel.

Alder til smågruppelærer følger naturlig samme mønster som erfaringsvariabelen, ut fra tabell A.5 i appendiks. Variabelen inntreer negativt signifikant til ett prosent i alle modellene. I (7) vil ett standardavviks økning i alder til smågruppelærer slå ut i 31.7% standardavvik lavere resultat på posttest. Inkludering av alder som kontroll gjør at både erfaring og kvadrert erfaring blir signifikante til respektive ett og fem prosent i modell (2). At koeffisientverdiene blir høyere positive er som forventet på grunn av den sterke positive korrelasjonen mellom erfaring og alder, og i modell (1) fanger erfaring opp noe av den negative effekten av alder.

At vi observerer en negativ effekt av alder kan forklares med at eldre lærere på grunnskolenivå kan oppleve jobben mer krevende. Barn i alderen syv til ni år har mye energi og både behøver og krever mer oppmerksomhet fra lærer. En naturlig tanke er at de eldste lærerne vil finne det mer utfordrende å jobbe i klasser med svært unge elever.

Smågruppeundervisning er en særegen form å drive undervisning på, det er dermed ikke utenkelig at yngre lærere vil være mer fleksibel til å tilpasse seg undervisningsformen i 1+1.

²² Angir sannsynligheten for type I-feil (forkasting av sann nullhypotese) ved hypotesetesting, og velges som det minste nivået for å forkaste nullhypotese.

Eldre lærere har ofte opparbeidet seg egne rutiner i løpet av karrieren som kan gjøre det vanskeligere å tilpasse seg. Likevel må resultatet tolkes med forsiktighet siden modellen vår kan være utsatt for uobserverbar støy, som potensielt påvirker estimatet.

I alle modellene er koeffisienten for pretest robust for inkludering av samtlige lærerkarakteristikker. Mye tyder derimot på at modell (1)- (6) er utsatt for utelatte variabler som er av betydning for posttest. Det trekkes i retning av å benytte den fullspesifiserte modellen der vi kontrollerer for initiale nivåforskjeller på tvers av skolene. Skolefaste effekter øker dessuten modellens forklaringskraft fra 0.442 til 0.478. Samlet gjør dette at spesifisering (7) med faste effekter prefereres framfor (6).

6.2 Alternative mål på lærerkompetanse

I hovedmodellen vår finner vi ingen separat effekt av lærernes studiepoeng i matematikk. Basert på tidligere litteratur kjenner vi til at ulike studier som benytter ulike mål på lærerkompetanse finner forskjellige resultat. Det motiverer for å forsøke og fange opp lærerkompetanse med alternative mål. I kapittel 6.2.1 inkluderer vi hvor mange år lærerne har hatt med matematikk på videregående skole til hovedmodellen og i kapittel 6.2.2 inkluderer vi kun et summert ledd for samlet antall studiepoeng for lærerne for å måle matematikkkompetanse.

6.2.1 Matematikk på videregående skole

Tabell 6.2: Utvidelse - Betydning av antall år matematikk på VGS

VARIABLER	(1)	(2)
std_pretest	0.658*** (0.0166)	0.660*** (0.0169)
erfarsmå	0.0457*** (0.0113)	0.0140 (0.0197)
erfarsmå ²	-0.000674** (0.000285)	-0.000186 (0.000441)
aldersmå	-0.0278*** (0.00510)	-0.0153** (0.00768)
poengsmå	0.00120 (0.00182)	-0.000787 (0.00189)
erfstor	0.0238*** (0.00908)	0.0176* (0.0101)
erfstor ²	-0.00104*** (0.000260)	-0.000740** (0.000303)
poengstor	0.000734 (0.00256)	-0.00242 (0.00285)
interstudie	-6.82e-05* (3.75e-05)	-1.87e-05 (4.07e-05)
mattesmåvgs		-0.0454 (0.0438)
mattestorvgs		0.146*** (0.0426)
Konstant	0.504*** (0.193)	0.156 (0.259)
Skolefaste effekter	Ja	Ja
Observasjoner	2,183	2,133
R ²	0.478	0.476

Robust standardavvik i parentes

*** p<0.01, ** p<0.05, * p<0.1

I dette delkapittelet ønsker vi å se om resultatene fra hovedanalysen er robust overfor inkludering av antall år lærerne har hatt med matematikk på videregående skole.

Motivasjonen er å sjekke om vi observerer en effekt hvis lærerne har hatt mye matematikk på videregående, når vi kontrollerer for antall studiepoeng.

Inkluderingen gjør at vi får testet en tilleggshypotese om at lærere som har mer matematikk enn obligatorisk krav på videregående har en større interesse for faget og potensielt underviser matematikk av høy kvalitet. Dersom hypotesen er sann vil vi observere en positiv effekt av variabelen på prestasjon.

Modell (1) i tabell 4 er hovedmodellen i tabell 3. Modell (2) er tilsvarende modell med inkluderte variabler som fanger opp små – og storgruppelærernes omfang av matematikk på videregående. Variabelen inngår ikke-kontinuerlig og tar verdien én hvis læreren har ett år matematikk, to hvis to år matematikk og følgelig tre hvis læreren har tre år med matematikk.

Vi finner i modell (1) signifikante effekter av alle inkluderte forklaringsvariabler med unntak av poengvariablene. I modell (2) observerer vi en positiv effekt av antall år på videregående for storgruppelæreren og motsatt effekt for smågruppelæreren, men effekten er ikke-signifikant. Inkludering av variablene reduserer signifikansnivået til erfaringsvariablene for både små – og storgruppelæreren.

For å gjennomføre en simultan test for å teste om kontrollvariablene er statistisk signifikant lik null, konstrueres følgende nullhypotese:

$$H_0: \text{mattevgssm\AA}=\text{mattevgstor}=0 \text{ mot}$$

$$H_A: \text{mattevgssm\AA}=\text{mattevgstor}\neq 0$$

Vi forkaster nullhypotesen på 1% og kan ikke si at variablene simultant ikke forklarer noe av variasjonen av posttestresultater.

Vi utfører også en tosidig test for variablene separat, der nullhypotesen er at variabelen er lik null mot en alternativhypotese om at den er ulik null. Vi beholder klart nullhypotesen for matematikk på vgs for smågruppelæreren. Koeffisientverdien til variabelen for storgruppelæreren er klart signifikant som gir indikasjon på at variabelen forklarer noe av variasjonen i testscore.

Det framkommer i tabell 4 at inkludering av matematikkvariablene for videregående fjerner den signifikante effekten av erfaring for smågruppelæreren. Variablene reduserer også betydningen til *aldersm\AA* og erfaringen til storgruppelæreren. Dette tyder på at erfaringsvariablene er sensitive for inkluderingen og fanger opp noe av den samme variasjonen som de to variablene.

Forklaringskraften til modellen reduseres med 0.002 og med bakgrunn i resonnementet over trekker det i retning av at variablene er overflødige.

6.2.2 Alternativt mål på studiepoeng

Tabell 6.3: Robusthetstest - Alternativt mål på studiepoeng

VARIABLER	(1)	(2)
std pretest	0.658*** (0.0166)	0.659*** (0.0166)
erfarsmå	0.0457*** (0.0113)	0.0457*** (0.0109)
erfarsmå ²	-0.000674** (0.000285)	-0.000746*** (0.000259)
aldersmå	-0.0278*** (0.00510)	-0.0262*** (0.00504)
poengsmå	0.00120 (0.00182)	
erfstor	0.0238*** (0.00908)	0.0203** (0.00883)
erfstor ²	-0.00104*** (0.000260)	-0.000937*** (0.000253)
poengstor	0.000734 (0.00256)	
interstudie	-6.82e-05* (3.75e-05)	
sumpoeng		-0.00248*** (0.000874)
Konstant	0.504*** (0.193)	0.662*** (0.180)
Skolefaste effekter	Ja	Ja
Observasjoner	2,183	2,183
R ²	0.478	0.477

Robust standardavvik i parentes
 *** p<0.01, ** p<0.05, * p<0.1

I hovedmodellen i tabell 3 finner vi at studiepoengvariablene til begge lærerne inngår positivt i modell (4) og (5) før samarbeidsindikatoren inkluderes. Dette kan gi indikasjoner på at summen av disse to variablene kan slå ut positivt på elevprestasjon når summen inngår som eneste mål på lærerutdanning.

Tabell 5 viser en modell der studiepoengvariablene for begge lærerne og interaksjonsleddet er erstattet med en sumvariabel. Variabelen inngår signifikant negativt til ett prosent.

Modellen gir ikke det resultatet indikasjonen tyder på, men funnet er heller ikke overraskende. Årsaken kommer av at interaksjonsleddet i modell (1) er et multiplikativt ledd som tar en maksimumsverdi på over 14 000 når to lærere med det høyeste nivået for studiepoeng betraktes. Sumvariabelen tar derimot en verdi lik 240 for disse lærerne. For modell (1) blir den negative effekten lik 0.85476 og i modell (2) blir tilsvarende effekt lik 0.6324. Vi ser altså at sumvariabelen ikke slår ut positivt på elevprestasjon.

Inkluderte variabler for smågruppelæreren er klart robuste overfor inkluderingen. For storgruppelæreren vil den negative effekten av erfaring inntre senere i karrieren enn i modell

(1), med redusert signifikans på det lineære leddet. En slik endring i storgruppelærers erfaring kan komme av at vi i modell (2) kun har inkludert en variabel for studiepoeng som inntreer negativt.

Vi konkluderer med at å benytte et alternativt mål på studiepoeng i hovedmodellen ikke endrer funnene våre.

6.3 Kort oppsummering

Oppsummert finner vi en positiv, signifikant effekt av pretest. Ergo vil elevprestasjon før tiltaket med smågruppeundervisning være av betydning for elevprestasjon etter tiltaket. Vi finner en positiv, avtakende korrelasjon mellom posttest og erfaring i yrket for både små- og storgruppelærer.

Hver for seg finner vi ingen signifikant effekt av studiepoeng for de to lærerne på elevprestasjon. Effekten av lærernes samlede antall studiepoeng er derimot av signifikant betydning for elevprestasjon. Resultatene gir indikasjoner på at erfaring er av positiv påvirkning på elevprestasjon, men vi ikke ser en tydelig effekt av studiepoeng i matematikk.

Inkludering av de alternative målene på utdanning, matte på videregående skole og summert variabel på studiepoeng, gir heller ikke et klart svar på effekt av lærerkompetanse på elevprestasjon. Resultatene gir indikasjoner på at matematikk på vgs for storgruppelæreren er korrelert med posttest, men inkluderte interessevariabler i hovedmodellen er ikke robuste overfor inkluderingen.

Vår modell rapporterer altså at det finnes korrelasjon mellom observerbare lærerkarakteristikk og elevprestasjon. Likevel eksisterer det et endogenitetsproblem vi ikke er i stand til å kontrollere for. Det vil dermed være nyttig å gjennomføre en sensitivitetsanalyse der vi angriper problemstillingen fra andre vinkler for å se om smågruppelærers erfaring er robust.

7. Sensitivitetsanalyse

I dette delkapittelet ønsker vi å undersøke om resultatene fra hovedmodellen er robuste når alternativ value added spesifikasjonen benyttes, delutvalg analyseres og utvidelser av modellen gjennomføres. Value added presenteres først, deretter studerer vi om hovedmodellen for utvalget samlet endrer seg når vi kjører regresjoner for de to trinnene separat. Tilslutt pålegges ulike kontrollvariabler i hovedanalysen for å undersøke om resultatene er upåvirket av inkluderingen.

7.1 Value added spesifikasjon

Første del av sensitivitetsanalysen er en value added spesifikasjon der resultatene rapporteres i tabell 6. Det som skiller denne fra hovedmodellen er at vi pålegger en restriksjon om at koeffisienten foran pretest settes lik én, med hensikt i å redusere problemene knyttet til estimeringen av PFU. Det er hensiktsmessig å se om estimatene for lærererfaring og utdanning er robuste når vi til en viss grad, med pålagt restriksjon, kontrollerer for uobserverbare faktorer som påvirker elevprestasjon.

Fra hovedmodellen rapporteres en klar, positiv signifikant koeffisient for pretest med effekt på 0,658. I test om at denne koeffisienten er lik én, forkastes hypotesen på 1% signifikansnivå. Konsekvensen av restriksjonen er at vi i value added *underestimerer* endringen i testscore, *gitt* absoluttverdi på pretest og verdier på forklaringsvariablene.

For å forklare analytisk benytter vi en forenklet variant av modellen:

Hovedmodellen gir:

$$\begin{aligned} Posttest_{igsk} = & \textit{konstant} + 0,658Pretest_{igsk} \\ & + \textit{øvrige forklaringsvariabler}_{igsk} \end{aligned} \quad (12.1)$$

Flytter Pretest på venstresiden:

$$\begin{aligned} Posttest_{igsk} - 0,658|Pretest_{igsk}| \\ = & \textit{konstant} + \textit{øvrige forklaringsvariabler}_{igsk} \end{aligned} \quad (12.2)$$

Når vi multipliserer et vilkårlig tall med et tall mindre enn én vil vi få en mindre verdi enn initialverdien på det vilkårlige tallet. Setter deretter inn gjennomsnittsverdiene til de standardiserte pre- og posttestene for behandlingsskolene samlet, -0.0568641 og 0.0577864.

Venstresiden er, med koeffisientverdi fra MKM og absoluttverdi på standardisert pretest, 0.02037.

VA gir:

$$\begin{aligned} Posttest_{igsk} - 1|Pretest_{igsk}| & \qquad \qquad \qquad (13) \\ & = konstant + \text{\textit{øvrige forklaringsvariable}}_{igsk} \end{aligned}$$

Setter inn de samme verdiene som ovenfor og finner at venstresiden, med restriktiv koeffisient, er 0.00092. Avviket og følgelig feilestimeringen i dette tilfellet er 0.01945, og sammenlignet med standardavviket for behandlingsskolene er avviket på 2.4%.

Modell (7) i tabell 6 rapporterer resultatene fra VA-spesifikasjonen, der tilsvarende forklaringsvariabler som i modell (7) tabell 3 er inkludert. Vi finner en signifikant effekt av erfaringsfunksjonen og alderen til smågruppelærerne. Erfaringen til storgruppelærerne inntreer kun med statistisk signifikant betydning i det kvadrerte leddet og de øvrige variablene har ingen signifikant effekt.

Tabell 7.1: Avhengig variabel er forskjell i standardisert testscore for 2. og 3. trinn samlet

VARIABLER	Value added metode (VA)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
erfarmsmå	-0.00227 (0.00427)	0.0219*** (0.00517)	0.0245*** (0.00573)	0.0339*** (0.00647)	0.0419*** (0.00781)	0.0424*** (0.00810)	0.0697*** (0.0123)
erfarmsmå ²	-0.000109 (0.000137)	-0.000272** (0.000138)	-0.000377*** (0.000184)	-0.000676*** (0.000201)	-0.000628*** (0.000224)	-0.000639*** (0.000229)	-0.00103*** (0.000313)
aldersmå		-0.0190*** (0.00243)	-0.0193*** (0.00246)	-0.0199*** (0.00258)	-0.0266*** (0.00341)	-0.0268*** (0.00347)	-0.0393*** (0.00543)
poengsmå			-0.000248 (0.000667)	0.00122* (0.000724)	-0.000409 (0.000780)	-0.000174 (0.00154)	-0.000717 (0.00207)
erfstor				-0.0233*** (0.00688)	-0.0297*** (0.00735)	-0.0301*** (0.00743)	0.0153 (0.00986)
erfstor ²				0.000549*** (0.000186)	0.000585*** (0.000199)	0.000596*** (0.000202)	-0.000894*** (0.000278)
poengstor					0.00222*** (0.000855)	0.00259 (0.00221)	-0.00107 (0.00287)
interstudie						-5.63e-06 (3.18e-05)	-5.53e-05 (4.29e-05)
Konstant	0.205*** (0.0273)	0.724*** (0.0734)	0.739*** (0.0906)	0.841*** (0.114)	1.082*** (0.144)	1.071*** (0.160)	0.955*** (0.208)
Skolefaste effekter							
Observasjoner	Nei 3,819	Nei 3,634	Nei 3,261	Nei 2,632	Nei 2,183	Nei 2,183	Ja 2,183
R ²	0.004	0.019	0.020	0.036	0.052	0.052	0.120

Robust standardavvik i parentes

*** Signifikant på 1% signifikansnivå (p<0.01)

** Signifikant på 5% signifikansnivå (p<0.05)

* Signifikant på 10% signifikansnivå (p<0.1)

Erfaring

Som i hovedmodellen støtter også fortegnene på erfaringsfunksjonene i value added spesifikasjonen opp mot tidligere litteratur. Vi finner endrede koeffisientverdier for erfaring fra hovedmodellen som en følge av den ugyldige restriksjonen pålagt koeffisienten foran pretest.

I tabell 6 vil et år erfaring for en smågruppelærer uten erfaring påvirke forskjellen i testscore med 0.06867, altså rapporteres et positivt VA.

I likhet med hovedmodellen vil erfaringsfunksjonen inngå som en ikke-lineær funksjon og effekten av erfaring vil flate ut etter 33.8 år. Sammenliknet med hovedmodellen, som rapporterte utflating etter 33.9 år, finner vi at toppunktet for begge spesifikasjonene er tilnærmet likt, men vi kan ikke konkludere med at de følger den samme kurven.

For storgruppelæreren er effekten av erfaring av en lavere størrelsesorden enn i MKM. I et tilfelle der rapportert erfaring er lik null gir VA en positiv effekt lik 0.01441 av en økning i erfaring lik ett år, mens MKM viser en positiv effekt på 0.045026. Direkte sammenligning av effektene krever innsetting av valgt verdi på pretestresultat.

Studiepoeng i matematikk fra høyere utdanning

Inkluderte variabler for studiepoeng inntreer negativt, ikke-signifikant i VA-spesifikasjonen.

Variabelen som fanger opp studiepoeng hos smågruppelæreren er kun signifikant i modell (4), når erfaringen til storgruppelæreren legges til. Til kontrast er variabelen for storgruppelæreren signifikant fram til inkluderingen av interaksjonsleddet. Det framkommer fra modell (5) at *poengstor* fanger opp noe av den svake positive effekten til *poengsmå*. Videre vil interaksjonsleddet påvirke de to variablene i modell (6), og vi observerer ingen signifikante effekter av studiepoeng i modell (6) og (7).

En økning i koeffisientverdien for erfaring og alder hos smågruppelæreren kan forklares med at value added overestimerer effekten av pretestresultater slik at negativ effekt fra hovedmodellen forsterkes i VA.

Hvis vi hadde valgt en VA-spesifikasjonen som hovedmodell ville den foretrukne modellen vært modell (5) og ikke (7). Årsaken til dette er at vi i (5) observerer flere signifikante effekter enn det som rapporteres i modell (7).

Hovedmodellen er derimot ikke robust overfor en VA-spesifikasjon. Samtlige av variablene får økte estimerte standardavvik og endrede koeffisientverdier. Spesifikasjonen pålegger en

svært streng restriksjon på pretest, og konsekvensen er at øvrige estimat blir påvirket. Resultatene i tabell 3 er mer forventningsrett og tillater påvirkning fra uobserverbare elev- og familiekarakteristikk mellom testene. Selv om de uobserverbare faktorene skaper problemer med inferens, er hovedmodellen å foretrekke framfor value added spesifikasjonen.

7.2 Trinnene separat

I kapittel 6 rapporterte vi resultater samlet for de to trinnene i utvalget. For å undersøke om effektene er heterogene for klassene gjennomføres regresjonene separat. Resultatene er rapportert i tabell A.3.1 og A.3.2 i appendiks.

Tabell A.3.1 for 2. trinn viser en robust effekt av pretest når trinnene betraktes separat. Effekten av smågruppelærerens erfaring følger identisk mønster som i hovedregresjonen, men rapportert effekt er betraktelig sterkere. Aldersvariabelen inntreffer med sterkere negativ effekt.

Erfaringen til storgruppelæreren endrer fortegn på både det lineære og det kvadrerte leddet. I hovedmodellen rapporteres en signifikant positiv, men avtakende effekt av erfaringsfunksjonen. For 2. trinn ser vi derimot en signifikant negativ, avtakende effekt. Tolkningen blir motsatt, slik at en storgruppelærer med mindre enn tolv års erfaring²³ vil ha negativ effekt på posttestscore. Påvirkningen fra smågruppelærernes erfaring vil til sammenligning ha positiv effekt fram til lærerne har i overkant av 24 år med erfaring.

Studiepoengvariabelen til smågruppelærer, *poeng2*, inngår med samme fortegn, men effektene er sterkere for 2. trinn. Tilsvarende variabel for storgruppelærer har også samme fortegn som i hovedmodellen og er robust for 2. trinn. Ingen av variablene inngår derimot signifikant i tabell A.3.1.

Interaksjonsleddet utgjør den største forskjellen mellom hovedmodellen i de to tabellene. Variabelen går fra å være signifikant til ti prosent i tabell 3 til å være signifikant til ett prosent for 2. trinn. Da er en økning i den ene lærerens studiepoeng, gitt nivået til den andre, av betydning for 2. trinn og variabelen inngår med positiv effekt lik 0.00289.

For 2. trinn observerer vi en predikert posttestscore på 23.1 i et tilfelle hvor en elev har prestert med 26 poeng på pretest, smågruppelæreren er 25 år, har ett år med erfaring og 60 studiepoeng i matematikk og tilsvarende karakteristikk gjelder for storgruppelæreren. Til

²³ Her benyttes samme utregning som i kapittel 6.1.

sammenlikning vil tilsvarende situasjon i det samlede utvalget gi predikert posttestresultat lik 16.4. Effekten er altså sterkere for 2. trinn.

Det er kun estimatet for pretest og studiepoeng for storgruppelæreren fra hovedmodellen som er robust når kun 2. trinn betraktes. Endringen i begge erfaringsfunksjonene er av en betydelig størrelse og effekten av erfaring tidlig i lærerkarrieren er i underkant av et standardavviks økning i posttestscore. Toppunktet av erfaring inntreffer langt tidligere enn for trinnene samlet og for 2. trinn observeres en betraktelig effekt av de første årene med erfaring før en rask utjevning av effekten. For storgruppelæreren er påvirkningen av erfaring tilnærmet lik, men til kontrast er fortegnene motsatte.

Storgruppelæreren vil i starten av sin karriere påvirke sine elever negativt fram til læreren har tolv år med erfaring. Det betyr at lærere med mindre enn tolv år fartstid i yrket har en negativ effekt på elevenes posttestresultat. Funnet strider mot tidligere litteratur for elevprestasjon og lærererfaring, slik at effekten framstår som kontraintuitivt. Hvis det skal spekuleres i hvorfor resultatene er slik, kan det være hensiktsmessig å fokusere på klasseromssituasjon og klassetrinn. Elevene i 2. trinn er mellom syv og åtte år og har følgelig et kortere konsentrasjonsspenn sammenlignet med eldre elever. Det kan tenkes at en lærer uten erfaring, eller med lite erfaring, ikke har opparbeidet seg den nødvendige verktøykassen som kreves for å øke elevenes prestasjoner.

Resultater for 3. trinn rapporteres i tabell A.3.2, modell (7). Separat for 3. trinn er heller ikke resultatene fra hovedmodellen robuste, unntaket er koeffisienten foran pretest. Erfaring hos smågruppelærer inngår signifikant til ti og ett prosent, med like fortegn som i hovedmodellen, men med sterkere effekt.

Effekten av erfaring til storgruppelærer er som for 2. trinn negativ, men avtakende.

Signifikansnivået til funksjonen er redusert fra ett til henholdsvis fem og ti prosent. Det rapporteres en utflating etter 21 år av erfaring på tredjeklassingers prestasjon.

Smågruppelæreren har positiv påvirkning fram til 14 år med erfaring er opparbeidet.

Resultatene tyder på effekten for 3. trinn av smågruppelærers erfaring er lavere sammenliknet med 2. trinn.

Ingen av de øvrige forklaringsvariablene har tolkning som statistisk signifikante for elevprestasjon i 3. trinn. Inkludering av skolefaste effekter reduserer antall signifikante variabler fra seks til fem, og variablene inngår med sterkere signifikans i modell uten faste effekter.

Smågruppelærers alder inngår i modell (7) svakt positivt, men er ikke av signifikant betydning. I hovedmodellen viste vi til en sterk negativ korrelasjon mellom alder og posttest. At vi nå observerer en ikke-signifikant positiv effekt kan være et resultat av at prestasjonsnivå til elever i 3. trinn ikke varierer med smågruppelærernes alder.

Vi finner ingen signifikant effekt av studiepoeng, uavhengig om læreren underviser i små- eller storgruppen, og effekt av samarbeidsindikatoren er heller ikke av betydning. Fortegnene på koeffisientene er lik som i hovedmodellen, men usikkerheten bak estimatene er større når vi ser på trinnene separat og variablene inngår ikke-signifikant.

Vi konkluderer med at effektene er heterogene for trinnene og hovedmodellen er ikke robust overfor analyse av trinnene hver for seg.

7.3 Utvidelse

I dette delkapittelet ønsker vi å undersøke om resultatene i tabell 3 er robuste. Vi inkluderer først antall uker elevene er ute i smågruppene. Deretter inkluderes tidligere utelatte ekstremverdier i studiepoengvariablene før vi tilslutt inkluderer et interaksjonsledd mellom erfaringen til begge lærerne som en alternativ proxy for samarbeid.

Hensikten med våre inkluderinger i dette kapitelet er å kartlegge hvorvidt betydningen av smågruppelærers erfaring fra hovedmodellen er robust når flere variabler legges til. Dersom erfaringsvariabelen viser seg å være robust, kan det være en indikasjon på at inkluderte mål på smågruppelærers erfaring faktisk fanger opp den sanne effekten av erfaring, slik at vår valgte modellspesifikasjon framstår som en rimelig tilnærming til produktfunksjonen for utdanning.

7.3.1 Antall uker i smågruppen

Tabell 7.2: Utvidelse - Betydning av antall uker i smågruppen

VARIABLER	(1)	(2)
std_pretest	0.658*** (0.0166)	0.656*** (0.0177)
erfarsmå	0.0457*** (0.0113)	0.0458*** (0.0113)
erfarsmå ²	-0.000674** (0.000285)	-0.000680** (0.000287)
aldersmå	-0.0278*** (0.00510)	-0.0277*** (0.00509)
poengsmå	0.00120 (0.00182)	0.00119 (0.00182)
erfstor	0.0238*** (0.00908)	0.0237*** (0.00917)
erfstor ²	-0.00104*** (0.000260)	-0.00103*** (0.000263)
poengstor	0.000734 (0.00256)	0.000707 (0.00256)
interstudie	-6.82e-05* (3.75e-05)	-6.76e-05* (3.76e-05)
antalluker23		0.00117 (0.00811)
Konstant	0.504*** (0.193)	0.490** (0.202)
Skolefaste effekter	Ja	Ja
Observasjoner	2,183	2,159
R ²	0.478	0.475

Robust standardavvik i parentes
 *** p<0.01, ** p<0.05, * p<0.1

Det finnes en mulighet for at antall uker elevene har vært ute i smågruppeundervisning kan ha betydning for elevprestasjon. Vi ønsker derfor å undersøke om resultatene våre er robuste når denne variabelen inkluderes.

Variabelen er en gjennomsnittsvariabel for trinnene samlet der ukeantall lik null og over 18 er ekskludert og resultatene rapporteres i tabell 7, modell (2).

Det framkommer at forklaringsvariablene i hovedmodellen er robuste overfor inkludering av variabelen, som følge av at koeffisientverdiene og standardavvikene endres marginalt.

Variabelen har ikke signifikant effekt på predikert posttest, og ved gjennomføring av test om at koeffisienten er lik null beholder vi hypotesen på 88%.

Med bakgrunn i diskusjonen ovenfor trekker det i retning av at antall uker ikke forklarer noe av variasjonen i posttestresultat og de inkluderte forklaringsvariabler påvirkes heller ikke av variabelen. Vi forkaster tilleggshypotesen om at antall uker i smågruppen er av betydning for predikert resultat.

7.3.2 Inkludering av ekstremverdier på studiepoeng

Tabell 7.3: Robusthetstest - Inkludering av ekstremverdier for studiepoeng i matematikk fra høyere utdanning

VARIABLER	(1)	(2)
std_pretest	0.658*** (0.0166)	0.666*** (0.0144)
erfarsmå	0.0457*** (0.0113)	0.0318*** (0.00747)
erfarsmå ²	-0.000674** (0.000285)	-0.000700*** (0.000182)
aldersmå	-0.0278*** (0.00510)	-0.0162*** (0.00411)
posmåekstrem		0.000177 (0.000712)
erfstor	0.0238*** (0.00908)	-0.00282 (0.00764)
erfstor ²	-0.00104*** (0.000260)	-5.44e-06 (0.000224)
postorekstrem		-0.000452 (0.00123)
interstuekstrem		3.24e-06 (1.35e-05)
poengsmå	0.00120 (0.00182)	
poengstor	0.000734 (0.00256)	
interstudie	-6.82e-05* (3.75e-05)	
Konstant	0.504*** (0.193)	0.301* (0.158)
Skolefaste effekter	Ja	Ja
Observasjoner	2,183	2,832
R ²	0.478	0.484

Robust standardavvik i parentes
 *** p<0.01, ** p<0.05, * p<0.1

I hovedanalysen inkluderer vi en variabel som forklarer hvor mange studiepoeng i matematikk lærerne i behandlingsskolene har. Denne variabelen ekskluderer alle ekstremverdier i intervallet [0,10] og over 140. I dette delkapittelet undersøker vi om resultatene fra hovedanalysen er robuste når ekstremverdiene inkluderes.

Resultatene er rapportert i tabell 8, modell (2). Koeffisientene foran pretest og kvadrert erfaringsledd for smågruppelæreren er robuste. De øvrige forklaringsvariablene endrer koeffisientverdi, i betydelig, men ulik grad. Signifikansnivået til storgruppelærerens erfaring går fra klart signifikant til ikke-signifikant. Effekten av erfaring hos smågruppelæreren har uendret signifikansnivå, men koeffisientverdien er redusert.

Ingen av de inkluderte variablene med ekstremverdier er statistisk signifikant i forklaringen av posttestresultat. Studiepoeng for storgruppelæreren og interaksjonsleddet har endrer fortegn når vi tillater ekstremverdier, mens fortegnet for smågruppelæreren endres ikke.

Når vi tester variablene kjører vi først en test på hver av variablene etterfulgt av en simultan test. I alle de enkle testene beholdes nullhypotese om at variablene hver for seg er lik null. I den simultane testen konstrueres nullhypotesen slik:

$$H_0: \text{posmåekstrem} = \text{postorekstrem} = \text{interstuekstrem} = 0 \text{ mot}$$

$$H_A: \text{posmåekstrem} = \text{postorekstrem} = \text{interstuekstrem} \neq 0$$

Nullhypotesen beholdes til 88% i den simultane testen, mens de øvrige hypotesene beholdes til 71% og høyere signifikansnivå.

Vi konkluderer med at tilnærmet halvparten av forklaringsvariablene ikke er robuste når ekstremverdier inngår. Variablene for studiepoeng er heller ikke signifikant i modellen og forklarer ikke med statistisk grunnlag noe av variasjonen i posttestresultat.

Dermed kan vi ikke si at hovedmodellen er robust overfor inkluderingen, og resultatene tyder på at inkludering av ekstremverdiene påvirker regresjonslinjen til populasjonen. Hvordan linjen påvirkes avhenger av hvilken studiepoengvariabel vi studerer og hvilke verdier den andre læreren har.

Ved en antakelse om at de to lærerne har 90 studiepoeng hver, som tilsvarer halvannet år med matematikk, vil lærerne uten ekstremverdier ha negativ effekt på sine elever lik 0.3784.

Tilsvarende lærere, med ekstremverdier inkludert, vil ha en positiv effekt lik 0.0001494.

Resultatene tyder på at de ni lærerne med over 165 studiepoeng i matematikk har en positiv påvirkning på klassene sine.

Det er to lærere i datasettet som har rapportert 240 studiepoeng i matematikk.

Syv lærere har mellom 165 og 240 studiepoeng i matte.

7.3.3 Inkludering av interaksjonsledd for erfaring

Tabell 7.4: Robusthetstest - Interaksjonsledd mellom erfaring som samarbeidsindikator

VARIABLER	(1)	(2)
std_pretest	0.658*** (0.0166)	0.657*** (0.0167)
erfarsmå	0.0457*** (0.0113)	0.0510*** (0.0123)
erfarsmå ²	-0.000674** (0.000285)	-0.000495* (0.000285)
aldersmå	-0.0278*** (0.00510)	-0.0275*** (0.00508)
poengsmå	0.00120 (0.00182)	-0.000898 (0.00129)
erfstor	0.0238*** (0.00908)	0.0268*** (0.00908)
erfstor ²	-0.00104*** (0.000260)	-0.000886*** (0.000276)
poengstor	0.000734 (0.00256)	-0.00311*** (0.00110)
interstudie	-6.82e-05* (3.75e-05)	
intererfaring		-0.000617** (0.000308)
Konstant	0.504*** (0.193)	0.554*** (0.183)
Skolefaste effekter	Ja	Ja
Observasjoner	2,183	2,183
R ²	0.478	0.478

Robust standardavvik i parentes
 *** p<0.01, ** p<0.05, * p<0.1

I hovedmodellen inngår interaksjonsleddet mellom studiepoeng til begge lærerne i modellen svakt negativt signifikant til ti prosent. Å anvende interaksjonsleddet som en proxy for samarbeid er ikke nødvendigvis en god tilnærming, og en potensielt bedre proxy vil være et interaksjonsledd mellom erfaring.

Resultatene rapportert i tabell 9 viser tydelig at koeffisienten til pretest, begge de lineære erfaringsvariablene og alder er robuste når definisjonen for samarbeid endres.

Hovedmodellen rapporterer positiv effekt av studiepoeng for begge lærerne, i motsetning er fortegnene for begge variablene endret i modell (2). De kvadrerte erfaringsvariablene inngår fortsatt signifikant, men den negative effekten er redusert. Resultatet er ikke overraskende siden interaksjonsleddet er negativt signifikant og fanger opp noe av effekten til de kvadratiske funksjonene for erfaring.

For å illustrere deriveres PFU med hensyn på erfaring når vi tar utgangspunkt i smågruppelæreren:

$$\frac{\partial Posttest_{i,gsk}}{\partial erfarsm\grave{a}_{i,gsk}} = 0.0510 - 0.00099erfarsm\grave{a} - 0.000617erfstor$$

Når vi deriverer PFU ser vi at effekten av en endring i variablene vil avhenge av erfaringen til begge lærerne. Hvis begge initialt har mye erfaring vil den marginale endringen i posttestscore bli lav som følge av at de to siste leddene i derivasjonsuttrykket inntreer negativt. For to lærere som har gjennomsnittlig erfaring, lik tolv år og 13 år for henholdsvis små- og storgruppelærer, vil predikert elevprestasjon øke med 3,1 prosent av standardavvik på posttest når vi ser på en marginal økning i smågruppelærerens erfaring.

Den samlede effekten av lærernes gjennomsnittlige erfaring er her estimert til 0.6427, som er høyere sammenlignet med effekten rapportert kapittel 6.1 på 0.5736.

Endringen i koeffisientverdiene til studiepoengvariablene kan forklares med at interaksjonsleddet i modell (1) inngår negativt. Når vi ekskluderer variabelen vil følgelig de to variablene fange opp effekten og vi observerer negative koeffisienter foran variablene i modell (2). Samlet effekt av gjennomsnittlig studiepoeng hos lærerne i de to modellene er ulike, og vil i modell (2) ha en sterkere negativ effekt lik 0.119.

Sett i lys av hvordan de to samarbeidsindikatorne er definert, er modell (1) relativt robust for den alternative samarbeidsindikatoren. Dette funnet støtter opp under at inkludering av en samarbeidsindikator som faktisk fanger opp den reelle graden av samarbeid er krevende. Det er vanskelig å kontrollere for individuelle synspunkt for hvordan et samarbeid fungerer og graden av det faktiske samarbeidet. I vår modell og med vårt datamateriale er vi ikke i stand til å inkludere mål for det uobserverbare samarbeidet.

8. Oppsummering og diskusjon

I analysen har vi betraktet hvordan bestemte lærerkarakteristikk kan knyttes til elevprestasjon. For å gi en tolkning av resultatene vil det være naturlig å se de i sammenheng med tidligere empiriske studier på lærerkarakteristikk og smågruppeundervisning. Siden det benyttes et unikt datasett fra et særegent prosjekt som fremdeles pågår, forventer vi ikke nødvendigvis at alle resultatene våre er lik funn fra tidligere studier.

Slik som Harris og Sass (2011) finner også vi positiv avkastning av erfaring på elevprestasjon. I vår analyse finner vi at effekten er positiv fram til 33.9 år for smågruppelærerne og fram til 11.4 år for storgruppelærerne. Resultatet vårt sammenfaller med Harris og Sass' som finner at avkastningen er størst de fem første årene av karrieren og er av positiv, men svakere betydning, i årene etter.

Vi finner at lærererfaring er av svært ulik betydning for de to lærertypene i prosjektet. Tilsvarende dokumenterer Kraft og Papay (2014) at avkastningen av erfaring varierer mellom individuelle lærere og mellom lærergrupper. Vårt funn for ulikt toppunkt for erfaring virker derfor ikke urimelig. Mueller (2013) rapporterer at lærererfaring ikke er av betydning i store klasser. Noe av det samme mønsteret ser vi for storgruppelæreren i vår analyse med toppunkt for erfaring etter 11.4 år, som er svært tidlig sammenliknet med smågruppelæreren. Forøvrig er vi klar over at våre funn bærer preg av målefeil i variablene for storgruppelæreren.

Fra tidligere empirisk forskning er det utfordrende å finne en generell sammenheng mellom erfaring og elevprestasjon. Mye trekker derimot i retning av at lærererfaring er av størst betydning tidlig i karrieren, og at effekten videre trolig er positiv, men avtakende ved generering av mer erfaring. Våre analyser føyer seg dermed inn i rekken av studier som lener seg mot disse resultatene.

I hovedmodellen vår anvender vi små- og storgruppelærers samlede antall studiepoeng i matematikk som forklaringsvariabler for elevprestasjon, men vi finner ingen signifikant effekt. Variablene er inkludert for å måle om matematikkutdanning korrelerer med elevprestasjon, noe vi ikke finner tegn til. Et slikt resultat stemmer overens med Harris og Sass (2011) som viser til at det ikke er en sammenheng mellom produktivitet i klasserommet og læreres utdanningsnivå målt ved blant annet opptakseksamen til universitetet. Funnene sammenfaller gitt at opptakseksamen fanger opp effekten av lærernes utdanningsnivå.

Som et forsøk på å teste påstanden om at lærere med lik matematikkompetanse trolig samarbeider bedre og genererer større elevprestasjon inkluderes interaksjonsledd mellom de to studiepoengvariablene. Vår modell rapporterer negativ, signifikant effekt på posttest av samlet studiepoeng. Tolkningen av leddet er at to lærere med mange studiepoeng har en sterkere negativ effekt på elevprestasjon enn to lærere med lite poeng.

Den negative tilleggseffekt vi finner av utdanning kan tenkes å være en konsekvens av at interaksjonsleddet er drevet av tilfeldigheter. Trolig vil et valg om å ta mye matematikkutdanning i større grad avhenge av interesser, preferanser og andre uobserverbare kjennetegn ved lærerne i prosjektet. En potensiell effekt av bedret samarbeid ved en økning i antall studiepoeng fanges derfor ikke opp av modellen.

Enzi (2017) finner, basert på lærerresultat fra videregående skole, at lærerne som befinner seg i øvre kvartil i fordelingen av valgte evnemål, kan assosieres med større grad av effektivitet i klasserommet. Den positive retningen på våre separate poengvariable sammenfaller med en slik vag tolkning. Imidlertid benytter Enzi et annet mål på kompetanse og våre variable er ikke signifikante.

Resultatet kan kobles opp mot vår modellutvidelse hvor vi inkluderer hvor mange år matematikk 1+1-lærerne har hatt på vgs. Vi finner derimot ingen effekt av å inkludere matematikk fra videregående skole når vi samtidig kontrollerer for studiepoeng. Variablene fanger derfor trolig opp samme informasjon om lærerne, slik at et mønster for kompetanse er vanskelig å finne.

Det er for utdanningsnivå og kompetanse for lærere i skolen langt vanskeligere å trekke en generell konklusjon for betydningen på elevprestasjon. Årsaken ligger i at det benyttes ulike mål for utdanning i forskjellige analyser.

I vår analyse er det også utfordrende å finne et mål på matematikkompetanse som ikke korrelerer med individuelle, uobserverbare kjennetegn. At vi finner en negativ effekt av mye matematikkutdanning underbygges ikke fra humankapitalteorien, hvor man vil forvente en positiv avkastning på elevprestasjon av høyt utdannede lærere. Det kan indikere at mange lærere med mye utdanning i matematikk har uobserverbare kjennetegn som bidrar negativt til elevenes læringsutbytte.

Det kan se ut til at smågruppeundervisning er av betydning for elevprestasjon også i 1+1 prosjektet. For øvrig har ikke vi i vår analyse sammenliknet med posttestscore for elevene i

kontrollgruppen, og vi har derfor ikke grunnlag til å uttale oss om smågruppetiltaket har effekt i form av en endring i forskjell i testscorene mellom elever i behandling- og kontrollskoler. Gitt den positive, signifikante effekten av erfaring til smågruppelærer er det derimot et godt grunnlag for å tolke denne faktoren til å ha betydning for elevprestasjon.

8.1 Styrker, svakheter og videre forskning

I de fleste empiriske studier er formålet å finne kausale effekter. Det finnes uendelig mange faktorer som påvirker elevprestasjon, og en isolering av disse er vanskelig. Fordelen med et eksperiment er at man til en viss grad kan sikre *ceteris paribus*²⁴ forutsetningen.

Et randomisert kontrollert eksperiment benyttes for å studere effekten av en politikkendring. Utfordringen uten eksperiment er at politikken ikke er eksogen for de som utsettes for den, slik som ved en klassestørrelsesreduksjon, og det kan ikke kontrolleres for seleksjonsskjevhet.

Med et eksperiment vil den endogene klassestørrelsen fjernes dersom fordelingen til eksperimentet er tilfeldig. Da vil ikke foreldre, lærere eller skoleledelse kunne selekere seg til disse klassene.

Skolene som er i behandlingsgruppen ansetter selv et årsverk til skolen og én utfordring er at lærere som søker seg til disse jobbene utgjør en selektert gruppe. Vi observerer ingen systematiske forskjeller mellom små- og storgruppelæreren. Problemet ligger ikke i de observerbare forskjellene mellom lærerne, men i de potensielle uobserverbare forskjellene mellom dem. Denne typen seleksjon har vi ikke mulighet å kontrollere for, siden vi ikke har informasjon om uobserverbare kjennetegn. Dersom det er spesielle individkarakteristikker som avgjør om du blir lærer eller ikke har vi systematisk seleksjon i modellen. Estimaten blir ikke forventningsrette som følge av at vi har et endogenitetsproblem i data²⁵.

I hovedmodellen benyttes interaksjonsledd mellom studiepoeng som samarbeidsindikator, mens i sensitivitetsanalysen inngår erfaring som mål på samarbeid. Ingen av leddene fanger opp samarbeidet mellom lærerne i tilstrekkelig grad. Effekten av samarbeid kan ikke måles på samme måte som inkluderte lærerkarakteristikker, og bruk av studiepoeng og erfaring som potensielle proxyer viser seg å ikke være tilstrekkelig. En eventuell annen tilnærming er å

²⁴ *Ceteris paribus* er latinsk og betyr *alt annet likt*. Forutsetningen gjør at effekten fra en variabel ses når øvrige forklaringsvariabler holdes konstant.

²⁵ Brudd på forutsetning IV, presentert i kapittel 5.

inkludere spørsmål i spørreundersøkelsen på lærernivå, slik at hver enkelt små- og storgruppelærer rapporterer sin oppfattelse av graden av samarbeid.

Lærernes pedagogiske utdanning kan tenkes å påvirke hvordan de formidler matematikk. Et videre forskningsmoment kan være å innhente observasjoner for lærernes pedagogiske fag og se på forskjeller mellom effekten på studiepoeng i matematikk. Det er muligens slik at lærere med høy pedagogisk kunnskap har evne til å benytte et mer effektivt undervisningsopplegg.

Aggregeringen av smågruppelærerne er utført på klassenivå. Som følge av at vi sammenligner lærere innen en skole i vår analyse kan vi få problemer hvis lærerne er like, men har ulik utvikling. Noen lærere er av høy kvalitet fordi de har mye erfaring, mens andre lærere med tilsvarende erfaring er av lav kvalitet. Det er en mulighet for at lærere forbedrer seg til ulike hastigheter og uten lærerfaste effekter er vi ikke i stand til å kontrollere for dette. Å inkludere lærerdummier vil skape problemer med for få frihetsgrader²⁶ siden antallet observasjoner måles ved andelen av lærere.

En kritisk forutsetning vi har benyttet oss av i analysen er at de skolefaste effektene ikke korrelerer med inkluderte forklaringsvariabler og restleddet. Hvis det innen kommunen eksisterer områder med ulike sosioøkonomiske forhold, vil elevprestasjonene trolig påvirkes av hvilket område de bor i og det skolespesifikke leddet vil da korrelere med forklaringsvariablene.

Prosjektet har per dags dato ingen tidshorisont i andre variabler enn i testene. Muligheten for transformasjon eller differensiering er ikke eksisterende, og for videre analyser anbefales det å inkludere elev- og familiekarakteristikker, potensielle bedre mål på samarbeid og benytte andre alternative estimeringsmetoder.

²⁶ Frihetsgrader er antall avvik fra gjennomsnittet som kan variere fritt og måler i hvor stor grad en modell kan variere. Formel for beregning: $df=n-(k+1)$. n =antall observasjoner, k =antall estimerte parametere og 1 representerer konstantleddet.

9. Konklusjon

Analysen i denne oppgaven er gjennomført med utgangspunkt i et tverrsnittsdatamateriale med individuelle testresultater for elevene i 1+1 prosjektet før og etter tiltaket med smågruppeundervisning. Vi har benyttet informasjon for i skoleåret 2016/17 for å teste hypotesen om at lærererfaring og utdanning er av betydning for elevprestasjon. Inkluderte variabler for lærerkarakteristikker har vi delt opp i kategoriene små- og storgruppelærer.

Vi finner ved minste kvadraters metode at begge lærernes erfaring er av betydning for elevprestasjon, men funnet er kun forventningsrett for smågruppelærerne. For antall studiepoeng separat for lærerne finner vi ingen signifikant effekt, mens for det samlede antall studiepoeng finner vi en negativ effekt. To resultater som føyer seg inn i rekken av funn fra tidligere forskning på feltet, der ingen klar betydning er kartlagt.

Fra den alternative value added spesifikasjonen finner vi lik effekt som i hovedmodellen for smågruppelærers erfaring, men ikke for storgruppelærer. Høyere estimerte verdier tyder på at value added overestimerer effekten av pretest, som følge av at vi pålegger en for sterk restriksjon på variabelen. Det er derfor tvilsomt at modellen bidrar med større innsikt i hvordan lærerkarakteristikker er av betydning for elevprestasjon.

Vi har videre undersøkt om elevprestasjon i 2. og 3. trinn påvirkes ulikt av de inkluderte lærerkarakteristikkene. Det vi finner er at for 2. trinn vil toppunktet for lærererfaring i smågruppen inntre langt tidligere. For 3. trinn er effekten lik som i hovedmodellen, men sterkere. Resultatet indikerer derfor at effekt av smågruppelærers erfaring er av betydning for trinnene separat, og at effekten er sterkere jo yngre elevene er.

For storgruppelærer finner vi en motsatt effekt av erfaring for både andre og tredje trinn. Et resultat som strider mot tidligere funn, men det er viktig å presisere at vi ikke er i stand til å kontrollere for storgruppelærer på tilsvarende måte som læreren i smågruppen.

Vi har også forsøkt å måle hvordan samarbeidet mellom de to lærerne er av betydning for elevenes prestasjon. Det vi finner er at det ikke er mulig å inkludere en korrekt proxy for samarbeid, som igjen åpner opp for videre forskning på effektene fra 1+1 prosjektet i framtiden.

Referanser

- Alborz, A., Pearson, D., Farrell, P., & Howes, A. (2009). The impact of adult support staff on pupils and mainstream schools: A systematic review of evidence. *London: Department for Children, Schools, and Families.*
- Anderson, C., Beuchert, L., Skyt Nilsen, H., Thomsen, Kjærgaard, M. (2015), The effect of Teacher Aides in the Classroom: Evidence from a Randomized Trial. Manuscript
- Blatchford, P., Russell, A., & Webster, R. (2011). *Reassessing the Impact of Teaching Assistants: How Research Challenges Practice and Policy.* Abingdon: Routledge
- Chetty, R., Friedman, J., & Rockoff, J. (2013). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. Chetty, R., Friedman, J., & Rockoff, J. (2013). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *NBER Working Paper Series*, 19423. doi:10.3386/w19423
- Darling-Hammond, L. (2000). How teacher education matters. Darling-Hammond, L. (2000). How teacher education matters. *Journal of teacher education*, 51(3), 166-173.
- Dobbie, W., & Fryer, R. G. (2013). Getting Beneath the Veil of Effective Schools: Evidence From New York City. *American Economic Journal: Applied Economics*, 5(4), 28-60. doi:10.1257/app.5.4.28
- Enzi, B. (2017). *The Effect of Pre-Service Cognitive and Pedagogical Teacher Skills on Student Achievement Gains: Evidence from German Entry Screening Exams (No. 243).* Ifo working paper.
- Fryer Jr, R. G. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. Fryer Jr, R. G. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *The Quarterly Journal of Economics*, 129(3), 1355-1407.
- Hanushek, E. A. (2002). Publicly provided education. *Handbook of public economics*, 4, 2045-2141.
- Hanushek, E. A. (2008). Education production functions. *The New Palgrave Dictionary of Economics.* Basingstoke: Palgrave Macmillan.
- Hanushek, E. A. (2011). The Economic Value of Higher Teacher Quality. *Economics of Education Review*, 30(3), 466-479. doi:10.1016/j.econedurev.2010.12.006

- Hanushek, E. A., & Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, 46(3), 607-668.
doi:10.1257/jel.46.3.607
- Hanushek, E., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. Hanushek, E., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17(4), 267-321. doi:10.1007/s10887-012-9081-x
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of public economics*, 95(7-8), 798-812.
- Kinsler, J. (2012). Assessing Rothsteins critique of teacher value-added models. *Quantitative Economics*, 3(2), 333-362. doi:10.3982/QE132
- Kraft, M. A., & Papay, J. P. (2014). Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational evaluation and policy analysis*, 36(4), 476-500.
- Looney, J. (2011). Developing High-Quality Teachers: teacher evaluation for improvement. *European Journal of Education*, 46(4), 440-455. doi:10.1111/j.1465-3435.2011.01492.x
- Mincer, J. (1984). Human Capital and Economic Growth. *Economics of Education Review*, 3(3), 195-205. doi:10.1016/0272-7757(84)90032-3
- Mincer, J. (1996). Economic development, growth of human capital, and the dynamics of the wage structure. Mincer, J. (1996). Economic development, growth of human capital, and the dynamics of the wage structure. *Journal of Economic Growth*, 1(1), 29-48. doi:10.1007/BF00163341
- Mueller, S. (2013). Teacher experience and the class size effect - experimental evidence. *Journal of public economics*, 98, 44-52. doi:10.1016/j.jpubeco.2012.12.001
- NTNU. (2018). Praktisk-pedagogisk utdanning heltid. Hentet fra: <https://www.ntnu.no/studier/prped>
- Regjeringen. (2013). PISA 2012: Svakere resultater i matematikk og naturfag. Hentet fra: <https://www.regjeringen.no/no/aktuelt/pisa-2012-svakere-resultater-i-matematik/id747180/>

Riis, C., & Moen, E. R. (2012). Arbeidsmarkedet. *Moderne mikroøkonomi* (2. utg., s. xxx-xx). Oslo: Gyldendal Norsk Forlag AS.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education finance and Policy*, 6(1), 43-74.

Statistisk Sentralbyrå. (2016). Utgifter til utdanning per elev/student. Hentet fra:

<https://www.ssb.no/nasjonalregnskap-og-konjunkturer/artikler-og-publikasjoner/utgifter-til-utdanning-i-norge?tabell=286297>

Sweetland, S. R. (1996). Human capital theory: Foundations of a field of inquiry. *Review of educational research*, 66(3), 341-359.

Todd, P. E., & Wolpin, K. I. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *Economic Journal*, 113(485), F3-F33.
doi:10.1111/1468-0297.00097

Webster, R., Blatchford, P., & Russell, A. (2013). Challenging and Changing How Schools Use Teaching Assistants: Findings from the Effective Deployment of Teaching Assistants Project. *School Leadership & Management*, 33(1), 78-96.
doi:10.1080/13632434.2012.724672

Wooldridge, J.M. (2016). *Introductory econometrics: A modern approach* (6. utg). Boston: Cengage Learning

A. Appendiks

A.1 Deskriptiv statistikk

A.1.1 Andre trinn

Tabell A.1.1: Deskriptiv statistikk 2. trinn

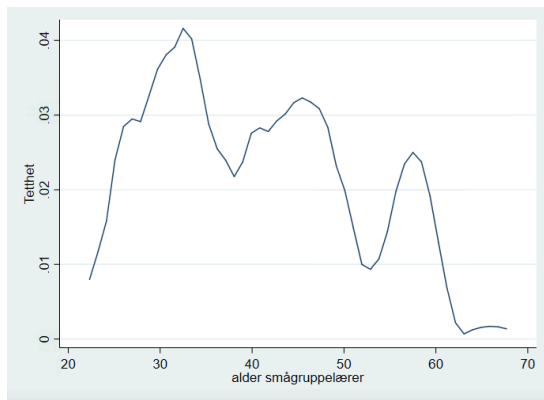
Variabel	Gjennomsnitt	Standardavvik	Min	Max	N
Posttest					
Hele utvalget	14.10184	4.856417	0	24	6 569
Behandling	14.36946	4.726176	0	24	3 183
Kontroll	13.85027	4,963295	0	24	3 386
Pretest					
Hele utvalget	27.11991	8.297735	0	41	8 198
Behandling	26.60094	8.320769	0	41	3 413
Kontroll	27.49007	8.262239	0	41	4 785
VA(standardisert)					
Hele utvalget	-0.0212797	0.8572917	-3.923656	3.86512	6 382
Behandling	0.0954553	0.8318808	-3.732908	3.709488	3 112
Kontroll	-0.1323744	0.8664318	-3.923656	3.86512	3 270

A.1.2 Tredje trinn

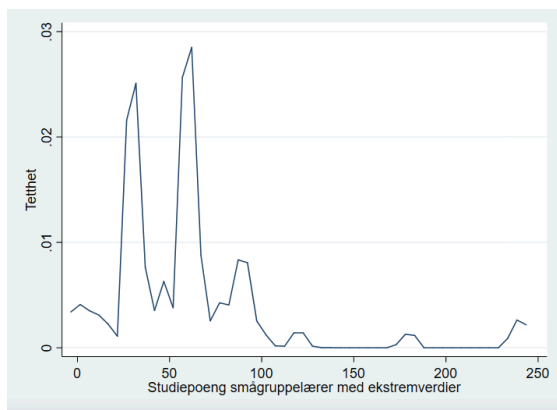
Tabell A.1.2: Deskriptiv statistikk 3.trinn

Variabel	Gjennomsnitt	Standardavvik	Min	Max	N
Posttest					
Hele utvalget	15.51135	5.37883	0	28	6 694
Behandling	15.8408	5.295877	0	27	3 015
Kontroll	15.24137	5.431707	0	28	3 679
Pretest					
Hele utvalget	25.29616	8.754216	0	42	7 864
Behandling	24.79944	8.834122	0	42	3 186
Kontroll	25.63446	8.68407	0	42	4 678
VA(standardisert)					
Hele utvalget	-0.0318307	0.8105789	-4.106485	3.625594	6 388
Behandling	0.0823324	0.8110057	-3.848887	3.625594	2 923
Kontroll	-0.1281363	0.7977272	-4.106485	3.352272	3 465

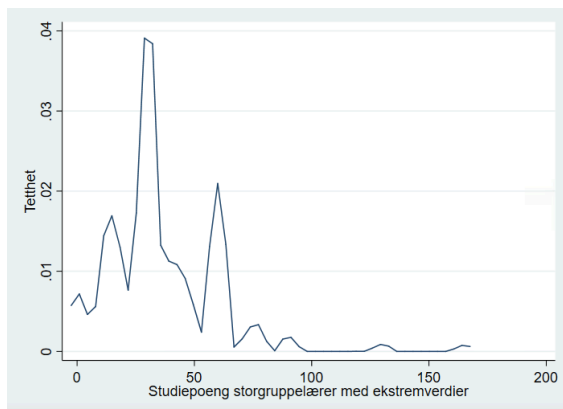
A.2 Tetthetsfunksjoner



Illustrasjon over tetthetsfunksjonen for smågruppelærers alder.



Illustrasjon over tetthetsfunksjonen for smågruppelærers studiepoeng i matematikk når alle observasjoner er inkludert.



Illustrasjon over tetthetsfunksjonen for storgruppelærers studiepoeng i matematikk når alle observasjoner er inkludert.

A.3 Resultater trinnvis

A.3.1 Andre trinn

Tabell A.3.1: Avhengig variabel er posttest standardisert Minste kvadraters metode (MKM) for 2.trinn							
VARIABLER	(1)	(2)	(3)	(4)	(5)	(6)	(7)
std_pretest	0.638*** (0.0172)	0.646*** (0.0176)	0.651*** (0.0186)	0.640*** (0.0219)	0.636*** (0.0238)	0.637*** (0.0237)	0.668*** (0.0249)
erf2	-0.00914 (0.00558)	0.00679 (0.00634)	-0.0186** (0.00895)	0.000580 (0.0109)	0.00845 (0.0138)	0.00354 (0.0145)	1.047*** (0.203)
erf2 ²	0.000177 (0.000192)	9.45e-05 (0.000193)	0.00138*** (0.000357)	0.000577 (0.000428)	0.000919 (0.000560)	0.00107* (0.000577)	-0.0217*** (0.00611)
alder2		-0.0138*** (0.00283)	-0.0121*** (0.00285)	-0.0147*** (0.00309)	-0.0172*** (0.00421)	-0.0164*** (0.00429)	-0.181*** (0.0265)
poeng2			0.00106 (0.000750)	0.00149* (0.000876)	0.000682 (0.000962)	-0.000751 (0.00190)	0.00715 (0.00460)
erfstor				-0.0291*** (0.0105)	-0.00641 (0.0105)	-0.00228 (0.0114)	-1.295*** (0.282)
erfstor ²				0.000985*** (0.000308)	2.73e-05 (0.000327)	-0.000101 (0.000356)	0.0552*** (0.0121)
poengstor					0.00190** (0.000956)	-0.000425 (0.00287)	0.000722 (0.00637)
interstu2						3.21e-05 (3.70e-05)	0.00289*** (0.000503)
Konstant	0.222*** (0.0345)	0.616*** (0.0891)	0.550*** (0.105)	0.767*** (0.140)	0.563*** (0.183)	0.636*** (0.203)	0.817 (0.527)
Skolefaste effekter	Nei	Nei	Nei	Nei	Nei	Nei	Ja
Observasjoner	1,937	1,840	1,672	1,299	1,108	1,108	1,108
R ²	0.426	0.436	0.437	0.431	0.416	0.417	0.485

Robust standardavvik i parentes
 ***Signifikant på 1% signifikansnivå (p<0.01)
 **Signifikant på 5% signifikansnivå (p<0.05)
 * Signifikant på 10% signifikansnivå (p<0.1)

A.3.2 Tredje trinn

Tabell A.3.2: Avhengig variabel er posttest standardisert Minste kvadraters metode (MKM) for 3.trinn							
VARIABLER	(1)	(2)	(3)	(4)	(5)	(6)	(7)
std_pretest	0.643*** (0.0191)	0.643*** (0.0193)	0.633*** (0.0210)	0.666*** (0.0216)	0.672*** (0.0233)	0.673*** (0.0231)	0.686*** (0.0226)
erf3	0.00938* (0.00563)	0.0366*** (0.00759)	0.0407*** (0.00823)	0.0414*** (0.00965)	0.0160 (0.0110)	0.00598 (0.0118)	0.0725* (0.0420)
erf3 ²	-0.000480*** (0.000173)	-0.000696*** (0.000179)	-0.000926*** (0.000219)	-0.00101*** (0.000243)	-0.000200 (0.000270)	5.55e-05 (0.000290)	-0.00264*** (0.000958)
alder3		-0.0195*** (0.00367)	-0.0180*** (0.00373)	-0.0155*** (0.00405)	-0.0221*** (0.00478)	-0.0209*** (0.00480)	0.0122 (0.0179)
poeng3			0.000803 (0.000965)	0.00274*** (0.00105)	-0.00177 (0.00113)	-0.00725** (0.00285)	0.00212 (0.0127)
erfstor				-0.0106 (0.00850)	-0.0439*** (0.00952)	-0.0427*** (0.00947)	-0.227** (0.112)
erfstor ²				0.000200 (0.000215)	0.00101*** (0.000239)	0.000975*** (0.000237)	0.00549* (0.00304)
poengstor					0.00670*** (0.00142)	0.000730 (0.00319)	0.00794 (0.0124)
interstu3						0.000127** (6.07e-05)	-0.000252 (0.000277)
Konstant	0.123*** (0.0365)	0.628*** (0.103)	0.518*** (0.133)	0.424** (0.165)	1.071*** (0.211)	1.335*** (0.243)	1.616 (1.132)
Skolefaste effekter	Nei	Nei	Nei	Nei	Nei	Nei	Ja
Observasjoner	1,882	1,794	1,589	1,333	1,075	1,075	1,075
R ²	0.421	0.429	0.420	0.467	0.489	0.491	0.534

Robust standardavvik i parentes
 ***Signifikant på 1% signifikansnivå (p<0.01)
 **Signifikant på 5% signifikansnivå (p<0.05)
 * Signifikant på 10% signifikansnivå (p<0.1)

A.4 Korrelasjonsmatriser

A.4.1 For trinnene samlet

	std_posttest	std_pretest	erfarsmå	erfarsmå ²	aldsmå	poengsmå	erfstor	erfstor ²	poengstor	interstudie
std_posttest	1									
std_pretest	0.645	1								
erfarsmå	-0.0372	0.00314	1							
erfarsmå ²	-0.0490	0.00583	0.942	1						
aldsmå	-0.0614	0.0495	0.839	0.779	1					
poengsmå	0.0659	0.0200	-0.276	-0.251	-0.354	1				
erfstor	-0.0592	0.0483	0.112	0.184	0.120	-0.192	1			
erfstor ²	-0.0561	0.0355	0.118	0.191	0.126	-0.194	0.958	1		
poengstor	0.0688	0.00458	-0.120	-0.0729	-0.121	0.0724	-0.254	-0.285	1	
interstudie	0.0915	0.00599	-0.217	-0.191	-0.294	0.598	-0.284	-0.286	0.776	1

A.4.2 Andre trinn

	std_posttest	std_pretest	erf2	erf2 ²	alder2	poeng2	erfstor	erfstor ²	poengstor	interstu2
std_posttest	1									
std_pretest	0.629	1								
erf2	0.0428	-0.0212	1							
erf2 ²	0.0285	-0.0598	0.955	1						
alder2	-0.00614	0.0451	0.739	0.695	1					
poeng2	0.00812	-0.0253	-0.212	-0.225	-0.320	1				
erfstor	-0.000401	0.0630	0.0579	0.0987	0.116	-0.0327	1			
erfstor ²	-0.00522	0.0509	0.0694	0.136	0.133	-0.106	0.949	1		
poengstor	0.0538	0.0257	-0.366	-0.322	-0.340	0.0343	-0.0522	-0.0812	1	
interstu2	0.0563	-0.00416	-0.326	-0.316	-0.422	0.526	-0.106	-0.138	0.808	1

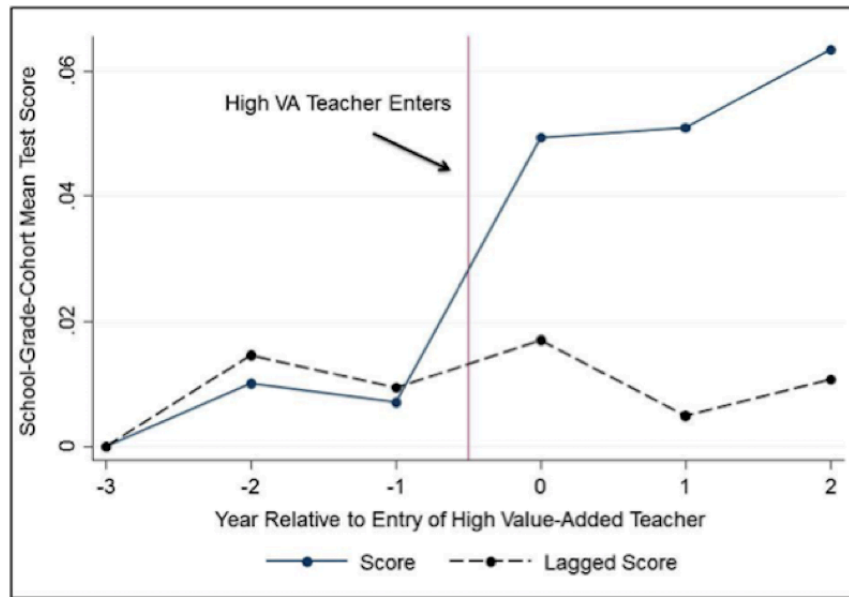
A.4.3 Tredje trinn

	std_posttest	std_pretest	erf3	erf3 ²	alder3	poeng3	erfstor	erfstor ²	poengstor	interstu3
std_posttest	1									
std_pretest	0.663	1								
erf3	-0.0968	0.0176	1							
erf3 ²	-0.0917	0.0339	0.957	1						
alder3	-0.105	0.0513	0.896	0.833	1					
poeng3	0.135	0.0811	-0.337	-0.284	-0.395	1				
erfstor	-0.102	0.0294	0.117	0.178	0.108	-0.305	1			
erfstor ²	-0.0873	0.0185	0.116	0.167	0.111	-0.241	0.966	1		
poengstor	0.0942	-0.0267	0.132	0.0970	0.114	0.141	-0.517	-0.538	1	
interstu3	0.154	0.0263	-0.118	-0.139	-0.177	0.751	-0.537	-0.497	0.702	1

A.4.4 Value added spesifikasjon

	VA	erfarsmå	erfarsmå ²	aldsmå	poengsmå	erfstor	erfstor ²	poengstor	interstudie
VA	1								
erfarsmå	-0.0474	1							
erfarsmå ²	-0.0646	0.942	1						
aldsmå	-0.131	0.839	0.779	1					
poengsmå	0.0535	-0.276	-0.251	-0.354	1				
erfstor	-0.127	0.112	0.184	0.120	-0.192	1			
erfstor ²	-0.108	0.118	0.191	0.126	-0.194	0.958	1		
poengstor	0.0754	-0.120	-0.0729	-0.121	0.0724	-0.254	-0.285	1	
interstudie	0.100	-0.217	-0.191	-0.294	0.598	-0.284	-0.286	0.776	1

A.5 Figur VA



Figur A.5: Grafisk framstilling av gjennomsnittlig elevprestasjon før og etter inntreden av høy VA lærer