

MASTEROPPGAVE

Masterstudium i skolerettet utdanningsvitenskap med fagfordypning i matematikk og matematikkdidaktikk

Mai 2020

Effekter av økt lærertetthet på elevers matematikkprestasjoner i
ungdomsskolen

Kvantitativ metode

Marie Vellesen Løken

OSLOMET

OsloMet – storbyuniversitetet

Fakultet for lærerutdanning og internasjonale studier

Institutt for grunnskole- og faglærerutdanning

Forord

Denne oppgaven er skrevet som en avslutning på masterutdanning i Skolerettet utdanningsvitenskap med matematikk og matematikdidaktikk ved OsloMet.

Gjennom masterutdanningen har jeg tilegnet meg verdifull kunnskap for yrket jeg nå skal gå inn i. Utdanningen har bidratt til at jeg i læreryrket ønsker å holde meg oppdatert på forskning knyttet til skolen og matematikdidaktikk. Dette er noe jeg anser som viktig for å kunne begrunne pedagogiske og faglige valg i arbeidshverdagen.

Siden jeg tidligere har studert ved Norges Handelshøyskole har jeg ønsket å benytte den metodiske kunnskapen herfra i masteroppgaven. Oppgaven har dermed vært en fin måte å knytte sammen de to utdanningene.

Jeg vil rette en stor takk til min veileder, Bjørn Smestad. Han har gitt veldig verdifulle innspill, tilbakemeldinger og upåklagelig veiledning.

I tillegg ønsker jeg å takke professor ved Norges Handelshøyskole, Katrine Vellesten Løken, for innspill på metodisk tilnærming og god hjelp i bruk av Stata, slik at jeg klarte å bygge gode regresjonsmodeller.

Til slutt vil jeg også takke min ektemann som har vært en viktig støttespiller det siste året. Takk for at du holdt ut!

Marie Vellesten Løken, Oslo mai 2020

Sammendrag

Gjennom kvantitativ metode søker oppgaven å undersøke forskningsspørsmålet «*Hva har økt lærertetthet å si for elevenes prestasjoner i matematikk?*». Jeg undersøker om det er en sammenheng mellom ungdomsskoleelevers prestasjoner i matematikk og målet jeg bruker på lærertetthet, gruppestørrelse 2, i årene 2009-2019 med bruk av longitudinale data fra Skoleporten. Analysene er gjort ved bruk av regresjonsmetodene OLS og fasteffekt. Mitt hovedbidrag til forskningsfeltet er at det i Norge ikke tidligere er gjort noen fasteffektanalyse på sammenhengen mellom elevers prestasjoner og lærertetthet med bruk av gruppestørrelse 2. Det er heller ikke tidligere gjennomført noe longitudinal forskning på gruppestørrelse 2. Hovedfunnet fra analysene er at jeg aldri finner signifikant positiv effekt av økt lærertetthet i ordinær undervisning på elevers prestasjoner. Dette er i samsvar med tidligere norsk forskning på økt lærertetthet. Derimot finner jeg signifikant negative effekter av økt lærertetthet i flere av analysene, spesielt for skoler med gruppestørrelse over medianen på 18,3 elever. Jeg finner også at økt lærertetthet er mer negativt for elever med lav støtte hjemmefra. I oppgaven diskuterer jeg først i detalj hvordan funnene mine relateres til internasjonal og norsk relevant forskning. Deretter drøfter jeg implikasjoner av de ulike funnene for politikkkutformingene i skolen og matematikkundervisningen.

Abstract

Title: Effects of increased teacher density on students' mathematics achievements in secondary school.

This thesis uses a quantitative method to explore the research question “*Does higher teacher density affect student achievements in mathematics?*”. In the thesis I investigate whether there is a correlation between secondary school student's achievements in mathematics and the variable for teacher density, “gruppestørrelse 2”. The thesis is based on longitudinal data from “Skoleporten” from 2009-2019. The analyses use OLS regression methods and fixed effect regressions. My contribution to the research field is first of all that no previous study has investigated the connection between student achievements and “gruppestørrelse 2” using fixed effect analyses. Secondly, no longitudinal research on “gruppestørrelse 2” has been conducted before. The results show that I never find positive significant effects of higher teacher density. This complies with previous Norwegian research on increased teacher density. In several of the analysis I find a significant negative effect of increased teacher

density, especially for average group size over a median of 18,3 students. I also find that increased teacher density is more negative for students with low support from home. In the discussion I consider these findings and how they relate to relevant Norwegian and international research. In addition, I look at the implications of the findings for school politics and teaching mathematics.

Innholdsfortegnelse

Forord	II
Sammendrag	III
Abstract	III
Liste over figurer og tabeller	VIII
1 Innledning	1
1.1 Bakgrunn for oppgaven.....	1
1.2 Forskningsspørsmål.....	2
1.3 Oppgavens oppbygning.....	3
1.4 Begrepsavklaringer.....	4
1.5 Avgrensninger	5
2 Empirisk forskning	7
2.1 Komparativt overblikk	8
2.2 Forskning på lærertetthet internasjonalt	8
2.2.1 STAR-prosjektet	9
2.2.2 Forskning på data fra TIMSS-undersøkelser	15
2.2.3 Øvrig forskning internasjonalt	18
2.2.4 Oppsummering av internasjonal forskning	20
2.3 Forskning på lærertetthet i Norge	20
2.3.1 Forskning på klassestørrelsesreduksjon	21
2.3.2 Forskning på gruppestørrelse 1	25
2.3.3 Forskning på gruppestørrelse 2	26
2.3.4 Pågående forskning på lærertetthet i Norge	28
2.4 Relevant matematikdidaktisk forskning.....	31
3 Metode	35
3.1 Begrunnelse for valg av metode.....	35
3.2 Datainnsamling.....	36
3.3 Variabler.....	38
3.4 Bearbeiding av datasett	40
3.5 Metode for regresjonsanalyse.....	43

3.5.1 Generelt om regresjonsanalyse og sentrale begreper	44
3.5.2 Multipl linear regresjon	45
3.5.3 Fasteffekt.....	48
3.6 Regresjonsmodeller brukt i analysen.....	49
3.6.1 OLS-regresjoner	49
3.6.2 Fasteffektmodeller.....	53
3.7 Svakheter ved metoden og datasettet	54
4 Resultater.....	58
4.1 Deskriptive data	58
4.2 OLS-regresjoner	61
4.2.1 OSL-regresjon med kontroll kun for år.....	61
4.2.2 OLS-regresjon med flere kontroller	63
4.2.3 OLS-regresjoner med støtte fra lærer som utfallsvariabel	65
4.3 Fasteffektregresjoner	66
4.3.1 Fasteffektregresjoner med karakterer som utfallsvariabler.....	66
4.3.2 Fasteffektregresjon med støtte fra lærer som utfallsvariabel	68
4.4 Analyser av heterogenitet mellom ulike grupper.....	68
4.4.1 Heterogenitet mellom gruppestørrelse over eller under medianen	68
4.4.2 Heterogenitet mellom jenter og gutter	71
4.4.3 Heterogenitet mellom grupper med ulik støtte hjemmefra	73
5 Drøfting og sammenlikning	75
5.1 Bidrag til forskningsfeltet	75
5.2 Diskusjon av funn.....	78
5.2.1 Funn 1: Negativ effekt av økt lærertetthet på prestasjoner	79
5.2.2 Funn 2: Signifikant effekt kun for gruppestørrelser større enn medianen	81
5.2.3 Funn 3: Effekter er i større grad signifikante på eksamenskarakterer enn standpunkt	82
5.2.4 Funn 4: Økt lærertetthet er mer negativt for elever med lav støtte hjemmefra.....	84
5.2.5 Funn 5: Ingen entydige ulikheter i effekt mellom jenter og gutter	86
5.3 Sammenlikning av funn mot annen norsk forskning	86
5.4 Sammenlikning av funn mot internasjonal forskning.....	90
5.5 Konsekvenser for politikktutforming og matematikkundervisning.....	92
6 Oppsummering	98
7 Referanseliste	102

8 Vedlegg	109
8.1 <i>Veiledning fra Utdanningsdirektoratet for hvordan gruppestørrelse 2 kalkuleres</i>	109
8.2 <i>Svaralternativene på indeksen støtte hjemmefra og støtte fra lærer i elevundersøkelsen</i>	112

Liste over figurer og tabeller

<i>Tabell 1: Statistikk over gruppestørrelse 1 og gruppestørrelse 2 fra 2009-2019, hentet fra datasett</i>	<i>5</i>
<i>Tabell 2: Klassifisering av skoletyper</i>	<i>41</i>
<i>Tabell 3: Fordeling av skoletyper og endelig antall skoler i analyse</i>	<i>42</i>
<i>Tabell 4: Gjennomsnitt og standardavvik for de ulike karakterene før standardisering</i>	<i>61</i>
<i>Tabell 5: Resultater fra regresjonsmodell 1 for standpunktkarakterer</i>	<i>62</i>
<i>Tabell 6: Resultater fra regresjonsmodell 1 for karakterer på skriftlig eksamen</i>	<i>62</i>
<i>Tabell 7: Resultater fra regresjonsmodell 1 for karakterer på muntlig eksamen</i>	<i>63</i>
<i>Tabell 8: Resultater fra regresjonsmodell 2 for standpunktkarakterer</i>	<i>63</i>
<i>Tabell 9: Resultater fra regresjonsmodell 2 for karakterer på skriftlig eksamen</i>	<i>64</i>
<i>Tabell 10: Resultater fra regresjonsmodell 2 for karakterer på muntlig eksamen</i>	<i>64</i>
<i>Tabell 11: Resultater fra regresjonsmodell 1 for indeksen støtte fra lærer</i>	<i>65</i>
<i>Tabell 12: Resultater fra regresjonsmodell 2, for indeksen støtte fra lærer</i>	<i>65</i>
<i>Tabell 13: Resultater fra fasteffektregresjonsmodell for standpunktkarakterer</i>	<i>66</i>
<i>Tabell 14: Resultater fra fasteffektregresjonsmodell for karakter på skriftlig eksamen</i>	<i>67</i>
<i>Tabell 15: Resultater fra fasteffektregresjonsmodell for karakter på muntlig eksamen</i>	<i>67</i>
<i>Tabell 16: Resultater fra fasteffektregresjonsmodell for indeksen støtte fra lærer</i>	<i>68</i>
<i>Tabell 17: Resultater fra fasteffektregresjonsmodell for standpunktkarakter, høy og lav gruppestørrelse</i>	<i>69</i>
<i>Tabell 18: Resultater fra fasteffektregresjonsmodell for karakter på skriftlig eksamen, høy og lav gruppestørrelse</i>	<i>69</i>
<i>Tabell 19: Resultater fra fasteffektregresjonsmodell for karakter på muntlig eksamen, høy og lav gruppestørrelse</i>	<i>70</i>
<i>Tabell 20: Resultater fra fasteffektregresjonsmodell for indeksen støtte fra lærer, høy og lav gruppestørrelse ..</i>	<i>70</i>
<i>Tabell 21: Resultater fra fasteffektregresjonsmodell med heterogenitet mellom gutter og jenter, høy og lav gruppestørrelse</i>	<i>72</i>
<i>Tabell 22: Resultater fra fasteffektregresjonsmodell med heterogenitet mellom høy og lav støtte hjemmefra</i>	<i>74</i>
<i>Figur 1: Illustrasjon av minste kvadraters metode</i>	<i>46</i>
<i>Figur 2: Utviklingen av gruppestørrelse 2 2009-2019</i>	<i>58</i>
<i>Figur 3: Utviklingen av gruppestørrelse 2 i de ulike fylkene fra 2009-2019</i>	<i>59</i>
<i>Figur 4: Histogrammet viser fordelingen av gruppestørrelse 2 samlet over alle årene</i>	<i>60</i>

1 Innledning

1.1 Bakgrunn for oppgaven

Lærertetthet er et sentralt tema i skolepolitikken i Norge. Fra høsten 2019 ble siste fase i implementeringen av lærernormen gjennomført. Lærernormen er en reform som stiller krav til at det ikke skal være flere enn 15 elever i gjennomsnitt per lærer i ordinær undervisning på 1.-4. trinn, og ikke flere enn 20 elever på 5.-10. trinn. Hensiktene med å satse på økt lærertetthet i ordinær undervisning er å bedre den tilpassede opplæringen og styrke elevenes grunnleggende ferdigheter. Dette skal igjen gi elevene bedre læringsutbytte, og dermed bedre prestasjoner (NOU 2019:23, 2019).

Reduksjon av klassestørrelse eller økning i lærertetthet har vært, og til dels er, en kontroversiell politisk tilnærming for å bedre prestasjoner i skolen, mye på grunn av at kostnadene relatert til dette er veldig store. Noen mener det ikke er et kostnadseffektivt tiltak fordi det er dyrt og usikkerheten om det har positive effekter for elever er stor. Andre mener at det er bra for elevene, spesielt dem fra lavere sosioøkonomisk bakgrunn, og at det dermed er verdt kostnadene og bør være et fokus i skolepolitikken.

Innføringen av lærernormen i Norge har skjedd uten at det eksisterer norske studier som bekrefter at dette gir positive effekter på prestasjoner. Mye av forskningen en har støttet seg på baserer seg på utenlandske studier, hvor mange også ligger relativt langt tilbake i tid. Spesielt har det amerikanske STAR-prosjektet fra 1980-tallet vært hyppig referert til i debatten rundt innføring av lærernormen.

Eksisterende forskning på lærertetthet i Norge kan deles inn i tre faser ut fra hvilket mål på lærertetthet forskningen har sett på. Disse henger sammen med hvordan skolepolitikken over tid har gått gjennom ulike skift, som har endret måten man har tenkt rundt lærertetthet på. Den første fasen er forskning på reduksjon i klassestørrelse. Denne fasen pågikk helt frem til det kom en lovendring i 2003, der regelen om maksimal klassestørrelse ble endret til pedagogisk forsvarlig gruppestørrelse (Kunnskapsdepartementet, 2004; Opplæringslova, 1998). Innenfor denne fasen er det mye forskning. Den neste fasen er knyttet til forskning på gjennomsnittlig gruppestørrelse 1 i undervisningen. Definisjon av gruppestørrelse 1 og 2 følger i 1.4 Begrepsavklaringer. Totalt innenfor denne forskningen er det, så vidt jeg vet, gjennomført kun to studier, der begge av dem kun inkluderer gruppestørrelse 1 som en del av en mer

omfattende forskningsrapport (Grøgaard, Helleand & Lauglo, 2008; Opheim, Grøgaard & Næss, 2010). Den siste fasen vokste frem som følge av politisk interesse rundt lærernormen. Lærertetthet i ordinær undervisning, beregnet etter gruppestørrelse 2, rapporteres fra skoleåret 2018 inn fra hver enkelt skole til GSI. På grunn av den store politiske interessen har også Utdanningsdirektoratet beregnet gruppestørrelse 2 tilbake i tid. Meg bekjent er det ingen som foreløpig har utgitt noen forskning der de bruker gruppestørrelse 2 med dataene som er beregnet tilbake i tid som en longitudinal studie. Den eneste norske studien som bruker målet gruppestørrelse 2 på lærertetthet er utgitt av Kirkebøen, Kotsadam, Raaum, Andresen og Rogstad (2017), men de bruker kun data fra skoleåret 2011/2012.

Grunnen til at det er viktig å være oppmerksom på disse tre fasene, er at forskning på klassestørrelse og gruppestørrelse ikke er likt bygd opp eller ser på nøyaktig det samme. Når en forsker på klassestørrelse er det kun endringer i antall elever en studerer. Forskning på gruppestørrelse ser på den samlede lærertettheten, hvor både antall elever og lærere per klasse kan endre seg. Dette har betydning både for metodisk tilnærming og hvordan resultater og underliggende årsaker forklares.

Årsaken til at jeg ønsker å se nærmere på lærertetthet i ordinær matematikkundervisning, målt etter gruppestørrelse 2, er fordi jeg anser forskningsfeltet som svært mangelfullt. For å kunne ta gode beslutninger for skolen, mener jeg det behøves mer forskning med et solid forskningsbasert fundament. Dette gjelder både for hvordan skolepolitikken utformes videre, og hvordan lærere i matematikk legger opp undervisningen for å sikre at en henter ut positive effekter og reduserer eventuelle negative effekter som økt lærertetthet kan medføre.

1.2 Forskningsspørsmål

Forskingsspørsmålet i oppgaven er «*Hva har økt lærertetthet å si for elevenes prestasjoner i matematikk?*».

For å svare på dette vil jeg gjennom kvantitativ analyse undersøke om det er en sammenheng mellom elevenes prestasjoner i matematikk og målet jeg bruker på lærertetthet, gruppestørrelse 2.

For å ha et datagrunnlag til å kunne besvare forskningsspørsmålet har jeg samlet data på skolenivå fra elever som fullfører 10. klasse fra 2009-2019 i hele Norge i ett datasett. Datasettet består for hvert år av informasjon om antall elever og antall lærere ved skolene både samlet og skilt etter kjønn, gjennomsnittlig standpunkt karakter i matematikk, engelsk skriftlig, og norsk hovedmål skriftlig samlet og skilt etter kjønn. Den samme informasjonen har jeg hentet ut for eksamen. Datasettet inneholder også tall på gruppestørrelse 2 for hver skole. I tillegg har jeg hentet ut to indekser fra Elevundersøkelsen.

For å besvare forskningsspørsmålet har jeg valgt å benytte regresjonsanalyse. Jeg har bygd to ulike OLS-regresjonsmodeller (Ordinary Least Squares) og en fasteffektmodell. Dette beskrives nærmere i kapittel 3 Metode.

Mitt hovedbidrag til forskningsfeltet er at det i Norge ikke tidligere er gjort noen fasteffektanalyse på sammenhengen mellom prestasjoner og lærertetthet med gruppestørrelse 2 som mål. Det er heller ikke gjennomført noe longitudinal forskning på gruppestørrelse 2.

1.3 Oppgavens oppbygning

I kapittel 2 vil jeg presentere relevant empirisk forskning på lærertetthet og effekter av klassestørrelse. Først vil det bli gitt et komparativt overblikk, før jeg deretter går mer konkret inn på eksisterende internasjonal forskning. Den internasjonale forskningen er knyttet til effekter av klassestørrelse og er delt i to deler: STAR-prosjektet og øvrig internasjonal forskning. Deretter presenterer jeg den norske forskningen på effekter av klassestørrelse og lærertetthet. I gjennomgangen av forskning vil jeg i tillegg til å se på funnene også gå inn på metodene på de studiene der det er relevant for min oppgave, som bakgrunn for eget valg av metode.

I kapittel 3 beskrives metoden for analysen. Her vil jeg forklare de metodiske verktøyene jeg har benyttet, samt hvordan datasettet jeg har samlet inn er bygd opp og bearbeidet. Deretter vil jeg i kapittel 4 presentere resultatene fra regresjonsanalysen, for til slutt i kapittel 5 drøfte funnene i lys av forskningsspørsmålet og utvalgt forskning på området.

1.4 Begrepsavklaringer

I oppgaven kommer jeg til å bruke en del sentrale begreper. De viktigste vil jeg allerede definere her. En del av disse vil også bli gjennomgått i mer detaljer i kapitlene om metode og empirisk forskning.

Eksogenitet og endogenitet er to begreper som går igjen i oppgaven. En mer dyptgående definisjon av dem gis i 3.5.2 Multippel lineær regresjon. En enkel måte å forklare eksogenitet på er at en variabel er eksogen når den er uavhengig av andre variabler i en forklaringsmodell. Det vil si at den ikke blir påvirket av noen av de andre variablene i modellen. Endogenitet er det motsatte. En variabel er endogen når den påvirkes av en eller flere av de andre variablene i forklaringsmodellen. I en analyse bør en etterstrebe høyest mulig grad av eksogenitet (Wooldridge, 2016).

De to målene på lærertetthet som brukes i norsk skole er sentrale begreper i oppgaven, gruppestørrelse 1 og gruppestørrelse 2. Før 2003 var det klassestørrelse som ble brukt som mål på lærertetthet. Gruppestørrelse 1 som mål på lærertetthet beregnes ved se på forholdet mellom antall elevtimer og antall lærertimer. $Gruppestørrelse\ 1 = \frac{\text{totalt antall elevtimer}}{\text{lærertimer totalt}}$.

Fra skoleåret 2018/2019 skal lærertetthet også innrapporteres etter beregningene av gruppestørrelse 2. Denne er begrenset til kun å gjelde ordinær undervisning. Da målet på lærertetthet ut fra gruppestørrelse 1 inneholder timer til spesialundervisning og timer til særskilt norskopplæring for språklige minoriteter i totalt antall elevtimer, blir tallet kunstig lavt om en for eksempel har to elever som har spesialundervisning med en egen lærer. Hele klassen deles da på to ressurspersoner, selv om den ene gruppen egentlig kun har to elever.

Gruppestørrelse 2 som mål på lærertetthet beregnes på følgende måte: $Gruppestørrelse\ 2 = \frac{(\text{totalt antall elevtimer} - (\text{timetall spesialundervining} + \text{timetall særskilt norsk}))}{(\text{ordinære undervisningstimer} + \text{timer til oppdeling til samiske språkalternativer og annen målform})}$

Lærertetthet i ordinær undervisning er forholdet mellom elevtimer subtrahert timer til spesialundervisning og særskilt norskopplæring, og ordinære undervisningstimer addert med oppdeling til samiske språkalternativer. Tallet vil trolig gi en mer realistisk indikasjon på antall elever per lærer i et faktisk klasserom, da ressursene til spesialundervisningen og undervisning i særskilt norsk er trukket fra.

For å illustrere hva denne forskjellen faktisk utgjør har jeg i Tabell 1 under gjengitt tallene for årene 2009-2019.

8.-10. trinn	Gruppestørrelse 1	Gruppestørrelse 2
2009/10	14,4	18,4
2010/11	14,4	18,7
2011/12	14,5	18,9
2012/13	14,6	18,7
2013/14	14,3	18,1
2014/15	14,2	18,0
2015/16	14,2	17,9
2016/17	14,2	18,1
2017/18	14,4	18,3
2018/19	14,3	18,3

Tabell 1: Statistikk over gruppestørrelse 1 og gruppestørrelse 2 fra 2009-2019, hentet fra datasett

Vi ser fra Tabell 1 at gruppestørrelse 1 er mye lavere enn gruppestørrelse 2. I gjennomsnitt skiller det omtrent 4 elever per lærer.

1.5 Avgrensninger

Jeg har gjort noen avgrensninger i oppgaven. I utformingen av datasettet har jeg måttet gjøre noen avgrensninger for å få et godt datasett å bygge analysene på. Jeg vil nevne noen av de mest sentrale under, ellers beskrives avgrensningene og årsaken til hvorfor de er gjort nærmere i kapittel 3 Metode.

En av avgrensningene i datasettet er at jeg kun har sett på fagene matematikk, norsk og engelsk. Årsaken til dette er at prosjektet ville blitt svært omfattende om jeg skulle sett på alle fagene. I tillegg har jeg valgt en master med fordypning i matematikk og matematikdidaktikk, og dermed er det dette faget som er hovedfokuset i mine analyser. Norsk og engelsk er i hovedsak tatt med for å ha et sammenlikningsgrunnlag.

En annen avgrensning i datasettet er at jeg kun har sett på offentlige norske skoler med

avgangselever fra 10. trinn, og ikke private skoler. Dette er fordi det er flere faktorer som er annerledes ved private skoler som kunne påvirket analysene (se kapittel 3 Metode).

I tillegg vil jeg avgrense meg til hovedsakelig å se på forhold i skolen og ikke utenfor skolen. Denne begrensningen oppstår naturlig, da mitt datasett ikke inneholder data om elevenes bakgrunn.

Et forhold det også er viktig å presisere er at forskningen min i stor grad handler om faglige resultater. Det kan også finnes andre ting, som ikke så lett kan måles, som kan være konsekvenser av økt lærertetthet, både positive og negative. Dette er imidlertid ikke fokus i min analyse.

2 Empirisk forskning

I dette kapitlet vil jeg gå gjennom empirisk forskning som er relevant for forskningsspørsmålet mitt, både ut fra tema i forskningen og analysemetoder som tidligere er benyttet.

Første del omhandler internasjonal forskning på effekter av klassestørrelse. Delkapitlet er strukturert slik at jeg først gjennomgår relevant forskning fra STAR-prosjektet, deretter forskning på TIMSS-data, for til slutt å se på andre relevante internasjonale studier. Den internasjonale forskningen jeg har gjennomgått er relevant for at jeg skal kunne sammenlikne mine funn mot internasjonal forskning, samt vurdere hvorvidt internasjonale funn kan overføres til norske forhold. I tillegg er det viktig å gjøre en omfattende gjennomgang for å vise at resultatene fra den internasjonale forskningen ikke har entydige gjennomgående resultater. Årsaken til at jeg har valgt å bruke en del plass på forskning fra STAR-prosjektet er at flere beslutningstakere også i Norge henviser til dette prosjektet når de skal vise til positive effekter på klassestørrelse. Prosjektet er dominerende innenfor litteraturen på effekter av klassestørrelse, og nesten all forskning på klassestørrelse og lærertetthet refererer til dette prosjektet og forskning på data fra det.

I neste del gjennomgår jeg den relevante norske forskningslitteraturen på klassestørrelse og lærertetthet. Gjennomgangen er strukturert etter de tre fasene for denne forskningen, henholdsvis forskning på klassestørrelse, gruppestørrelse 1 og gruppestørrelse 2. Den norske forskningen er svært relevant å sammenlikne med mine funn, da forskningen er gjennomført i samme land med de samme kulturelle, økonomiske og sosiale betingelsene.

Ved flere av de norske og internasjonale studiene vil jeg i tillegg til å gjennomgå funnene også gå nøye inn på valg og beskrivelse av metode. Ved kvantitative analyser er dette sentralt, da det har betydning for hvordan resultatene tolkes og hvilke styrker og svakheter ulike forskningsmetoder har. Som gjennomgangen vil vise er det metodiske svakheter ved mye av den tidlige forskningen på dette temaet. Dette vil jeg bruke aktivt i diskusjon og sammenlikning av mine funn opp mot annen forskning i kapittel 5.

Avslutningsvis vil jeg også gå gjennom utvalgt forskning på matematikkdiridaktikk som blir viktig for diskusjon av mulige årsaker og konsekvenser av funnene i min analyse.

2.1 Komparativt overblikk

Årlig gir OECD (Organisasjonen for økonomisk samarbeid og utvikling) ut rapporten Education at a Glance som er en statistisk informasjonskilde om skolesystemene i de ulike landene (OECD, 2019). Denne rapporten sammenlikner skolene i OECD-landene på en rekke ulike områder. I forbindelse med denne oppgaven er det spesielt indikatoren D2 i rapporten, som gir statistikk om forholdet mellom lærer, elev og klassestørrelse, som er interessant. Tabell D2.2 i rapporten fra 2019 viser at lærertettheten, altså antall elever per lærer, i gjennomsnitt er 13 elever på ungdomstrinnet i OECD-landene. Tallene rapporten bygger på er fra 2017 (OECD, 2019). Det vil si at Norge allerede ligger langt under gjennomsnittet med 9 elever per lærer på ungdomstrinnet sammenliknet med OECD-landene. I OECD-rapporten ser de på antall elever opp mot antall lærere, altså hvor mange elever det er per ansatt lærer. Det tar ikke hensyn til hvor mange undervisningstimer elevene har. Det regnes derfor ut på en annen måte enn Utdanningsdirektoratet regner ut gruppestørrelse 1, hvor de som tidligere nevnt tar hensyn til timeantall og deler antall elevtimer på antall lærertimer. Som vi ser fra Tabell 1, er tallene for gruppestørrelse 1 vesentlig høyere enn OECD sine utregninger. For begge utregningene er det viktig at en ikke sammenlikner dette med klassestørrelse, vi snakker her om hvor mange elever det er i gjennomsnitt per lærer, ikke hvor mange det er i hver klasse.

2.2 Forskning på lærertetthet internasjonalt

Internasjonalt har det vært forsket mye på og debattert offentlig hva konsekvensene er for elever av store versus små klasser. For mine analyser er det vesentlig å kartlegge hvilke metoder den internasjonale forskningen bruker og hvilke funn den har. Dette er for å eventuelt kunne avgjøre hvor sammenliknbare og overførbare resultatene er. I all hovedsak omhandler den internasjonale forskningen reduksjon av klassestørrelse (CSR= class size reduction). Forskningen på klassestørrelse internasjonalt er svært omfattende. Jeg har derfor valgt å ta utgangspunkt i en kartlegging Kunnskapssenteret for utdanning gjennomførte på litteraturen om effekter av endringer i klassestørrelse (Kunnskapssenteret, 2015) for å redusere omfanget av forskningsartikler internasjonalt. I tillegg har jeg gjort egne søk etter annen relevant internasjonal forskning og nyere internasjonal forskning.

I rapporten påpeker Kunnskapssenteret at lærertetthet (engelsk: teacher density) ikke er et begrep som brukes i den internasjonale forskningen, og at reduksjon av klassestørrelser er hovedvinklingen i den internasjonale forskningen. Til tider blir også begrepet elev-lærer-rate (engelsk: pupil-teacher ratio) brukt.

På grunn av omfattende forskning på området, har jeg i hovedsak i den internasjonale forskningen sett på forskning etter 2010. De siste 10 årene har dessuten metoden innenfor utdanningsøkonometri blitt stadig mer avansert og nøyaktig. Forskningen tar derfor i større grad hensyn til utfordringene i å isolere effekten av klassestørrelse på elevers prestasjoner fra alt annet som også påvirker elevene. I litteraturen brukes kausalitet som begrep når en har funnet en direkte effekt av klassestørrelse. Det vil si effekten av klassestørrelse på elevers prestasjoner når alt annet er likt. I løpet av gjennomgangen vil jeg komme tilbake til flere ulike metoder for å komme så nærme som mulig en kausal sammenheng mellom klassestørrelse og elevers prestasjoner.

2.2.1 STAR-prosjektet

Et av de største randomiserte kontrollerte eksperimentene internasjonalt på effekten av reduksjon av klassestørrelse på prestasjoner ble gjennomført i perioden 1985-1989 i Tennessee, under navnet Project Student/Teacher Achievement Ratio (STAR). Det er skrevet et stort antall artikler ut fra datamaterialet også i de senere årene. I prosjektet deltok barn fra det siste året i barnehagen til og med 3. trinn på barneskolen. Ideelt sett skulle jeg gjerne sett at denne studien var gjennomført på ungdomstrinnet, men dessverre finnes det ingen randomiserte kontrollerte studier av vesentlig omfang her. Årsaken til at jeg likevel har brukt så mye plass på STAR-prosjektet er at det er en viktig del av forskningsgrunnet innenfor effekter av klassestørrelsesreduksjon og bortimot all forskning på klassestørrelse henviser til STAR-litteraturen. Fra prosjektet er det i hovedsak de positive effektene av reduksjon i klassestørrelse som henvises til i annen forskning. Dette gjelder uansett hvilket trinn det forskes på i de ulike studiene.

I prosjektet deltok over 11.500 barn fra 79 ulike barnehager og skoler. I hver av barnehagene ble barna tilfeldig delt inn i én av tre intervensjonsgrupper. Den første var en liten gruppe, fra 13-17 elever. Dette omtales i prosjektet som tiltaksgruppen. Den andre gruppen var normal størrelse, på 22-25 elever og omtales som kontrollgruppen. Den tredje gruppen var normal størrelse, men hadde alltid med en lærerassistent tilstede i klassen i tillegg til den vanlige

læreren. Etter fullføring av barnehagen gjennomførte elevene seks ulike standardiserte tester som målte prestasjoner i ulike dimensjoner. Testene inneholdt blant annet lesing, lytting og forståelse, matematikk og ordgjenkjenning. I tillegg til de standardiserte testene gjennomførte elevene en selvbilde- og motivasjonstest, som skulle dekke de ikke-kognitive ferdighetene¹.

Forskningen fra STAR-prosjektet refereres til hyppig selv i dag. Det har skjedd en betydelig metodisk fremgang på feltet siden tidlig på 90-tallet og frem til i dag. Det er derfor viktig for min oppgave å vise til styrker og mangler både ved den tidlige og nyere forskningen, for å kunne vurdere hvor robuste funnene i ulike studier er. Under vil jeg gjennomgå et utvalg av de metodisk gode studiene, som også påpeker flere mangler ved en del tidligere studier og konklusjoner. For disse vil jeg først oppsummere hva som er hovedfunnet i studien, før jeg beskriver kritikken deres til tidligere studier og hvordan de har forsøkt å løse det metodisk. Etter gjennomgangen av relevante artikler, oppsummerer jeg hovedutfordringene ved bruk av STAR-prosjektet og hvorfor det ikke bør tillegges så stor vekt som det har fått ved vurdering av effekter av klassestørrelsesreduksjon.

Schanzenbach (2006) gjør en systematisk gjennomgang av de fleste artikler publisert med utgangspunkt i STAR-data før 2007. Hun konkluderer med bakgrunn i disse at redusert klassestørrelse har positiv effekt på de ulike testene. Som nevnt har det i etterkant av 2007 vært publisert flere artikler som er uenige med konklusjonene fra den tidlige forskningen. For å få frem nyanser og metodiske fremskritt har jeg valgt å fokusere på artikler gjort på STAR-datamateriale etter 2010. Dette også for å begrense omfanget.

Hadde selve prosjektet blitt utført ideelt, ville det vært svært enkelt å estimere effektene av liten klassestørrelse ved kun å sammenlikne prestasjonene til de i små klasser med de i større, slik det er blitt gjort i mange av de tidlige studiene. Men i senere tid har det vist seg at det var flere utfordringer med implementeringen av prosjektet og det har kommet nye vinklinger på forskningen, som bruker mer avanserte metoder for å finne den kausale effekten av klassestørrelse på elevers prestasjoner.

Flere tidlige studier på STAR-prosjektet konkluderte med at minoriteter og elever med lav sosioøkonomisk status er dem som får størst effekt av redusert klassestørrelse (Finn, 2002;

¹ Her har jeg oversatt non-cognitive skills til ikke-kognitive ferdigheter. I STAR-prosjektet er det denne terminologien som er brukt.

Finn & Achilles, 1990). Dette er motsatt av hva Ding og Lehrer (2011) finner. De finner en større positiv effekt for elever som presterer godt enn for de som ikke presterer like godt. Ding og Lehrer (2011) ser kun på dataene fra første klasse. I studien ser de på betydningen av at en ser på mange korrelerte utfall for det samme klassestørrelsestiltaket. Forskerne mener at tidligere forskning ikke tok hensyn til at elevenes resultater på de ulike testene korrelerer, og at dette kunne føre til at en fikk for mange signifikante resultater. Ding og Lehrer (2011) benyttet derfor en regresjonsmodell som tar høyde for dette. Videre problematiserer de at de tidligere studiene nesten bare har fokusert på estimering av gjennomsnittseffekter. Det vil være nødvendig å få en større forståelse for hvilke grupper av elever som får størst fordeler eller ulemper av mindre klassestørrelser. Derfor benytter forfatterne kvantilregresjon for å se om effektene av tiltaket er konsentrert i ulike deler av resultatfordelingen. Dette gjør at de kan undersøke mer detaljert den heterogene effekten redusert klassestørrelse har på mål for kognitive og ikke-kognitive prestasjoner. Funnene viser at elever som presterer høyere på de standardiserte testene får større utbytte av redusert klassestørrelse. De finner ingen støtte for at minoritets elever eller elever med lav sosioøkonomisk status får mer utbytte av redusert klassestørrelse. Samlet sett sier forfatterne at de finner sprikende resultater på hvor effektivt tiltaket med redusert klassestørrelse i førskolen er. Dette fordi de finner noe positiv effekt på kognitive resultater for elever som presterer godt på disse testene, men ingen effekt på ikke-kognitive ferdigheter for noen av elevene (Ding & Lehrer, 2011).

En studie som også finner forskjell i effekter mellom ulike grupper er Shin (2012). I artikkelen undersøker Shin hvorvidt minoritetsbarn har en større fordel av små klasser enn majoritetsbarn. I analysen bruker forskeren avanserte økonomiske modeller for å studere både homogene og heterogene kausale effekter av klassestørrelse på prestasjoner på tvers av skoler. Funnene er at minoritets elevene har større fordel av redusert klassestørrelse enn majoritets elever. Studien finner ikke bevis for at den kausale minoritetsforskjellen er heterogen på tvers av skoler i et gitt år (Shin, 2012). Dette er mer i tråd med den tidligere forskningen (Finn, 2002; Finn & Achilles, 1990), og står i motsetning til funnene fra Ding og Lehrer (2011) som ikke finner noen effekt for denne gruppen.

Wilde, Finn, Johnson og Muennig (2011) finner sprikende resultater når de undersøker effekten mindre klassestørrelse fra barnehage til 3. klasse har på inntekt, arbeidsstatus og uførhet. I studien har de koblet data fra STAR-prosjektet mot registerdata. De påpeker at en av årsakene til at det brukes så store midler på intervensjoner i tidlig utdanning er fordi det er

tiltak som trolig reduserer sosiale ulikheter på inntekt og helse senere i livet. Spørsmålet blir hvorvidt det å redusere klassestørrelse er et tiltak som bidrar til dette. Derfor ønsker de å se om dataene fra STAR-prosjektet der klassestørrelse ble redusert forbedret ansettelsesrater, inntekt og helse for svarte relativt til hvite frem til 2008². Da er majoriteten av elevene mellom 18 og 28 år. De finner, som mye tidligere forskning ved bruk av gjennomsnitt, at elevene som deltok i STAR med redusert klassestørrelse gjennomsnittlig forbedret kognitive resultater og hadde høyere gjennomføringsgrad på videregående. Når de ser på undergrupper finner de at gruppene som dro mest fordel av små klasser var elever med lav sosioøkonomisk status og minoritets elever. I tillegg finner de at inntekt og ansettelsesrate øker for svarte som var i små klasser, og da spesielt svarte menn. Inntekt øker også for hvite menn som var i små klasser, men redusert klassestørrelse reduserer både ansettelsesrate og inntekt for hvite kvinner (Wilde et al., 2011). De finner ikke at redusert klassestørrelse gir lavere grad av uførhet for noen av gruppene. Totalt kan vi si at de finner signifikante positive effekter av liten klassestørrelse for minoritets elever, men ellers sprikende funn for andre grupper.

Konstantopoulos (2011) finner små eller ingen effekter av klassestørrelse på prestasjoner. I studien prøver han å estimere effekten av små klasser på matematikk- og leseresultater for hver av skolene fra barnehagen og fram til og med 3. klasse. Deretter kombinerer han alle skolenes estimater for å kalkulere et overordnet vektet sammenliknbart estimat. Resultatene viser at en del av elevene har positive effekter av å være i mindre klasser, mens en annen del av elevene har negative effekter. Fra dette konkluderer han at noen elever har stor nytte av små klasser, mens for andre er det ikke fordelaktig, og for noen er det en ulempe. Effektene av liten klasse var inkonsistente og varierte signifikant på tvers av trinn på alle skolene. Dette indikerer en interaksjon mellom liten klasse og skole som vil si at det er andre ting ved skolene som er viktige for om små klasser fungerer eller ikke. Denne forskningen påpeker videre hvordan en del tidligere forskning på STAR-data som finner positive effekter av små klassestørrelser, og ser på gjennomsnittlige effekter, ikke ser hele bildet. Hadde en gjort det samme i denne studien ville en også fått positiv gjennomsnittseffekt av tiltaket, men det skjuler at noen skoler har negativ eller ingen effekt. Konstantopoulos stiller spørsmål om hvorvidt den positive effekten da kan tilskrives andre mekanismer som opererer sammen med klassestørrelse, for eksempel hvorvidt noen skoler klarer å tilpasse læringen bedre til mindre grupper (Konstantopoulos, 2011).

² I artikkelen opereres det med begrepene black og white, uten at det klart defineres hva som ligger i disse begrepene.

Konstantopoulos og Li (2012) viser at når en kontrollerer for forrige års prestasjoner og deltakelse i små klasser, er effektene av dette årets små klasser ikke signifikant. De drøfter på hvilken måte en skal sikre validiteten til dataene, og belyser flere sentrale problemer med tidligere forskning. For det første påpeker de at ønsket om å finne positive resultater av tiltaket gjør at en kan overtolke datafunnene, slik Hanushek (1999) og Milesi og Gamoran (2006) tidligere har påpekt. For det andre kritiserer de igjen funnene fra tidligere forskning som finner gjennomsnittlige fordeler av små klasser i de første skoleårene. Disse forskerne finner også de samme resultatene dersom de ser på den samlede effekten av små klasser. Men dersom de bruker en mer avansert metode der de undersøker differensielle «tilleggsverdi»-effekter av små klasser på tvers av fordelingen av prestasjoner, finner de ikke disse signifikante effektene. Funnene viser at når en kontrollerer for forrige års prestasjoner og deltakelse i små klasser, er effektene av små klasser ikke signifikant.

Konstantopoulos og Sun (2014) har hatt en litt annen tilnærming til analyse av STAR-dataene. De har undersøkt hvorvidt effekten av lærerens praksis i klasserommet har større påvirkning på resultater i små eller store klasser. Lærere har ulik måte å drive sin praksis på i klasserommet. Effekten av denne ulikheten kaller de lærereffekt. Fordi læreren bruker mesteparten av sin tid i klasserommet på disse trinnene, ønsker de å undersøke om effekten av lærerens praksis er større i mindre klasser. Derfor ser de på om klassestørrelse interagerer med og påvirker lærerens praksis i klasserommet, det vil si om effekten læreren har på elevers prestasjoner er forskjellige i små eller større klasser. Lærereffekt måles i analysen ut fra feilledet, og et større feilledd innebærer at en større andel av prestasjonene skyldes ulik praksis blant lærere. De ser på effektene separat for de fire årene fra barnehagen til 3. klasse i STAR-prosjektet. Resultatene er svake, og det er ikke noe systematisk mønster (Konstantopoulos & Sun, 2014). Det kan muligens være slik at det er mer variasjon i effekten av lærere i små klasser enn store, men at det ikke noe klart mønster i funnene. En tolkning er at dette kan være at de fleste lærere ikke endrer nevneverdig på undervisningen om den er for 15 eller 30 elever. De etterspør dermed en studie der en legger føringer for endring i undervisning, for å se om en kan finne effekter og mønster.

Mueller (2013) ser isteden på hvordan erfaringen læreren har kan være en modererende faktor for effekten av reduksjon av klassestørrelse på elevers prestasjoner. Bakgrunnen for analysen er at det å redusere klassestørrelse gir store utgifter så det er derfor viktig å sikre at denne investeringen i ulike klasserom og med ulike lærere faktisk gir den ønskede effekten.

Lærererfaring definerer de ved hjelp av en binær variabel som er 1 hvis læreren har mindre enn 3 års erfaring og har verdi 0 dersom læreren har mer. Mueller ser altså kun på en enkelt egenskap ved læreren, erfaring, mens Konstantinopoulos og Sun ser på «hele læreren» i sin studie. Mueller finner effekt av reduksjon i klassestørrelse for erfarne lærere. OLS-estimatene for matematikk viser at elevene i små klasser med erfarne lærere i gjennomsnitt får 0,16 standardavvik bedre resultat på eksamenene i matematikk enn de i vanlig klasse med erfaren lærer. Denne effekten finner de ikke for uerfarne lærere. Videre i analysen viser de at det ikke er forskjell på nye og erfarne lærere i store klasser, bare i små. Dette tolker de dithen at de erfarne lærerne trolig gir høyere undervisningskvalitet i små klasser versus de uerfarne (Mueller, 2013).

Oppsummert viser studier etter 2010 basert på STAR-data sprikende resultater på hvorvidt redusert klassestørrelse har påvirkning på elevers prestasjoner. Gjennomgangen viser at vi de siste ti årene har fått flere studier med et sterkere metodisk design enn tidligere, og som tar inn over seg kritikken som er kommet av de første studiene som rapporterte positiv effekt. Fordi det kun er dette datamaterialet som eksisterer av en så omfattende randomisert kontrollert studie, gjør det at prosjektet er blitt «overforsket» i forhold til hvor mye vekt ett prosjekt bør tillegges. Når konklusjonene basert på prosjektet spriker i så stor grad gjør det at troverdigheten til selv sterke metodiske studier reduseres. Mueller (2013) poengterer nettopp dette, og etterspør andre randomiserte kontrollerte studier. Mangelen på liknende studier gjør at resultatene fra dette prosjektet blir brukt til å argumentere med av beslutningstakere i ulike situasjoner. Derfor blir resultatene overtolket og en trekker mer ut fra STAR-prosjektets resultater enn det som er gunstig for bare ett enkeltstående prosjekt.

En annen svakhet som påpekes ved STAR-prosjektet er overrepresentasjonen av store skoler i prosjektet. Det ble tatt en beslutning om å kun innlemme skoler som har minst tre klasser per trinn (Hanushek, 1999). Dette ble gjort for å kunne randomisere etter normal, liten klassestørrelse og normal med ekstra ressurs på alle deltakende skoler. Det er i utvalget både en overrepresentasjon av urbane skoler og skoler med høy grad av minoritets elever. Mens gjennomsnittlig andel minoritets elever i aldersgruppen lå på 21 % i delstaten, er det i utvalget 37 % minoritets elever (Hanushek, 1999; Mueller, 2013). Det kan derfor stilles spørsmål ved hvorvidt datasettet gir et representativt utvalg av populasjonen.

Til slutt vil jeg peke på et generelt problem med denne type studier, og det er at lærerne og elevene vet at de deltar i tiltaksgruppen og dermed kan endre atferd deretter. Lærerne kan selv påvirke hvordan de underviser i små grupper. Siden det er funnet mange positive effekter i disse studiene sammenliknet med kvasiexperimentelle studier kan det være en potensiell utfordring ved STAR-prosjektet.

2.2.2 Forskning på data fra TIMSS-undersøkelser

Videre vil jeg rette fokuset mer direkte mot eldre elever og forskning på klassestørrelse fra fjerde trinn og oppover. Som tidligere nevnt er årsaken til at jeg har brukt så mye plass på STAR-prosjektet at all forskning på området, uansett trinn, henviser til denne litteraturen. Hovedfokuset videre vil ligge på internasjonale studier knyttet til data fra Trends in International Mathematics and Science Study (TIMSS). Grunnen til at jeg fokuserer på TIMSS er at dette omhandler blant annet matematikk, og at studiene er designet med RD-design, som er anerkjent som en god forskningsmetode (Wooldridge, 2016).

TIMSS er en internasjonal undersøkelse som gjennomføres hvert fjerde år i matematikk og naturfagene. Undersøkelsen gir grunnlag for å kunne sammenlikne på tvers av landegrenser, og det deltar ofte opp mot 65 land i undersøkelsen. Undersøkelsen måler elevenes kompetanser i matematikk og naturfag på 4./5.-trinn og 8./9.-trinn. Hvilket trinn varierer litt fra skolesystem til skolesystem. I tillegg samles det inn informasjon via spørreskjemaer fra elever, lærere og skoleledere (Universitetet i Oslo, 2019).

Wößmann og West (2006) ser på effekten av klassestørrelse på prestasjoner i 11 forskjellige skolesystemer i verden. Dataene er hentet fra TIMSS-undersøkelsen fra 1994/1995 og tester et representativt utvalg elever i to etterfølgende klasser med alder rundt 13 år i 38 land. Forskerne utnytter i sine analyser nettopp det at databasen inneholder informasjon om klassestørrelse på to påfølgende trinn på hver skole, samt gjennomsnittlig klassestørrelse i hver klasse på hver skole. Som beskrevet nærmere i metodekapitlet kan enkle regresjonsmodeller føre til forventningsskjevne estimater fordi det er endogene faktorer som er forskjellige mellom elever i ulike klasser. Ved å benytte fasteffekt for skole kan en i større grad kontrollere for dette, noe Wößmann og West (2006) har gjort her. Forskerne peker på at foreldre kan velge at deres barn går på skoler med større eller mindre klassestørrelser på bakgrunn av barnets resultater. Hvordan dette fungerer varierer selvsagt fra land til land, men det blir et problem når en skal estimere effekten av klassestørrelse, da det kan tilsløre det

kausale forholdet mellom klassestørrelse og elevprestasjoner hvis elevene har ulik bakgrunn i klasser med ulik størrelse. Dette problemet kontrollerer de for ved bruk av fasteffektmetoden, hvor de ser på effektene innad på en skole og mellom skoler. Denne metoden skal sikre at potensiell endogenitet på klassestørrelsesvariabelen ses på innad i skolene, og endogenitetsproblemene blir dermed mindre når en kontrollerer for skolefasteffekter enn ved bruk av for eksempel OLS.

Resultatene viser at de kun finner klassestørrelseseffekter i land der lærerlønnen er lav. Av de 11 landene er det kun i Hellas og Island de finner signifikant positiv effekt av mindre klassestørrelse. Begge landene har lav lærerlønn og forskerne finner at de, sammenliknet med de andre landene, har en lærerstab med et lavere utdanningsnivå. I studien måler de kvalitet ut fra hva slags utdanning læreren har, og legger dermed til grunn at land med lavt utdannede lærere har lavere kvalitet på lærerne. I de øvrige åtte landene finner de enten ingen signifikant effekt eller effekter av ubetydelig størrelse. Ifølge Wößmann og West (2006) viser resultatet at svaret på om klassestørrelse påvirker prestasjoner kan å ha noe med hvilket skolesystem en undersøker, og at effekter en finner i ett land ikke er overførbart til et annet. Der de tror reduksjon av klassestørrelse er nyttig ut fra deres forskning, er i land der kvaliteten på lærerstaben i utgangspunktet er lav. Det kan forklare hvorfor det er forskjeller mellom landene. Eksempelvis er Singapore et av landene i studien som ikke viser noen effekt av lavere klassestørrelse, men som har høyt prestasjonsnivå og høy kvalitet på lærerstaben. I konklusjonen fremmer forskerne, basert på studien, at det trolig er mer hensiktsmessig å ansette lærere med høy kompetanse, fremfor kvantitet. Da dette er mer en hypotese enn noe de kan teste direkte ønsker forfatterne mer forskning på kvalitet på lærer versus klassestørrelse i ulike land.

Altinok og Kingdon (2012) sin artikkel tar opp mye av den samme problematikken som Wößmann og West (2006) og prøver også å svare på kritikken metaanalyser ofte får med tanke på endogenitetsproblemer i klassestørrelsesvariabelen. I artikkelen brukes data fra TIMSS 2003 for 47 land og fasteffektmetoder for å utelukke de endogene problemene. De måler effekten av klassestørrelse ved å relatere forskjellen i en elevs prestasjoner på tvers av fag til forskjellen i elevens klassestørrelse på tvers av fag. Resultatene viser at det er en statistisk signifikant negativ effekt av redusert klassestørrelse i 14 av landene, men at denne effekten er liten i de aller fleste tilfellene. De kontrollerer både for elevers prestasjoner i hver av fagene, fagspesifikke lærerkarakteristikker, og for målefeil. Likevel holder resultatene. I

tillegg finner Altinok og Kingdon (2012) de samme resultatene som Wößmann og West (2006), at effektene av klassestørrelse er mindre i land med høyere kvalitet på lærerstaben. De måler kvalitet på samme måte som Wößmann og West (2006).

Flere forskere har valgt å fokusere på hvordan klassestørrelse påvirker elevers prestasjoner i mindre utviklede land. En av disse er Breton (2014), som har undersøkt TIMSS-data fra Colombia. Colombia, som de fleste andre utviklingsland, presterte langt under gjennomsnittet på de matematiske testene på TIMSS 2007. De fikk en gjennomsnittscore på 355, mens gjennomsnittet internasjonalt var 500. Fordelingen av testresultatene viser at 95 % av de colombianske elevene scorer under det internasjonale gjennomsnittet på 500. Ved bruk av dataene fra Colombia undersøkte Breton hvordan klassestørrelse påvirker elevenes prestasjoner. Utvalget var 2361 elever i 66 klasser. Etter kontroll for ulike karakteristikk som for eksempel kjønn, datamaskin hjemme, antall bøker hjemme, andel familier med høy inntekt på skolen og lærererfaring og dens utdanning finner han i sine regresjonsanalyser at større klasser har ugunstig effekt på elevers prestasjoner. En økning i klassestørrelse fra 20 til 53 elever reduserer resultatene i gjennomsnitt per elev med 80 poeng, dvs. 2,4 poeng for hver ekstra elev i en klasse. Breton mener at det mest sannsynlig er den kumulative effekten av klassestørrelse vi ser her, altså opphopning av effekten fra 1. til 4. klasse. Det pekes på foreldres lave utdanningsnivå og inntekt som årsaken til at prestasjonene bedres så markant med færre elever i klassen. Trolig får elevene i Colombia lite hjelp hjemme og mer tid med læreren gir dermed stor effekt. Resultatene i denne studien er trolig lite overførbare til situasjonen i andre, mer utviklede land der vi snakker om å redusere klassestørrelse fra 30 elever og nedover, mens det her er snakk om å gå fra mellom 50-80 og ned til 30. Effekten av klassestørrelse er mye større i Colombia og prestasjonene mye lavere enn i utviklede land.

Konstantopoulos og Shen (2016) er spesielt interessant for min oppgave, både grunnet funnene i artikkelen og metoden de benytter. De bruker data fra TIMSS-undersøkelsen fra både 2003 og 2007 når de undersøker sammenhengen mellom klassestørrelse og prestasjoner på offentlige skoler på Kypros på 4. trinn og 8. trinn. Resultatene indikerer at det ikke er et signifikant forhold mellom klassestørrelse og matematiske prestasjoner på 8. trinn, men de finner små positive effekter på 4. trinn.

For å gjennomføre selve analysen bruker de et design kalt regresjonsdiskontinuitet (RD-design) for å undersøke de kausale effektene mellom klassestørrelse og prestasjoner. På

tidspunktet for undersøkelsen hadde Kypros et vedtak som sikret at maksimum klassestørrelse var 30 elever. Dette gjør at en kan bruke RD-design for å beregne kausale klassestørrelseeffekter som ligger nær diskontinuitetspunktene. Ideen er å sammenlikne klassestørrelser på 29 og 30 elever med klasser som har 31 og 32 elever og dermed blir delt i to, med 15-16 elever i hver klasse. Siden det er tilfeldig om en klasse når 30 eller 31 elever vil alt annet i gjennomsnitt være likt for klasser på 30 og klasser på 15/16 elever bortsett fra stor forskjell i klassestørrelsen. En svakhet ved studien er at estimatene på klassestørrelseeffekter ikke tar opp i seg at skolene kan endre andre ting når klassestørrelsene endres rundt regelen om 30 elever. Hvis en har skolefasteffekter kunne en sammenlikne klasser rundt 30 elever innad på samme skole og dermed bedre kontrollert for forskjeller på skolens lærerstab og lærerpraksis.

Li og Konstantopoulos (2016) bruker samme metoden som Konstantopoulos og Shen (2016) når de skal se på effektene av klassestørrelse på 4. trinns elevers matematikkprestasjoner for 14 europeiske land ut fra TIMSS-data fra 2011. Alle de 14 landene har regel om maksimum klassestørrelse, som gjør at RD-design som metode egner seg. De sammenliknet i tillegg resultatene fra metodene over med OLS, og i de fleste tilfeller var resultatene samsvarende. Forfatterne finner noen positive sammenhenger for Romania og Slovakia, men finner at det overordnet ikke er noen systematiske klassestørrelseeffekter på tvers av landene. Resultatene indikerer at i de fleste europeiske land er ikke reduksjon i klassestørrelse et effektivt virkemiddel for å forbedre matematikkprestasjoner i 4. klasse.

2.2.3 Øvrig forskning internasjonalt

Selv om mye av forskningen på klassestørrelsesreduksjon er gjort med utgangspunkt i STAR- og TIMSS-data, er det også en del andre internasjonale studier som er relevante for min analyse.

I kjølvannet av STAR-prosjektet ble det i Florida innført en ny policy for reduksjon i klassestørrelse. Denne bestemte at det ikke skulle være mer enn 22 elever fra 4.-8. trinn og 25 fra 9.-12. trinn. Chingos (2012) ser på effektene av denne statlige innføringen av reduksjon i klassestørrelse for 4.-8. klasse i årene 2004 til 2009. I Chingos metode sammenlikner han avvikene fra tidligere prestasjonstrender i distrikt som en vet ble pålagt å redusere klassestørrelse mot avvik fra tidligere trender i distrikter som fikk samme mengde ressurser, men som hadde en klassestørrelse så nærme det nye målet fra før at de nå kunne bruke

ressursene slik de ønsket. Skoledistriktene fikk altså like mye penger i forbindelse med reduksjon i klassestørrelse, selv om det ene distriktet ikke trengte å redusere klassestørrelse. Han bruker videre det samme komparative designet til å sammenlikne skoler som hadde ulik påvirkning av tiltaket, det vil si om de ble pålagt å redusere klassestørrelse eller ikke. Resultatene fra regresjonsanalysene både på distriktsnivå og skolenivå indikerer at mandatet med reduksjon i klassestørrelse i Florida hadde liten eller ingen effekt på elevers prestasjoner i matematikk og lesing. Ifølge Chingos (2012) er overraskelsen i resultatene at det på skolenivå ikke er noe bevis for positive effekter av klassestørrelsesreduksjon. Flere av estimatene var derimot negative og statistisk signifikante. Han finner ingen bevis for heterogene effekter på noen av nivåene.

Hattie er en av vår tids mest refererte forskere når det kommer til klassestørrelse. I boka hans fra 2009 oppsummerer han over 800 metaanalyser og effekten av ulike initiativ for å bedre prestasjoner, og viser han at å redusere klassestørrelsen gir svært lav effekt (Hattie, 2009). Effekttørrelsen finner han til å være mellom 0,1-0,2 standardavvik. Dette kalkuleres ut fra å dele endringen i elevers testresultater på prestasjonenes standardavvik. Hattie (2005) finner at lærere selv tror at de kan bedre elevers prestasjoner om klassestørrelsen er mindre, men at utallige studier har vist at de positive effektene er svært små. Han peker på dette som en utfordring ved forskningen på reduksjon i klassestørrelse; selv om mange finner positive effekter, også studier som er klassifisert som metodisk gode, blir det lite problematisert at effekten av tiltaket ofte er veldig liten. Hattie konkluderer med at gitt de store kostnadene ved å redusere klassestørrelse, er det ikke et effektivt virkemiddel for å bedre prestasjonene.

Et annet viktig poeng i Hattie (2009) er at disse resultatene ikke bør generaliseres på tvers av land, spesielt ikke mellom lavt og høyt utviklede land. I studier fra høyt utviklede land har en ofte sett på reduksjoner fra 25 til 15 elever per klasse, mens det i utviklingsland ofte er snakk om reduksjon fra store klasser på 30-80 ned til 20-30 elever. I store klasser på 30-80 elever observerer Hattie at læring er mer knyttet til mekanisk innlæring, mens det i klasser på 20-30 i større grad kan tas i bruk pedagogiske verktøy som grupper og mer variert undervisning. Endringene i hvilke pedagogiske verktøy som benyttes er derfor mindre ved reduksjon fra 25 til 15 elever enn for samme reduksjon fra veldig store grupper. Effektene av reduksjoner fra 30-80 til 20-30 elever per klasse kan derfor umiddelbart ikke generaliseres til å gjelde på reduksjoner fra 25 til 15 elever per klasse. Det innebærer dermed at resultater fra utviklingsland ikke er direkte overførbare til høyt utviklede land, og omvendt.

Hanushek og Wößmann (2017) peker på en relatert problemstilling i sin gjennomgang av forskning på ressurser i skolen og elevers prestasjoner. Mønsteret på tvers av landegrenser er at klassestørrelse kun er en relevant variabel der hvor lærerkvaliteten er lav. Spørsmålet de stiller seg er da om ressursene i de landene som finner klassestørrelseseffekter egner seg bedre til å bruke på å forbedre lærerkvaliteten enn å redusere klassestørrelse. Totalt konkluderer de at ut fra internasjonal forskning, på tvers og innad i de ulike landene, har en lite som underbygger at klassestørrelse er en stor driver for elevers prestasjoner. Dette fordi studier som bruker ulike metoder for å identifisere kausale klassestørrelseseffekter gjennomgående ikke finner sterke effekter av klassestørrelse i de fleste land (Hanushek & Wößmann, 2017).

2.2.4 Oppsummering av internasjonal forskning

Oppsummert virker effektene av reduksjon i klassestørrelse i den internasjonale forskningen å være små og til dels ikke-eksisterende i majoriteten av tilfellene. Studier etter 2010 basert på STAR-data viser sprikende resultater på hvorvidt redusert klassestørrelse har påvirkning på elevers prestasjoner. Forskningen med bakgrunn i TIMSS-data finner at det er store forskjeller i lands skolesystemer og at landets demografi, økonomi og satsning på skole har stor påvirkning på hvorvidt en finner effekter av klassestørrelse. I land som er sammenliknbare med Norge finner en ikke effekter av klassestørrelse, eller så er effektene svært marginale. I sin litteraturanalyse finner Hattie (2009) at hvis det er en effekt, er gjennomsnittseffekten av klassestørrelsesreduksjon så liten som mellom 0,1 og 0,2 standardavvik. Disse anses ifølge han som spesielt små dersom en sammenlikner med effekten av andre mulige utdanningsintervensjoner. Som vi ser ut fra den internasjonale forskningen er det kun forskning på klassestørrelse og ikke på lærertetthet i undervisningen. Videre skal jeg gå inn på den norske forskningen gjennomført på effektene av klassestørrelse og lærertetthet på prestasjoner.

2.3 Forskning på lærertetthet i Norge

I gjennomgangen av norsk forskning på lærertetthet ser vi at det i den tidligste forskningen er mest fokus på klassestørrelse, slik som i den internasjonale forskningen generelt. Dette fokuset dominerer i forskningen frem til innføringen av ny læreplan i 2006. Frem til 2006 var styringen i norsk skole karakterisert ved at den var sterkt sentralisert. Det var strenge regler og reguleringer som påvirket antall elever per lærer, blant annet at det maksimum skulle være 30

elever i en klasse på ungdomstrinnet og 28 elever på barnetrinnet (Kunnskapsdepartementet, 2004). Regelen om maksimum klassestørrelse i grunnskolen ble opphevet i 2003, og i stedet ble det innført en bestemmelse om pedagogisk forsvarlig gruppestørrelse (Kunnskapsdepartementet, 2004; Opplæringslova, 1998). Etter 2003 ble det dermed mindre strenge krav til organiseringen i klasserommet og skolene fikk bestemme mer selv, spesielt ble dette gjeldende etter innføringen av Kunnskapsløftet i 2006. De lokale skolene og administrasjonene skulle ha mer autonomi. Dermed skjer det også en dreining i den norske forskningen fra klassestørrelse mot mer forskning på lærertetthet.

I denne delen vil jeg først gjennomgå norsk forskning på effekter av klassestørrelse, deretter forskning knyttet til gruppestørrelse 1 som mål på lærertetthet, for tilslutt å se på forskning knyttet til gruppestørrelse 2.

2.3.1 Forskning på klassestørrelsesreduksjon

I rapporten *Læringsmiljø og utvikling* fra 2002 oppsummeres Birkemoprojektet «Fra ressurs til resultat» (Birkemo, 2002). Siden prosjektet er gjennomført før lovendringen fra regelen om maksimal klassestørrelse ble endret, er fokuset i forskningen på klassestørrelse. Rapporten tar utgangspunkt i en studie der utvalget er elever på 9. og 10. trinn. Elevene følges gjennom to år og gjennomgikk fire tester mellom 1998-1999. Rapporten ser på hvordan ressursinnsats i skolen samvarierer med læringsutbyttet ulike elever i skolen får. Datainnsamlingen gikk ut på at elevene en gang hvert halvår i de to årene fikk en matematikkprøve, en norskprøve og spørreskjemaer for holdninger og sosiale ferdigheter. I tillegg samlet de i løpet av de to årene inn informasjon om alder, kjønn, sosioøkonomiske- og kulturelle forhold, samt undervisning og læringsmiljø. Prosjektet inkluderte 58 skoler og ca. 2500 elever. Birkemo finner ingen gjennomgående sammenheng mellom klassestørrelse og prestasjonsnivå. Han finner riktignok en liten positiv effekt i matematikk og en liten positiv effekt på gjennomsnittlig karakternivå, men for norskfaget finner han ingen effekter. I rapporten skriver Birkemo dette om funnene: «I den grad det kan spores noen positiv faglig effekt av små klasser, er effekten så svak at den neppe har noen praktisk betydning» (Birkemo, 2002, s. 222). Rapporten konkluderer med at det ut fra dette utvalget ikke er mulig å dokumentere at ressursbruk til for eksempel redusert klassestørrelse har innvirkning på elevenes resultater over tid. Birkemo mener at økte ressurser til klassestørrelsesreduksjoner trolig ikke er et hensiktsmessig middel for å bedre elevenes prestasjoner.

Bonesrønning (2003) ser på sammenhenger og mønstre mellom klassestørrelse og elevens prestasjoner. Han peker på at det hittil i forskningen på klassestørrelse ikke er noe konsensus, blant annet fordi de kausale effektene av klassestørrelse er vanskelig å etablere fordi elever ikke tilfeldig tildeles klasser av ulik størrelse. I sin kvantitative studie der han ser på kausale effekter av klassestørrelse på elevenes prestasjoner bruker han forutsetningen at vi i Norge hadde en regel for maksimal klassestørrelse på 30 elever per klasse for å utvikle et instrument for faktisk klassestørrelse. Han ser på de kausale effektene av klassestørrelse ved å bruke variasjonen i klassestørrelse generert av interaksjonen mellom innskriving i skolen og maksimal størrelse på en klasse. Dette er samme RD-designet som jeg beskrev for den internasjonale forskningen på CSR i kapittel 2.2.2. Utvalget bestod av utvalgte ungdomsskoler i Norge. Han bruker her resultatene fra matematikkprøvene i Birkemoprojektet, som han var en del av. Bonesrønning finner også små effekter. Noen elevgrupper presterer bedre i matematikk i mindre klasser, men disse effektene er vanskelig å forklare da det ikke finnes noen tydelige felles karakteristikk for disse elevgruppene som binder dem sammen. Videre i analysen kommer det frem at fordelene av mindre grupper er større på skoler med elever med høy sosioøkonomisk status, spesielt da med en høy andel elever der foreldrene ikke er skilte (Bonesrønning, 2003). Men fordi utvalget er lite når det deles opp i mindre grupper er det vanskelig å trekke sterke konklusjoner fra dette.

I 2008 gjennomførte Leuven, Oosterbeek og Rønning en omfattende kvasiekperimentell studie som så på effekten av størrelsen på klassen på prestasjonene til elevene i Norge. De brukte en omfattende database fra Statistisk sentralbyrå som dekket to hele årskull av elever på ungdomstrinnet som deltok i landsomfattende tester i skoleårene 2001-2003. Det spesielle med denne studien er at på grunn av denne omfattende datainnsamlingen vet de faktisk klassestørrelse på hver skole gjennom disse tre årene. Dette vil gi mer presise estimater enn studier som bare har data for hvert årskull, men ikke for hver klasse innenfor hvert årskull. Grunnlaget for analysene ble basert på to ulike og uavhengige tilnærminger. Den første ble gjennomført basert på regelen om maksimum klassestørrelse, på samme måte som Bonesrønning (2003), som instrument for klassestørrelse. Den andre tilnærmingen ser på variasjonen i faktisk klassestørrelse som kan tilskrives demografisk variasjon. Dette ga de samme resultatene, nemlig at de finner ingen eller minimal effekt av endringer i klassestørrelse. Fordi de bruker flere identifiseringsmetoder kan de utelukke effekter helt ned til 1,5 % av et standardavvik for endring i én elev i klassestørrelse i tre påfølgende år (Leuven, Oosterbeek & Rønning, 2008).

I en analyse av nasjonale prøver på 5. trinn og 8. trinn fra 2007 og 2008 gjennomført av Bonesrønning og Vaag Iversen for Utdanningsdirektoratet, finner de liten effekt av økt lærertetthet på elevenes prestasjoner på nasjonale prøver. Forskerne bruker multipl regressjonsanalyse som metode i rapporten. De finner som sagt liten effekt av økt lærertetthet på elevenes prestasjoner, bortsett fra at de finner effekt for jenter på 5. trinn som kommer fra familier med lav sosioøkonomisk status. Disse jentene gjør det signifikant bedre med økt lærertetthet. Denne effekten finner de ikke på 8. trinn. I rapporten fra 2010 finner de også en signifikant negativ effekt for 5. trinns elever som går på en 1.-10.-trinnskole, men denne effekten på prestasjonene finner de ikke for 8. trinn. I rapportene sår de imidlertid usikkerhet rundt hva dette funnet egentlig forteller oss (Bonesrønning & Vaag Iversen, 2008, 2010).

I doktorgraden til Vaag Iversen ser han på effekten av klassestørrelse opp mot læringsutbytte ved å undersøke bruken og effekten av skolens ressurser i L97 og LK06, og ser om det er forskjell mellom disse (Vaag Iversen, 2013). Han spiller videre på funnene fra Bonesrønning og Vaag Iversens tidligere arbeid og undersøker hvorvidt redusert klassestørrelse er et effektivt verktøy for å bedre prestasjoner i L97, med fokus på elever med lav sosioøkonomisk status. I den første artikkelen i avhandlingen, som han skriver sammen med Bonesrønning, utnytter de at Norge på den tiden hadde maksimal klassestørrelsesregel, og anvender derfor RD-design som metode. De finner at det å redusere klassestørrelse trolig ikke er et effektivt virkemiddel for å øke den gjennomsnittlige elevens prestasjoner, men at det kan brukes som et verktøy som kan være sosialt utjevne for elever med foreldre med lav sosioøkonomisk status (Vaag Iversen & Bonesrønning, 2013). Det vil si at elever med foreldre med lav sosioøkonomisk status presterer bedre i små klasser enn i store. Dermed blir forskjellene her mindre. Dette er i tråd med flere av de overnevnte internasjonale studier (Finn, 2002; Finn & Achilles, 1990; Wilde et al., 2011). Denne effekten finner de ikke for barn med foreldre med høy sosioøkonomisk status. De finner at dette også gjelder for barn med skilte foreldre. Dette funnet står i kontrast til funnet Bonesrønning (2003) har, der elever som ikke har skilte foreldre er de som har størst fordel av mindre grupper. I Vaag Iversen og Bonesrønning (2013) sin studie er effekten av å endre klassestørrelsen for elever med foreldre med lav sosioøkonomisk status og elever med skilte foreldre størst for klasser mellom 11 og 18 elever. Dersom klassestørrelsen er over 18 elever finner de ingen signifikante negative effekter ved å øke klassestørrelsen til og med 28 elever for noen av gruppene.

Falch, Sandsør og Strøm (2017) ser på effekten av mindre klasser på elevenes fremtidige inntekt og utdanning. De bruker samme informasjon som flere av de overnevnte studier når det gjelder regelen med maksimum klassestørrelse og sammenlikner dette opp mot informasjon om individuelle elever og skoleregisterdata fra 1982-2011. I og med at denne maksimum klassestørrelsesregelen var gjeldende, gjorde dette det mulig å sammenlikne skolekull der det akkurat var så mange elever at de opprettet en ekstra klasse mot skolekull der det var akkurat så mange elever at de ikke opprettet denne ekstra klassen. Dette skaper variasjon i dataene på klassestørrelse som vil være uavhengig av elevbakgrunn og individuelle forutsetninger, og dette bruker forskerne til å måle og isolere effekten av endringer i klassestørrelse. Dermed kan de undersøke hvorvidt klassestørrelseseffekter i ungdomsskolen avhenger av karakteristikk ved miljøet som skolene og elevene opererer i. Det er da naturlig å bruke RD-design slik mange av studiene gjengitt over har gjort. Gjennom OLS og bruk av faste effekter, på blant annet kull og skole, estimerer de kausale langtidseffekter på inntekt og utdanning til elevgruppen. De bruker registerdata og kvasiexperimentell metode for å estimere hvorvidt effekten av klassestørrelse avhenger av eksterne forutsetninger som kvalitet på lærer, skoledistrikt, lokale skattemessige begrensninger og forhold i arbeidsmarkedet. Resultatene viser at klassestørrelseseffekten ikke er avhengig av disse faktorene, og ikke spiller inn på inntekt og utdanning på lang sikt (Falch et al., 2017). Selv ikke når de kontrollerer for kjennetegn ved kommunene.

Leuven og Løkken (2020) har i likhet med Falch, Sandsør og Strøm undersøkt langtidseffektene av klassestørrelse i grunnskolen, og hvilken innvirkning klassestørrelsen har på hva slags inntekt elevene får som voksen, karriere og hvor lang utdanning elevene tar. De utnytter, som mange av studiene over, reglene for maksimum klassestørrelse, og isolerer variasjonen i inntekt og utdanning som kun er konsekvens av klassestørrelse. Statistisk sentralbyrå har en omfattende registerdatabase, og Leuven og Løkken trekker ut data for skolekohortene som fullførte grunnskolen fra 1979-2003. De så da på elever som hadde nådd en alder opp til 48 år. Den omfattende databasen gjør det mulig å observere den økonomiske utviklingen til elevene. Dette er den største norske langtidstudien på klassestørrelse. Denne langtidstudien finner ingen positive effekter av liten klassestørrelse på inntekt eller utdanning, det vil si at det hverken påvirker hvorvidt du tar høyere utdanning eller tjener mer (Leuven & Løkken, 2020). Dette er i samsvar med funnene til Falch et al. (2017) på omtrent samme datagrunnlaget. De norske funnene står i kontrast til en del av den internasjonale forskningen som finner langtidseffekter for ulike grupper (Wilde et al., 2011).

I den skandinaviske forskningen er det funnet effekter på lang sikt. En skulle tro at Sverige og Norge, som tilsynelatende har relativt like skolesystem med offentlige skoler med ønske om like muligheter på tvers av sosiale klasser, et velfungerende arbeidsmarked og et velutviklet velferdssamfunn finner like resultater, men slik er det altså ikke. I studien Fredriksson, Öckert og Oosterbeek (2013) ser de på elever fra 11-13 år, og bruker et liknende metodisk oppsett som de norske forskerne (Falch et al., 2017; Leuven & Løkken, 2020). De finner effekter på inntekt og utdanning i alderen 27-42 år. En mangel ved denne studien er at de bruker kun 10 % av utvalget for kohortene, og gjør en del metodiske grep for å sikre eksogenitet, noe som fører til at utvalget kun blir 6000 elever og 191 skoler. I Falch et al. (2017) inkluderer de alle norske skoler i sin analyse. Utvalget utgjør da over 50 000 elever, noe som sikrer en større robusthet i analysen deres enn i Fredriksson et al. (2013). Spørsmålet er hvorfor forskerne finner forskjellige effekter, når de skandinaviske landene ofte anses som svært like. Et argument kan være at skoledistriktene i Norge ofte er mye mindre enn de i Sverige. Dette avvises av Falch et al. (2017) da de ikke finner at effekter av klassestørrelse i Norge er avhengig av størrelse på skoledistrikt. Falch et al. (2017) undersøker flere kilder til heterogenitet i klassestørrelseseffekter ved å interagere klassestørrelse med indikatorer knyttet til lærerkvalitet, skolevalg, skoledistrikt, lokale skatteregler og arbeidsmarkedet med samme rammeverk. De finner ikke at klassestørrelseseffekter avhenger av noen av disse eksterne føringene. Forskerne mener at resultatene står i skarp kontrast til funnene fra STAR-eksperimentet og de kvasiekperimentelle funnene fra Sverige, som begge finner signifikante positive effekter på lang sikt. Mangelen på interaksjonseffekter av de eksterne kildene til heterogenitet indikerer at det må være andre forskjeller mellom landene i lærerpraksis og utdanningskultur som kan være forklaringen på hvorfor en får så ulike resultater.

2.3.2 Forskning på gruppestørrelse 1

I en NIFU-rapport fra Grøgaard et al. (2008), der de gjennomfører en analyse av ulikhet i elevers prestasjonsnivå i 4., 7. og 10. trinn, bruker de et omfattende datasett satt sammen av flere registre i samarbeid med Statistisk Sentralbyrå. Ett av temaene de har undersøkt er effekten av endringer i gruppestørrelse 1, men dette er kun ett av mange temaer i rapporten, og ikke hovedfokuset. Datasettet inneholder blant annet resultater fra nasjonale prøver etter fjerde og sjuende trinn i skoleåret 2004-2005, i tillegg til eksamensresultater og standpunkt karakterer. Dataene er hentet ut fra Grunnskolens informasjonssystem (GSI). Forskerne har i tillegg tilgang til sensurert data i GSI som normalt sett ikke er tilgjengelig for alle grunnet personvern hensyn. Dette kobles i datasettet opp mot registeropplysninger om

elevens kjønn og sosioøkonomiske status. Tallene på lærertetthet er hentet fra GSI og regnet ut fra gruppestørrelse 1.

Analysene knyttet til lærertetthet viser at økt lærertetthet gir dårligere prestasjoner i norskfaget på 7. trinn, altså finner de en signifikant negativ effekt. Deres tolkning av denne effekten er at det trolig settes inn flere lærere som en respons på for eksempel egenskaper ved elever som krever flere lærere og som fører til prestasjonsproblemer. Samtidig peker de på et manglende sammenlikningsgrunnlag innad i skolen, tilstanden kunne vært verre dersom flere lærere ikke ble satt inn. Denne effekten holder selv om de kontrollerer for enkeltvedtak og andel elever med spesialundervisning. I og med at den negative effekten holder til tross for disse kontrollvariablene, diskuterer de hvorvidt det er mulig at lærertettheten kan bli for stor, og klarer ikke utelukke dette alternativet. Men de tror mest på tolkningen rundt økt ressursinnsats grunnet prestasjonsproblemer (Grøgaard et al., 2008).

I en sammensatt NIFU-rapport undersøker Opheim et al. (2010) betydningen av skoleressurser, undervisningsformer og læringsmiljø for elevers prestasjoner på 5., 8. og 10. trinn. Rapporten baserer seg på data fra GSI og elevundersøkelsen for 5. og 8. trinn i årene 2007-2009, samt 10. trinn vår 2009. I tillegg har de individdata om sosioøkonomiske forhold knyttet til eleven. Resultatene knyttet til lærertetthet viser uklare og sprikende effekter på elevers prestasjoner ved økt lærertetthet. Blant annet finner de en klar negativ effekt av økt lærertetthet på prestasjoner fra den nasjonale prøven i matematikk på 5. trinn, men ikke på prøvene i lesing og engelsk. På 8. trinn finner de signifikant positiv effekt av økt lærertetthet på prestasjoner på den nasjonale prøven i norsk på 8. trinn, men ikke på de to andre. I rapporten konkluderes det med at disse sprikende og lite gjennomgående funnene indikerer at effekten av lærertetthet er lite stabil (Opheim et al., 2010). Som Grøgaard et al. (2008) er også dette en omfattende rapport som tar for seg en rekke temaer. Siden lærertetthet ikke er hovedtema går de ikke detaljert inn på resultatene knyttet til disse funnene.

2.3.3 Forskning på gruppestørrelse 2

Gruppestørrelse 2 som mål på lærertetthet er et relativt nytt begrep, og det er derfor foreløpig ikke publisert mer enn én studie (Kirkebøen et al., 2017) som tar for seg effektene av dette på prestasjoner. Bakgrunnen for studien er at det i statsbudsjettet i 2012 ble bevilget 1,5 milliarder kroner over en periode på 4 år til 600 nye lærerstillinger på ungdomstrinnet. Dette

ble fordelt på 166 ungdomsskoler i 98 kommuner som fikk øremerkede midler til å ansette tilsammen 573 årsverk. Hvor mange årsverk de ulike skolene ble tildelt ble bestemt ut fra elevtallet på ungdomstrinnet, og hver skole fikk inntil 5 årsverk. Bevillingen gjaldt fra skoleåret 2013/2014 frem til og med skoleåret 2016/2017. Kravene til de utvalgte skolene var at de måtte ha en gruppestørrelse i ordinær undervisning på over 20 elever og grunnskolepoeng som lå under landssnittet. Lærerstillingene kunne kun brukes til å øke lærertettheten i ordinær undervisning, og et krav var å ansette lærere med godkjent utdanning. Skolene hadde pedagogisk frihet til å selv bestemme hvordan undervisningen skulle organiseres (Utdanningsdirektoratet, 2017c).

Kirkebøen, Kotsadam, Raaum, Andresen og Rogstad utførte en studie på gruppestørrelse 2 som mål på lærertetthet på vegne av Utdanningsdirektoratet (Kirkebøen et al., 2017). Målet med deres arbeid var å finne ut om lærertettheten i ordinær undervisning har økt og om den har gitt prestasjonsresultater. I analysen bruker de også RD-design, regresjonsdiskontinuitet. De tar utgangspunkt i skoler som på starttidspunktet har nesten samme gjennomsnittlig grunnskolepoeng og gruppestørrelse i ordinær undervisning. Sammenlikningsgruppen fikk ikke tilskuddet. De bruker de to faktorene gruppestørrelse i ordinær undervisning (regnet etter gruppestørrelse 2) og elevenes prestasjoner, som er målt etter skriftlig eksamen på 10. trinn og nasjonale prøver på 9. trinn. Etter innføring av tilskuddet kan en over tid se om disse skolene presterer ulikt.

Resultatene de finner viser at gruppestørrelsen går ned med 2 elever per lærer (ca. 10 %) i ordinær undervisning etter innføring av tilskuddet, men at dette ikke har effekt på elevprestasjonene. I konklusjonen finner de heller ikke at det er positiv effekt for noen undergrupper av elever, slik andre studier finner. I beste fall finner de at gjennomsnittlig eksamenskarakter går opp 0,06 karakterpoeng, noe som er veldig lite. Rapporten for Utdanningsdirektoratet bestod også av en kvalitativ del som ble gjennomført av Andresen og Rogstad fra Fafo. De finner gjennom intervju med skoleleder, lærere og elever i de skolene som har fått tilskudd generelt positive holdninger til økt lærertetthet. Lærerne opplever en effekt på deres arbeidshverdag, på elevenes læringsutbytte og læringsmiljø, selv om de kvantitative studiene ikke støtter dette. De finner at elevene og lærerne synes mindre grupper fungerer bra, og at nivådeling ikke oppleves som stigmatiserende. Rapporten drar dermed i to ulike retninger i den kvantitative og kvalitative delen, og det er manglende samsvar mellom aktørenes opplevelse, og funn fra effektanalysen (Kirkebøen et al., 2017). Den kvantitative

delen av denne studien er den metodisk mest solide norske studien gjort på lærertetthet, og den eneste studien, meg bekjent, som har brukt gruppestørrelse 2 som et mål på lærertetthet. De har også testet mange ulike kilder til heterogenitet i effekter, men ser ingen tegn til bedre prestasjoner.

Solheim og Opheim (2019) gjennomgår mye av den norske forskningen på reduksjon av klassestørrelse og lærertetthet, og presiserer at det i all hovedsak er klassestørrelsesreduksjon det er forsket på i Norge. Det er først i 2017 i Kirkebøen m.fl. de har funnet nyere forskning hvor forskerne ikke bruker klassestørrelsesregelen. Derfor vet vi lite fra forskningen om konsekvensene av økt lærertetthet på andre alternative metoder for å redusere antall elever per lærer, eksempelvis sette inn en ekstra kompetent ressurs i et klasserom. I artikkelen ser Solheim og Opheim på begrensningene ved tidligere forskning, spesielt da klassestørrelsesforskning, og foreslår mer fleksible måter å implementere økt lærertetthet på. De behandler spørsmålet «hvorfors trenger en å rekonseptualisere hvordan vi ser på reduksjon av elev-lærer-raten». Hovedproblemet er at en har sett på klassestørrelse som en statisk faktor. Men med dagens trender i klasserommet endres dette fra dag til dag, med for eksempel åpent klasserom og nivådeling i enkelte fag. I nyere tider ser vi en større fleksibilitet i klasserommet og størrelsene på gruppene, og dette krever en mer fleksibel forskningstilnærming. For eksempel trekker de frem det å være to lærere og samundervise som en kostnadseffektiv måte å øke lærertettheten og kunne gi elevene individuell tilpasset opplæring, men at dette er lite forsket på (Solheim & Opheim, 2019).

Oppsummerende kan en si at forskningen i Norge i svært liten grad finner effekter av endringer i klassestørrelse hverken på kort eller lang sikt. Dette er i samsvar med en del av den internasjonale forskningen som sier at land som har god lærerkvalitet sjelden finner effekter av redusert klassestørrelse. Etter innføring av pedagogisk forsvarlig gruppestørrelse i 2003 og innrapportering og beregning av både gruppestørrelse 1 og 2 fikk mer fokus, har forskningen hatt en dreining bort fra klassestørrelseseffekter og over på effekt av økt lærertetthet. Debatten fremover blir hvorvidt resultatene fra tidligere forskning er overførbare fra det ene området til det andre.

2.3.4 Pågående forskning på lærertetthet i Norge

I tillegg til den publiserte forskningen på lærertetthet er det også noen pågående prosjekter som er relevant for mitt forskningsspørsmål, og som jeg ønsker å gjennomgå kort.

Nordisk institutt for studier av innovasjon, forskning og utdanning (NIFU) skal på vegne av Utdanningsdirektoratet gjennomføre evalueringen av implementeringen av lærernormen (NIFU, 2019). Denne evalueringen skal gi innsyn i hvilken grad reglene om lærertetthet følges og om kompetansekravene til lærerne følges. I tillegg skal de undersøke struktur- og styringseffekter av normen, for eksempel hvordan normen påvirker klassestørrelse og hvordan organiseringen fungerer. De skal også se på effekten av normen på rekrutteringen til yrket og på elevprestasjoner. Evalueringen skal bestå av en kvantitativ del og en kvalitativ del med intervju og observasjon. Prosjektet startet i juni 2019 og varer til mai 2022 (NIFU, 2019).

LÆREEFFEKT er midler som skal finansiere forskning som omhandler økt lærertetthet på 1.-4. trinn. Innunder her er det mange prosjekter som er finansiert av forskningsrådet. Bare i 2018 ble det tildelt 12,4 millioner kroner til prosjekter knyttet til dette utgangspunktet for forskningen (Forskningsrådet, 2020). Foreløpig er det tre pågående studier knyttet til dette.

Den første av de tre er ledet av Opheim, og er et prosjekt som går fra 2015 til 2021. Prosjektet kalles ofte 1+1 prosjektet. Studien går på å studere undervisning av matematikk i små grupper, der de skal se på effektene av en randomisert kontrollert intervensjonsstudie (RCT-studie). Formålet er å se på effektene av økt lærertetthet på elevenes ferdigheter i matematikk på 2.-4. trinn, og se på om det å undervise matematikk i mindre grupper øker elevenes ferdigheter. 161 skoler, fordelt på 10 kommuner, deltar i prosjektet. I RCT-studien deles de deltakende skolene i to grupper, en kontrollgruppe og en forsøksgruppe. Forsøksgruppen får midler til å ansette en ekstra lærerressurs i en fireårs periode, mens kontrollgruppen ikke får dette.

Studien gjennomføres på en slik måte at en kan se effektene av små grupper i matematikkundervisning ut fra om eleven har hatt smågruppeundervisning i ett, to eller tre år, i tillegg til at en kan sammenlikne effekter av smågruppeundervisning blant elever med forskjellig alder (Forskningsrådet, 2015a). Forskningsprosjektet er designet metodisk slik at det ikke er så avhengig av at klassestrukturen er stabil over tid og kan dermed fortsette etter innføringen av lærernormen.

Det pågår også et tilsvarende prosjekt kalt Two Teachers ledet av Solheim som skal se på effekten av økt lærertetthet på leseundervisningen (Forskningsrådet, 2015b). Dette er det

andre finansierte prosjektet via LÆREEFFEKT. Bakgrunnen for studien er at de mener forskningen viser at det å sette inn en ekstra ressurs ikke har noen effekt på læringsutbytte til elevene, og de ønsker å se på hva som er effekten om en legger føringer for hva den ekstra ressursen skal brukes til. Studien går fra 1.-4. trinn. De randomiserer på to nivå innad i skolen, og i studien deltar det 150 skoler, totalt 300 klasser i 9 fylker i Sør-Norge. Først trekker de en klasse per skole til å få ekstra ressurser, deretter en klasse på samme skole til å være kontrollklasse. Deretter trekkes alle skolene tilfeldig inn i 3 kategorier. Kategori 1 har fått en ekstra ressurs, men ikke fått føringer for hvordan praksis skal utføres og ressursen utnyttes. Kategori 2 har fått en ekstra ressurs, og forpliktet seg til å «arbeide systematisk med kompetanseutvikling knyttet til lese- og skriveopplæringen på 1. til 2. trinn» (Forskningsrådet, 2015b). Kategori 3 er lik som kategori 2, men har i tillegg fått en del føringer for hvordan den ekstra ressursen skal brukes. «Som et resultat av ny lærernorm for grunnskolen fra høsten 2019, har intervensjonsdelen blitt justert ned fra opprinnelig 4 til 2 år, og intervensjonen ble avsluttet etter skoleåret 2017/18. Datainnsamlingen vil imidlertid fortsette som planlagt i prosjektperioden.» (Forskningsrådet, 2015b).

I det tredje prosjektet finansiert av LÆREEFFEKT skal et knippe norske og internasjonale forskere, med prosjektleder Sandsør i spissen, undersøke utdanningsressurser og elevprestasjoner. Forskningen foregår mellom 2016-2020. Målet for prosjektet er å gi en kausal forståelse om hvordan økt lærertetthet på 1.-4 trinn påvirker elevprestasjonene i Norge. De skal se på effekten av økt ressursbruk på prestasjonene i norsk, engelsk og matematikk ved å se hvordan elevene presterer på 5. trinn. I tillegg er det et mål å få kunnskap om hvordan myndighetene kan påvirke tilbudet som blir gitt av de lokale myndighetene. Studien skal utnytte at myndighetene i 2015 ga de 100 skoledistriktene med lavest lærertetthet ekstra ressurser til økt lærertetthet for å styrke tidlig innsats og elevenes læring på 1.-4.-trinn. Metoden skal bruke dette til å identifisere de kausale effektene av lærertetthet for elevenes prestasjoner. De skal også se på om midlene brukes til det de skal i de ulike skoledistriktene, og om forskjeller i egenskaper ved kommunene har noe å si for bruken av ressursene. De bruker i hovedsak data fra GSI og kommunedata fra SSB. Dette prosjektet samarbeider tett med de to andre prosjektene finansiert under LÆREEFFEKT-prosjektet. Det er vanskelig å finne informasjon om hvorvidt dette prosjektet ble stoppet etter innføringen av lærernormen, men det er foreløpig ikke kommet ut noe arbeidsnotat eller andre rapporter fra gruppen.

Når Norge omsider hadde satt i gang tre ulike intervensjonsstudier på tidlig innsats i skolen som forsket på lærertetthet og ikke på reduksjon i klassestørrelse, kom lærernormen midtveis i perioden og til dels forstyrrer en del av intervensjonsstudiene. Trolig er en 2-års periode for lite til å trekke gyldige slutninger i Two Teachers og for å kunne se de kausale sammenhenger mellom lærertetthet og prestasjoner i Sandsør sitt prosjekt. Dette er beklagelig sett fra et forskningsperspektiv, da en ikke får anledning til å fullføre studier som kunne gitt verdifulle svar på effekten av lærertetthet på elevers prestasjoner under norske forhold.

2.4 Relevant matematikdidaktisk forskning

I diskusjonen av resultatene fra analysene mine i kapittel 5, er det behov for å trekke inn deler av den matematikdidaktiske forskningen for å kunne forklare mulige årsaker bak mine funn. Denne forskningen vil i korthet bli gjennomgått under.

Gode diskusjoner gjennom helklassesamtaler i matematikk anses som en viktig del av effektiv matematikkundervisning (Cobb, Boufi, McClain & Whitenack, 1997; Nathan & Knuth, 2003). Det å sammen diskutere kognitivt krevende matematiske problemer bidrar til å fremme konseptuell forståelse og utvikler ferdigheter innenfor tenkning, resonnering og problemløsning (Cobb et al., 1997). Dette har vært mer fremtredende etter at lærerrollen endret seg fra at læreren overfører sin kunnskap, til at læreren fasiliterer mer for elevenes læring i matematikk (Stigler & Hiebert, 2009). Lærerens rolle i helklassediskusjonen vil være å bygge på elevenes forslag, resonnementer og innspill, fremfor å kun «fremme» ønskede løsningsmetoder som korrekte eller som følger en standard prosedyre (Fennema et al., 1996). Med dette som bakgrunn har Stein, Engle, Smith og Hughes (2008) utviklet et rammeverk kalt 5 praksiser, som skal støtte opp om at læreren kan legge opp til gode helklassediskusjoner. I diskusjonen skal elevenes mange mulige løsninger trekkes frem slik at klassen sammen i diskusjonen klarer å se matematiske sammenhenger mellom de ulike løsningene. En fordel er at elevene her både kan utvikle sine egne strategier, men at elevene også deler strategiene sine med hverandre (Stein et al., 2008). Stylianides (2019) undersøker i sin studie ungdomsskoleelevers konstruksjon av matematiske bevis, både skriftlig og muntlig, og finner at elevene i større grad konstruerer argumenter som møter kriteriene for et bevis muntlig foran klassen og i helklassediskusjon enn dersom de gjør det skriftlig. Gode diskusjoner vil trolig støtte elevenes læring i matematikk, da det hjelper å innlemme elevene i matematikdiskursen og utvikler den enkelte elevs evne til å resonnere og tenke kreativt.

Brousseaus Topaze-effekt stammer fra den franske matematikdidaktikken, og kalles også traktkommunikasjon (Brousseau, 2002). Effekten går ut på at læreren gir elevene så mye hjelp og hint i kommunikasjonen i klasserommet at det virker begrensende for elevenes læring og faglige utbytte. Dette fordi samtalen eller oppgaven tømmes for det matematiske innholdet og læringspotensialet, og kommunikasjonen får mer preg av «gjett hva læreren tenker på». Når læreren gir så mye hjelp og hint, snevres oppgaven til slutt så mye inn, at elever får servert fasiten. Det vil si at det lærere oppfatter som nyttig hjelp i virkeligheten kan frata elevene utfordringene og det matematiske innholdet i oppgaven (Skott, Hansen & Jess, 2008).

Begrepet «Grit» vil si at elevene har det vi på norsk vil kalle «ståpåvilje», indre driv eller pågangsmot (Bettinger, Ludvigsen, Rege, Solli & Yeager, 2018). Disse elevene vil ha personlighetstrekk som indikerer at de evner å arbeide mot langsiktige mål med pågangsmot og lidenskap (Duckworth & Quinn, 2009). En studie med fokus på matematikkundervisning i 1. klasse studiespesialiserende, har gjennom en randomisert kontrollert studie funnet effekter av å manipulere elevens grit og tro på egen evne til å lære. Effektene finner de kun for elever som før tiltaket har lite grit eller lave prestasjoner i matematikk. Etter tre uker med intervensjon finner de positive effekter for denne gruppen på elevenes grit og prestasjoner i matematikk (Bettinger et al., 2018).

Vi skiller mellom sosiale normer og normer som er spesifikke for matematikken, sosiomatematiske normer. De sosiale klasseromnormene omhandler den alminnelige deltakerstrukturen innad i klasserommet (Yackel & Cobb, 1996). Eksempel på slike normer er plassering i klasserommet og håndsopprekking. Dette er normer som er uavhengig av fag. Sosiomatematiske normer er normer spesielt knyttet til matematikkundervisningen (Yackel & Cobb, 1996). De definerer sosiomatematiske normer som en normativ forståelse for hva som anses som matematisk sofistikert, matematiske effektivt og matematisk elegant. På samme måte er hva som teller som en akseptabel matematisk forklaring og begrunnelse innad i klassen, en sosiomatematisk norm (Yackel & Cobb, 1996). Disse normene utvikles gjennom deltakelse. Utviklingen av den normative forståelse for hva som er akseptabelt utvikles i interaksjonen i klasserommet, gjennom klasseromdialogen og helklassesamtalen. Kommunikasjon er nøkkelen. Utvikling av sosiomatematiske normer en kontinuerlig prosess. Gjennom diskusjoner kan det endres hva som anses som effektive og elegante løsninger og begrunnelser (Kleve & Ånestad, 2016). Læreren må legge til rette for diskusjoner.

Gee (2015) skiller mellom primærdiskurs og sekundærdiskurs. Vi tilegner oss primærdiskurs når vi vokser opp, og i de nære relasjonene. Dermed dominerer hjemmespråket i primærdiskursen, den er en verdidiskurs. Primærdiskursen er en base hvor vi senere enten tilegner oss eller slår fra oss senere diskurser. Det er i møte med omverdenen at primærdiskursen blir utfordret. Det vil si i andre sosiale settinger utenfor hjemme, altså mer offentlig sfære, der barn møter andre sosiale språk, andre teksttyper osv. at sekundærdiskurser blir dannet. Skolediskurser er sekundærdiskurser. Noen barn innlemmes rett inn i skolediskursen, mens andre er ikke blitt forberedt på dette hjemmefra, og vil ha mye større vanskeligheter med å tilpasse seg diskursen. Dette er tett knyttet til sosioøkonomisk status.

Forskning på sosiale ulikheter viser at foreldrenes sosioøkonomiske status påvirker elevprestasjoner. Zevenbergen (2001) finner at dette sterkt påvirker om du presterer godt eller ikke i matematikk. Studien viser at barn i høyere sosiale lag virker å ha en fordel av hvordan de snakker sammen og bruker språket hjemme, når de senere skal innlemmes i matematikdiskursen. Gee (2015) finner akkurat det samme som Zevenbergen (2001). Han hevder at «kvaliteten» på barns primærdiskurs vil påvirke tilegning av sekundærdiskurs. Barns primærdiskurs vil være sterkt preget av foreldrenes diskurs. Foreldre med høyere utdanning vil ofte bringe elementer av slik diskurs inn i barnas primærdiskurs. Dette kaller Gee (2015) tidlig låning (early borrowing). Slike barns primærdiskurs vil ligge tettere opp mot skolediskursen. På denne måten er de predisponert for å tilegne seg skolens diskurs, og dette vil gi de en fordel fremfor barn som ikke har dette. Dersom en har primærdiskurs som ligger tett opptil skolediskursen, vil en kunne begynne å lære med en gang og skolen føles mindre fremmed enn for dem der den ikke samsvarer. Dette vil reprodusere forskjeller. Ut fra dette kan vi si at elevene starter på skolen med ulik førforståelse.

I boken *The Teaching Gap* studerer Stigler og Hiebert (2009) TIMSS-data og videostudier av undervisning fra ulike land. På bakgrunn av funnene presenterer de et program for hvordan endring i undervisning kan implementeres i skolen. Funnene fra studien viser at undervisningen varierte mye på tvers av kulturer, men svært lite innad i en kultur. Videre påpeker de at det å endre en undervisningskultur i et land er en svært omfattende prosess som er kulturelt betinget, og noe som hele landet samlet må enes om for at endringen skal skje. Et viktig element i en slik endring er at lærerne må være inkludert i en slik prosess. De må føle eierskap for at implementeringen skal skje og for at undervisningen endres. De peker på det at

lærere i liten grad er involvert i å utvikle undervisningsreformer som en årsak til at reformer ikke gi de ønskede forbedringene, i tillegg til at de fleste pedagogiske reformer ikke endrer dypt innebygde kulturelle modeller for undervisning (Stigler & Hiebert, 2009). Siden undervisning er kulturelt betinget, og motstandsdyktig for endringer, finner forskerne at det ofte er slik at lærere underviser slik de selv ble undervist da de gikk på skolen (Stigler & Hiebert, 2009).

3 Metode

I metodedelen vil jeg vise hvordan de kvantitative analysene er gjennomført. Først vil jeg begrunne valg av metode, deretter forklare datainnsamlingsprosessen og bearbeidingsprosessen for så å gå detaljert inn i det metodiske knyttet til OLS-regresjon og fasteffektregresjon. Til slutt vil jeg påpeke svakheter ved min metode og datasettet jeg har benyttet.

3.1 Begrunnelse for valg av metode

Ved valg av metode er det forskningsspørsmålet som er styrende. Det finnes tre ulike hovedtyper innen forskningsdesign; eksplorativt, deskriptivt og kausalt (Ghauri & Grønhaug, 2005). Siden jeg ønsker å komme så nær som mulig en kausal sammenheng, er det ut fra mitt forskningsspørsmål «*Hva har økt lærertetthet å si for elevenes prestasjoner i matematikk?*» naturlig å velge kausalt design. Dette fordi formålet med oppgaven er å kartlegge sammenhengen mellom to variabler og se om variabelen lærertetthet har påvirkning på variablene tilknyttet prestasjoner i matematikk. Fordi jeg har valgt et kausalt design, legger dette føringer for valg av metode. For å kunne nærmest mulig si noe om en kausal effekt fordrer dette at en velger en kvantitativ metode (Ghauri & Grønhaug, 2005).

Videre må en med bakgrunn i design og metode velge datakilde. Vi skiller i hovedsak mellom primær- og sekundærdata (Christoffersen & Johannessen, 2012). Primærdata er data en henter inn selv, eksempelvis via spørreundersøkelser og eksperimenter. Sekundærdata er data samlet inn eksternt og har ikke forskningsspørsmålet som opprinnelig formål for innsamlingen, eksempelvis statistiske databaser. For å kunne si noe om kausal sammenheng trengs det store mengder data og jeg har derfor valgt å bruke sekundærdata. Ulempen med dette er at dataene ikke er direkte knyttet til mitt forskningsspørsmål selv om jeg har valgt de mest relevante datakildene jeg kunne finne. Hvorvidt sekundærdataene jeg bruker er pålitelige og av god kvalitet gjennomgås i 3.7 Svakheter ved metoden og datasettet. Hvordan innhenting av data fra sekundære kilder til min oppgave foregikk, utdypes i 3.2 Datainnsamling.

For å besvare forskningsspørsmålet er det vesentlig at en strukturerer datasettet. Det finnes tre hovedgrupper for hvordan en strukturerer data: tverrsnitt, tidsserie og paneldata (Wooldridge, 2016). Variablene jeg har hentet fra den sekundære kilden er paneldata. Paneldata er variabler

hvor vi har observasjoner for hver enhet/individ målt på ulike tidspunkt over tid (Wooldridge, 2016). Dette kalles også longitudinale data.

Regresjonsanalyser er det vanligste analyseverktøyet som brukes for å se på sammenhengen, eller korrelasjonen, mellom en avhengig variabel og en eller flere uavhengige variabler. I mitt tilfelle kan dermed regresjonsanalyser brukes til å undersøke om det er en kausal sammenheng mellom lærertetthet og prestasjoner i matematikk. Det er samtidig noen viktige forutsetninger som må ligge til grunn for at regresjonsanalyser skal kunne vise kausale sammenhenger. Disse forutsetningene gjennomgås i detalj i 3.5.2 Multippel lineær regresjon.

Med bakgrunn i datamateriale jeg har tilgang til, og hva andre forskere har brukt i prosjekter med liknende problemstillinger, anser jeg OLS-regresjonsanalyse, samt fasteffektregresjonsanalyse som best egnet til å si noe om kausal sammenheng.

Grunnen til at jeg bruker både OLS og fasteffekt er at det er vanlig innenfor økonometri å sammenlikne resultatene i andre modeller mot resultatene fra OLS. Dette er fordi OLS er den mest brukte metoden for å kunne si noe om sammenhenger mellom den avhengige variabelen og en eller flere uavhengige variabler, og et standard referansepunkt da modellen er den beste lineære estimatoren (se 3.5.2 Multippel lineær regresjon for en grundig gjennomgang av antagelsene). Årsaken til at en kan sammenlikne mellom modellene er at de har samme utgangspunkt ved å minimere kvadratiske feilledd. Ved å gjennomføre analyser med begge metodene gir det også grunnlag for å kommentere på endogenitetsproblemer i modellene. Dersom det er store forskjeller mellom OLS-resultatene og fasteffekt-resultatene, kan det tyde på at en har endogenitetsproblemer i OLS-estimeringen. I en fasteffektanalyse kan en i større grad enn OLS kontrollere for variabler som er kildene til endogenitet, selv om det alltid vil være utfordrende å kontrollere for alt. Dette kommer jeg tilbake til i 3.6.1 OLS-regresjoner.

3.2 Datainnsamling

Datasettet som er grunnlaget for regresjonsanalysene er basert på sekundærdata, i hovedsak hentet fra Skoleporten og Grunnskolens Informasjonssystem (GSI). Hensikten med Skoleporten er at skoler, skoleeiere og andre interessenter skal få tilgang til pålitelige nøkkeltall for grunnopplæringen. Dette skal kunne brukes som et verktøy for vurdering av kvalitet i grunnopplæringen (Utdanningsdirektoratet, 2019). Alle tallene i Skoleporten er

basert på tallene som blir rapportert inn til GSI (Utdanningsdirektoratet, 2017a). Tallene i GSI rapporteres inn fra hver skole innen 1. oktober hvert år, og det rapporteres inn over 1000 opplysninger per grunnskole. Eksempler på data er elevtall, ressurser, spesialundervisning og målform. I mitt datasett har jeg i hovedsak hentet ut data ved hjelp av Skoleportens rapportverktøy, da dette er mer brukervennlig, men har supplert med data på lærertetthet fra GSI.

Gjennom rapporteringsverktøyet i Skoleporten har jeg hentet ut ulike variabler tilbake til 2009 for alle skoler i Norge. For hvert av årene har jeg hentet ut informasjon om antall elever og antall lærere ved skolene både samlet og skilt etter kjønn, gjennomsnittlig standpunktkarakter i matematikk, engelsk skriftlig, samt norsk hovedmål skriftlig, både samlet og skilt etter kjønn. I tillegg har jeg hentet ut den samme informasjonen knyttet til eksamen. Det vil si gjennomsnittlig eksamenskarakterer i matematikk skriftlig og muntlig, engelsk skriftlig og muntlig, samt norsk hovedmål skriftlig og muntlig, både samlet og skilt etter kjønn. Alle elever har standpunktkarakter i alle tre fagene, men en elev kommer enten opp i norsk skriftlig, engelsk skriftlig eller matematikk skriftlig til eksamen i tiende klasse. Det vil si at gjennomsnittlig eksamenskarakter baseres på det utvalget av elever som ble trukket opp til hvert av fagene. Dersom ingen av elevene, eller så få at det er fare for gjenkjennelse, kommer opp i et gitt fag vil det ikke registreres noe gjennomsnitt i dette året for denne skolen. I analysen bruker jeg bare de årene jeg har datapunkter på gjennomsnittlig karakter som observasjoner. Antall observasjoner i de ulike analysene vil derfor variere noe.

I datasettet er det også hentet ut data på indekser og indikatorer som er relevante for å til en viss grad kunne si noe om sosioøkonomisk status. Hvilke indikatorer jeg har hentet ut er beskrevet nærmere under kapittel 3.3 Variabler.

Det er ulike måter å måle lærertetthet på, noe jeg vil gjennomgå detaljert under 3.3 Variabler. Det finnes to ulike mål på lærertetthet ut fra beregning etter gruppestørrelse 1 og 2. I GSI er gruppestørrelse 2 kun oppgitt nøyaktig i undervisningsåret 2018/2019, da det er innført en ny måte å beregne dette på. Likevel er det slik at det er mulig å kunne regne seg tilbake i tid ut fra definisjonen av gruppestørrelse 2. I utregningen av gruppestørrelse 2 kreves det informasjon som delvis er prikket/sensurert. Jeg kontaktet derfor Utdanningsdirektoratet og de har gjennomført et uttrekk av gruppestørrelse 2 for meg tilbake til 2005. Gjennom kommunikasjonen fikk jeg også vite at skolene ofte rapporterte inn dataene til GSI for hele 1.-

10.trinn når det gjelder spesialundervisning, og Utdanningsdirektoratet har derfor gjort et estimat for 8.-10. trinn tilbake i tid når de har beregnet lærertetthet ut fra gruppestørrelse 2 for 8.-10. trinn.

Det er noe variasjon i datasettet for hvilke år som er med. I hovedsak er det fra 2009-2019 dataene gjelder for, med noen få unntak. Det første unntaket er for norsk, hvor det kun er data fra 2010-2019. Det gjelder både norsk standpunkt, skriftlig og muntlig norsk eksamen. 2009 var ikke tilgjengelig i Skoleporten. Det andre unntaket er elevundersøkelsen, der det kun er data for skoleårene 2015-2019. Dette er fordi undersøkelsen ble endret fra og med 2015, og den nye utformingen gjør at den ikke tester nøyaktig de samme indikatorene og indeksene før og etter 2015.

3.3 Variabler

I dette delkapitlet vil jeg forklare de sentrale variablene jeg benytter for lærertetthet, prestasjoner og resultater fra elevundersøkelsen.

Som unik identifiseringsmarkør har jeg benyttet organisasjonsnummer for hver skole for å unngå å forveksle skoler med samme navn. For hver skole har jeg lagt inn variabler for hvilken kommune og hvilket fylke skolen ligger i. I tillegg har jeg, som beskrevet under i 3.4 Bearbeiding av datasett, laget en variabel *skole* der jeg sorterer skolene inn etter hvilken skoletype de tilhører.

Variabelen jeg bruker som mål på lærertetthet er gruppestørrelse 2. Gruppestørrelse 2 er beregnet på følgende måte: *Gruppestørrelse 2* =

$$\frac{(\text{totalt antall elevtimer} - (\text{timetall spesialundervining} + \text{timetall særskilt norsk}))}{(\text{ordinære undervisningstimer} + \text{timer til oppdeling til samiske språkalternativer og annen målform})}$$

Lærertetthet i ordinær undervisning er forholdet mellom elevtimer subtrahert timer til spesialundervisning og særskilt norskopplæring, og ordinære undervisningstimer addert med oppdeling til samiske språkalternativer. Tallet vil trolig gi en mer realistisk indikasjon på antall elever per lærer i klasserommet, da ressursene til spesialundervisningen og undervisning i særskilt norsk er trukket fra. Utdanningsdirektoratet har publisert en detaljert forklaring av hvordan denne beregningen gjennomføres. Denne ligger som vedlegg til oppgaven.

Variabelen *antall elever 10. trinn* indikerer hvor mange elever som etter Opplæringslova §2-1 har rett til grunnskoleopplæring på dette trinnet og benytter denne retten ved en grunnskole (Opplæringslova, 1998). Tallet registreres 1. oktober hvert skoleår. *Antall gutter 10. trinn* og *antall jenter 10. trinn* skiller den overnevnte variabelen etter kjønn, og utgjør to nye variabler.

Antall lærere opplyser hvor mange lærere som underviser på skolen. Lærer er ifølge Skoleportens indikatorveiledning (Utdanningsdirektoratet, 2019) definert som undervisningspersonell registrert i GSI. Utdanningsdirektoratet presiserer at assistenter ikke er inkludert i dette tallet. Variabelen *antall mannlige lærere* og variabelen *antall kvinnelige lærere* skiller den overnevnte variabelen etter kjønn, og utgjør to nye variabler.

Variablene knyttet til gjennomsnittlig standpunktkarakterer og eksamenskarakterer i de ulike fagene gis etter skalaen 1-6, der 6 er beste karakter, og er et gjennomsnittlig mål på elevenes sluttkompetanse. Det er én variabel for gjennomsnittlig standpunktkarakter i hver av de tre fagene norsk hovedmål skriftlig, engelsk skriftlig og matematikk. For eksamen har jeg variabler for gjennomsnittlig eksamenskarakterer i matematikk skriftlig og muntlig, engelsk skriftlig og muntlig, samt norsk hovedmål skriftlig og muntlig. Alle disse variablene er hentet ut både samlet for alle elevene og skilt etter kjønn.

Årlig gjennomføres det en undersøkelse i skolen der elevene får rapportere inn sin mening om læring og trivsel i skolen (Utdanningsdirektoratet, 2018a). Dette kalles elevundersøkelsen. Undersøkelsen er frivillig for elevene å delta i. Informasjonen fra undersøkelsen brukes av skoler, kommuner og stat til å forbedre skolen. Hver høst er det en del obligatoriske temaer som går igjen, to av disse er *støtte fra lærer* og *støtte hjemmefra*. Dette er de to indeksene jeg har hentet ut i datasettet. Datasettet består både av resultater fra hver av indikatorene og indekser der alle resultatene fra indikatorene under hver indeks er snittet. Innenfor indeksen *støtte fra lærer* er indikatorene 1) Opplever du at lærerne dine bryr seg om deg, 2) Opplever du at lærerne dine har tro på at du kan gjøre det bra på skolen, 3) Opplever du at lærerne behandler deg med respekt, 4) Når jeg har problemer med å forstå arbeidsoppgaver på skolen, får jeg god hjelp av lærerne, og 5) Lærerne hjelper meg slik at jeg forstår det jeg skal lære. Innenfor indeksen *støtte hjemmefra* er indikatorene, 1) Hjemme viser de interesse for det jeg gjør på skolen, 2) Jeg får god hjelp til leksene mine hjemme, 3) Hjemme oppmuntrer de voksne meg i skolearbeidet. Dette er hentet ut i datasettet også for hvert kjønn. Skalaen som

er brukt til å presentere resultatene er en likert-skala fra 1-5 der 5 er det mest positive svaret. Svaralternativene ligger som vedlegg til oppgaven.

3.4 Bearbeiding av datasett

Data fra kildene beskrevet under kapittel 3.2 ble importert og sammenstilt i Excel. Siden flere skoler har samme navn, er organisasjonsnummeret brukt som unik identifiseringsmarkør til hver skole. Dette gjør det mulig å kombinere de ulike datasettene, og unngå at det oppstår feil når for eksempel flere skoler har samme navn. Datasettet inneholder data fra alle offentlige skoler i Norge. Private skoler er dermed ikke inkludert i datasettet. Årsaken til at jeg har valgt å utelate disse er fordi private skoler ofte har et annet budsjett, ulik elevsammensetning, ikke samme rapporteringsansvar i GSI og trolig flere ressurser per elev enn offentlige skoler. Dette kunne derfor påvirket analysene. Totalt er det 3258 offentlige skoler i datasettet.

Alle offentlige skoler i Norge inngår i datasettet, men jeg er kun interessert i at datasettet skal inneholde skoler som har elever med avgangskarakterer fra 10. trinn. Derfor var det i utviklingen av datasettet nødvendig å kategorisere skolene i ulike typer.

I alt definerte jeg 10 ulike skoletyper, som vist i Tabell 2. Fordi det ikke allerede eksisterer en slik klassifisering måtte jeg for hver skole gå manuelt inn i Skoleporten og undersøke hva slags skoletype den enkelte skolen skulle klassifiseres som. For analysene er det kategoriene 1, 2, 4, 7, 8, 9 og 10 som er aktuelle. Dette er alle skoler som har elever med avgangskarakterer fra 10. trinn, og utgjør totalt 1019 skoler. Kode 3 inneholder alle 1.-7. skoler, og varianter av ulike trinnkombinasjoner innenfor intervallet 1.-7. eks. skoler av typen 1.-4. trinn. Kode 5 er alle utgåtte skoler. Kode 6 er skoler som ikke hører under koden til de andre skoletypene og utgjør i hovedsak spesialskoler.

Koder	Skoletype
1	8.-10. trinn
2	1.-10. trinn
3	1.-7. trinn
4	5.-10 trinn
5	Utgått
6	Annet, eksempelvis spesialskole som ikke har data
7	1.-13. trinn
8	6.-10. trinn
9	7.-10. trinn
10	8.-13. trinn

Tabell 2: Klassifisering av skoletyper

I arbeidet med klassifiseringen oppdaget jeg at to av skolene som står klassifisert i Skoleporten som skoletype 2 er registrert feil. Dette var mulig å oppdage ved å søke opp duplikater i organisasjonsnumrene, hvor jeg fant at Huseby barneskole i Trøndelag, og Gjerpen barneskole i Telemark var registrert to ganger, både som 1.-10. trinn skole og 8.-10. trinn. Disse ble endret til skoletype 3, da det ikke er registrert karakterer på skolene.

I etterkant kontrollerte jeg at kategoriseringen var gjort korrekt, ved å teste at alle skoler i kategori 1, 2, 4, 7, 8, 9 og 10 hadde registrerte avgangskarakterer. Kontrollen viste at kategoriseringen stemte.

Koder	Skoletype	Antall
1	8.-10. trinn	456
2	1.-10. trinn	534
4	5.-10 trinn	16
7	1.-13. trinn	4
8	6.-10. trinn	3
9	7.-10. trinn	4
10	8.-13. trinn	2
Samlekategori for 3, 5 og 6	Varianter av 1.-7. skoler, utgåtte skoler, kategorien annet	2239
	Sum alle skoler	3258
	Sum relevante skoler	1019
	Skoler utelatt på grunn av krav til datasett	188
	Endelig antall skoler brukt i analysen	831

Tabell 3: Fordeling av skoletyper og endelig antall skoler i analyse

Fra Tabell 3 ser vi at det totalt er 1019 skoler i Norge som har avgangselever fra 10. trinn i perioden 2009-2019. Av dette er det 188 som ikke tilfredsstillt kravene jeg videre definerte for utvalget i datasettet, og er dermed utelatt. Kravene til datasettet er at det er registrert avgangskarakterer i matematikk, engelsk og norsk både ved eksamen og i standpunkt. I tillegg er det stor usikkerhet i gruppestørrelser knyttet til skoler med svært få elever, så skoler med i snitt mindre enn 10 elever totalt på 10. trinn over hele analyseperioden ble fjernet.

Hovedgrunnen til å fjerne disse er at ved skoler som er veldig små kan små endringer i gruppestørrelse gi store utslag i prosentvis endring i gruppestørrelsesvariabelen. Det kan også bli mye støy i gjennomsnittskarakterene når det legges sammen for få individer. I tillegg er det mange grunner til at de minste skolene kan være annerledes, det er mer sannsynlig at elevene har undervisning på tvers av klassetrinn, og det er mer sannsynlig at de ligger i rurale områder. På grunn av dette valgte jeg å se på en mer homogen gruppe av skoler slik at jeg sammenlikner relativt like typer skoler. Siden jeg kun utelater veldig små skoler vil utvalget mitt dekke de aller fleste elevene i Norge. I tillegg satte jeg inn en forutsetning om at gruppestørrelse måtte være mellom 10 og 40 elever. Det for å fjerne eventuelle feilrapporteringer av gruppestørrelse 2 fordi noen verdier er veldig små eller veldig store og derfor utenfor hva en kan forvente av normale gruppestørrelser. I den endelige analysen er det 831 skoler som tilfredsstillt disse kravene.

Jeg overførte datasettet fra Excel til statistikkprogrammet Stata versjon 15 for å utføre analysene. Det første jeg gjorde var å sette opp datasettet i det en kaller long-format. Det vil si at jeg har én observasjon for hver skole og for hvert år under hverandre, med tilhørende avgangskarakterer og andre variabler for hver skole i hvert år. Så lagde jeg variabler for gjennomsnittlig gruppestørrelse over alle skolene for hvert år og for hvert fylke for hvert år. I starten av resultatdelen i 4.1 deskriptive data, har jeg figurer som beskriver dette.

Videre standardiserte jeg utfallsvariablene mine. Det vil si at de har gjennomsnitt 0 og standardavvik 1. Da vil tolkningen og sammenlikningen med ulike utfall være enklere. En enhets endring i gruppestørrelse 2 kan da tolkes som % endring i standardavviket til utfallsvariabelen. En mer formell definisjon følger i 3.5.1.

I 3.6 Regresjonsmodeller brukt i analysen, kommer jeg tilbake til hvilke kontrollvariabler jeg har brukt i mine analyser, men allerede i bearbeiding av dataene omdefinerte jeg en av kontrollvariablene. Årsaken til at jeg bearbeidet variabelen, var for å få mer nøyaktige kontrollvariabler i analysene. For å kunne bruke kjønn på lærer som kontrollvariabel lagde jeg variabelen *andel kvinnelige lærere* ved å dele antall kvinnelige lærere på totalt antall lærere.

Fra elevundersøkelsen hentet jeg ut indeksen *støtte hjemmefra*. I utgangspunktet var tanken å bruke dette som kontrollvariabel, men grunnet at det kun er data for 4 år fra elevundersøkelsen, mister jeg for mange observasjoner hvis jeg bruker denne som kontrollvariabel. Derfor utviklet jeg heller denne variabelen til å fungere som proxy for *støtte hjemmefra* som brukes i en heterogenitetsanalyse. For å se om det er ulik effekt av lærertetthet mellom de som får mye støtte hjemmefra og de som får lite, valgte jeg å dele inn disse i to grupper etter medianen. Jeg lagde en ny variabel som er 1 hvis en skole er over medianen på scoren for *støtte hjemmefra* og 0 hvis en skole er under medianen på scoren for *støtte hjemmefra*, altså en dummy-variabel.

3.5 Metode for regresjonsanalyse

For å besvare forskningsspørsmålet vil jeg som nevnt i 3.1 Begrunnelse for valg av metode bruke to ulike analysestrategier, multippel lineær regresjon og fasteffektregresjon. I dette kapitlet vil jeg gjennomgå disse metodene i tillegg til noen overordnede begreper knyttet til regresjonsanalyser.

3.5.1 Generelt om regresjonsanalyse og sentrale begreper

Regresjonsanalyse er en statistisk analyseteknikk hvor en undersøker hvordan gjennomsnittsverdien på en avhengig variabel Y varierer med en eller flere uavhengige variabler (Johannessen, Christoffersen & Tuft, 2011). En bivariat lineær regresjonslinje kan uttrykkes ved $Y = \beta_0 + \beta_1 * X_1$ (Christophersen, 2013). Uttrykket for regresjonslinjen er en modell som illustrerer tendensen i forholdet mellom avhengig variabel Y og uavhengig variabel X_1 . Det vil si at avhengig variabel er en funksjon av uavhengig variabel, hvor β_1 er regresjonskoeffisienten og β_0 er der regresjonslinjen krysser y-aksen. β_0 er altså en konstant (Johannessen et al., 2011).

X kan være både en kontinuerlig variabel (for eksempel lærertetthet, gjennomsnittlig antall elever per lærer i ordinær undervisning) eller en dummyvariabel (for eksempel kjønn). For en kontinuerlig variabel tolkes koeffisienten som forskjellen i Y når X øker med én enhet. Dersom Y er elevenes gjennomsnittskaracter, vil koeffisienten vise endringen i gjennomsnittskaracter når antall elever per lærer øker med én enhet. En dummyvariabel tar enten verdien 1 eller 0. Eksempelvis kan den settes til 1 hvis jente og 0 hvis gutt. Koeffisienten foran X kan da tolkes som forskjellen i gjennomsnittskaracter (Y) mellom jenter og gutter.

Videre vil jeg definere noen begreper. Parameter er en deskriptiv størrelse som beskriver en populasjon eller en sannsynlighetsfordeling (Wooldridge, 2016). I oppgaven vil β_1 være en viktig parameter, da denne beskriver hva som skjer med prestasjonene til elevene dersom gruppestørrelse 2 (X_1) øker med én enhet. Standardavvik er også en viktig parameter i oppgaven. Denne beskriver spredningen av verdiene i datasettet (Christophersen, 2013). Effektstørrelse beskriver grad av sammenheng mellom X og Y (Wooldridge, 2016). Tolkningen av effektstørrelse avhenger av hvilken funksjonsform variablene har. Det vil si om de er kontinuerlige variabler med mange verdier, eller dummyvariabler med kun verdier 1 eller 0. Det vil også ha betydning om kontinuerlige variabler er lineære eller på logaritmisk form.

For å gjøre det enklere å tolke effektstørrelsen har jeg standardisert utfallene i mitt datasett. Dersom variablene viser til den skalaen de er målt på, er koeffisientene ustandardiserte. Når variablene transformeres, kalles koeffisientene standardiserte (Johannessen et al., 2011). I mitt datasett har jeg standardisert karakterene til å ha gjennomsnitt 0 og standardavvik 1. Dette

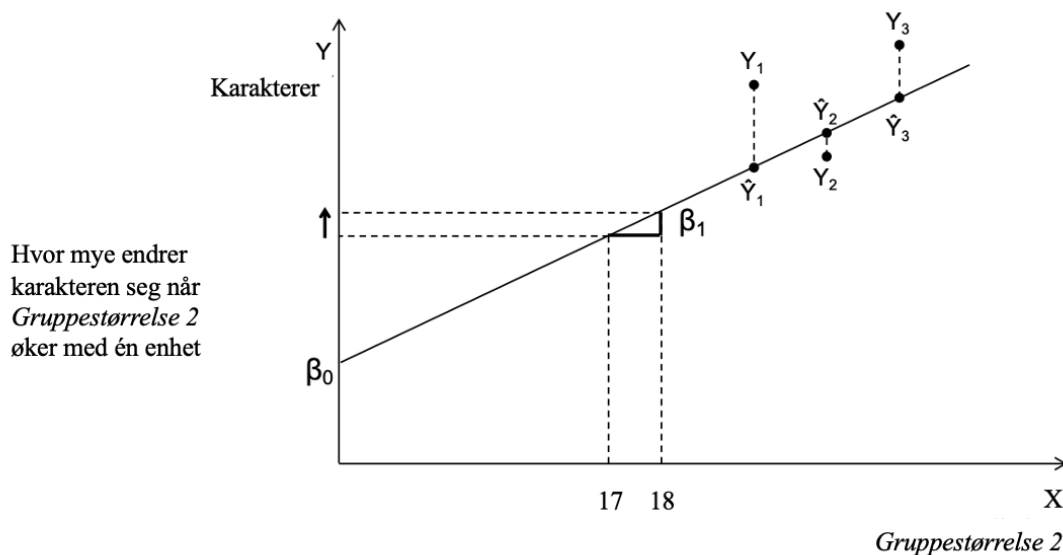
gjør det lettere å tolke og sammenlikne. Standardavviket gir oss en beregning på variasjonen rundt gjennomsnittet. Det uttrykker spredningen langs variabelens skala (Christophersen, 2013). Ved å standardisere karakterene til å ha gjennomsnitt 0 og standardavvik 1, vil β_i kunne tolkes som prosent av et standardavvik. For eksempel vil $\beta_1 = 0,17$ bety at en økning med 1 i X-verdi henger sammen med en økning på 0,17 i Y-verdi. En annen måte å si dette på er at sammenhengen mellom Y (standardisert karakter) og X (uavhengig variabel) er 17 % av et standardavvik. Elevenes resultater måles dermed som avvik fra gjennomsnittet.

Populasjonen i denne oppgaven utgjør alle elever med avgangskarakterer på 10. trinn i Norge mellom 2009 og 2019. Som nevnt har jeg ikke tatt med privatskoler og heller ikke skoler med gjennomsnittlig mindre enn 10 elever per trinn over hele analyseperioden, eller med gruppestørrelser per trinn på mindre enn 10 eller større enn 40. Det gjør at jeg har gjort et utvalg av populasjonen. For å tolke resultatene jeg får fra regresjonsanalysen er det da sentralt å finne ut om resultatet er statistisk signifikant (Johannessen et al., 2011). For å kunne tolke dataene riktig må en bestemme et signifikansnivå. Dette angir den maksimale sannsynligheten som aksepteres for at et resultat kan være tilfeldig (Christophersen, 2013). I min oppgave bruker jeg 0,01, 0,05 og 0,1 som signifikansnivåer. I tabeller vil dette vises ved at et signifikansnivå på under 1 % angis med tre stjerner (***). Det vil si at det er over 99 % sannsynlighet for at variabelen har reell innvirkning på utfallet. Et signifikansnivå på under 5 % vil angis ved to stjerner, og 10 % signifikans med én stjerne. I min oppgave vil jeg angi i tabeller funn også på 10 % signifikansnivå, men jeg vil ikke vektlegge funnene i analysen, da det er knyttet for stor usikkerhet til dem. Årsaken til at jeg definerer det og markerer det, er at en del av de andre studiene jeg ser på forholder seg til dette nivået i analysen.

3.5.2 Multippel lineær regresjon

I analysene vil jeg se på sammenhengen mellom to eller flere variabler og derfor vil jeg bruke multippel lineær regresjon. Da vil en eksplisitt kunne kontrollere for mange faktorer samtidig, og studere hvordan de påvirker den avhengige variabelen. Jeg vil først beskrive metoden og antagelsene, og ut fra dette vurdere hvordan antagelsene fungerer for mitt forskningsspørsmål og eventuelle svakheter ved datasettet. Dette er et sentralt tema i all statistisk empirisk forskning, da validiteten til resultatene avhenger av at disse antakelsene i så stor grad som mulig er oppfylt. Jeg har derfor valgt å utdype metoden tilstrekkelig til at lesere som ikke har inngående kjennskap til denne kan få forståelse av grunnlaget for vurderingene.

En av de vanligste metodene innenfor multippel regresjon, og som jeg kommer til å anvende, er minste kvadraters metode, også kalt OLS-regresjon. Metoden er illustrert i Figur 1. En OLS er en estimeringsmodell som sikrer at summen av de vertikale avstandene mellom datapunktene og regresjonslinjen er minst mulig (Thrane, 2017). Metoden går ut på å velge linjen som gjør summen av kvadrerte avvik mellom observert verdi Y og forventet verdi \hat{Y} minst mulig (Christophersen, 2013). Avstandene kvadreres altså før de summeres. Koeffisientene gir da den regresjonslinjen som er best tilpasset observasjonene fra datasettet. Alle estimerer er estimert med usikkerhet, dette fanges opp i feilleddet. Jo mer avstand mellom Y og \hat{Y} , jo mer usikkerhet blir det i estimatet. Figur 1 er en illustrasjon av minste kvadraters metode og bruker gruppestørrelse 2 som mål på lærertetthet som X variabel og avgangskarakterer som Y variabel.



Figur 1: Illustrasjon av minste kvadraters metode

Minste kvadraters metode bygger på fem antakelser, som kalles Gauss-Markov antakelsene (Wooldridge, 2016). Dersom disse fem antakelsene er oppfylt, vil OLS være den beste lineære forventningsrette estimatoren.

En OLS kan skrives på formen $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + u$, der $\beta_0, \beta_1, \dots, \beta_k$ er koeffisienter og u er feilleddet (Wooldridge, 2016). Den første antakelsen sier at modellen må

være lineær i koeffisientene, men ikke i variablene. I tillegg er det en del av antakelsen at regresjonsmodellen må være korrekt angitt. Det vil si at alle variablene må være med i modellen, eksempelvis at man ikke systematisk utelukker noen variabler fra modellen.

Den andre antakelsen er at utvalget må være tilfeldig. Vi må ha et tilfeldig utvalg med n observasjoner $\{(x_{i1}, x_{i2}, \dots, x_{ik}, y_i) : i = 1, 2, \dots, n\}$ (Wooldridge, 2016).

Antakelse tre handler om at det er ingen perfekt kollinearitet. Det vil si at ingen uavhengig variabel må være perfekt korrelert med andre uavhengige variabler, men det betyr ikke at det ikke kan være noe korrelasjon, det må bare ikke være perfekt korrelert.

Den fjerde antakelsen handler om eksogenitet. Antakelsen er at feilledet u har en forventet verdi på null uansett verdien av de uavhengige variablene $E(u|x_1, x_2, \dots, x_k) = 0$. Det vil si at antakelsen svikter når minst én uavhengig variabel er korrelert med noe i feilledet (Wooldridge, 2016). Disse variablene er da endogene, og når det er tilfelle får vi det som kalles forventningsskjevne estimater. Det er tre potensielle kilder til slik endogenitet. Den første er utelatte variabler som da inngår som en del av feilledet, som egentlig er relevante i regresjonen. En har ekskludert en relevant variabel. Dette kan igjen føre til spuriøse eller undertrykte sammenhenger. En spuriøs sammenheng vil si at det er en korrelasjon som ser ut som et årsaks-virkningsforhold, men som ikke er det (Wooldridge, 2016). En annen kilde til endogenitet er omvendt kausalitet. Forventningsskjevhet skjer når den avhengige variabelen i seg selv påvirker en eller flere av de uavhengige variablene. Omvendt kausalitet vil si at vi ikke bare fanger opp effekten av den uavhengige variabelen, men estimatet får tilbakeslagseffekt fra den avhengige variabelen. Målefeil er den siste formen for endogenitet. Litteraturen skiller mellom systematiske feil og tilfeldige feil, dette er nært knyttet til begrepene validitet og reliabilitet. God forskningsskikk er avgjørende, siden eksogenitet vanskelig kan testes fordi utelatte variabler og målefeil som regel ikke kan observeres (Wooldridge, 2016). Det er viktig å være nøye på hvilke variabler en tester for og hvilke en utelater, og vurdere om det kan være endogene variabler i regresjonsmodellen.

Den femte antagelsen sier at variansen til feilledet u skal være konstant.

$Var(u|x_1, x_2, \dots, x_k) = \sigma^2$. Dette kalles homoskedastisitet (Wooldridge, 2016). Antagelsen sier at variansen er konstant for alle verdier av de uavhengige variablene. Denne antagelsen er

det enkelt å løse opp i, da statistikkprogrammer har en enkel måte å kontrollere for heteroskedastisitet i feilledet.

I kapittel 3.6 kommer jeg tilbake til om disse antagelsene er oppfylt for min analyse.

3.5.3 Fasteffekt

Den andre analysemetoden jeg vil bruke er fasteffekt. Fasteffektmodellen gjør ifølge Allison (2009) det mulig å kontrollere for variabler som ikke kan eller ikke har blitt målt. I en årrekke har kritikken mot kvantitativ forskning vært at en kan utelate variabler, noe som igjen kan føre til spuriøse eller undertrykte sammenhenger. Fordelen med fasteffekt er at denne baserer estimeringen kun på endringer innad i hver enhet, ved å holde noe fast. Enkelt sagt er prinsippet bak modellen at en bruker hver enhet som ens egen kontrollgruppe (Allison, 2009). I fasteffektanalysen bruker jeg dermed hver skole som sin egen kontrollgruppe.

Fasteffektmodellen stiller to krav til dataene. Den avhengige variabelen må måles for hver enhet minst to ganger ved forskjellige anledninger. Det er vesentlig at disse målingene kan sammenliknes direkte (Allison, 2009). Det vil si at dataene kan tolkes på samme måte, og er målt etter samme kriterier. I datasettet denne oppgaven bygger på har vi paneldata, som oppfyller dette kravet. Det andre kravet er at den predikerende variabelen, altså X , må endre verdi ved flere anledninger i utvalget.

Ifølge Woolridge er styrken til fasteffektmodellen at den åpner for å kontrollere for uobserverte kilder til heterogenitet som ikke forandrer seg over tid, men som kan påvirke eller forskyve predikatoren eller utfallsvariabelen (Wooldridge, 2016). Fasteffekt undersøker forholdet mellom predikatoren og utfallsvariablene innenfor en enhet (i min analyse vil enheten være skole). Hver enhet har individuelle egenskaper som kan påvirke predikatoren, men som ikke endrer seg over tid, eller endrer seg veldig sakte. Eksempel på dette kan være at skolen ligger i samme kommune hvert år eller fysisk infrastruktur på skolebygget som ikke endrer seg i perioden vi studerer. Fasteffektmodellen gjør det mulig å isolere slike uobserverte kilder, slik at en kun ser på det som forandrer seg.

For å se hvordan en fasteffektmodell fungerer kan vi ta utgangspunkt i en generell modell med én uavhengig variabel, hvor $Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$, $t = 1, 2, \dots, T$. Her er Y_{it} den

avhengige variabelen, der t er tid og i er enhet, X_{it} representerer en uavhengig variabel med koeffisient β_1 , og u_{it} er feilledet. α_i er den uobserverbare effekten for hver enhet som vi ønsker å eliminere (Wooldridge, 2016). Dersom vi nå for hver enhet i ser på gjennomsnittet av likningen over tid, får vi $\bar{Y}_i = \beta_1 \bar{X}_i + \alpha_i + \bar{u}_i$, hvor $\bar{Y}_i = T^{-1} \sum_{t=1}^T Y_{it}$. Siden α_i er fast over tid, vil denne være lik i begge likningene. Hvis vi subtraherer de to likningene fra hverandre for hver t får vi: $Y_{it} - \bar{Y}_i = \beta_1 (X_{it} - \bar{X}_i) + u_{it} - \bar{u}_i$, for $t = 1, 2, \dots, T$. Her ser vi at den uobserverbare effekten α_i har forsvunnet. Det vil si at resterende variasjon kun kan forklares med variasjon innad på en skole og ikke mellom skoler. «Utregning av estimatene baserer seg på gjennomsnittssentrering, der en fjerner forskjellen mellom enheter, og kun ser på variasjonen i X og Y rundt deres gjennomsnitt på enhetsnivå» (Finckel, 2007, s. 480).

Siden en har paneldata i en fasteffektanalyse, er det vesentlig å gjøre standardavvikene robuste. Det betyr at standardavvikene tar hensyn til seriekorrelasjon mellom skoler over tid. Hvis det er noe uobserverbart med skolene som er systematisk over tid, vil dette føre til seriekorrelasjon i feilledene. For å unngå dette må en klustre regresjonene på skolenivå i fasteffektregresjonen. Det vil si at man grupperer feilledene fordi det er seriekorrelasjon. OLS standardavvik er forventningsrett når residualene er uavhengige og likt distribuert. Men når residualene er korrelert på tvers av observasjoner, kan OLS standardavvik være forventningsskjeve og enten over- eller underestimere variabiliteten til koeffisientestimatene (Wooldridge, 2016). I mine fasteffektregresjoner har jeg derfor klustret på skolenivå for å kontrollere for denne seriekorrelasjonen.

3.6 Regresjonsmodeller brukt i analysen

I dette delkapitlet vil jeg beskrive hvordan jeg har definert regresjonsmodellene brukt i analysen for hver av analysemetodene. I 3.6.1 vil jeg også gjennomgå i hvilken grad OLS-regresjonene mine holder opp mot Gauss-Markov-antakelsene for minste kvadraters metode.

3.6.1 OLS-regresjoner

I den første regresjonsmodellen gjør jeg en enkel OLS-regresjon der jeg ønsker å se på hvordan gjennomsnittlig lærertetthet (ut fra gruppestørrelse 2) påvirker variabelen standpunkt i matematikk. I regresjonen ser jeg dermed på sammenhengen mellom Y på venstre siden, der Y er standpunkt karakter i matematikk, og X på høyresiden, der X er gruppestørrelse 2. Tolkningen av β_1 vil da gi forskjellen i Y -variabelen når gruppestørrelse 2 øker med én

enhet. For eksempel hva skjer med Y når gruppestørrelse 2 øker fra 15 til 16 elever per lærer. Formelen kan skrives som

$$Y_{it} = \beta_0 + \beta_1 * \text{gruppestørrelse } 2_{it} + \beta_2 * \text{år}_t + \varepsilon_{it},$$

der $t = 2009, \dots, 2019$ og $i = \text{skoleid}$

I modellen er ikke β_2 en konstant koeffisient for alle år, men skal tolkes som en vektor av koeffisienter med en verdi for hvert av årene t . Dette er skrevet slik for å forenkle beskrivelsen av modellen.

I likningen har jeg også lagt inn kontroll for år. Denne kontrollen fanger opp trender i karakterene over tid. Eksempelvis kan det være ulikt karaktersnitt på en skole fra ett år til et annet, som kan skyldes at eksamen er ulik fra et år til et annet.

Denne regresjonen gjennomfører jeg også når avhengig variabel Y er norsk standpunkt, engelsk standpunkt, matematikkeksamen skriftlig og muntlig, norskeksamen skriftlig og muntlig, og engelskeksamen skriftlig og muntlig.

I den andre regresjonsmodellen gjennomfører jeg også OLS-regresjoner, men her legger jeg inn flere kontrollvariabler i tillegg til år. Jeg kontrollerer for type skole, antall lærere på skolen, andel kvinner på skolen og hvilken kommune skolen er i. Årsaken til at jeg har valgt akkurat disse kontrollene, er at de kontrollerer godt for en del karakteristikk ved skolene og kommunene, samt at det er disse jeg har tilgang til i datasettet. Eksempelvis ved å kontrollere for kommuneeffekter i OLS-regresjonen utnyttes kun variasjonen innad i kommunene, slik at jeg ser på effekten av skolevariabler uavhengig av kjennetegn ved kommunene. De overnevnte kontrollene på skolenivå er de samme variablene som flere sammenliknbare studier benytter seg av (Bonesrønning & Vaag Iversen, 2008; Falch et al., 2017; Grøgaard et al., 2008; Leuven et al., 2008; Opheim et al., 2010). Studiene til Opheim et al. (2010) og Grøgaard et al. (2008) kontrollerer for andel elever med spesialundervisning, men for mine analyser er dette ikke nødvendig da dette er en del av beregningen av gruppestørrelse 2. I tillegg har alle disse studiene også inkludert variabler for sosioøkonomiske forhold på individnivå, eksempelvis foreldres inntekt, foreldres utdanningsnivå, arbeidsmarkedsstatus, familiestruktur og etnisk bakgrunn. Dette skulle jeg gjerne også hatt data på, men det finnes

dessverre ingen gode tilgjengelige datakilder som kunne kombineres med mitt datasett for dette.

Regresjonsmodellen kan dermed beskrives på følgende måte:

$$Y_{it} = \beta_0 + \beta_1 * \text{gruppestørrelse}_{it} + \beta_2 * \text{år}_t + \beta_3 * \text{skoletype}_g + \beta_4 * \text{antall lærere}_{it} + \beta_5 * \text{kvinneandel}_{it} + \beta_6 * \text{kommune}_i + \varepsilon_{it},$$

der $t = 2009, \dots, 2019$ og $i = \text{skoleid}$ og $g = \text{skoletype}$ (1, 2, 4, 7, 8, 9, 10)

Som i modellen over er ikke koeffisientene β_2 , β_3 , og β_6 konstante koeffisienter, men skal tolkes som en vektor av koeffisienter med en verdi for hvert av årene t , hver skoletype g og hver kommune. Dette er skrevet slik for å forenkle beskrivelsen av modellen.

Det er her samme tolkning av β_1 som over, men tolkningen er nå betinget på flere kontroller. Tolkningen er fortsatt endring i Y når gruppestørrelse 2 endres med én enhet. De øvrige koeffisientene gir sammenhengen mellom de andre uavhengige variablene og Y . Eksempelvis hvor mye Y endres når antall lærere på skolen øker med én enhet. I tillegg til å benytte karakterer som avhengig variabel, har jeg i den andre modellen også benyttet score på indeksen *støtte fra lærer* i elevundersøkelsen som Y i en egen analyse.

For OLS-modellene over skal jeg nå se nærmere på antagelsene for OLS og vurdere om disse holder i min studie. Antagelse 1 var at variablene må være lineære i parameterne. Det vil si at det er en lineær sammenheng mellom gruppestørrelse 2 og elevers prestasjoner. Dette kan jo hende ikke er tilfelle. For eksempel kan det være viktigere å gå fra 20 til 15 elever enn å gå fra 15 til 10 elever. Men det er utenfor rammen av denne oppgaven å gå i dybden på ikke-lineære effekter da dette gjør regresjonene mer kompliserte (Wooldridge, 2016). Det at jeg har utelatt de aller minste og største gruppene gjør at det er mer sannsynlig at sammenhengen er tilnærmet lineær siden vi ikke får voldsomme utslag i veldig store eller veldig små grupper.

For antagelse 2, at utvalget er tilfeldig, har jeg med alle skoler som tilfredsstillt kravene til å være med i analysen. For antagelse 3, at ingen uavhengige variabler kan være perfekt korrelerte, er dette oppfylt. Gruppestørrelse 2 er ikke perfekt korrelert med år, antall lærere,

andel kvinnelige lærere, eller kommunene. Jeg velger å vurdere antagelse 5 om homoskedastiske feilledd først, fordi antagelse 4 krever mer drøfting. Denne antagelsen er enkel å håndtere fordi Stata har en innebygd kommando som løser opp i dette og tar hensyn til at feilleddet også kan være heteroskedastisk.

Antagelse 4, at feilleddet u har en forventet verdi på null uansett verdien av de uavhengige variablene er den som er mest utfordrende med OLS-analysen. Å kontrollere for karakteristikk ved kommune, skole og år er en teknikk som gjør at en kan kontrollere for forstyrrende variabler som kan skape spuriøse sammenhenger (Wooldridge, 2016). Dette vil kunne hjelpe på antagelsen, men det er stor sannsynlighet for at det er andre uobserverbare variabler som både kan påvirke gruppestørrelse og elevers prestasjoner. I mitt eksempel vil det for eksempel være et problem at jeg ikke har en variabel for sosioøkonomisk bakgrunn. Sosioøkonomisk bakgrunn kan både påvirke gruppestørrelser og prestasjoner, for eksempel fordi foreldre bosetter seg i nærheten av skoler med lavere gruppestørrelse eller skolene selekterer elevene i grupper etter sosioøkonomisk bakgrunn. Siden sosioøkonomisk bakgrunn også kan ha en direkte sammenheng med elevers prestasjoner, kan sammenhengen mellom gruppestørrelse 2 og elevers prestasjoner være spuriøs hvis vi ikke kontrollerer for sosioøkonomisk bakgrunn. Selv om vi hadde hatt en god variabel for sosioøkonomisk bakgrunn kan det og være andre uobserverbare variabler som for eksempel elevers evner, motivasjon eller ulike egenskaper ved lærere på ulike skoler som kan ha en sammenheng med både gruppestørrelse og elevers prestasjoner.

Som følge av dette blir en svakhet ved OLS-estimatene at det kan oppstå endogenitetsproblemer. Det vil si at den uavhengige variabelen gruppestørrelse 2 ikke er helt uavhengig av en av de andre variablene i modellen. Som jeg har vært inne på over har jeg forsøkt å kontrollere etter beste evne, men det er ikke sikkert dette er tilstrekkelig for å sikre eksogenitet. Dette fører til at det kan være vanskelig å avdekke den kausale sammenhengen mellom den uavhengige variabelen gruppestørrelse 2 og utfallene, Y .

Så selv om jeg kontrollerer for noen karakteristikk ved skole og kommune har jeg likevel ikke data på individnivå, og ikke all data for hva som kan være ulikt på skoler. For å kunne si noe om kausal effekt må jeg derfor kunne kontrollere bedre for at skoler er ulike. Dette motiverer valget av å bruke fasteffektanalyse i stedet for OLS, som bedre kontrollerer for at skoler med høy og lav gruppestørrelse kan være ulike, og i høyere grad sikrer eksogenitet.

Dette er en modell der jeg ser på hver skole isolert og ser på hvordan gruppestørrelse 2 varierer innad på hver skole i årene 2009-2019, mot variasjonen i karakterene til elevene innad i den samme skolen for de samme årene. Det skal mye til at det er stor variasjon i andre forhold innad på en skole over tid og en får dermed bort mye av skjevheten ved å gjøre dette. For eksempel vil elevsammensetningen på en skole være ganske lik over ulike år, i alle fall over en relativt kort periode.

3.6.2 Fasteffektmodeller

Fasteffektmodellen jeg har benyttet kan skrives som

$$Y_{it} = \beta_1 * \text{gruppestørrelse } 2_{it} + \alpha_i + \lambda_t + u_{it},$$

der $t = 2009, \dots, 2019$ og $i = \text{skoleid}$

α viser den faste effekten for hver skole, mens λ viser den faste effekten for hvert år.

Fasteffektmodellen ser dermed på variasjonen innad i en skole over tid. I min modell ønsker jeg å se på endring i lærertetthet over tid mot endringen i karakterer over tid. Årsfasteeffekter fanger eksempelvis opp ting som er uobserverbare som er likt mellom skoler i ulike år.

Eksempelvis kan det fange opp om eksamen er vanskeligere fra ett år til et annet. Det vil si at jeg tar ikke bort variasjonen, men jeg tar bort det som er likt mellom skolene i det samme året.

Fasteffektmodellen er bygd opp relativt likt som den andre OLS-modellen. Y_{it} er karakterer eller indeksen *støtte fra lærer*, og β_1 gir effekten av gruppestørrelse på utfallene. Her har jeg kontrollert for de faste effektene ved skolen, år, skoletype, antall lærere på skolen og andel kvinnelige lærere.

Tolkningen av fasteffektparameteren β_1 er lik tolkningen til minste kvadrats metode. Hva skjer med karakteren i skriftlig matematikk når gruppestørrelse 2 øker med en enhet. Siden vi nå har kontrollert bedre for at skoler er forskjellige er vi nærmere en kausal tolkning, at én enhet økning i gruppestørrelse 2 forårsaker den eventuelle endringen i karakteren i matematikk. Selv om vi er nærmere en kausal sammenheng vil jeg komme tilbake til noen svakheter ved metoden og alternative forklaringer til at effekten er kausal i 3.7.

3.7 Svakheter ved metoden og datasettet

I dette underkapittelet diskuteres reliabiliteten og validiteten til datasettet, samt hvordan jeg har forsøkt å kontrollere for potensielle feilkilder og spuriøse sammenhenger.

Reliabilitet i denne sammenheng omhandler hvor pålitelig dataene er, hvor de kommer fra, hvordan de er samlet inn og hvordan de brukes (Christoffersen & Johannessen, 2012).

Dataene brukt i analysen er hentet fra Skoleporten og GSI som er driftet av

Utdanningsdirektoratet. Det er trolig lite usikkerhet knyttet til gjennomsnittet i karakterer på de ulike skolene, da disse rapporteres inn etter standardiserte rutiner og det er få til ingen tolkninger involvert i rapporteringen. Det kan så klart problematiseres hvorvidt eksempelvis standpunktkarakterer avspeiler elevenes faktiske prestasjon, altså om alle lærere er helt objektive og setter karakterer på samme grunnlag. Dette er det vanskelig å etterprøve, så jeg må anta at lærere i stor grad setter «riktige» karakterer. Jeg kommer imidlertid tilbake til dette ved diskusjon av funn, og hvorvidt lærere setter karakterer ulikt etter ulik gruppestørrelse.

Fra elevundersøkelsen er det lite usikkerhet knyttet til innrapporteringen. En kan også her problematisere hvorvidt elevene svarer «sant» på spørsmålene de stilles, enten fordi de krysser vilkårlig eller ikke har forstått spørsmålene på rett måte. Her må jeg anta at fagpersoner som har utarbeidet elevundersøkelsen har tatt tilstrekkelig høyde for dette, og utformet undersøkelsen etter god praksis for spørreundersøkelser, samt stole på at elevenes svar representerer deres reelle oppfatning.

Imidlertid kan det være noe usikkerhet knyttet til hvordan gruppestørrelse 2 er rapportert og beregnet, da det er en komplisert beregning og mange innrapporteringer som ligger bak det endelige tallet for hver skole. Den første usikkerheten ligger i måten gruppestørrelse 2 er rapportert for 1.-10.-trinnskoler før 2017 fordi lærertetthet ikke ble regnet ut på denne måten tidligere. For 1.-10.-trinnskolene er dataene blitt rapportert samlet, og her har Utdanningsdirektoratet måttet gjøre et estimat på hvor mange dette gjaldt på ungdomsskolen. Dette gjør at det for denne gruppen er knyttet noe mer usikkerhet til denne estimeringen. Den andre usikkerheten handler om hvordan timer til spesialundervisning og særskilt norskopplæring er rapportert inn fra skolene, da disse beregnes ut fra gjennomsnitt. Eksempelvis er det slik at dersom eleven er tildelt mellom 1-75 timer spesialundervisning, beregnes det her et snitt på 57 timer per elev i GSI-rapporteringen. Slik er det også for de andre intervallene (Utdanningsdirektoratet, 2019b). Om vi ser på enkelttilfeller vil det dermed

være eksempler hvor den reelle gruppestørrelsen ikke er helt riktig, for eksempel når en elev faktisk har hatt 20 timer spesialundervisning og det beregnes 57 timer. Men siden GSI-rapporteringen er basert på snittet må vi anta at det over tid og med mange observasjoner vil jevne seg ut, samt at tilfellene med over- og underestimering er tilfeldig fordelt på tvers av skoler. Enhver slik tilfeldig feil kan føre til økt standardfeil og dermed er det vanskeligere å estimere signifikante effekter. Siden spesialundervisning utgjør en veldig liten del av formelen for gruppestørrelse 2, vil dette trolig ha lite å si.

Generelt vil det alltid være usikkerhet knyttet til innrapportering av data, men i så stor skala vi snakker om her vil dette trolig ikke ha stor innvirkning på relabiliteten til hele datasettet samlet. Jeg må derfor stole på de dataene som er rapportert inn til Utdanningsdirektoratet, da jeg ikke har annen informasjon tilgjengelig.

Validitet eller gyldighet handler om hvorvidt dataene klarer å måle det de har til hensikt å måle, slik at tolkningen av resultatene kan generaliseres (Christoffersen & Johannessen, 2012). En styrke ved disse dataene, som gjør de mer generaliserbare, er at jeg har observasjoner for alle skoler i landet i datasettet. Dermed kan jeg med større sikkerhet si at det gjelder for hele landet, siden det er et representativt utvalg. Samtidig har jeg tatt noen valg med tanke på utvalget for å unngå målefeil ved å droppe veldig små og veldig store gruppestørrelser. Dette kan dermed ha en liten innvirkning på representativiteten.

I den andre regresjonsmodellen forsøker jeg å kontrollere for karakteristikk ved skolen, men dette er fortsatt mangelfullt, ettersom mye informasjon om skolene og elevene ikke ligger i datasettet. Eksempelvis klarer jeg ikke kontrollere for den sosioøkonomiske bakgrunnen til elevene. Dette kan potensielt føre til at koeffisientene mine viser spuriøse sammenhenger mellom gruppestørrelse 2 og karakterene. Det kan også føre til at reelle sammenhenger blir skjult av at jeg ikke kontrollerer for dette.

I fasteffektanalysen kontrollerer jeg bedre for dette, fordi det er mindre sannsynlig at dette varierer mye innad i skolen over kort tid. Likevel kan det hende at sosioøkonomisk bakgrunn på elevene varierer fra ett år til et annet, og er korrelert med eksamenskarakter. For eksempel kan det komme mange flere innvandrere samtidig som gruppestørrelse 2 synker, og da vil koeffisienten min som fanger opp effekten av gruppestørrelse 2 på elevenes resultater kanskje heller fange opp effekten av at det er flere innvandrere på skolen. En kommer ikke

helt unna at det kan være andre uobserverbare variabler som endres over tid. Men det er en stor styrke at jeg har mange skoler over mange år, slik at hvis det skjer små endringer, og disse ikke er systematiske, burde disse kun føre til støy i analysene mine. Det må dermed være ganske store, systematiske endringer over tid for at mine estimater for gruppestørrelse på elevers prestasjoner er spuriose i fasteffektmodellen.

Et annet element som er viktig å påpeke er at gruppestørrelse 2 som mål gjelder for skolen generelt, og ikke spesifikt for matematikk. Det er dermed et spørsmål hvorvidt oppgitt gjennomsnittlig gruppestørrelse er representativ for hvordan det faktisk er i matematikkundervisningen. Det kan tenkes at noen skoler prioriterer å ha små grupper i mer praktiske fag som mat og helse, og kunst og håndverk. Dette kunne potensielt vært med på å vanne ut en eventuell effekt av endring i gruppestørrelsen når vi ser på matematikk spesifikt. Et argument mot at dette er tilfellet, er at i den kvalitative delen i Kirkebøen et al. (2017) oppgir skolene som har fått ekstra ressurser, at de hovedsakelig brukes på fagene norsk, engelsk og matematikk. Derfor kan det kanskje heller være omvendt, at gjennomsnittlig gruppestørrelse i matematikk faktisk er lavere enn snittet på skolen tilsier.

Uansett er dette noe en bør være oppmerksom på ved tolkning av resultatene. For å illustrere dette kan vi ta utgangspunkt i følgende hypotetiske eksempel på en skole som får tilført flere lærere og alt annet holdes likt. Dersom alle disse blir brukt til økt lærertetthet i kunst og håndverk vil det ikke være noen endring for matematikkundervisningen selv om gruppestørrelse 2 for skolen går ned. I en regresjonsanalyse vil vi da ikke finne noen sammenheng mellom gruppestørrelse 2 og matematikkarakterer, siden sistnevnte ikke påvirkes av at gruppestørrelse 2 endres. Vi kan da risikere å feiltolke at lærertetthet ikke betyr noe for prestasjoner selv om en sammenheng faktisk skulle eksistere. Tilsvarende kan vi ende opp med å overvurdere effekten av økt lærertetthet, dersom alle de nye lærerne blir brukt i matematikk. Lærertettheten i matematikk vil i praksis da øke betydelig, men endringen i gruppestørrelse 2 blir mye mindre siden den ser på snittet av alle fag. Dersom det er en sammenheng mellom prestasjon og lærertetthet vil det se ut som kun en liten endring i gruppestørrelse 2 får stor effekt på prestasjoner. Men i realiteten har lærertettheten i matematikk økt betydelig, slik at vi risikerer å overvurdere hvor stor effekten faktisk er. Utfordringen er uansett at med måten gruppestørrelse 2 er rapportert på, kan vi ikke vite i hvilken grad det første eller andre eksempelet her utspiller seg. Ideelt sett skulle vi derfor hatt tall på lærertetthet eller gruppestørrelse fag for fag, men dette er altså ikke mulig med måten

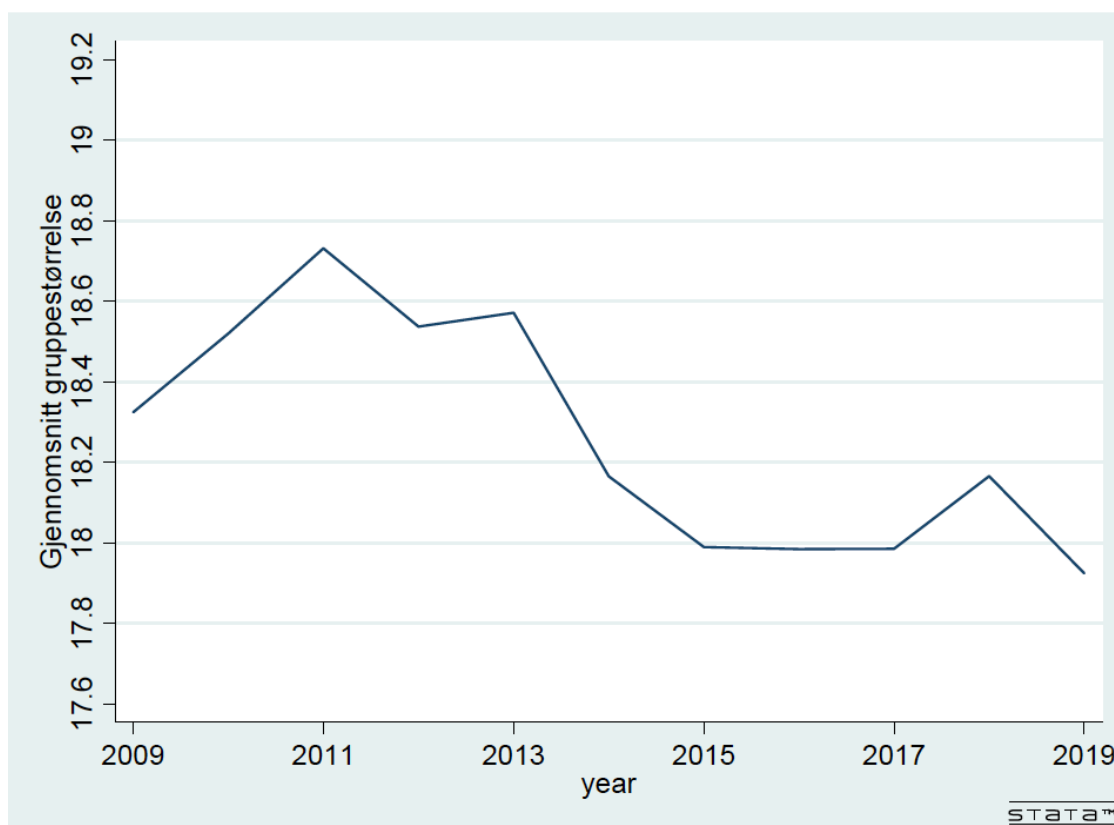
tallene rapporteres på i dag. Vi må derfor gjøre en antakelse om at endringer i gruppestørrelser i hovedsak fordeler seg på samme måte på tvers av fag og skoler.

4 Resultater

Under vil jeg presentere resultatene og funn jeg skal bruke i analysene. Før jeg går inn på resultatene vil jeg først presentere noen deskriptive data som er nyttig for forståelsen av datasettet jeg bygger analysen på.

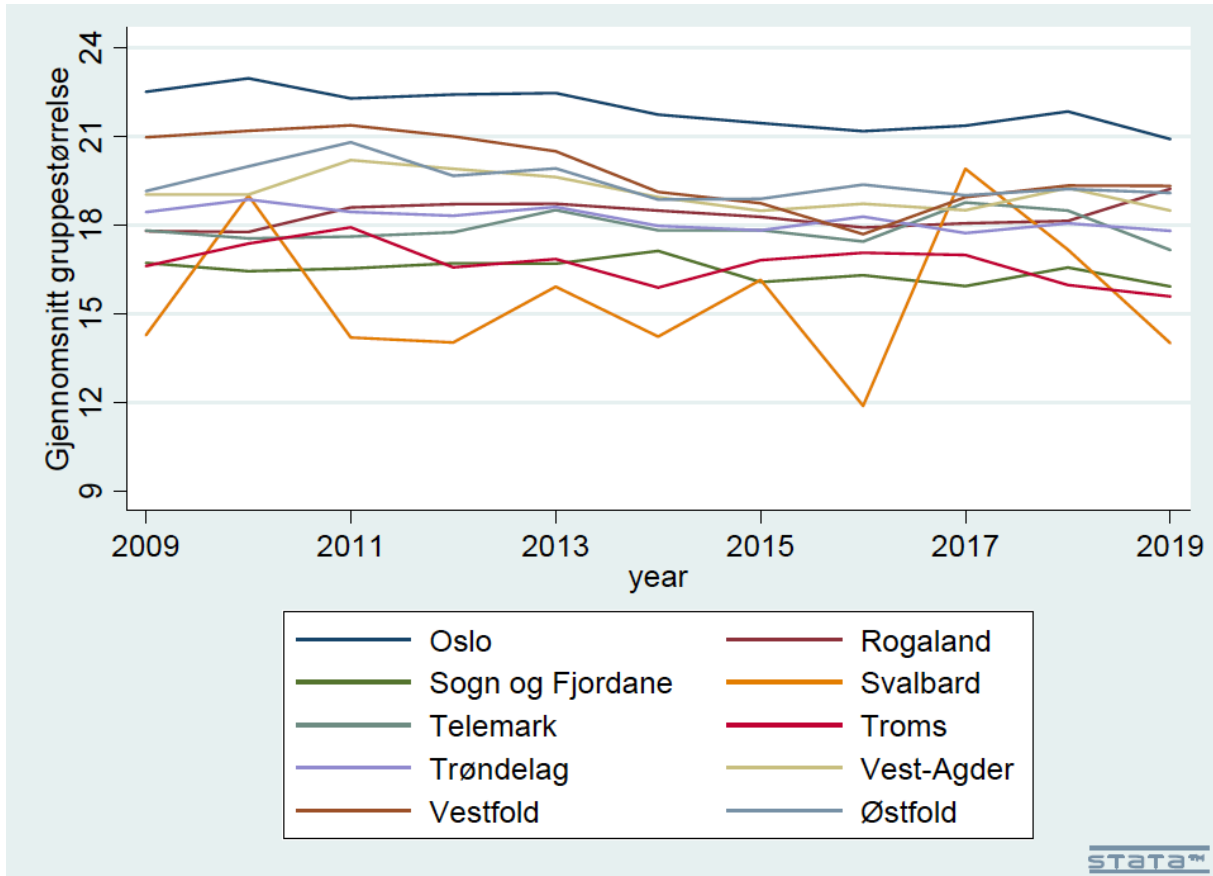
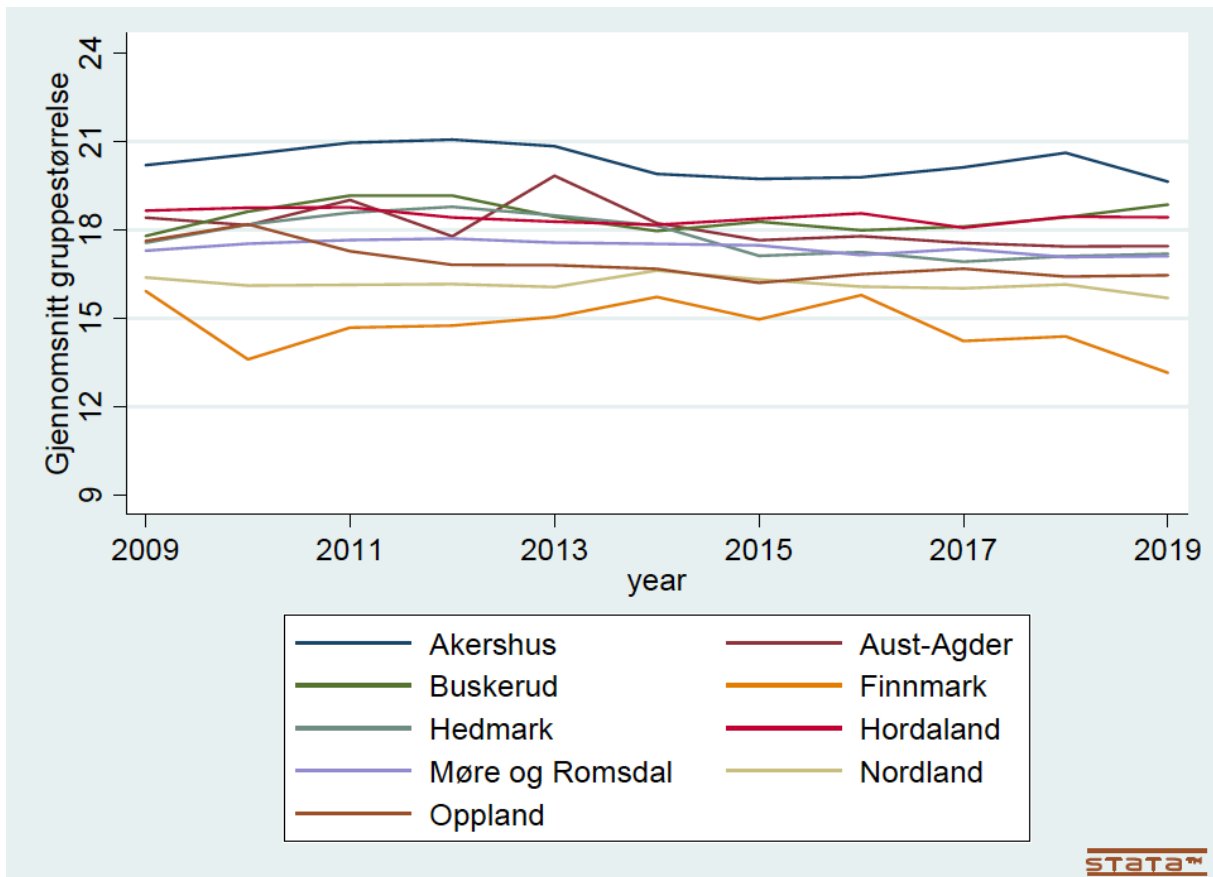
4.1 Deskriptive data

Grafen i Figur 2 viser utviklingen i gjennomsnittlig gruppestørrelse 2 fra 2009 til 2019. Grafen viser en klar nedadgående trend, spesielt et fall fra 2013 til 2015. Det er også en del svingninger fra år til år. Det er verdt å merke seg at y-aksen er avkortet, og at diagrammet dermed gir et overdrevet inntrykk av endringene i gruppestørrelse 2.



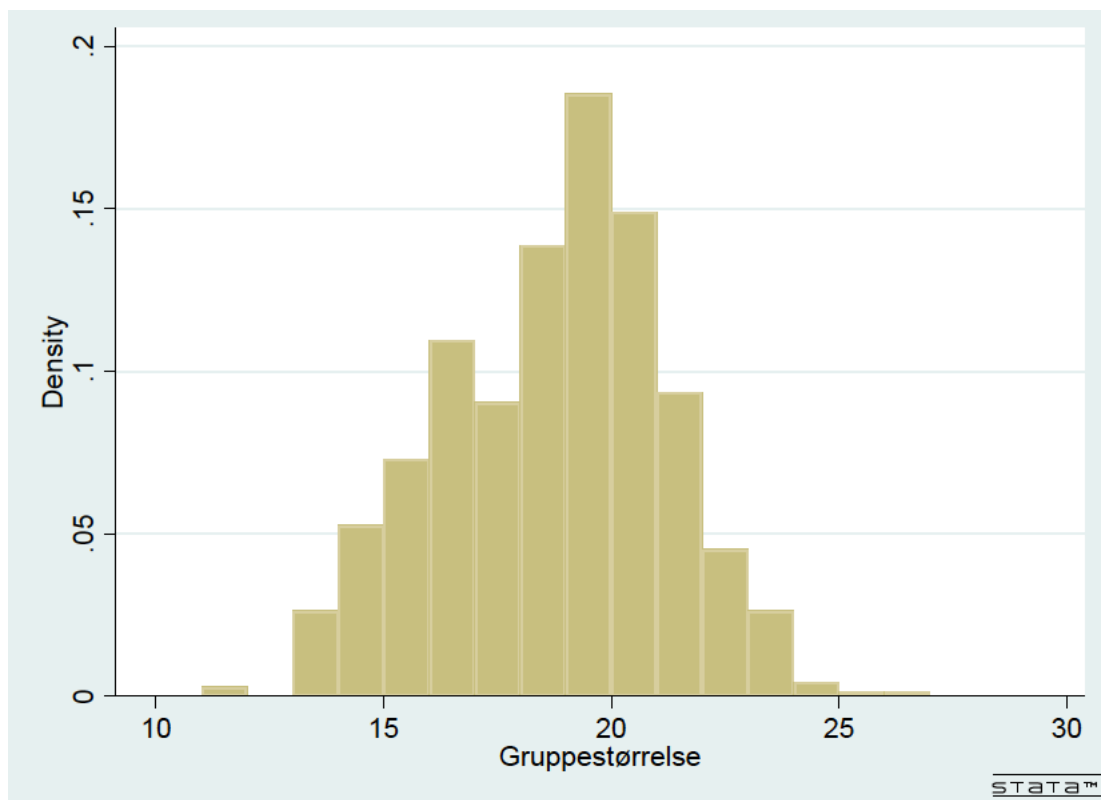
Figur 2: Utviklingen av gruppestørrelse 2 2009-2019

I grafene i Figur 3 representerer hver linje ett fylke, hvor vi ser at det er store forskjeller mellom fylkene når det gjelder nivået på gruppestørrelse 2. Grafene er delt i to av praktiske årsaker, da det var vanskelig å lese dem med alle fylkene i en og samme graf. Mens Oslo ligger høyest med et snitt på mellom 21 og 23 elever per lærer, ligger Finnmark nederst med rundt 15 elever per lærer.



Figur 3: Utviklingen av gruppestørrelse 2 i de ulike fylkene fra 2009-2019

Histogrammet i Figur 4 viser fordelingen av gruppestørrelse 2 samlet over alle årene. Tetthet (density) vil her summeres til 1, og representerer totalen, altså 100 %. Vi ser at $19 - 19,9$ er den mest vanlige gruppestørrelsen, i nesten 20 % av skole-år-observasjonene. Medianen i utvalget er 18,3 elever. Medianen kommer jeg til å benytte senere når jeg deler opp utvalget i undergrupper for å analysere mulig heterogenitet mellom gruppestørrelser.



Figur 4: Histogrammet viser fordelingen av gruppestørrelse 2 samlet over alle årene

Som følge av funn senere i resultatdelen, har det vært viktig i den deskriptive delen å se på korrelasjonen mellom gruppestørrelse 2 og størrelse på skolen. Jeg finner at det er sterk korrelasjon mellom disse. Gjennomsnittlig størrelse på tvers av alle skoler over alle årene er 68 elever på 10. trinn. For skoler som i snitt har gruppestørrelse 2 mindre enn 18,3 elever er dette snittet 43 elever, mens det for skoler med gruppestørrelse 2 større enn 18,3 elever i snitt er 81 elever. Det er dermed en sterk korrelasjon mellom gruppestørrelse 2 og størrelse på 10. klassetrinnet på skoler.

Tabell 4 viser gjennomsnittet og standardavviket for karakterene i de ulike fagene før standardisering.

	Gj.snitt	Std.avvik	Antall obs.	Min.	Max.
Matematikk standpunkt	3,54	0,35	8516	1,7	4,8
Engelsk skriftlig standpunkt	3,89	0,30	8510	2,5	5,2
Norsk hovedmål standpunkt	3,83	0,29	8498	2,3	5,1
Matematikk skriftlig eksamen	3,24	0,49	6529	1,2	5,0
Engelsk skriftlig eksamen	3,71	0,37	6352	2,0	5,2
Norsk skriftlig eksamen	3,43	0,34	6288	2,0	5,0
Matematikk muntlig eksamen	4,09	0,47	4817	1,7	5,8
Engelsk muntlig eksamen	4,45	0,41	5263	2,7	5,9
Norsk muntlig eksamen	4,42	0,45	5159	2,0	5,9
Indeks: støtte fra lærer	3,97	0,24	4639	2,6	4,8
Indeks: støtte hjemmefra	3,98	0,23	4639	2,9	4,8

Tabell 4: Gjennomsnitt og standardavvik for de ulike karakterene før standardisering
n= 831 skoler, antall observasjoner er antall skoler*antall år

Gjennomsnittet for standpunktkarakterer i matematikk over alle årene er 3,54, mens det er litt lavere på matematikk skriftlig eksamen med et gjennomsnitt på 3,24. For matematikk muntlig eksamen er snittet 4,09. Det er også verdt å merke seg at standardavvikene på standpunkt er litt lavere enn for eksamenskarakterene. Dette skyldes trolig at eksamen er en engangsmåling av kompetanse, mens standpunkt settes ved hjelp av flere vurderinger over tid. Det blir dermed større variasjon i karakterene ved eksamen enn ved standpunkt.

4.2 OLS-regresjoner

Under vil jeg gå gjennom resultatene for de to OLS-regresjonene.

4.2.1 OSL-regresjon med kontroll kun for år

I den enkle OLS-regresjonen ser jeg på sammenhengen mellom lærertetthet og variablene standpunkt i matematikk, norsk hovedmål, engelsk skriftlig, matematikkeksamen skriftlig og muntlig, norskeksamen hovedmål skriftlig og norskeksamen muntlig, samt engelskeksamen skriftlig og muntlig. Jeg kontrollerer for år slik at jeg kun sammenlikner mellom skoler innad i samme året. Dette er fordi det kan være årlige variasjoner i karakterer som er uavhengig av målet på læring. For eksempel kan eksamen være enklere ett år enn et annet. Disse årlige

variasjonene vil jeg da kontrollere bort. Fra Tabell 5 ser vi i kolonne 2 koeffisienten β_1 , som gir forskjellen i karakterer dersom en øker gruppestørrelse med én enhet. I kolonne 2 vises standardfeilen, som er målet for usikkerheten. Alle koeffisientene er estimert med usikkerhet på grunn av variasjon i data.

	β_1	Standardfeil	Antall obs.
Matematikk standpunkt	0,013***	0,003	8354
Engelsk skriftlig standpunkt	0,028***	0,003	8349
Norsk hovedmål standpunkt	0,013***	0,003	8344

Tabell 5: Resultater fra regresjonsmodell 1 for standpunktkarakterer.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

Ut fra resultatene ser vi at det er positiv korrelasjon mellom gruppestørrelse 2 og matematikk standpunktkarakter. Dersom gruppestørrelsen i ordinær undervisning øker med én enhet vil elevene i gjennomsnitt ha 1,3 % av et standardavvik bedre karakterer i standpunkt i matematikk. En positiv korrelasjon mellom gruppestørrelse og karakterer når vi ikke kontrollerer for annet enn år er ikke uventet, da det ikke kontrolleres for variabler som størrelse på kommunen, hvilken kommune en er i, kjønn, type skole, antall lærere med mer. For eksempel har Osloskolen landets beste karakterer, men også de største gruppestørrelsene. Jeg finner liknende resultater som for matematikk standpunkt også for engelsk og norsk standpunktkarakterer.

	β_1	Standardfeil	Antall obs.
Matematikk skriftlig eksamen	0,043***	0,003	6470
Engelsk skriftlig eksamen	0,048***	0,004	6291
Norsk skriftlig eksamen	0,038***	0,004	6235

Tabell 6: Resultater fra regresjonsmodell 1 for karakterer på skriftlig eksamen.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

Vi ser at de samme tendensene går igjen for skriftlig eksamen i Tabell 6, men her er det en sterkere positiv korrelasjon for eksamen i matematikk, engelsk og norsk enn for standpunktkarakterer. Dersom gruppestørrelsen i ordinær undervisning øker med én enhet vil elevene i gjennomsnitt ha 4,3 % av et standardavvik bedre karakterer på skriftlig eksamen i matematikk.

	β_1	Standardfeil	Antall obs.
Matematikk muntlig eksamen	0,019***	0,004	4795
Engelsk muntlig eksamen	0,031***	0,004	5236
Norsk muntlig eksamen	0,025***	0,004	5141

Tabell 7: Resultater fra regresjonsmodell 1 for karakterer på muntlig eksamen.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

Som vi ser er resultatene fra muntlig eksamen i Tabell 7 i de ulike fagene ganske tilsvarende de for standpunkt og skriftlig eksamen, og koeffisientene ligger omtrent midt i mellom.

4.2.2 OLS-regresjon med flere kontroller

Jeg har deretter gjennomført en liknende OLS-regresjon som over hvor jeg i tillegg til år kontrollerer for type skole, antall lærere på skolen, andel kvinnelige lærere på skolen og hvilken kommune en bor i. Som tidligere nevnt, er dette kontroller som til en viss grad kan kontrollere for kjennetegn ved skole og sted. Grunnen til at de er valgt ut er fordi det var disse som var tilgjengelig i datasettet, og at dette er variabler som andre studier også bruker som kontrollvariabler (Bonesrønning & Vaag Iversen, 2008; Falch et al., 2017; Grøgaard et al., 2008; Leuven et al., 2008; Opheim et al., 2010).

	β_1	Standardfeil	Antall obs.
Matematikk standpunkt	0,005	0,004	8354
Engelsk skriftlig standpunkt	0,012***	0,004	8349
Norsk hovedmål standpunkt	0,008**	0,004	8344

Tabell 8: Resultater fra regresjonsmodell 2 for standpunktkarakterer.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

Resultatene fra standpunktkarakterene, vist i Tabell 8, viser nå at den positive sammenhengen i matematikk ikke lenger er signifikant. Det er fortsatt en signifikant positiv korrelasjon i engelsk og norsk, men disse er mindre enn i den enklere regresjonsmodellen med kun kontroll for år. Hvis en sammenlikner med resultatene i Tabell 4 ser vi at dersom gruppestørrelsen øker med én enhet vil elevene i gjennomsnitt få 2,8 % av et standardavvik bedre karakterer i

engelsk skriftlig standpunkt. Etter å ha lagt inn kontroller gjør nå elevene det kun 1,2 % av et standardavvik bedre.

	β_1	Standardfeil	Antall obs.
Matematikk skriftlig eksamen	0,014***	0,004	6470
Engelsk skriftlig eksamen	0,008*	0,004	6291
Norsk skriftlig eksamen	0,011**	0,005	6235

Tabell 9: Resultater fra regresjonsmodell 2 for karakterer på skriftlig eksamen.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

Fra Tabell 9 finner jeg fremdeles en signifikant positiv sammenheng på skriftlig eksamen, men størrelsen er sterkt redusert sammenliknet med når vi kun kontrollerte for år i Tabell 6 over. β_1 går vesentlig ned sammenliknet med koeffisientene i regresjonsmodell 1. For skriftlig eksamen i matematikk viste Tabell 6 at elevene i gjennomsnitt får litt over 4 % av et standardavvik bedre karakterer dersom gruppestørrelse øker med én enhet. Etter å ha lagt inn kontroller gjør nå elevene det kun litt over 1 % av et standardavvik bedre. Det vil si at sammenhengen mellom økning i gruppestørrelse 2 og bedre eksamenskarakterer svekkes mye når en legger inn kontrollene type skole, antall lærere på skolen, andel kvinnelige lærere på skolen og hvilken kommune en er i.

	β_1	Standardfeil	Antall obs.
Matematikk muntlig eksamen	0,013**	0,006	4795
Engelsk muntlig eksamen	0,015***	0,005	5236
Norsk muntlig eksamen	0,010*	0,005	5141

Tabell 10: Resultater fra regresjonsmodell 2 for karakterer på muntlig eksamen

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

Som vi ser i Tabell 10, er koeffisientene for muntlig eksamen også lavere enn tidligere når jeg kun kontrollerer for år. Sammenlikner vi for eksempel resultatene for engelsk muntlig eksamen fra Tabell 7, viste denne at dersom gruppestørrelsen øker med en enhet vil elevene i gjennomsnitt ha litt over 3 % av et standardavvik bedre karakterer. Etter å ha lagt inn kontroller reduseres denne til kun 1,5 % av et standardavvik. Signifikansnivået svekkes også noe for både matematikk og norsk.

Generelt kan vi si at den positive korrelasjonen ser ut til å svekkes når en inkluderer kontroller for andre faktorer som kan ha betydning for elevens prestasjoner. Selv om koeffisienten er mye mindre finner vi fremdeles positive korrelasjoner for alle estimatene, hvor de fleste også er statistisk signifikante. Det vil si at høyere gruppestørrelse 2 er positivt korrelert med bedre karakterer på standpunkt og eksamener for de fleste av fagene.

4.2.3 OLS-regresjoner med støtte fra lærer som utfallsvariabel

Jeg har også gjennomført de samme OLS-regresjonen, men nå med indeksen *støtte fra lærer* fra elevundersøkelsen som utfallsvariabel (Y). *Støtte fra lærer* bruker jeg som en kontinuerlig utfallsvariabel for å se om endringer i gruppestørrelse endrer opplevd støtte fra læreren. Jeg har brukt indeksen som viser gjennomsnitt av hva elevene har svart på de 5 indikatorene i analysene, og for å kunne sammenlikne resultater har jeg standardisert denne variabelen på samme måte som karakterer.

	β_1	Standardfeil	Antall obs.
Indeks: <i>støtte fra lærer</i>	-0,008**	0,004	4566

Tabell 11: Resultater fra regresjonsmodell 1 for indeksen *støtte fra lærer*.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

Resultatene i Tabell 11, fra regresjonsmodellen der jeg kun kontrollerer for år, viser at elevene svarer signifikant mer negativt på elevundersøkelsen knyttet til indeksen *støtte fra lærer* når gruppestørrelsen øker. Hvis gruppestørrelse 2 øker med én enhet rapporterer elevene 0,8 % av et standardavvik mindre støtte fra lærer. Her kan det igjen være flere forhold ved skolene som påvirker resultatene, for eksempel at skoler som har større gruppestørrelser ligger i byene. Videre vil jeg derfor kontrollere for noe av dette ved å legge inn kontroller for type skole, antall lærere på skolen, andel kvinnelige lærere på skolen og hvilken kommune en bor i.

	β_1	Standardfeil	Antall obs.
Indeks: <i>støtte fra lærer</i>	0,006	0,005	4566

Tabell 12: Resultater fra regresjonsmodell 2, for indeksen *støtte fra lærer*.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

Dersom en legger inn de samme kontrollene som tidligere finner jeg nå ingen signifikant sammenheng med indeksen *støtte fra lærer*, som vist i Tabell 12. Det kan tyde på at det er forskjeller i gruppestørrelse 2 mellom ulike kommuner eller ulike skoler, som har vært med på å drive korrelasjonen jeg fant i regresjonen i Tabell 11. Siden den signifikante sammenhengen forsvinner med de nevnte kontrollvariablene, kan vi konkludere med at gruppestørrelse 2 ikke har noen signifikant sammenheng med opplevd støtte fra lærer.

4.3 Fasteffektregresjoner

Fra resultatene over ser vi at det er nødvendig, for å kunne si noe om årsakssammenhengen mellom gruppestørrelse 2 og utfallene, å kunne kontrollere bedre for at skoler er ulike. Dette kan vi gjøre ved å se på karakterene innad på en skole over tid, noe fasteffektanalyser tillater oss å gjøre. Hittil har jeg kjørt OLS-regresjoner uten og med en rekke kontrollvariabler for å vise seleksjonen. I fasteffektmodellen vil en del av denne seleksjonen automatisk kontrolleres for. Jeg vil derfor bruke begrepet effekt på prestasjoner av gruppestørrelse 2 videre i oppgaven for å indikere at vi nå er nærmere en kausal sammenheng.

I OLS-modell 2 kontrollerte jeg for år, type skole, antall lærere på skolen, andel kvinnelige lærere på skolen og hvilken kommune en er i. For fasteffektregresjonene er det kun år, skoletype, antall lærere og andel kvinnelige lærere som er relevant å kontrollere for, siden samme skole ligger i samme kommune for alle årene. De aller fleste skoler er også samme type skole over tid, men noen skoler endrer skoletype så derfor bør vi ha med denne kontrollvariabelen.

4.3.1 Fasteffektregresjoner med karakterer som utfallsvariabler

I Tabell 13 viser koeffisienten β_1 hva som skjer med standpunktkarakterene dersom gruppestørrelse 2 øker med én enhet.

	β_1	Standardfeil	Antall obs.
Matematikk standpunkt	0,001	0,004	8354
Engelsk skriftlig standpunkt	0,013***	0,004	8349
Norsk hovedmål standpunkt	0,006	0,004	8344

Tabell 13: Resultater fra fasteffektregresjonsmodell for standpunktkarakterer.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

For standpunkt i matematikk og norsk viser resultatene at innad på samme skole, dersom gruppestørrelsen øker fra ett år til et annet, vil det statistisk sett ikke gjøre at elevene får hverken bedre eller dårligere resultater. Det er altså ingen signifikant effekt av gruppestørrelse 2 på standpunkt karakterer i disse to fagene. For engelsk standpunkt vises en signifikant positiv effekt. Det vil si at dersom gruppestørrelsen øker med én enhet fra ett år til et annet vil elevene prestere 1,3 % av et standardavvik bedre i engelsk standpunkt. Hvis jeg sammenlikner med resultatene fra OLS-regresjonen med kontroller i Tabell 8, ser vi at også her var det engelsk som hadde en liknende positiv sammenheng, mens det var ingen sammenheng i matematikk standpunkt. Samtidig er størrelsen på effekten i engelsk standpunkt relativt liten også her.

	β_1	Standardfeil	Antall obs.
Matematikk skriftlig eksamen	0,008**	0,004	6470
Engelsk skriftlig eksamen	0,011***	0,004	6921
Norsk skriftlig eksamen	0,011**	0,005	6235

Tabell 14: Resultater fra fasteffektregresjonsmodell for karakter på skriftlig eksamen.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

For skriftlig eksamen viser resultatene presentert i Tabell 14 en signifikant positiv effekt mellom gruppestørrelse 2 og skriftlig eksamen i alle fagene. Dersom gruppestørrelsen øker med én enhet fra ett år til et annet innad på en skole, vil elevene prestere ca. 1 % av et standardavvik bedre på eksamen. Effekten er relativt lik på tvers av alle tre fagene. Den signifikante positive effekten er også omtrent på samme størrelse med de fra OLS-regresjonsmodellen med alle kontrollvariabler i Tabell 9.

	β_1	Standardfeil	Antall obs.
Matematikk muntlig eksamen	0,008	0,005	4795
Engelsk muntlig eksamen	0,013***	0,005	5236
Norsk muntlig eksamen	0,010*	0,005	5141

Tabell 15: Resultater fra fasteffektregresjonsmodell for karakter på muntlig eksamen.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

Fra Tabell 15 ser vi at tendensene også er de samme som for tilsvarende OLS-regresjon i Tabell 10, men at matematikk ikke lenger er signifikant. Det er kun for engelsk muntlig at

effekten av gruppestørrelse er signifikant, og størrelsen på effekten er omtrent den samme som for skriftlig eksamen.

4.3.2 Fasteffektregresjon med *støtte fra lærer* som utfallsvariabel

Jeg har også gjennomført fasteffektregresjonen med indeksen *støtte fra lærer* fra elevundersøkelsen som utfallsvariabel. Kontrollene er de samme som for modellene over, altså år, skoletype, antall lærere og andel kvinnelige lærere. Resultatene vises i Tabell 16.

	β_1	Standardfeil	Antall obs.
Indeks: <i>støtte fra lærer</i>	0,001	0,006	4566

Tabell 16: Resultater fra fasteffektregresjonsmodell for indeksen *støtte fra lærer*.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

Resultatene viser det samme som for OLS-regresjonen med alle kontroller, at endring i gruppestørrelse 2 ikke har signifikant effekt på indeksen *støtte fra lærer*.

4.4 Analyser av heterogenitet mellom ulike grupper

I de videre analysene ønsker jeg å se hva som driver resultatene, og undersøke hvorvidt det er heterogenitet mellom ulike undergrupper i datasettet. Undergruppene jeg har valgt å se nærmere på er skoler med gjennomsnittlig gruppestørrelse over versus under medianen, gutter versus jenter, og til slutt mellom skoler som i gjennomsnitt opplever stor versus liten grad av støtte hjemmefra, som en proxy for sosioøkonomisk bakgrunn.

I disse analysene velger jeg kun å se på matematikkarakterer og score på *støtte fra lærer* som utfallsvariabler. Dette fordi det er disse som er mest relevante for mitt forskningsspørsmål, og fordi sammenlikningsgrunnlaget med de andre fagene er dekket av de overnevnte analysene. Det er lite trolig at det ville vært store forskjeller mellom dem. Videre bruker jeg kun fasteffektmodellen beskrevet i 4.3 når jeg gjør analysene av heterogenitet mellom ulike grupper fordi dette er den modellen jeg anser som nærmest fanger opp årsakssammenhenger.

4.4.1 Heterogenitet mellom gruppestørrelse over eller under medianen

Jeg starter med å se på mulig heterogenitet i størrelse på gruppen fordi det kan være at sammenhengen mellom gjennomsnittlig gruppestørrelse og elevers prestasjoner ikke er

lineær. Det vil for eksempel si at det ikke er sikkert at det gir samme effekten å øke en gruppe fra 13 til 14 elever som å øke fra 20 til 21 elever. En måte å undersøke dette nærmere på er å dele utvalget i to, de over og under medianen for gruppestørrelser, og se på effekten separat for disse to gruppene.

Med den samme fastregresjonsmodellen som er brukt i hele delkapittelet, skal vi nå se på heterogenitet i gruppestørrelse 2 dersom gruppestørrelsen er henholdsvis større og mindre enn medianen på 18,3. Kontrollene i regresjonsmodellen er de samme som i den over, år, skoletype, antall lærere og andel kvinnelige lærere.

	β_1	Standardfeil	Antall obs.
Matematikk standpunkt $Gr_2 > 18,3$	0,009	0,006	4247
Matematikk standpunkt $Gr_2 \leq 18,3$	-0,004	0,009	4107

Tabell 17: Resultater fra fasteffektregresjonsmodell for standpunktarakter, høy og lav gruppestørrelse.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

I Tabell 17 ser vi fra koeffisienten β_1 hva som skjer med karakterene til elevene i matematikk dersom gruppestørrelse 2 øker med én enhet. Resultatene viser at det ikke påvirker matematikk standpunkt, hverken for gruppen over eller under medianen. Vi finner altså ingen signifikant effekt av gruppestørrelse 2 på matematikk standpunkt, og det er ikke forskjeller mellom skoler med høy og lav gruppestørrelse.

Deretter gjennomfører jeg samme regresjonsmodell for skriftlig eksamen i matematikk, for de to undergruppene over og under medianen.

	β_1	Standardfeil	Ant. obs.
Matematikk skriftlig eksamen $Gr_2 > 18,3$	0,012* ³	0,006	3810
Matematikk skriftlig eksamen $Gr_2 \leq 18,3$	0,013	0,010	2660

Tabell 18: Resultater fra fasteffektregresjonsmodell for karakter på skriftlig eksamen, høy og lav gruppestørrelse.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

³ P er her lik 0,06, altså et signifikansnivå på 6 %.

Resultatene i Tabell 18 viser at i undergruppen med gruppestørrelse 2 under 18,3 elever, har endringer i gruppestørrelse ingen signifikant effekt på skriftlig eksamenskarakter i matematikk. For gjennomsnittlig gruppestørrelse over 18,3 viser resultatene en signifikant positiv effekt på eksamenskarakter i matematikk, men kun på 10 %-nivå. I utgangspunktet vektlegger jeg ikke resultater på 10 %-nivå, men i dette tilfellet er effekten på eksamenskarakter i matematikk svært nærme å være på 5 % nivå og vil bli nevnt i drøftingen.

	β_1	Standardfeil	Ant. Obs.
Matematikk muntlig eksamen $Gr_2 > 18,3$	0,023***	0,007	2954
Matematikk muntlig eksamen $Gr_2 \leq 18,3$	-0,006	0,016	1841

Tabell 19: Resultater fra fasteffektregresjonsmodell for karakter på muntlig eksamen, høy og lav gruppestørrelse.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

I Tabell 19 ser vi at for muntlig eksamen finner vi en signifikant positiv effekt for gjennomsnittlig gruppestørrelse over medianen. Disse elevene gjør det 2,3 % av et standardavvik bedre på muntlig eksamen i matematikk om gruppestørrelse 2 øker med én enhet. Det er ingen signifikant effekt for gjennomsnittlig gruppestørrelse under medianen.

Som nevnt i 4.1 Deskriptiv data, er det en sterk korrelasjon mellom store skoler og høy gjennomsnittlig gruppestørrelse 2. Derfor kan en si at for store skoler og høy gruppestørrelse ser det altså ut til at det er en positiv effekt av å ha større gruppestørrelse for muntlig eksamenskarakterer i matematikk, men ikke skriftlig eksamen eller standpunkt. For gjennomsnittlig gruppestørrelse under medianen finner jeg ingen effekter for noen av karakterene. For disse virker det ikke som en økning i gruppestørrelse har effekt på prestasjoner i matematikk.

	β_1	Standardfeil	Ant. Obs.
Indeks: støtte fra lærer $Gr_2 > 18,3$	-0,001	0,010	2191
Indeks: støtte fra lærer $Gr_2 \leq 18,3$	0,020	0,013	2375

Tabell 20: Resultater fra fasteffektregresjonsmodell for indeksen støtte fra lærer, høy og lav gruppestørrelse.

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

I regresjonsmodellen med *støtte fra lærer* som utfallsvariabel finner vi heller ingen forskjeller for høy eller lav gjennomsnittlig gruppestørrelse. Som Tabell 20 viser er det ingen signifikant effekt av gruppestørrelse 2 på opplevd støtte fra lærer for grupper større eller mindre enn 18,3.

Det er også verdt å merke seg at jeg aldri, uansett metode eller størrelse på gruppene, finner en negativ effekt av gruppestørrelse 2 på karakterer, altså at mindre gruppestørrelser er positivt for elevers prestasjoner.

4.4.2 Heterogenitet mellom jenter og gutter

Jeg ønsker også å undersøke hvorvidt jenter og gutter responderer ulikt på endringer i gruppestørrelse. Siden jeg har hentet inn karakterer og svar på elevundersøkelsen både samlet og delt på kjønn, kan jeg analysere resultatene for gutter og jenter hver for seg. Jeg har benyttet samme fasteffektanalysemodell som før, og tar kun for meg karakterer i matematikk og indeksen *støtte fra lærer*. Jeg ser på resultatene for både gutter og jenter over og under medianen for gjennomsnittlig gruppestørrelse. Tabell 21 oppsummerer resultatene fra disse analysene, for henholdsvis gutter og jenter.

	β_1	Std.feil	Antall obs.
Gutter: Matematikk standpunkt, Gr. 2 >18,3	0,008	0,007	4094
Jenter: Matematikk standpunkt, Gr. 2 >18,3	0,013**	0,007	4113
Gutter: Matematikk standpunkt, Gr. 2 ≤18,3	0,023*	0,012	3233
Jenter: Matematikk standpunkt, Gr. 2 ≤18,3	-0,006	0,011	3199
Gutter: Matematikk skriftlig eksamen, Gr. 2 >18,3	0,006	0,007	3401
Jenter: Matematikk skriftlig eksamen, Gr. 2 >18,3	0,018**	0,007	3355
Gutter: Matematikk skriftlig eksamen, Gr. 2 ≤18,3	0,017	0,014	1882
Jenter: Matematikk skriftlig eksamen, Gr. 2 ≤18,3	0,015	0,014	1846
Gutter: Matematikk muntlig eksamen, Gr. 2 >18,3	0,038***	0,012	1199
Jenter: Matematikk muntlig eksamen, Gr. 2 >18,3	0,010	0,014	1099
Gutter: Matematikk muntlig eksamen, Gr. 2 ≤18,3	-0,005	0,048	385
Jenter: Matematikk muntlig eksamen, Gr. 2 ≤18,3	0,008	0,037	340
Gutter: Indeks støtte fra lærer, Gr. 2 >18,3	-0,006	0,010	2149
Jenter: Indeks støtte fra lærer, Gr. 2 >18,3	0,007	0,010	2138
Gutter: Indeks støtte fra lærer, Gr. 2 ≤18,3	0,012	0,016	1892
Jenter: Indeks støtte fra lærer, Gr. 2 ≤18,3	0,063	0,017	1841

Tabell 21: Resultater fra fasteffektregresjonsmodell med heterogenitet mellom gutter og jenter, høy og lav gruppestørrelse

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

Det første en kan merke seg er at antall observasjoner er ulikt for gutter og jenter og ligger lavere enn for det totale utvalget, spesielt mye lavere for muntlig eksamen. Grunnen til dette er at på mange skoler er det kun noen elever som blir trukket opp i eksamen i et fag og dermed vil det for noen år være kun jenter eller kun gutter som tar eksamen. I tillegg anonymiserer Skoleporten resultatene dersom det er så få elever i en klasse at det er fare for gjenkjennelse. Når jeg splitter i gutter og jenter blir dette gjeldende for en noe større andel av observasjonene.

Fra Tabell 21 ser vi at resultatene varierer noe for gutter og jenter på begge eksamensformene i matematikk, og indeksen *støtte fra lærer*, men hovedbildet er at de fleste estimatene ikke er signifikante. Det er en positiv effekt for jenter av økt gruppestørrelse 2 hvis man allerede er i store grupper, da både effekten for standpunkt og skriftlig eksamen er positiv signifikant på 5

%-nivå. Samtidig er det en positiv effekt på muntlig eksamen for gutter. For gruppestørrelse under medianen finner jeg fortsatt ingen effekter av å øke gruppestørrelse, hverken for gutter eller jenter. Siden effektene på gutter og jenter for gjennomsnittlig gruppestørrelse over medianen er noe sprikende er jeg derfor varsom med å legge for mye vekt på disse resultatene og konkluderer med at det ikke er entydige ulikheter i hvordan gutter og jenter responderer på endringer i gruppestørrelse.

4.4.3 Heterogenitet mellom grupper med ulik støtte hjemmefra

Til sist ønsker jeg å undersøke hvorvidt jeg kan finne heterogenitet mellom grupper med ulik sosioøkonomisk bakgrunn. Jeg har dessverre ikke noen variabler på individnivå som kan si noe om sosioøkonomisk bakgrunn, men indeksen *støtte hjemmefra* i elevundersøkelsen kan brukes som en proxy for dette. Grunnen til dette er at elever som opplever mye støtte hjemmefra oftere har høy sosioøkonomisk bakgrunn, mens det er motsatt for dem med mindre støtte hjemmefra (Nilsen & Bergem, 2016).

I utgangspunktet ønsket jeg å bruke indeksen *støtte hjemmefra* som en kontrollvariabel i modellene over, på samme måte som for eksempel antall lærere. Siden denne indeksen i elevundersøkelsen kun finnes for noen av årene, mister jeg veldig mange av datapunktene hvis jeg skulle brukt denne som kontrollvariabel. I stedet bruker jeg den altså til å dele utvalget i undergrupper, for å se om det er heterogenitet i analysene mine etter hvor mye hjelp en har hjemmefra. Jeg gjør altså kun analysen for de fire årene jeg har resultater fra indeksen på elevundersøkelsen. Jeg har valgt å dele utvalget i to etter medianen, hvor skoler med score på indeksen over medianen utgjør én gruppe og skoler med score under medianen utgjør den andre. Tabell 22 under oppsummerer resultatene av denne analysen.

	β_1	Standardfeil	Antall obs.
Matematikk standpunkt, Lav støtte	0,016**	0,008	1934
Matematikk standpunkt, Høy støtte	0,007	0,007	2533
Matematikk skriftlig eksamen, Lav støtte	0,025***	0,008	1333
Matematikk skriftlig eksamen, Høy støtte	0,023***	0,007	2064
Matematikk muntlig eksamen, Lav støtte	0,030***	0,010	1181
Matematikk muntlig eksamen, Høy støtte	0,009	0,008	1915
Indeks: støtte fra lærer, Lav støtte	-0,006	0,009	1904
Indeks: støtte fra lærer, Høy støtte	-0,003	0,006	2662

Tabell 22: Resultater fra fasteffektregresjonsmodell med heterogenitet mellom høy og lav støtte hjemmefra

* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$. $n = 831$ skoler, antall observasjoner er antall skoler*antall år

Resultatene i Tabell 22 viser at den positive effekten av gruppestørrelse 2 på karakterer i matematikk, som vi har sett spesielt for gruppestørrelser over medianen, for det meste er drevet av elever som har lav støtte hjemmefra. Både for standpunkt og muntlig eksamen er det kun signifikant effekt for de med lav støtte hjemmefra. For skriftlig eksamen er den derimot tilnærmet lik og signifikant for begge. Det er ingen signifikante effekter på indeksen *støtte fra lærer* hverken for de med lav eller høy støtte hjemmefra.

5 Drøfting og sammenlikning

I dette kapitlet skal jeg drøfte betydningen av mine funn og sammenlikne dem opp mot den tidligere gjennomgåtte internasjonale og norske forskningen. Gjennom drøftingen av mine funn og sammenlikningen med tidligere forskning skal jeg forsøke å besvare forskningsspørsmålet: «*Hva har økt lærertetthet å si for elevenes prestasjoner i matematikk?*». Kapitlet er bygd opp slik at jeg først kommenterer hva som er mine fire bidrag til forskningsfeltet. Deretter diskuterer jeg hovedfunnene, og mulige bakenforliggende årsaker til disse. Videre sammenlikner jeg funnene mot annen norsk og internasjonal forskning, for til slutt å si noe om hvilke konsekvenser funnene mine kan ha for matematikkundervisningen og politikktutforming.

Et forhold det er viktig å presisere er at forskningen min i stor grad handler om faglige resultater. Det kan også finnes andre ting, som ikke så lett kan måles, som kan være konsekvenser av større lærertetthet, både positive og negative. Dette er ikke fokus i min analyse.

5.1 Bidrag til forskningsfeltet

Under vil jeg gjennomgå og diskutere oppgavens fire viktigste bidrag til forskningsfeltet. Oppsummert er dette kombinasjonen av fokus på lærertetthet, bruk av gruppestørrelse 2, en bred gjennomgang og sammenlikning av litteratur, samt en metode som kontrollerer bedre for endogenitet enn det en del liknende studier har benyttet.

Det første bidraget er at jeg fokuserer på forskning på lærertetthet. Det finnes kun tre norske studier som bruker gruppestørrelse, og ikke klassestørrelse, som mål på lærertetthet. Hovedfokuset i både internasjonal og nasjonal forskning har vært å se på effekten av klassestørrelse og om dette er korrelert med prestasjoner. Etter dreiningen i 2003 fra maksimum klassestørrelse til pedagogisk forsvarlig gruppestørrelse (Kunnskapsdepartementet, 2004) tok det tid før det skjedde en endring på forskningsfeltet i Norge. De første studiene som bruker gruppestørrelse 1 ble publisert først i 2008. Forskningsfeltet er relativt nytt, og dermed er tilskudd til forskning på området viktig.

Mitt andre bidrag til forskningsfeltet er at jeg bruker gruppestørrelse 2 som mål på lærertetthet i mine analyser. Det er så vidt jeg vet ikke publisert noen studier i Norge som har brukt

gruppestørrelse 2 i en fasteffektregresjonsmodell. Det er heller ingen studier med like omfattende longitudinale data som ser på gruppestørrelse 2 opp mot prestasjoner på eksamen og standpunkt, slik jeg gjør. Gruppestørrelse 2 er et mye bedre mål på lærertetthet enn gruppestørrelse 1 og klassestørrelse fordi det viser antall elever per lærer i ordinær undervisning. Det vil si at ressurser til spesialundervisning og undervisning i særskilt norsk ikke regnes med. Bruk av gruppestørrelse 2 vil altså gi mer presise analyser for å undersøke effekter av lærertetthet. Som gjennomgangen av internasjonale studier i kapittel 2 viste, er det heller ikke vanlig internasjonalt å forske på lærertetthet i betydningen gruppestørrelse. Disse studiene ser kun på klassestørrelse, og det er endringer i klassestørrelse de viser til når de bruker begrepet «student teacher ratio». Den eneste norske studien gjort med utgangspunkt i gruppestørrelse 2 som variabel er Kirkebøen et al. (2017). Som tidligere nevnt så de på effektene av dette i en 4-årsperiode, der 166 ungdomsskoler fikk øremerkede midler til å ansette ekstra lærerressurser. De bruker dermed gruppestørrelse 2, men kun for 4 år. En av årsakene til at det ikke har vært forsket på paneldataene som inneholder gruppestørrelse 2 tidligere, er at det kun nylig har blitt beregnet gruppestørrelse 2 tilbake i tid fra Utdanningsdirektoratet. Direktoratet har hele tiden samlet inn alle dataene som gjør det mulig å beregne dette, men det har ikke vært et politisk fokus før rundt 2015, og de har derfor ikke innført denne måten å beregne det på tidligere. I den norske forskningen fremover vil trolig kun gruppestørrelse 2 bli brukt som mål på lærertetthet.

Det tredje bidraget er at jeg har gått gjennom store mengder litteratur innenfor forskning på lærertetthet i Norge, norsk forskning på effekter av klassestørrelse og internasjonal forskning på effekter av klassestørrelse. En av hensiktene med å gjøre en så omfattende gjennomgang av både det metodiske og funnene i denne forskningen er for å kunne sammenlikne mot mine funn, men også å trekke frem en viktig diskusjon knyttet til sammenlikningsgrunnlaget for det empiriske forskningsgrunnlaget for oppgaven. Spørsmålet er hvorvidt forskning på effekten av klassestørrelse er overførbart til forskning på lærertetthet. Dette er usikkert, siden det foreløpig er gjennomført så lite forskning med utgangspunkt i både gruppestørrelse 1 og 2. I og med at sammenlikningsgrunnlaget blir svært lite uten å diskutere mine funn opp mot forskning på klassestørrelseeffekter, blir det naturlig å se mine funn opp mot både forskning på lærertetthet og klassestørrelse. Jeg tar derfor forbehold om at å sammenlikne og overføre resultater mellom de ulike måtene å måle lærertetthet er usikkert, men nødvendig, grunnet for lite forskning med lærertetthet i fokus. Dersom det på sikt viser seg å ikke være overførbart, bør lærertetthet som forskningsfelt satses ytterligere på fremover. Som nevnt i avsnittet om

pågående norsk forskning er det flere prosjekter som skal se nærmere på gruppestørrelse. Med den nye endringen fra regjeringen som trådte i kraft høsten 2019 med maksimale gruppestørrelser i ordinær undervisning, vil det overraske om det ikke kommer mer forskning på dette de neste årene. Diskusjonen om overførbarhet mellom forskningsfeltene lærertetthet og klassestørrelse er ikke en diskusjon jeg kan ha alene, men et problem det bør rettes lys mot i videre forskning på lærertetthet.

Et siste bidrag er at jeg bruker en metode som kontrollerer bedre for endogenitet enn mange andre metoder. Fordi lærernormen er såpass fersk vil det ta noen år før man kan bruke endringer i lærertetthet grunnet denne reformen. Fasteffekt, som er metoden jeg bruker, er ut fra de longitudinale dataene vi har på gruppestørrelse 2 den beste metoden jeg kunne finne. I sammenheng med dette er det viktig å diskutere hvorvidt variabelen lærertetthet er endogen. Er den korrelert med andre variabler, som egenskaper ved elevene, lærerne og skolen? I tidligere forskning på effekten av klassestørrelsesreduksjon trekkes dette ofte frem som et problem med forskningen, og at det må arbeides mer for å sikre eksogenitet (Leuven et al., 2008). Endogenitetsproblemer kan gi forventningskjevne OLS-estimerer på effekten av lærertetthet på prestasjoner, fordi det er mange faktorer som kan påvirke både lærertetthet og prestasjoner som det er vanskelig å kontrollere for. For eksempel kan det være lavere lærertetthet fordi elever er mer motiverte for læring og sitter mer stille (dermed trenger de mindre lærerressurser), samtidig som mer motiverte elever kan presentere bedre på tester. Hvis en da ikke kontrollerer for ulik motivasjon til elevene i OLS-regresjonen kan en tilskrive effekter til lærertetthet på elevers prestasjoner som egentlig skyldes motivasjon. Dette er årsaken til at jeg har benyttet fasteffektanalyse, nettopp for å få en mer eksogen variasjon i analysen. Fasteffektanalysen skal sikre at potensiell endogenitet på gruppestørrelse 2-variabelen ses på innad på skolene, og endogenitetsproblemene blir da mindre når en kontrollerer for skolefasteffekter. For eksempel vil en da kontrollere for at noen skoler har mer motiverte elever enn andre typer skoler. Dermed antar en at den gjenværende variasjonen i gruppestørrelse 2 på prestasjoner faktisk reflekterer effekten av lærertetthet og ikke motivasjonen til elevene. Det er viktig å merke seg at antagelsen her er at elevers motivasjon (og andre kjennetegn ved skolen og elevene) innad på en skole er konstant over tid. I metodedelens over, påpekte jeg nærmere utfordringer med fasteffektanalysen. Samtidig er det gode argumenter for at en kommer betydelig nærmere en kausal effekt med fasteffektmetoden enn med standard OLS-regresjon.

Som jeg har nevnt ved flere anledninger er RD-metoden en god metode for å sikre eksogen variasjon i gruppestørrelser. Endringen som trådte i kraft høsten 2019 med maksimum gruppestørrelse vil i fremtiden gjøre det mulig å bruke denne metoden, fordi en da kan sammenlikne gruppestørrelser rundt de maksimale gruppene. Til nå er det ikke tilstrekkelig med data tilgjengelig etter endringen til å kunne gjøre en slik analyse. Dermed er fasteffekt hittil metoden en må bruke for å sikre mest mulig eksogen variasjon dersom en skal bruke longitudinale data.

5.2 Diskusjon av funn

Overordnet viser mine analyser at jeg ikke finner noen entydig og gjennomgående signifikant sammenheng mellom gruppestørrelse 2 og karakterer eller opplevd støtte fra lærer. Spesielt er det viktig å merke seg at jeg aldri, uansett metode eller gruppestørrelse, finner en signifikant negativ sammenheng mellom gruppestørrelse 2 og karakterer. Det vil si at jeg aldri finner positiv effekt av økt lærertetthet i ordinær undervisning på prestasjoner. Derimot finner jeg signifikant negativ effekt av økt lærertetthet i flere av analysene, spesielt for de med gruppestørrelse over medianen på 18,3 elever.

I dette delkapittelet vil jeg ta for meg og diskutere de fem viktigste delfunnene fra mine analyser, som belyser hvorvidt lærertetthet har betydning for elevenes prestasjoner. Det første funnet er at jeg, som nevnt, finner negativ effekt av lærertetthet på prestasjoner. Det andre er at jeg kun finner effekten for høy gjennomsnittlig gruppestørrelse. Det tredje funnet er at denne signifikante effekten i større grad er gjeldende på eksamen enn på standpunkt. Det fjerde at jeg, noe overraskende, finner størst negativ effekt av økt lærertetthet for elever i lavere sosioøkonomiske grupper, målt med *støtte hjemmefra* som proxy. Og til sist at jeg ikke finner noen gjennomgående ulikhet i effekt mellom gutter og jenter.

Ser vi resultatene samlet tyder funnene mine på at økt lærertetthet ikke bidrar til en gjennomgående forbedring i prestasjoner eller opplevd støtte fra lærer. Dette funnet vil jeg diskutere mer i detalj opp mot norsk og internasjonal forskning i 5.3 og 5.4.

5.2.1 Funn 1: Negativ effekt av økt lærertetthet på prestasjoner

Litt overraskende finner jeg at økt lærertetthet har en signifikant negativ effekt på elevenes prestasjoner. Det er en signifikant negativ effekt av økt lærertetthet for alle tre fagene ved skriftlig eksamen og muntlig eksamen i engelsk.

En mulig årsak til den negative sammenhengen kan være en «gruppedynamikkeffekt» som dominerer over lærereffekten. Med gruppedynamikkeffekt mener jeg at en får en bedre gruppedynamikk og positive synergier ved å få inn flere elever i gruppen, både faglig og sosialt. Eksempelvis vil helklassesamtale i mindre grupper trolig være mindre produktiv (Cobb et al., 1997; Stein et al., 2008), da det er færre elever, noe som kan gi færre innspill til diskusjoner, løsninger og variasjoner i strategier. Siden dette er en viktig del av effektiv matematikkundervisning, er dette uheldig for elevene.

I matematikkundervisningen vil det i en gruppe også være elever på ulike nivåer. Ved gjennomsnittlig lav gruppestørrelse i matematikkundervisningen vil trolig den kontinuerlige utviklingen av de sosiomatematiske normene (Kleve & Ånestad, 2016; Yackel & Cobb, 1996) skje mye saktere fordi interaksjonen i klasserommet ikke har like mange deltakere og elever på alle prestasjonsnivåer. En fordel med høyere gruppestørrelse er at flere av elevene på de ulike nivåene vil kunne hjelpe hverandre, og til dels vil lærere kunne ha flere «hjelpelærere» i elever som har forstått fagstoffet. En del elever responderer også bedre på å bli hjulpet av likesinnede, og lærer mer av dette (Hattie, 2009). Det å ha flere elever som presterer i det øvre sjiktet kan også trekke prestasjonene til de øvrige elevene oppover, ved at de fungerer som et referansepunkt på hvordan gode prestasjoner ser ut. Dersom en da får inn flere elever i gruppen kan det ha positive synergier for de øvrige elevene, og dermed bidra til økt prestasjonsnivå.

Ved tolkning av funnene i lys av dette er det viktig å huske at antall lærere ikke er den eneste årsaken til at gruppestørrelsen endres. Gruppestørrelse 2 beregnes litt forenklet med antall ordinære elevtimer «over brøkstreken» og antall ordinære lærertimer «under brøkstreken». Hvis gruppestørrelse 2 øker, kan det altså både være fordi en får flere elever på samme antall lærere, eller færre lærere på samme antall elever. Gruppedynamikkeffekten kan bare være en forklaring dersom det er førstnevnte som ligger bak, altså at endring i antall elevtimer er årsaken. Det er derfor viktig å ikke tolke en negativ sammenheng som at en økning i antall lærere negativt påvirker prestasjonsnivå. Den økte lærertettheten kan ha en nøytral eller til og

med positiv effekt, men gruppedynamikkeffekten kan ha en sterkere effekt på resultatene og dermed føre til en samlet negativ sammenheng.

En annen forklaring er at elever i større grupper utvikler større selvstendighet, og at mer selvstendige elever presterer bedre. Med høy lærertetthet, hvor læreren kommer bortom hyppig og gir veldig mye støtte, kan elever bli vant til å få denne støtten så snart de støter på utfordringer. Når gruppestørrelsen er større har læreren mindre anledning til det, men det kan føre til at elevene utvikler større selvstendighet til å finne løsninger selv, og økt utholdenhet til å ikke gi opp så raskt når de møter utfordringer. De utvikler det man kaller grit (Bettinger et al., 2018). Dette er egenskaper som er viktige når en skal prestere på egenhånd, som eksamen er et godt eksempel på. Dermed kan lærere, ved å følge elevene tett, ende med å gjøre dem en bjørnetjeneste og gjøre dem for avhengig av å bli leiet frem til svarene.

En relatert hypotese er den såkalte Topaze-effekten, eller traktkommunikasjon (Skott et al., 2008). Det går ut på at det lærere oppfatter som nyttig hjelp, i virkeligheten ofte fratras elevene utfordringene og reduserer utforskende problemløsningsoppgaver til rutineoppgaver. Dette kan lettere skje når gruppestørrelsen reduseres, siden læreren på samme måte som over får mer tid til hver elev. Dersom dette bidrar til at elever som får mer oppfølging fra lærer får mindre utfordrende oppgaver, vil det i neste omgang gjøre at de utvikler seg mindre og presterer dårligere. Både Topaze-effekten og selvstendighetseffekten er knyttet til hvordan læreren legger opp og gjennomfører undervisningen. Dette vil jeg diskutere nærmere i kapittel 5.5, hvor jeg drøfter konsekvensene av mine funn for matematikkundervisningen.

En annen mulig forklaring er at det kan være en omvendt kausalitet mellom prestasjoner og lærertetthet. Det kan oppstå dersom en skole eller klasse opplever en negativ utvikling i prestasjonsnivå og setter inn flere lærere som respons på dette, men at de settes inn for sent til å hindre at prestasjonene faller ytterligere. Dermed kan det se ut som resultatene utvikler seg negativt selv om lærertettheten økes. Liknende effekter er funnet i for eksempel Opheim et al. (2010). Det som taler mot dette som forklaring er at, dersom økt lærertetthet skulle hatt positiv effekt, burde vi observert en positiv effekt i prestasjoner i de påfølgende årene. Dette ser vi altså ikke ut fra dataene vi har.

En relatert forklaring kan være at det settes inn ekstra ressurser i en klasse som respons på egenskaper ved én eller et fåtall elever. Dermed vil gruppestørrelsen reduseres, men uten at

elevene i klassen i realiteten får nytte av den ekstra læreren, foruten den eller de tiltaket gjelder. Dermed kan også en nedgang i prestasjoner blant flertallet av elevene, som skyldes andre årsaker, tilskrives en økning i lærertetthet. Dette skal i prinsippet gruppestørrelse 2 fange opp, men det kan hende at i en del tilfeller vil ikke alle slike tiltak registreres i henhold til reglementet, for eksempel, dersom skoler opplever press på å holde gruppestørrelse 2 nede.

De tre førstnevnte effektene, gruppedynamikk, selvstendighet og traktkommunikasjon, er de mest intuitive forklaringskandidatene til den negative effekten av økt lærertetthet jeg finner. Det er samtidig mange solide matematikdidaktiske og pedagogiske argumenter for at vi burde forvente å finne motsatt effekt. Eksempelvis vil læreren med færre elever få mer tid til å veilede den enkelte, det er lettere å tilpasse undervisningen til hver elev og elever vil i større grad føle seg «sett» i en mindre gruppe (Kirkebøen et al., 2017). Uten å gå i detaljer på disse, er det intuitivt naturlig å tro at disse skulle bidra til en positiv effekt på prestasjoner av økt lærertetthet. Det er derfor behov for ytterligere forskning som kan bekrefte hvorvidt funnene mine avdekker en reell sammenheng, og hva eventuelle årsaker kan være.

5.2.2 Funns 2: Signifikant effekt kun for gruppestørrelser større enn medianen

Etter heterogenitetsanalyser gjennomført som følge av funnet i 5.2.1 finner jeg, når jeg deler opp observasjonene i to etter medianen for gruppestørrelse, ikke signifikante effekter av økt lærertetthet for undergruppen under medianen. Det er kun for undergruppen med gjennomsnittlig gruppestørrelser over medianen jeg finner signifikant negative effekter. Det er derfor denne undergruppen som er driveren for den negative effekten diskutert i 5.2.1. Konkret finner jeg for disse signifikant negativ effekt for muntlig eksamen i matematikk og veldig nærme en signifikant effekt på 5 %-nivå på skriftlig eksamen i matematikk.

En implikasjon dersom hypotesen om gruppedynamikkeffekt diskutert i funn 5.2.1 stemmer, er at en stor gruppe med to lærere vil gi bedre resultater enn å dele gruppen i to med én lærer hver. Dette gir også datasettet jeg har brukt noe støtte for, der elever på skoler med gruppestørrelse 2 over 18,3 har signifikant bedre karakterer på muntlig eksamen enn elever på skoler med gruppestørrelse under 18,3.

Selv om det fremdeles ikke er en positiv observert effekt for noen av undergruppene, kan en likevel tolke dette som at de med gjennomsnittlig lavere gruppestørrelse responderer mindre negativt på økt lærertetthet, relativt til de store. For å forklare dette kan det være nødvendig å

splitte effekter av lærertetthet i drivere som kan virke i ulike retninger. La oss anta at gruppedynamikkeffekten diskutert over er tilnærmet konstant for alle endringer i gruppestørrelser vi ser på. Da kan det være en annen driver som virker i motsatt retning, og som i større grad har effekt når gruppestørrelsen blir liten nok.

En mulig driver og årsak kan knyttes til måten en lærer underviser på ved ulike gruppestørrelser. Tidligere forskning har vist at undervisningsstil og undervisningsmetoder kan endre seg ved store reduksjoner i klassestørrelse (Breton, 2014; Hattie, 2009). Disse studiene har riktignok sett på klassestørrelsesreduksjon fra svært store klasser til mellomstore klasser (fra 50-80 og ned til 25-30 elever). Hattie (2009) har senere stilt spørsmål ved om ytterligere reduksjoner vil ha liknende effekt, siden endringene i hvilke pedagogiske verktøy som benyttes er mindre ved reduksjon fra for eksempel 25 til 15 elever. Samtidig kan det være slik at når gruppestørrelsen blir liten nok vil læreren ha nye muligheter til å legge opp undervisningen på andre måter. Siden det er rimelig å anta at lærere legger opp undervisningen for å få størst mulig læringsutbytte, vil økt fleksibilitet til å tilpasse opplegget kunne resultere i bedre resultater enn det en klarer i større grupper.

Resultatene mine kan gi støtte for at en slik effekt finnes, men at den kun gjør seg gjeldende når gruppestørrelsen blir liten nok. Nøyaktig hvor dette brytningspunktet ligger er vanskelig å si ut fra mine analyser. Men siden gruppestørrelser under 18,3 responderer relativt sett bedre på endringer i lærertetthet, ligger det trolig et sted under dette punktet. Dersom dette stemmer, betyr det at for å få en positiv effekt av økt lærertetthet må det skje ganske store endringer i gruppestørrelse, slik at lærerne får mulighet til å gjøre reelle endringer i undervisningsmetoder.

5.2.3 Funn 3: Effekter er i større grad signifikante på eksamenskarakterer enn standpunkt

De negative effektene av økt lærertetthet diskutert over finner jeg i større grad på eksamenskarakterer enn for standpunktkarakterer. Det vil si at karakterene læreren selv setter ikke påvirkes i samme grad som karakterene som settes av ekstern sensor. For å forklare hvorfor jeg får disse resultatene kan det igjen være nødvendig å bryte ned sammenhengene i underliggende drivere som kan virke i ulike retninger.

Dersom vi fremdeles antar at gruppedynamikkeffekten som beskrevet i kapittel 5.2.1 eksisterer, vil denne altså isolert sett føre til at en økning i gruppestørrelse gir økt

karakternivå. Når denne effekten er mer synlig på eksamen, må det være en annen effekt som går i motsatt retning som spiller inn på standpunkt karakterene, men som ikke påvirker eksamenskaractersettingen i like stor grad. En slik årsak kan være at standpunkt og eksamen tester elevene på noe ulike grunnlag. Eksamen er av natur mer spisset, og tester elevene på enten skriftlige eller muntlige ferdigheter innen et avgrenset tema. Standpunkt har en bredere vurderingsform, og påvirkes av elevenes samlede prestasjon over tid samt innsats og motivasjon i faget. Med høyere lærertetthet har lærerne i snitt mer tid per elev, og i utgangspunktet mer tid på å observere hver enkelt elev. Dermed har de mulighet til å gjøre en grundigere vurdering av hver elev.

Det kan også være at lærere når gruppestørrelsen er mindre vektlegger ting ulikt ved karactersetting enn ved høyere gruppestørrelser. Ved større gruppestørrelse blir det gjerne mindre tid til observering av muntlige og praktiske prestasjoner. Lærere i slike grupper kan derfor potensielt velge å basere karactersettingen i større grad på skriftlige prøver, som likner mer på eksamen.

Så langt har jeg antatt at lærerens standpunkt vurdering er rent objektiv. Men det kan også tenkes at lærere når de får mer tid med hver enkelt elev utvikler økt sympati, og ubevisst er snillere i vurderingen fordi han eller hun kjenner bedre til hvorfor en elev for eksempel sliter med noe. Dermed vil økt lærertetthet i gjennomsnitt gjøre at standpunkt vurderingen blir snillere, og tar høyde for ting en ekstern sensor ikke vil gjøre.

Hvilke av de tre foreslåtte årsakene som er reelle blir spekulasjoner fra min side. Men dersom mer tid per elev i snitt gjør at lærerne gir en bedre karakter, vil det utlikne gruppedynamikkeffekten, og dermed gjøre at vi ikke finner de samme resultatene på standpunkt som eksamen.

På den andre siden kan en tenke seg at det er naturlige grunner til at eksamens karakterer og standpunkt karakterer avviker litt. Standpunkt karakterer skal ifølge Opplæringslova (1998) være et mål på sluttkompetanse. Imidlertid er det trolig at lærere i en viss grad utviser skjønn i noen situasjoner, som for eksempel at en elev som gjennom alle prøvene et år har fått karakter 3 og på siste tentamen får 2, vil nok læreren likevel gi karakter 3. På samme måte som dersom en elev har fått karakter 5 på alle prøver, men får karakter 6 på siste prøve, vil kunne få karakter 6 i standpunkt. På denne måten blir vurdering av standpunkt karakterer muligens litt

kunstig høy. Samtidig skulle dette da gjelde for alle elever, og ikke bare for gruppestørrelser over 18,3 slik jeg finner. Dermed er forklaringene over mer sannsynlig enn denne.

5.2.4 Funn 4: Økt lærertetthet er mer negativt for elever med lav støtte hjemmefra

Zevenbergen (2001) og Huang (2009) har i tidligere studier funnet at hjemmemiljøet har innvirkning på elevers prestasjoner. Elever med foreldre med lav sosioøkonomisk status presterer statistisk sett dårligere enn elever med foreldre med høy sosioøkonomisk status. Elevers læring er altså tett knyttet til hjemmeforhold. Andre studier har videre funnet at det er en sammenheng mellom sosial bakgrunn og støtte hjemmefra (Bakken, Frøyland & Sletten, 2016; Nilsen & Bergem, 2016), hvor elever med foreldre med høy sosioøkonomisk status får mer støtte til skolearbeidet hjemme. Jeg testet derfor hvorvidt elever med ulik sosioøkonomisk bakgrunn og støtte hjemmefra får ulik effekt av endringer i lærertetthet.

I mangel på andre og mer spesifikke variabler, og med bakgrunn i funnene fra Bakken et al. (2016) og Nilsen og Bergem (2016) benyttet jeg indeksen *støtte hjemmefra* som proxy for sosioøkonomisk bakgrunn. Ut fra denne delte jeg opp skolene i to grupper, ut fra hvem som scorer høyt og lavt på denne indeksen. Deretter undersøkte jeg om hver av gruppene har signifikant effekt på prestasjoner eller opplevd støtte fra lærer av endringer i gruppestørrelse 2. Tanken bak en slik analyse er å teste en hypotese om at det er elever som opplever lite støtte hjemmefra som får størst effekt av mer støtte hos læreren, noe økt lærertetthet kan gi.

Resultatene viser derimot at det kun er for gruppen med lav støtte hjemmefra vi finner signifikant negativ effekt av økt lærertetthet. En reduksjon i gruppestørrelse slår altså mer negativt ut over denne elevgruppen enn de som har høyere grad av støtte hjemmefra.

Disse funnene er ikke i samsvar med annen norsk forskning på området. Innenfor klassestørrelsesforskningen finner Bonesrønning og Vaag Iversen (2008) og Vaag Iversen (2013) enkelte positive effekter for elever med lav sosioøkonomisk status av reduksjon i klassestørrelse på lave trinn, altså motsatt av mine funn. I en annen artikkel av Bonesrønning (2003) finner han at reduksjon i klassestørrelse er fordelaktig for elever med høy sosioøkonomisk status. Forfatterne i alle tre studiene konkluderer selv med at funnene er sprikende, og sår tvil rundt hva funnene vil si. Med andre ord finner majoriteten av annen norsk forskning ingen tydelig ulikhet i effekt for ulike elevgrupper.

Når det gjelder den internasjonale forskningen er det i hovedsak STAR-prosjektets forskning som fokuserer på undergrupper av elevene i den forskningen jeg har sett på. Her ser vi det samme som i den norske klassestørrelsesforskningen, at de finner sprikende resultater for hvilke elevgrupper som får fordeler eller ulemper av reduksjon i klassestørrelse. Generelt er det få funn som finner gjennomgående positive effekter for noen elevgrupper.

Det er intuitivt å tenke at økt lærertetthet skulle gi elevene med lite støtte hjemmefra størst positiv effekt. Når dette ikke stemmer må det være en annen mekanisme som motvirker det positive ved økt lærertetthet. En mulig forklaring kan igjen være knyttet til gruppedynamikkeffekten, og at denne er viktigere for elever med lav støtte hjemmefra. Det kan tenkes at disse elevene har et større behov for forbilder på skolen fordi de ikke har det hjemme, og at det er medelevene deres som fungerer som forbilder. I så fall kan funnene mine tyde på at medelever har mer å si for elevenes prestasjoner for de med lav støtte hjemmefra enn det læreren har.

Vi vet fra forskningen til Gee (2015) og Zevenbergen (2001) at elevene kommer inn i matematikkundervisningen med ulik førforståelse og ulik «kvalitet» på primærdiskursen. Noen elever er predisponert hjemmefra til å lettere kunne tilegne seg skolediskursen og matematikkdiskursen. Det kan tenkes at for elever med lav støtte hjemmefra er det å være del av en større gruppe elever fra ulike sosiale lag viktigere for å innlemmes i skolediskursen, fordi de får økt sosial stimuli med flere elever og har flere forbilder. De blir altså mer utfordret når gruppestørrelsen øker, og presterer derfor bedre.

Det er en stor utfordring for læreren å få alle med i matematikkdiskursen på grunn av de ulike førforståelsene elevene har med seg. Ifølge Gee (2015) er det lærere som må gjøre matematikkdiskursen konkret og tilgjengelig for alle elevene. Det er ikke slik at det en gjør innenfor en diskurs, er selvsagt for alle som er med i diskursen. Kleve og Penne (2016) finner at det er utenforskap i skolen. Noen kommer aldri inn i fagdiskursen i det hele tatt, og i deres studier snakker heller ikke disse elevene som om de er innenfor diskursen (Kleve & Penne, 2016). Læreren er uten tvil viktig for at elever med lav støtte hjemmefra skal innlemmes i matematikkdiskursen, men funnene mine tyder ikke på at det nødvendigvis er slik at flere lærere gjør denne diskursen klarere for elevene. Det å ha en elevgruppe med et vidt spenn av elever på alle nivåer, og fra alle samfunnslag, kan muligens være mer prestasjonsfremmende enn økt lærertetthet.

5.2.5 Funn 5: Ingen entydige ulikheter i effekt mellom jenter og gutter

Analysene mine rundt hvorvidt det er forskjell mellom gutter og jenter av endringer i lærertetthet for deres prestasjoner viser et sprikende bilde. I enkelte regresjoner er effekten kun signifikant for jenter, i andre kun for gutter, mens for de øvrige er det ingen signifikante effekter hverken for jenter eller gutter. Det sprikende bildet gjør det vanskelig å trekke ut noen entydige funn, og jeg har derfor valgt å være varsom med å tillegge dette særlig stor vekt.

Tall fra Utdanningsspeilet viser at jenter i gjennomsnitt får høyere karakterer enn gutter i alle fag, utenom kroppsøving. Dette gjelder også eksamenskarakterer (Statistisk sentralbyrå, 2017; Utdanningsdirektoratet, 2017b, 2018b). Det er dermed et kjent faktum at jentene drar fra guttene i nesten alle fag.

Dersom læreren eller gruppestørrelse hadde vært en viktig driver for at gutter og jenter presterer ulikt, skulle vi forventet å ha sett ulik effekt av endringer i gruppestørrelse 2 for gutter og jenter. Selv om mine funn er sprikende, er hovedtolkningen at det ikke er noe tydelig bilde på hvilket kjønn som påvirkes mest. Gruppestørrelse ser dermed ikke ut til å være en sentral forklaring på kjønnsforskjeller i prestasjoner. Dersom dette stemmer er det et viktig funn i arbeidet med å redusere kjønnsforskjeller i skolen, siden løsningen da ikke ligger i å øke lærertettheten.

5.3 Sammenlikning av funn mot annen norsk forskning

Utover å se på mulige årsaker til funnene mine, er det viktig og interessant å sammenlikne hvordan funnene står seg opp mot annen norsk forskning på dette temaet. En sentral del av denne diskusjonen handler om hvordan ulike valg av metode og analyseverktøy i varierende grad kan bidra til å belyse ulike deler av forskningsspørsmålet. Da annen norsk forskning har brukt ulike måter å måle lærertetthet på har jeg strukturert gjennomgangen slik at jeg først sammenlikner mot forskningen på lærertetthet med gruppestørrelse 2, deretter mot gruppestørrelse 1 og til slutt mot forskningen på redusert klassestørrelse.

Hovedfunnet med at jeg aldri finner signifikant positiv effekt av økt lærertetthet støtter opp under tidligere norsk forskning på lærertetthet. Den eneste norske studien som bruker

gruppestørrelse 2, Kirkebøen et al. (2017), finner ikke signifikante sammenhenger mellom gruppestørrelse 2 og prestasjoner. Det vil si at de ikke finner noen effekt av økt lærertetthet.

For gjennomsnittlig gruppestørrelse over medianen, finner jeg en positiv effekt av større gruppestørrelse på matematikkarakteren på eksamen. På grunn av et annet design på metoden og en studie over kortere tid har ikke Kirkebøen et al. (2017) mulighet til å dele opp utvalget i mindre eller større kohorter basert på gruppestørrelse. Jeg kan derfor ikke sammenlikne funnet mitt for gruppestørrelse over medianen direkte med deres funn. De påpeker selv at de ikke klarer å si om det er noen grupper som får negative effekter av økt lærertetthet, og om det er andre grupper som får positive effekter slik at effekten utjevnes. Årsaken til at de ikke kan si dette sikkert ligger i RD-designets oppbygning. I analysen tar de utgangspunkt i skoler som på starttidspunktet har nesten like gjennomsnittlige grunnskolepoeng og gruppestørrelser i ordinær undervisning. Sammenlikningsgruppen fikk ikke tilskuddet. De bruker de to faktorene gruppestørrelse i ordinær undervisning og elevenes prestasjoner. Etter innføring av tilskuddet kan en over tid se om disse skolene presterer forskjellig. På grunn av at sammenlikningen ligger rundt diskontinuitetspunktet (gjennomsnittlig gruppestørrelse før tilskuddet), kan de ikke dele opp dataene i underutvalg slik jeg kan, og dermed ser de heller ikke de samme negative effektene jeg finner for enkelte utvalg i dataene mine. Samtidig, siden de har et diskontinuitetspunkt, er det veldig sannsynlig at skoler rett over og rett under diskontinuitetspunktet er like på alle dimensjoner. Selv om fasteffektmetoden også kontrollerer for alt som er likt ved skolen, gjelder det kun det som er fast, men ikke det som varierer over tid. Dette gir et godt eksempel på at de to analysemetodene har ulike styrker og svakheter.

Forskningen som bruker gruppestørrelse 1 som variabel finner i hovedsak lite gjennomgående effekter av gruppestørrelse 1 på prestasjoner (Grøgaard et al., 2008; Opheim et al., 2010). Selv om funnene deres i stor grad samsvarer med mine funn, skal jeg videre drøfte hvorfor bruken av variabelen gruppestørrelse 1 i disse to studiene er problematisk.

Hovedutfordringen med å bruke gruppestørrelse 1 som utregningsvariabel er at den ikke er begrenset til ordinær undervisning, slik gruppestørrelse 2 er. Målet på lærertetthet ut fra gruppestørrelse 1 inneholder timer til spesialundervisning og timer til særskilt norskopplæring for språklige minoriteter. Dette fører til at tallet blir kunstig lavt, for eksempel fordi en har to elever som har spesialundervisning og hele klassen da deles på to ressurspersoner, selv om den ene gruppen egentlig kun har to elever, mens den andre gruppen har alle de andre

elevene. Selv om en forsøker å kontrollere for andel elever med spesialundervisning og enkeltvedtak, blir dette mer upresist enn å bruke gruppestørrelse 2 som variabel. Årsaken til dette er at en ikke med sikkerhet kan si om denne variabelen reflekterer egenskaper ved elevene. Disse egenskapene er vi nødt til, på en pålitelig måte, å utelukke for å kunne finne effekten av lærertetthet på elevers prestasjoner. Selv om beregningen av gruppestørrelse 2 heller ikke er perfekt, slik forskning sjelden er innenfor empirisk skoleforskning, er vi her nærmere å utelukke egenskaper ved elevene og dermed ha et mer nøyaktig mål for faktisk gruppestørrelse. Dette gjør at funnene fra studiene over er mer usikre enn om de hadde benyttet gruppestørrelse 2 som mål.

Grøgaard et al. (2008), som bruker gruppestørrelse 1 som mål på lærertetthet, finner signifikant negativ effekt for norskarakterer dersom en øker lærertettheten, men ikke i andre fag. Opheim et al. (2010) finner også klar negativ effekt av økt lærertetthet på prestasjoner, men kun fra den nasjonale prøven i matematikk på 5. trinn og ikke for lesing og engelsk. Dette likner en del på mine funn, hvor jeg i større grad finner resultater på eksamen enn på standpunktkarakterer. Der Grøgaard et al. (2008) tilskriver den negative effekten de finner til at en setter inn ekstra ressurser som en respons på egenskaper ved elevene som krever flere lærere og fører til prestasjonsproblemer, kan ikke jeg heller utelukke dette med 100 prosent sikkerhet. Da metoden min kontrollerer for skolefasteeffekter tar en i alle fall ut det som er ulikt mellom elever mellom skoler. I mine resultater er den signifikante negative effekten derfor kun tilfelle hvis en setter inn ekstra ressurser som en respons på egenskaper ved elevene som varierer over år innad på den samme skolen. Selv om det kan være noe variasjon, kommer elevene i de samme skolene fra de samme geografiske områdene hvert år. Mens det kan være tilfeldigheter som gir variasjon over år vil det over tid jevne seg ut og være omtrent den samme elevgruppen per kull. Det at jeg ser på effekter over mange år er derfor en styrke ved analysen. Grøgaard et al. (2008) kontrollerer også for enkeltvedtak og andel elever med spesialundervisning, men effekten holder. Dette kan tyde på at enten klarer ikke gruppestørrelse 1-variabelen å fjerne tilstrekkelig med effekt av egenskaper ved elevene, eller så er det andre årsaker til at vi finner effekter som de ikke finner. En styrke ved min forskning kontra disse studiene er, i tillegg til å se på gruppestørrelse 2, nettopp at studien er longitudinal slik at jeg kan bruke data over mange år.

Et annet element ved Opheim et al. (2010) og Grøgaard et al. (2008) sine funn fra nasjonale prøver er at dette er en eksternt vurdert prøve, slik eksamen også er. Utfordringen for begge

studiene er at datasettet deres går over så kort tid, Grøgaard et al. kun over ett år, mens Opheim et al. over 3 år. Dette gjør at enkeltfunn ofte ikke vektlegges, fordi en ikke kan se effekter over tid og sammenstille mot eksamen- og standpunkt karakter. For å undersøke om disse funnene egentlig kan tillegges mer vekt, etterspør jeg forskning som ser på nettopp nasjonale prøver over tid for å se om disse funnene kan tillegges mer vekt enn de gjør i deres rapporter. I norsk forskning på lærertetthet og klassestørrelse er det generelt stor mangel på longitudinale studier. Dette er en svakhet ved norsk forskning på lærertetthet, i tillegg til at det er for lite forskning på lærertetthet ved bruk av gruppestørrelse 2.

Også opp mot den norske forskningen på effekt av klassestørrelsesreduksjon på prestasjoner stemmer til en viss grad funnene mine overens med forskningen. Forskningen spriker og noen finner marginale positive effekter for visse elevgrupper, men majoriteten konkluderer med liten til ingen effekt. Det vil si at forskningen i Norge på både effekten av lærertetthet og reduksjon i klassestørrelse i stor grad er i samsvar og få finner nevneverdig positive effekter.

Funnet med negativ sammenheng for gjennomsnittlig gruppestørrelse over medianen er vanskelig å sammenlikne mot norsk klassestørrelsesreduksjonsforskning, da disse studiene i stor grad benytter andre metoder som ikke enkelt lar seg sammenlikne med min. Metoden de fleste bruker utnytter regelen om maksimum klassestørrelse på 30 elever. Dette gjør at en kan bruke RD-designet for å beregne kausale klassestørrelseseffekter som ligger nær diskontinuitetspunktene. Ideen er å sammenlikne klassestørrelser på 29 og 30 elever med klasser som har 31 og 32 elever og dermed blir delt i to, i 15-16 elever. Siden det er tilfeldig om en klasse har 30 eller 31 elever, vil alt annet i gjennomsnitt være likt for klasser på 30 og klasser på 15/16 elever bortsett fra stor forskjell i klassestørrelsen. De bruker da regelen som et instrument i regresjonen. Dette er metodisk en veldig bra metode som har mange styrker, men en begrensning er at en ikke kan dele opp utvalget slik jeg kan med fasteffektmetoden. Det er da ikke mulig å se på klasser med 25 og 10 elever istedenfor fordi rundt disse størrelsene er det ingen diskontinuitet. Dermed er funnet med negativ sammenheng for gjennomsnittlig gruppestørrelse over medianen vanskelig å sammenlikne mot norsk klassestørrelsesreduksjonsforskning.

5.4 Sammenlikning av funn mot internasjonal forskning

Til nå har jeg sett på mulige årsaker bak funnene jeg har gjort, og sammenliknet hvordan mine funn står seg opp mot annen norsk forskning. Videre vil jeg sammenlikne mine funn med den internasjonale forskningen på effekter av klassestørrelse. Som tidligere nevnt er det kun effekter av klassestørrelse som er forsket på internasjonalt, da lærertetthet er et «norsk fenomen». Resultatene mine står ikke like godt i samsvar med den internasjonale forskningen som den gjør med den norske. Jeg vil derfor nedenfor sammenlikne funnene mine opp mot den mest sentrale internasjonale forskningen på området, og diskutere hvorvidt funnene herfra er overførbare til norske forhold. Først vil jeg ta for meg STAR-prosjektet, deretter studier fra TIMSS, og til slutt, ved hjelp av øvrige studiers resultater, peke på mulige årsaker til hvorfor vi ikke finner de samme effektene av lærertetthet eller klassestørrelsesreduksjon på prestasjoner i Norge.

I sammenlikningen av mine funn mot internasjonal forskning, er det naturlig å starte med STAR-prosjektet, og belyse hvorfor dette ikke kan sammenliknes direkte med mine funn. STAR-prosjektet har, som beskrevet i kapittelet om empirisk forskning, vært sentralt i mye av den internasjonale forskningen på området. Resultatene fra forskning på STAR-prosjektet er sprikende, men enkelte studier tyder på at det er noen positive effekter av mindre klasser, spesielt for enkelte elevgrupper. Det er likevel en del forhold som gjør at resultatene ikke er direkte overførbare til mine forhold. For det første er elevene i studien mye yngre enn elevene jeg har i min studie. Behovene til elever i ungdomsskolen kan være annerledes enn behovene elever i barneskolen har, og hvordan reduserte klassestørrelser eller økt lærertetthet bidrar til å møte disse behovene kan være ulike. På grunn av den kognitive forskjellen og forskjellen i det faglige nivået, vil måten undervisningen utføres på også være forskjellig. Dette kan gjøre at størrelsen på klassen eller antall lærere i mindre grad er relevant for hvor godt elevene lærer og presterer. For det andre ble STAR-prosjektet gjennomført som en intervensjonsstudie, mens jeg har studert de naturlige variasjonene i gruppestørrelse uten intervensjon. Når lærere og elever vet de deltar i et eksperiment kan de endre atferd fra hvordan de ellers ville agert, noe som kan ha skjedd i STAR-prosjektet. Dette vil ikke være tilfellet i min analyse, siden det ikke har vært et kjent eksperiment for deltakerne. For det tredje vil det være forskjeller både i skolesystem og kulturelle aspekter mellom Tennessee, hvor STAR-prosjektet ble utført, og Norge. Ulike måter å legge opp undervisningen på, og ulik grad av involvering av foreldrene, kan være eksempler på variabler som potensielt kan ha innvirkning på hvordan økt lærertetthet påvirker prestasjonsnivå. Med bakgrunn i disse ulikhetene bør en være varsom

med å sammenlikne funnene fra STAR-prosjektet med norske forhold. Resultatene fra min studie vil i så måte være mer direkte relevante for vurderinger av effektene for norsk skole.

Forskningen på TIMSS-dataene er nærmere forholdene i min oppgave fordi de fleste av dem ser på klassestørrelser for eldre elever. For det andre er dette data fra en regelmessig prøve, og ikke intervensjonseffekten en undersøker i disse studiene. Det gjør datagrunnlaget mer likt i utgangspunktet. Min forskning er i stor grad samsvarende med resultatene fra TIMSS-forskningen. Funnene derfra viser at effekten av klassestørrelse varierer i stor grad fra skolesystem til skolesystem, men at det i sammenliknbare land til Norge i liten grad finnes signifikante effekter av klassestørrelse på prestasjoner.

Selv om min studie og forskningen på TIMSS-data i liten grad finner signifikante og betydelige effekter, så er det flere andre internasjonale studier som gjør det. Dette gjelder også studier fra liknende land, som for eksempel Sverige (Fredriksson et al., 2013), som finner positive langsiktige effekter av mindre klasser. Et interessant spørsmål er da hva som kan være årsakene til at en finner effekter i andre land som ikke kommer til syne i Norge. Under vil jeg ved hjelp av andre studier peke på mulige årsaker til hvorfor vi ikke finner positiv effekt av lærertetthet eller klassestørrelsesreduksjon på prestasjoner i Norge.

En viktig årsak til at min og majoriteten av norsk forskning ikke finner positiv effekt av klassestørrelsesreduksjon eller endringer i lærertetthet er at vi i Norge har relativt høy lærerkvalitet. Høy lærerkvalitet pekes på som en viktig forklaringsvariabel i internasjonale studier for hvorfor en ikke finner effekt av klassestørrelse. Både Hattie (2009), Mueller (2013) og Wößmann og West (2006) peker på at det i all hovedsak er i land som anses å ha lav kvalitet på lærerstaben en finner store effekter av å redusere klassestørrelsen. Dette finner vi også igjen i flere artikler som ser på TIMSS-resultater i ulike land og på tvers av landene (Breton, 2014; Shen & Konstantopoulos, 2019). I tråd med forskningen gir det derfor mening at Norge i stor grad ikke finner gjennomgående positive effekter av hverken klassestørrelsesreduksjon eller økt lærertetthet. Fra regjeringen er det etter 2014 satt i gang flere tiltak for å bedre kvaliteten på både eksisterende og kommende lærere (Kunnskapsdepartementet, 2014). Ett eksempel er at alle lærere i kjernefagene matematikk, norsk og engelsk skal ha undervisningskompetanse i faget de underviser. Dette har ført til at en rekke lærere må etterutdannes for å få godkjenning til å undervise videre i disse fagene. For kommende lærere har grunnskolelærerutdanningen blitt 5-årig med integrert master fra

2017, samt at opptakskravet i matematikk har økt fra karakter 3 til 4 for å komme inn på lærerutdanningene. Dette, sammen med satsing på videreutdanning for lærere, kan ifølge forskningen være gode tiltak med tanke på å forbedre kvaliteten i skolen. Et aspekt som bør nevnes her, er at min studie ser på resultater fra 2009 og frem til i dag. Eventuelle effekter på økt lærerkvalitet fra regjeringens tiltak vil derfor kun delvis være reflektert i datasettet. Dersom hypotesen er at høy lærerkvalitet reduserer effekten av økt lærertetthet, og jeg allerede fra 2009 ikke finner effekt av økt lærertetthet, vil det bety at lærerkvaliteten i Norge allerede før regjeringens tiltak var god nok til at økt lærertetthet ikke gir signifikant positiv effekt.

Underveis har jeg problematisert hvorvidt funn fra internasjonale studier på effekter av klassestørrelse er overførbare til norske forhold. For det første varierer skolesystemer fra land til land, for det andre har ulike land kulturelle forskjeller. Disse to elementene sammen gjør at vi ikke kan direkte sammenlikne Norge med alle land. Vi kan, med forbehold i de to elementene, forsøke å trekke sammenhenger og funn ut fra studier som ser på liknende land som Norge. Det vil si land med relativ lik populasjon, sosioøkonomisk status, skolesystem og kultur. Dette er en av grunnene til at skandinaviske land ofte sammenliknes. Det er derfor interessant at enkelte studier fra Sverige, som nevnt i Fredriksson et al. (2013) over, finner positive effekter. Av forklaringene jeg har trukket frem over er det ikke betydelige forskjeller mellom Norge og Sverige på hverken lærerkvalitet, eksisterende lærertetthet eller hvor store klassestørrelsesreduksjoner en har studert. Det må derfor trolig være andre årsaker til at resultater fra Sverige er motstridende med mine og annen norsk forskning sine funn, enten kulturelle eller metodiske. Overordnet kan vi likevel oppsummere med at direkte overføring av tiltak og resultater fra ett land til et annet er usikkert, og bør unngås. Derfor er det viktig at det gjøres metodisk solide undersøkelser i det enkelte land.

5.5 Konsekvenser for politikkutforming og matematikkundervisning

Videre vil jeg drøfte konsekvensene av funnene mine, dersom de kan generaliseres til å gjelde for norsk skole. I hovedsak vil jeg ta utgangspunkt i den overordnede konklusjonen om at jeg i liten grad finner signifikante effekter av lærertetthet på hverken prestasjoner eller indeksen *støtte fra lærer*, og spesielt at effektene aldri er positive. Jeg vil drøfte konsekvensene på to nivåer: først konsekvenser for politikkutforming og organisering av

matematikkundervisningen, deretter for måten en som lærer driver matematikkundervisning på.

Fra høsten 2019 ble siste fase i implementeringen av lærernormen gjennomført. Lærernormen er en reform som stiller krav til at det ikke skal være flere enn 15 elever i gjennomsnitt per lærer i ordinær undervisning på 1.-4. trinn, og ikke flere enn 20 elever på 5.-10. trinn. Ifølge Opplæringslova (1998) har skolene en viss grad av autonomi for å gjennomføre dette på best mulig måte. Hensiktene med å satse på økt lærertetthet i ordinær undervisning er å bedre den tilpassede opplæringen og styrke elevenes grunnleggende ferdigheter. Dette skal igjen gi elevene bedre læringsutbytte, og dermed bedre prestasjoner (NOU 2019:23, 2019).

Det er vesentlig å stille seg spørsmålet om hva et slikt tiltak koster og hva som er den samfunnsøkonomiske nytten. Et lærerårsverk i 2019 er anslått til å koste gjennomsnittlig 745 000 kr per år (Kunnskapsdepartementet, 2018a). Fra lærernormen ble vedtatt og til full innføring høsten 2019 krevde dette ansettelse av 2 800 nye årsverk (Kunnskapsdepartementet, 2018b). Dersom en legger disse forutsetningene til grunn, vil dette føre til en kostnad på 2 milliarder kroner årlig. Etter gjennomlesing av statsbudsjett og veiledninger fra Kunnskapsdepartementet opereres det imidlertid ikke med helt klare tall på hva kostnadene ved lærernormen er. Fra deres side anslås det til å være mellom 1-1,5 milliarder kroner per år, litt avhengig av hvilken kilde en leser. Uansett om det er 1 eller 2 milliarder per år, er dette en betydelig årlig kostnad, og spørsmålet er hvorvidt den samfunnsøkonomiske nytten av tiltaket er stor nok, dersom det gir liten eller ingen effekt på matematikkompetansen til elevene. I en sammenstilling av en lang rekke internasjonale studier finner Hattie at *reduksjon i klassestørrelse* gjennomsnittlig har en så liten effekt som 0,13 standardavvik på elevenes prestasjoner (Hattie, 2005). Denne vil trolig være enda lavere i Norge, da svært få studier finner positive effekter. Til sammenlikning finner samme studie for eksempel at i gjennomsnitt gir tiltak for *tidlig intervensjon* effekt på hele 0,64 standardavvik på prestasjoner og *kvalitet på undervisning* gir effekt på 0,67 standardavvik på prestasjoner (Hattie, 2005). Studien fra Hattie er en sammenstilling av en rekke analyser på hva som påvirker elevens prestasjoner, hvor de ulike funnene er kategorisert i gruppene nevnt over, med flere. De nøyaktige størrelsene på effektene og hvordan hver kategori av tiltak er definert i Hatties studie er ikke det viktigste her. Det interessante er å observere at flere andre kategorier av tiltak potensielt kan ha vel så stor effekt som reduksjon i klassestørrelse. En viktig diskusjon

bør derfor være hvorvidt det finnes andre, mer kostnadseffektive tiltak vi burde satset på i skolen, fremfor økt lærertetthet.

Her er det også relevant å trekke inn funnet mitt om at elever med lav støtte hjemmefra ser ut til å ha større negativ effekt av redusert gruppestørrelse enn elever med høy støtte. Om det stemmer vil redusert gruppestørrelse faktisk bidra til å øke ulikheter mellom sosiale grupper. Sosiale ulikheter i skolen har stått på den politiske agendaen i årtier. Fellesskolen har vært et prinsipp i norsk og skandinavisk skole helt siden 1930-tallet. Det ble utviklet en demokratisk modell for utdanning som var svært inspirert av Deweys tanker (Telhaug, Mediås & Aasen, 2006). Med rapporten «For store forventninger? Kunnskapsløftet og ulikhetene i grunnskolekarakterer» kommer Bakken og Elstad (2012) med en evaluering av LK06. Målet i innføringen av LK06 var at Kunnskapsløftet skulle bidra til å utjevne sosiale forskjeller av prestasjoner, altså gjøre betydningen av familiebakgrunn mindre for elevenes prestasjoner. Funnene til Bakken og Elstad etter den kvantitative evalueringen er at læreplanen ikke har lyktes i dette, på noen områder er forskjellene blitt større (Bakken & Elstad, 2012). Til tross for stor satsning på fellesskolen og å redusere sosiale ulikheter, øker ulikhetene i skolen.

Mine funn tyder på at medelevene er viktigere for en elevs prestasjon enn det læreren er, og for elever med lav støtte hjemmefra synes dette å være enda viktige. Dersom dette stemmer er det å sikre demografisk utjevning, med god spredning på skolene på tvers av sosiale lag, viktigere for å sikre gode prestasjoner hos alle enn det økt lærertetthet vil gi. I den politiske diskusjonen bør derfor tiltak for å utjevne sosiodemografiske ulikheter mellom skoler vurderes på linje med investeringer i tiltak som lærernormen. Om ikke risikerer en å forsterke eksisterende ulikheter.

God kvalitet på lærere er et annet tiltak det virker som myndighetene anser som viktig for å bedre prestasjoner, i og med at de satser på å øke kompetansen på lærere, som også diskutert i 5.4. Begrunnelsen bak dette er at kompetente lærere har høyere kvalitet på undervisningen. Et sentralt spørsmål er da hvorvidt det også hadde vært mer nyttig å bruke pengene som er brukt til lærernormen til å øke kompetansen til lærere ytterligere, eller øke lønn til lærere, slik at flere av de dyktigste i samfunnet rekrutteres til lærerutdanning, og dermed øker kompetansen på den måten.

En annen bekymring er at lærernormen fører til økning i ukvalifiserte eller uerfarne lærere. I Norge var gjennomsnittlig erfaring i lærerstaben på ca. 15,5 år i 2014 (Kunnskapsdepartementet, 2014), og det er forventninger om at dette tallet vil synke som følge av lærernormen. Årsaken til at en økning i andel uerfarne lærere som følge av lærernormen kan være problematisk, er at en del internasjonale studier kun finner positiv effekt av økt lærertetthet for erfarne lærere. Mueller (2013) finner for eksempel klassestørrelseseffekt for erfarne lærere i små grupper, men ikke i store klasser, samt ingen effekter for uerfarne lærere for noen av gruppene. Dette tolkes dithen at erfarne lærerne trolig klarer omstillingen av lærerstil til små klasser i en viss grad, mens uerfarne lærere har samme lærerstil uansett og dermed ikke henter ut noen gevinst av å ha færre elever å holde styr på.

Dersom lærernormen fører til at en får inn mange uerfarne lærere, og funnet til Mueller er overførbare til norske forhold, vil en ikke få noen effekt av satsingen. Om de uerfarne lærerne i tillegg er mindre kompetente, kan det faktisk føre til dårligere resultater for alle elevene ved at de i snitt får «dårligere» lærere. Dersom en økt satsing på antall lærere skal ha en positiv effekt, må det derfor komme klarere retningslinjer for hvordan slik økt lærertetthet skal utnyttes, og at en i parallell sikrer at de nye lærerne som ansettes har tilstrekkelig god kompetanse.

Videre skal jeg se på konsekvenser av funnene mine for hvordan lærere gjennomfører selve matematikkundervisningen, og hva de potensielt kan gjøre for at økt lærertetthet skal gi en positiv effekt.

Først vil jeg ta utgangspunkt i de mulige forklaringene til at mine funn viser negative effekter av økt lærertetthet, spesielt for gruppestørrelser over medianen, som jeg drøftet i kapittel 5.2.1. Flere av disse var knyttet til hvordan læreren bruker den ekstra tiden per elev på en måte som virker mot sin hensikt. Ved at læreren er «for hjelpsom» og gir elevene for tett rettleiding hindrer han ubevisst elevene i å utvikle selvstendighet og utholdenhet til å løse utfordrende oppgaver. På samme måte kan mer hyppig kommunikasjon fra læreren forsterke den såkalte Topaze-effekten, hvor læreren bidrar til å redusere utforskende problemløsningsoppgaver til rutineoppgaver. Begge disse effektene forårsakes av lærerens gode intensjoner, men læreren er ikke tilstrekkelig bevisst på hvordan måten den rettleider elevene på påvirker dem. Dette er som lærer viktig å være klar over uansett. Men dersom

lærertettheten i norsk skole øker fremover, vil kjennskap til og bevissthet rundt dette bli enda viktigere.

Gruppedynamikkeffekten, som ble løftet som en annen viktig forklaring til en negativ sammenheng mellom lærertetthet og prestasjoner, kan også nøytraliseres gjennom måten lærere velger å legge opp undervisningen. Dersom større grupper presterer bedre, kan en i stedet for å dele klasser i to med én lærer hver heller beholde klassen samlet og ha to lærere. For eksempel kan én ha ansvaret for plenumsgjennomganger, mens den andre går rundt og hjelper elever. Da vil en potensielt både kunne hente ut de positive synergiene det synes å gi av å ha en stor gruppe, og gi elevene tettere og mer tilpasset oppfølging. Vi vet fortsatt lite om effekten av å faktisk utnytte potensialet i å være to lærere i undervisningen.

Forskningsprosjektet 1+1 ser nettopp på hvordan en utnytter det å være to lærere, så forhåpentligvis kommer det i kjølvannet av prosjektet flere anbefalinger for hvordan en kan utnytte økte ressurser på best mulig måte (Forskningsrådet, 2015a).

Momentene jeg har trukket frem over adresserer utfordringene ved at økt lærertetthet eller reduserte gjennomsnittlige gruppestørrelser fører til endringer i undervisning som kan gi utilsiktede negative effekter. Men reduksjon i gruppestørrelse kan også muliggjøre endringer i undervisningsmetode som isolert sett har positive virkninger. Mine analyser og drøftinger i kapittel 5.2.2 tyder på at for at dette skal skje må gruppestørrelsene være tilstrekkelig små, siden jeg ikke finner tilsvarende effekter ved gruppestørrelse over medianen på 18,3 elever. Her er det viktig å presisere at jeg i analysene ikke har funnet en samlet positiv effekt av økt lærertetthet for noen grupper, men at skolene under medianen for gruppestørrelse responderer mindre negativt på økt lærertetthet enn de over medianen.

En mulig årsak til at vi kun ser denne effekten på de minste gruppestørrelsene er at det da blir tilstrekkelig åpenbart for læreren at undervisningsmetoder kan endres, mens det i større grupper er mer styrt av kulturelle forhold. I en stor gruppe kan det også være at læreren ser mange utfordringer ved å endre undervisningsmetoder, eksempelvis knyttet til organisering, og dermed ikke endrer nevneverdig på sin undervisning. Det å endre en undervisningskultur i et land er en svært omfattende prosess som er kulturelt betinget, og noe som hele landet samlet må enes rundt for at endringen skal skje (Stigler & Hiebert, 2009). Det å bare øke lærertettheten, men gi få retningslinjer til hva ressursene skal brukes til, er lite prestasjonsfremmende med mindre endringen i gruppestørrelse er tilstrekkelig stor. Vi må

utvikle en mer langsiktig og kontinuerlig forbedringsmodell som vil øke kvaliteten på undervisningen og som faktisk utnytter ressursene økt lærertetthet gir, også for mindre endringer. Et viktig element, som også Stigler og Hiebert (2009) påpeker, er at lærerne må være inkludert i en slik prosess. De må føle eierskap. Dersom Norge fortsatt skal satse på økt lærertetthet kan det være nyttig å arbeide for å utvikle slike felles retningslinjer og en langsiktig kontinuerlig prosess for hvordan en utnytter økt lærertetthet på best mulig måte.

Som drøftet over viser internasjonale studier kun positiv effekt av økt lærertetthet for klasser med erfarne lærere, og at det kun er disse lærerne som i særlig grad tilpasser undervisningen når klassestørrelsen endres (Mueller, 2013). Dersom dette er tilfelle også i Norge blir det viktig å klare å utnytte disse lærernes erfaring og kompetanse på en effektiv måte. Et konkret tiltak kan derfor være at uerfarne lærere får tettere oppfølging på hvordan en skal legge opp undervisning etter størrelsen på klassene. Dette kan erfarne lærere på den enkelte skole lettere gjøre allerede nå, i parallell med at det utvikles mer overordnede retningslinjer og modeller for utnyttelse av økt lærertetthet.

6 Oppsummering

I min oppgave har jeg sett på sammenhengen mellom økt lærertetthet og elevenes prestasjoner i matematikk. Datagrunnlaget tar utgangspunkt i resultatdata fra Skoleporten på skolenivå for elever som fullfører 10. trinn i årene 2009-2019, samt lærertetthet målt etter gruppestørrelse 2. Gjennom OLS- og fasteffektregresjoner har jeg analysert effekter av endringer i gruppestørrelse 2 på karakterer på standpunkt og eksamen, samt elevenes score på indeksen *støtte fra lærer* i Elevundersøkelsen. Analysene har avdekket fem funn, som sammen besvarer forskningsspørsmålet i oppgaven «*Hva har økt lærertetthet å si for elevenes prestasjoner i matematikk?*».

Ser vi de fem funnene samlet, er det ingen entydige effekter av endringer i gruppestørrelse for noen elevgrupper. Spesielt er det viktig å merke seg at jeg aldri, uansett metode eller størrelse på gruppene, finner positiv effekt av økt lærertetthet. Det vil si at økt lærertetthet ikke ser ut til å ha en positiv effekt på prestasjoner i ungdomsskolen. Dette er i tråd med hva tidligere norsk forskning har funnet. Ser vi på funnene hver for seg har jeg likevel avdekket en del potensielle effekter av endringer i lærertetthet.

Mitt første funn er at jeg i flere analyser har funnet en signifikant negativ effekt av økt lærertetthet på karakterer. I analysen har jeg pekt på flere mulige årsaksforklaringer til hvorfor dette oppstår, hvor de tre mest intuitive er gruppedynamikkeffekt, grad av selvstendighet og traktkommunikasjon. Med gruppedynamikkeffekt mener jeg at en får positive synergier ved å få inn flere elever i gruppen, både faglig og sosialt. Denne effekten kan dominere over en eventuell positiv lærereffekt. Dersom denne hypotesen stemmer vil elever prestere bedre om det er to lærere i samme klasse, fremfor å dele gruppen i to med én lærer hver. En annen forklaring er at elever i grupper med lav lærertetthet utvikler større selvstendighet og utholdenhet fordi de må klare seg uten like tett oppfølging fra lærer. Den tredje forklaringen er at læreren ubevisst utøver traktkommunikasjon i undervisningen, og at dette blir mer tydelig når gruppestørrelsen reduseres. Ved for høy lærertetthet kan en av konsekvensene være at elevene får «for mye hjelp» og for hyppig, og dermed ikke utvikler tilstrekkelig kompetanse til problemløsning og evne til å tenke kreativt for å finne løsninger. Dette kan også bidra til at de presterer dårligere.

Mitt andre funn er at driveren av effekten i det første funnet er skoler med høy gjennomsnittlig gruppestørrelse. Heterogenitetsanalyser av undergrupper viser at det kun er for undergruppen med gjennomsnittlig gruppestørrelse over medianen på 18,3 at den signifikant negative effekten holder. Dette kan tyde på det er ulike drivere innad i variabelen gruppestørrelse 2 som virker i ulike retninger. Dersom gruppedynamikkeffekten beskrevet over er konstant for alle endringer i gruppestørrelse, har jeg diskutert at det trolig finnes en mekanisme som virker i motsatt retning, og som er sterkere når gruppestørrelsen er lav. En slik mekanisme kan være at lærere, når gruppestørrelsen blir lav nok, er i stand til å endre undervisningsmetoder slik at det isolert sett får en positiv effekt på elevenes prestasjoner. Dersom dette stemmer, betyr det at for å få en positiv effekt av økt lærertetthet må det skje ganske store endringer i gruppestørrelse, slik at lærerne får mulighet til å gjøre reelle endringer i undervisningsmetoder.

Det tredje funnet er at den negative effekten jeg finner for skoler med gruppestørrelse over medianen, kun gjelder på eksamenskarakterer og ikke standpunkt. Det vil si at karakterene læreren selv setter ikke påvirkes i samme grad som karakterene som settes av ekstern sensor. En mulig årsak kan være at eksamen og standpunkt måler ulike aspekter, hvor standpunkt også tar inn momenter som kun kan observeres over lengre tid. Det kan også være at lærere vektlegger ting ulikt ved karaktersetning når gruppestørrelse endres. Ved høy gruppestørrelse kan lærere potensielt velge å basere karaktersetningen i større grad på skriftlige prøver, som likner mer på eksamensformen for vurdering. En tredje årsak kan være at lærere, når de får mer tid med hver elev, ubevisst blir mer subjektiv og positiv i karaktersetningen på standpunkt. Disse tre årsakene vil isolert sett gi en positiv sammenheng mellom lærertetthet og standpunkt, og dermed nøytralisere en eventuell negativ sammenheng som funnet på eksamen.

Det fjerde funnet er at reduksjoner i gruppestørrelse slår mer negativt ut for elever med lav støtte hjemmefra. *Støtte hjemmefra* er her brukt som en proxy for sosioøkonomisk bakgrunn, som tidligere studier har vist henger tett sammen. Om dette stemmer bidrar faktisk økt lærertetthet til å øke forskjellene mellom sosioøkonomiske grupper. En forklaring på funnet mitt kan igjen være knyttet til gruppedynamikkeffekten, og at denne er viktigere for elever med lite støtte hjemmefra. Disse har et større behov for forbilder på skolen og trenger mer støtte for å innlemmes i matematikkdiskursen, siden de ikke får dette like naturlig inn fra hjemmet. Når jeg finner at elever med lav støtte responderer dårligere på redusert

gruppestørrelse kan det tyde på at medelevene er viktigere forbilder og bidragsytere til å innlemme de i matematikdiskursen enn læreren. I så fall kan det å ha en elevgruppe med et vidt spenn av elever på alle nivåer, og fra alle samfunnslag, være mer prestasjonsfremmende enn økt lærertetthet.

Det siste funnet er at jeg finner ingen entydige effektforskjeller av endringer i gruppestørrelse mellom jenter og gutter. Lærertetthet ser dermed ikke ut til å være en forklaring på kjønnsforskjeller i prestasjoner. Dersom dette stemmer er det et viktig funn i arbeidet med å redusere kjønnsforskjeller i skolen, siden løsningen da ikke ligger i å øke lærertettheten.

Ut fra resultatene har jeg konkludert med at det ikke er noen signifikante positive effekter på elevers prestasjoner av økt lærertetthet. Derimot kan det for gruppestørrelser over medianen, og elever med lav støtte hjemmefra, være signifikante negative effekter av økt lærertetthet på prestasjoner på eksamen. Videre har jeg pekt på hva konsekvensene av disse funnene kan være for politikktutforming og organiseringen av matematikkundervisningen, dersom funnene kan generaliseres til å gjelde for norsk skole.

En implikasjon av lærernormens økte lærertetthet er at vi bruker store summer på et tiltak som ikke resulterer i økte prestasjoner. En viktig samfunnsdiskusjon bør derfor være hvorvidt det finnes andre, mer kostnadseffektive tiltak vi burde satset på i skolen, fremfor økt lærertetthet. Dette kan for eksempel være enda mer satsning på økt kompetanse hos lærere, som deler av gjennomgått forskning peker på som et viktig premiss for å få effekt av økt lærertetthet, utover den direkte effekten økt kompetanse i seg selv gir.

Det bør også diskuteres hvorvidt tiltak for sosial utjevning bør prioriteres fremfor økt lærertetthet. Når mine funn antyder at økt lærertetthet kan bidra til å øke prestasjonsforskjeller mellom ulike sosiale grupper, kan det å sikre elevgrupper med et vidt spenn av elever på alle nivåer, og fra alle samfunnslag, være mer prestasjonsfremmende enn økt lærertetthet.

Flere av de mulige årsakene bak den manglende positive effekten av lærertetthet handler om hva læreren gjør i klasserommet. Dersom disse stemmer, bør det også få konsekvenser for hvordan matematikklærere legger opp sin undervisning, slik at de kan bidra til positive gevinster av lærernormen. For det første bør lærere være bevisste på å ikke utøve traktkommunikasjon eller gi elevene for tett oppfølging til at de utvikler selvstendighet og

utholdenhet. Dette blir spesielt viktig med økende lærertetthet, hvor læreren får mer tid til hver elev.

Videre bør de sørge for å utnytte den økte tiden per elev til å tilpasse undervisningen. Som antydnet i funn 2 kan være at lærere, når gruppestørrelsen blir lav nok, er i stand til å endre undervisningsmetoder tilstrekkelig slik at en isolert sett får en positiv effekt på elevenes prestasjoner. Samtidig bør dette skje ved å ha flere lærere i en felles gruppe heller enn å dele grupper i to med én lærer hver, for å sikre at man ikke mister de positive gruppedynamikkeffektene.

Et forhold det er viktig å presisere er at forskningen min i stor grad har fokusert på faglige resultater. Skolen har også andre formål som ikke kan måles i karakterer, og som kan påvirkes av endringer i lærertetthet, både i positiv og negativ retning. Dette har ikke vært fokus i min analyse.

Det er enkelte metodiske svakheter ved min analyse. I hovedsak ligger dette i datagrunnlaget, da jeg helst skulle hatt data på individnivå og elevdata om sosioøkonomisk bakgrunn. Dermed kunne jeg enda bedre kontrollert for om de sammenhengene jeg finner er kausale.

Fasteffektmetoden løser problemet med mangel på sosioøkonomisk bakgrunn til en viss grad, men faktiske data ville gjort resultatene enda mer valide og reliable. I tillegg har gruppedynamikkeffekten vært en viktig gjennomgående forklaring i diskusjonen av årsaker bak funnene mine. Dersom jeg hadde hatt data på klassestørrelse, og ikke bare gruppestørrelse, kunne jeg brukt dette som ekstra kontrollvariabel for å isolere ut gruppedynamikkeffekten. Da ville resterende effekt i variabelen gruppestørrelse 2 i større grad skyldes forhold ved læreren, og jeg ville i større grad kunne si hvorvidt lærernormen vil bidra til å øke prestasjonsnivået i skolen.

Endringen som trådte i kraft høsten 2019 med maksimum gjennomsnittlig gruppestørrelse vil i fremtiden bidra til videre forskning på området. Dette gjør det mulig å bruke RD-design i nye analyser, fordi en da kan sammenlikne gruppestørrelser rundt det maksimale punktet. Til nå er det ikke tilstrekkelig med data tilgjengelig etter endringen til å kunne gjøre en slik analyse, men om noen år vil dette være en mulighet.

7 Referanseliste

- Allison, P. D. (2009). *Fixed effects regression models*. Los Angeles: SAGE.
- Altinok, N. & Kingdon, G. (2012). New Evidence on Class Size Effects: A Pupil Fixed Effects Approach. *Oxford Bulletin of Economics and Statistics*, 74(2), 203-234. doi: 10.1111/j.1468-0084.2011.00648.x
- Bakken, A. & Elstad, J. I. (2012). *For store forventninger?: kunnskapsløftet og ulikhetene i grunnskolekarakterer* (NOVA-rapport, Bind nr. 7/12). Oslo: Norsk institutt for forskning om oppvekst, velferd og aldring.
- Bakken, A., Frøyland, L. R. & Sletten, M. A. (2016). *Sosiale forskjeller i unges liv. Hva sier ungdomsundersøkelsene?* Oslo: Norsk institutt for forskning om oppvekst, velferd og aldring.
- Bettinger, E., Ludvigsen, S., Rege, M., Solli, I. F. & Yeager, D. (2018). Increasing perseverance in math: Evidence from a field experiment in Norway. *Journal of Economic Behavior and Organization*, 146, 1-15. doi: 10.1016/j.jebo.2017.11.032
- Birkemo, A. (2002). *Læringsmiljø og utvikling*. Oslo: Universitetet i Oslo, Pedagogisk forskningsinstitutt i samarbeid med Unipub.
- Bonesrønning, H. (2003). Class Size Effects on Student Achievement in Norway: Patterns and Explanations. *Southern Economic Journal*, 69(4), 952-965. doi: 10.2307/1061660
- Bonesrønning, H. & Vaag Iversen, J. M. (2008). *Suksessfaktorer i grunnskolen: Analyse av nasjonale prøver 2007* (SØF-rapport 05/08). Trondheim: Senter for økonomisk forskning AS.
- Bonesrønning, H. & Vaag Iversen, J. M. (2010). *Prestasjonsforskjeller mellom skoler og kommuner: Analyse av nasjonale prøver 2008* (SØF-rapport 01/10). Trondheim: Senter for økonomisk forskning AS.
- Breton, T. R. (2014). Evidence that class size matters in 4th grade mathematics: An analysis of TIMSS 2007 data for Colombia. *International Journal of Educational Development*, 34(1), 51-57. doi: 10.1016/j.ijedudev.2013.04.003
- Brousseau, G. (2002). *Theory of Didactical Situations in Mathematics : Didactique des Mathématiques, 1970–1990*. Dordrecht: Springer Netherlands.
- Chingos, M. M. (2012). The impact of a universal class-size reduction policy: Evidence from Florida's statewide mandate. *Economics of Education Review*, 31(5), 543-562. doi: 10.1016/j.econedurev.2012.03.002

- Christoffersen, L. & Johannessen, A. (2012). *Forskningsmetode for lærerutdanningene*. Oslo: Abstrakt forlag.
- Christophersen, K.-A. (2013). *Introduksjon til statistisk analyse: regresjonsbaserte metoder og anvendelser*. Oslo: Gyldendal akademisk.
- Cobb, P., Boufi, A., McClain, K. & Whitenack, J. (1997). Reflective discourse and collective reflection. *Journal for Research in Mathematics Education*, 28(3), 258. doi: 10.2307/749781
- Ding, W. & Lehrer, S. F. (2011). Experimental estimates of the impacts of class size on test scores: robustness and heterogeneity. *Education Economics*, 19(3), 229-252. doi: 10.1080/09645292.2011.589142
- Duckworth, A. L. & Quinn, P. D. (2009). Development and Validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment*, 91(2), 166-174. doi: 10.1080/00223890802634290
- Falch, T., Sandsør, A. M. J. & Strøm, B. (2017). Do Smaller Classes Always Improve Students' Long-run Outcomes? *Oxford Bulletin of Economics and Statistics*, 79(5), 654-688. doi: 10.1111/obes.12161
- Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R. & Empson, S. B. (1996). A Longitudinal Study of Learning to Use Children's Thinking in Mathematics Instruction. *Journal for Research in Mathematics Education*, 27(4), 403-434. doi: 10.2307/749875
- Finckel, D. E. (2007). Linear Panel Analysis. I S. W. Menard (Red.), *Handbook of Longitudinal Research: Design, Measurement, and Analysis* (s. 475-504). Amsterdam: Elsevier.
- Finn, J. (2002). Class size reduction in grades K-3. I A. Molnar (Red.), *School reform proposals: The reseach evidence*. Greenwich, CT: Information Age Publishing.
- Finn, J. & Achilles, C. M. (1990). Answers and Questions about Class Size: A Statewide Experiment. *American Educational Research Journal*, 27(3), 557. doi: 10.3102/00028312027003557
- Forskningsrådet. (2015a). Small group Instruction in Mathematics for Pupils Level 1-4: Effects of a randomized controlled trial intervention study. Hentet 10. februar 2020 fra <https://prosjektbanken.forskningsradet.no/#/project/NFR/256217>
- Forskningsrådet. (2015b). Two Teachers in the Class: Increasing the Opportunities to Differentiate Literacy instruction. Hentet 10. februar 2020 fra <https://prosjektbanken.forskningsradet.no/#/project/NFR/256197>

- Forskningsrådet. (2020). LÆREEFFEKT - Lærertetthet og læringseffekt. Hentet 10. februar 2020 fra <https://www.forskningsradet.no/om-forskningsradet/programmer/lareeffekt/>
- Fredriksson, P., Öckert, B. & Oosterbeek, H. (2013). Long-Term Effects of Class Size. *Quarterly Journal Of Economics*, 128(1), 249-285. doi: 10.1093/qje/qjs048
- Gee, J. P. (2015). *Social Linguistics and Literacies: Ideology in Discourses*. New York: Routledge
- Ghuri, P. N. & Grønhaug, K. (2005). *Research methods in business studies: a practical guide*. Harlow: Financial Times Prentice Hall.
- Grøgaard, J. B., Helleand, H. & Lauglo, J. (2008). *Elevenes læringsutbytte: Hvor stor betydning har skolen? En analyse av ulikhet i elevenes prestasjonsnivå i fjerde, syvende og tiende trinn i grunnskolen og i grunnkurset i videregående (45/08)*. Oslo: NIFU STEP Norsk institutt for studier av innovasjon, forskning og utdanning.
- Hanushek, E. A. (1999). Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects. *Educational Evaluation and Policy Analysis*, 21(2), 143-163. doi: 10.3102/01623737021002143
- Hanushek, E. A. & Wößmann, L. (2017). School Resources and Student Achievement: A Review of Cross-Country Economic Research. I M. Rosén, K. Yang Hansen & U. Wolff (Red.), *Cognitive Abilities and Educational Outcomes*. Cham: Springer International Publishing.
- Hattie, J. (2005). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research*, 43(6), 387-425. doi: 10.1016/j.ijer.2006.07.002
- Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Huang, L. (2009). Social Capital and Student Achievement in Norwegian Secondary Schools. *Learning and Individual Differences*, 19(2), 320-325. doi: 10.1016/j.lindif.2008.11.004
- Johannessen, A., Christoffersen, L. & Tufte, P. A. (2011). *Forskningsmetode for økonomisk-administrative fag*. Oslo: Abstrakt forlag.
- Kirkebøen, L. J., Kotsadam, A., Raaum, O., Andresen, S. & Rogstad, J. (2017). *Effekter av satsing på økt lærertetthet (Rapporter nr. 39/2017)*. Hentet fra https://www.ssb.no/utdanning/artikler-og-publikasjoner/_attachment/332335?_ts=16044ba4410

- Kleve, B. & Penne, S. (2016). Learning subjects in School - being outsiders or insiders in the disciplinary discourse of mathematics and Language 1. *International Journal of Educational Research*, 78, 41-49.
- Kleve, B. & Ånestad, G. (2016). Læringspartner og sosiomatematiske normer som potensial for elevers læring. I B. Kleve & E. K. Hovik (Red.), *Undervisningskunnskap i matematikk* (s. 31-45). Oslo: Cappelen Damm akademisk.
- Konstantopoulos, S. (2011). How Consistent Are Class Size Effects? *Evaluation Review*, 35(1), 71-92. doi: 10.1177/0193841X11399847
- Konstantopoulos, S. & Li, W. (2012). Modeling Class Size Effects Across the Achievement Distribution. *International Journal of Sociology of Education*, 1(1), 5-26. doi: 10.4471/rise.2012.01
- Konstantopoulos, S. & Shen, T. (2016). Class size effects on mathematics achievement in Cyprus: evidence from TIMSS. *Educational Research and Evaluation*, 22(1), 86-109. doi: 10.1080/13803611.2016.1193030
- Konstantopoulos, S. & Sun, M. (2014). Are teacher effects larger in small classes? *School Effectiveness and School Improvement*, 25(3), 312-328. doi: 10.1080/09243453.2013.808233
- Kunnskapsdepartementet. (2004). *Kultur for læring*. (Meld. St. 030 (2003-2004)). Hentet fra <https://www.regjeringen.no/no/dokumenter/stmeld-nr-030-2003-2004-/id404433/>
- Kunnskapsdepartementet. (2014). *Lærerløftet - På lag for kunnskapsskolen*. Kunnskapsdepartementet. Hentet fra https://www.regjeringen.no/globalassets/upload/kd/vedlegg/planer/kd_strategiskole_w eb.pdf
- Kunnskapsdepartementet. (2018a). Budsjettspørsmål - lærerårsverk. Hentet 2. april 2020 fra <https://www.statsbudsjettet.no/Statsbudsjettet-2019/Budsjettsporsmal/Bevilgningssporsmal/Rdt55/?query=1%C3%A6rer%C3%A5rsverk&part=1&sort=date&Ball=true#hopp89416>
- Kunnskapsdepartementet. (2018b). Budsjettspørsmål - ressursnorm i grunnskolen. Hentet 16. april 2020 fra <https://www.statsbudsjettet.no/Statsbudsjettet-2018/Budsjettsporsmal/Bevilgningssporsmal/Sosialistisk-venstreparti115/?query=norm&part=1&sort=date&all=true#hopp86888>
- Kunnskapsdepartementet. (2015). Lærertetthet - en kartlegging. Hentet 10. november 2019 fra <http://docplayer.me/27746420-Laerertetthet-en-kartlegging-kunnskapssenter-for-utdanning-lysaker.html>

- Leuven, E. & Løkken, S. (2020). Long term impacts of class size in compulsory school. *Journal of Human Resources*, 55(1), 309-348.
- Leuven, E., Oosterbeek, H. & Rønning, M. (2008). Quasi-experimental Estimates of the Effect of Class Size on Achievement in Norway. *Scandinavian Journal of Economics*, 110(4), 663-693. doi: 10.1111/j.1467-9442.2008.00556.x
- Li, W. & Konstantopoulos, S. (2016). Class Size Effects on Fourth-Grade Mathematics Achievement: Evidence From TIMSS 2011. *Journal of Research on Educational Effectiveness*, 9(4), 503-530. doi: 10.1080/19345747.2015.1105893
- Milesi, C. & Gamoran, A. (2006). Effects of Class Size and Instruction on Kindergarten Achievement. *Educational Evaluation and Policy Analysis*, 28(4), 287-313.
- Mueller, S. (2013). Teacher experience and the class size effect - Experimental evidence. *Journal of Public Economics*, 98(C), 44-52. doi: 10.1016/j.jpubeco.2012.12.001
- Nathan, M. J. & Knuth, E. J. (2003). A Study of Whole Classroom Mathematical Discourse and Teacher Change. *Cognition and Instruction*, 21(2), 175-207. doi: 10.1207/S1532690XCI2102_03
- NIFU. (2019). Evaluering av norm for lærertetthet. Hentet 10. februar 2019 fra <https://www.nifu.no/projects/evaluering-av-norm-for-laerertetthet-21051/>
- Nilsen, T. & Bergem, O. K. (2016). Hjemmebakgrunn. I O. K. Bergem, H. Kaarstein & T. Nilsen (Red.), *Vi kan lykkes i realfag*. Oslo: Universitetsforlaget.
- NOU 2019:23. (2019). *Ny opplæringslov*. Oslo: Kunnskapsdepartementet.
- OECD. (2019). *Education at a Glance 2019: OECD Indicators*. Paris: OECD Publishing.
- Opheim, V., Grøgaard, J. B. & Næss, T. (2010). *De gamle er eldst? Betydning av skoleressurser, undervisningsformer og læringsmiljø for elevenes prestasjoner på 5., 8. og 10. trinn i grunnsopplæringen* (nr. 34/2010). Oslo: NIFU Step Norsk institutt for studier av innovasjon, forskning og utdanning.
- Opplæringslova. (1998). *Lov om grunnskolen og den vidaregåande opplæringa*. Hentet fra https://lovdata.no/dokument/NL/lov/1998-07-17-61#KAPITTEL_9
- Schanzenbach, D. W. (2006). What Have Researchers Learned from Project STAR? *Brookings Papers on Education Policy*, 2006/2007(9), 205-228.
- Shen, T. & Konstantopoulos, S. (2019). Estimating causal effects of class size in secondary education: evidence from TIMSS. *Research Papers in Education*. doi: 10.1080/02671522.2019.1697733

- Shin, Y. (2012). Do Black Children Benefit More From Small Classes? Multivariate Instrumental Variable Estimators With Ignorable Missing Data. *Journal of Educational and Behavioral Statistics*, 37(4), 543-574.
- Skott, J., Hansen, H. C. & Jess, K. (2008). *Delta: fagdidaktik*. Frederiksberg: Forlaget Samfundslitteratur.
- Solheim, O. J. & Opheim, V. (2019). Beyond class size reduction: Towards more flexible ways of implementing a reduced pupil–teacher ratio. *International Journal of Educational Research*, 96, 146-153. doi: 10.1016/j.ijer.2018.10.008
- Statistisk sentralbyrå. (2017). Karakterer ved avsluttet grunnskole. Hentet 15. november 2019 fra <https://www.ssb.no/utdanning/statistikker/kargrs>
- Stein, M. K., Engle, R. A., Smith, M. S. & Hughes, E. K. (2008). Orchestrating Productive Mathematical Discussions: Five Practices for Helping Teachers Move Beyond Show and Tell. *Mathematical Thinking and Learning*, 10(4), 313-340. doi: 10.1080/10986060802229675
- Stigler, J. W. & Hiebert, J. (2009). *The teaching gap: best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.
- Stylianides, A. (2019). Secondary students' proof constructions in mathematics: The role of written versus oral mode of argument representation. *Review of Education*, 7(1), 183-184. doi: 10.1002/rev3.3158
- Telhaug, A. O., Mediås, O. A. & Aasen, P. (2006). The Nordic Model in Education: Education as part of the political system in the last 50 years. *Scandinavian Journal of Educational Research*, 50(3), 245-283. doi: 10.1080/00313830600743274
- Thrane, C. (2017). *Regresjonsanalyse: en praktisk tilnærming*. Oslo: Cappelen Damm akademisk.
- Universitetet i Oslo. (2019, 14. oktober). Om TIMSS. Hentet 17. mars 2020 fra <https://www.uv.uio.no/ils/forskning/prosjekt-sider/timss-norge/TIMSS/om-timss/index.html>
- Utdanningsdirektoratet. (2017a). Grunnskolens Informasjonssystem. Hentet 6. november 2019 fra <https://gsi.udir.no/app/#!/view/units/collectionset/1/collection/77/unit/1/>
- Utdanningsdirektoratet. (2017b). Hovedfunn i karakterstatistikken for grunnskolen 2016-17. Hentet 5. november 2019 fra <https://www.udir.no/tall-og-forskning/finn-forskning/tema/karakterer/karakterer-i-grunnskolen-2016-17/>
- Utdanningsdirektoratet. (2017c). Hva vet vi om lærertetthet? Hentet 13. januar 2020 fra <https://tallogforskning.udir.no/innhold/hva-vet-vi-om-laerertetthet/>

- Utdanningsdirektoratet. (2018a). Elevundersøkelsen. Hentet 5. november 2019 fra <https://www.udir.no/tall-og-forskning/brukerundersokelser/elevundersokelsen/>
- Utdanningsdirektoratet. (2018b). Utdanningsspeilet Hentet 06. november 2019 fra <http://utdanningsspeilet.udir.no/2017/innhold/del-5/5-2-grunnskolepoeng-og-karakterer-etter-10-trinn/>
- Utdanningsdirektoratet. (2019). Skoleporten. Hentet 1. november 2019 fra <https://skoleporten.udir.no/>
- Vaag Iversen, J. M. (2013). *Educational governance: interactions between institutions and school resources*, (Doktoravhandling, Norges teknisk-naturvitenskapelige universitet). Hentet fra <https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/267456>
- Vaag Iversen, J. M. & Bonesrønning, H. (2013). Disadvantaged students in the early grades: will smaller classes help them? *Education Economics*, 21(4), 305-324. doi: 10.1080/09645292.2011.623380
- Wilde, E. T., Finn, J., Johnson, G. & Muennig, P. (2011). The Effect of Class Size in Grades K-3 on Adult Earnings, Employment, and Disability Status: Evidence from a Multi-center Randomized Controlled Trial. *Journal of Health Care for the Poor and Underserved*, 22(4), 1424-1435.
- Wooldridge, J. M. (2016). *Introductory econometrics: a modern approach*. Australia: Cengage Learning.
- Wößmann, L. & West, M. (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS. *European Economic Review*, 50(3), 695-736. doi: 10.1016/j.euroecorev.2004.11.005
- Yackel, E. & Cobb, P. (1996). Sociomathematical Norms, Argumentation, and Autonomy in Mathematics. 27(4), 458-477. doi: 10.2307/749877
- Zevenbergen, R. (2001). Language, social class and underachievement in school mathematics. I P. Gates (Red.), *Issues in mathematics teaching*. London: Routledge Falmer.

8 Vedlegg

8.1 Veiledning fra Utdanningsdirektoratet for hvordan gruppestørrelse 2 kalkuleres

Forklaring på beregning av gruppestørrelse 2

Stortinget har i forbindelse med behandlingen av statsbudsjettet for 2018 gjort følgende vedtak om lærernorm, jf. Innst. 2 S (2017-2018):

Stortinget ber regjeringen innføre en norm for lærertetthet på skolenivå. Målet er at det høsten 2018 skal være 1 lærer per 16 elever i 1.–4. klasse og 1 lærer per 21 elever i 5.–10. klasse, og fra høsten 2019 er målet at det skal være 1 lærer per 15 elever i 1.–4. klasse og 1 lærer per 20 elever i 5.–10. klasse. Normen skal evalueres underveis og sees i sammenheng med tiltak for å rekruttere et tilstrekkelig antall kvalifiserte lærere. Dagens lærerutdanning, herunder gjeldende opptakskrav, og regjeringens kompetansekrav skal legges til grunn. Det foretas en kvalitetssikring av kostnader knyttet til oppdaterte GSI-tall på skolenivå, slik at ingen kommuner skal tape på innføringen av normen. En justering av de samlede kostnadene legges inn i regjeringens forslag til revidert nasjonalbudsjett for 2018. Videre skal det utredes hvordan innfasingen av en norm kan gjennomføres uten fare for forsterket lærermangel i deler av landet.

Det fremgår av vedtaket i Stortinget at lærertetthetsnormen skal være på skolenivå.

Lærertetthetsnormen skal være en norm for forholdstallet mellom lærere og elever i ordinær undervisning (gruppestørrelse 2). Gruppestørrelse 2 er en indikator som viser antall elever per lærer i ordinær undervisning, hvor estimerte ressurser til spesialundervisning og undervisning i særskilt norsk ikke regnes med. Normen sier bare noe om forholdet mellom antall elever og lærere i en gjennomsnittlig undervisningssituasjon.

Formelen for gruppestørrelse 2 er definert som en brøk med følgende størrelser i teller og nevner:

$$\frac{\text{Elevtimer i ordinær undervisning} - \text{timetall spesialundervisning} - \text{timetall særskilt norsk}}{\text{Ordinære undervisningstimer} + \text{timer til oppdeling til samiske språkalternativer og annen målform}}$$

De enkelte størrelsene i uttrykket er definert på følgende måte:

Elevtimer i ordinær undervisning (hentes i ramme A på GSI-skjemaet):

antall elever 1. årstrinn * skolens timetallsfordeling 1. årstrinn + antall elever 2. årstrinn * skolens timetallsfordeling 2. årstrinn + antall elever 3. årstrinn * skolens timetallsfordeling 3. årstrinn osv.

Timetall spesialundervisning (hentes i ramme D på GSI-skjemaet):

antall elever med **1 - 75** timer spes. undervisning * **57** (gj.snitt antall timer) +
antall elever med **76-190** timer spes. undervisning * **134** (gj.snitt antall timer) +
antall elever med **191-270** timer spes. undervisning * **222** (gj .snitt antall timer) +
antall elever med **mer enn 270** timer spes. undervisning * **513** (gj.snitt antall timer)

Timetall særskilt norsk (hentes i ramme E i GSI-skjema):

(antall elever med særskilt norskopplæring 1. årstrinn + antall elever med særskilt norskopplæring 2. årstrinn + antall elever med særskilt norskopplæring 3. årstrinn + antall elever med særskilt norskopplæring 4. årstrinn)*(931/4) + (antall elever med særskilt norskopplæring 5. årstrinn + antall elever med særskilt norskopplæring 6. årstrinn + antall elever med særskilt norskopplæring 7. årstrinn)*(441/3) + (antall elever med særskilt norskopplæring 8. årstrinn + antall elever med særskilt norskopplæring 9. årstrinn + antall elever med særskilt norskopplæring 10. årstrinn)*(398/3)

Verdiene for særskilt norsk er fremkommet ved å ta det totale timetallet i norsk for de tre hovedtrinnene (931 timer for 1.-4. årstrinn, 441 timer for 5.-7. årstrinn og 398 timer for 8.-10 årstrinn) og dele på antall trinn.

Ordinære undervisningstimer (hentes fra ramme B i GSI-skjema): B18

Timer til oppdeling til samiske språkalternativer og annen målform (hentes fra ramme B i GSI-skjema): B9, B10, B11

Disse feltene inneholder lærertimer til opplæring i og på samisk språk, og lærertimer til elever med rett til annen målform enn kommunens hovedmål.

Gruppestørrelse 2 på 1.-4., 5.-7., og 8.-10. årstrinn

I GSI samles ikke timefordelingen av vedtak inn per årstrinn. Dermed må man bruke andelen elever med vedtak om spesialundervisning på de ulike hovedtrinnene til å estimere totalt antall timer til spesialundervisning på henholdsvis 1.-4., 5.-7., og 8.-10. årstrinn. Dette er nødvendig for å kunne beregne gruppestørrelse 2 for hovedtrinnene.

Beregning av undervisningsårsverk

Behovet for undervisningsårsverk som vises er et estimat, der antallet ekstra årstimer til undervisning som er nødvendig for å oppfylle normen gjøres om til årsverk basert på timetallet i årsrammen til undervisning (741 for 1.-4. trinn og 5.-7. trinn, 656 for 8.-10. trinn).

I kalkulatoren finner man først antall årstimer til undervisning som er nødvendig for at skolen skal oppfylle norm til lærertetthet. Deretter deler man dette årstimetallet på enten 741 for barnetrinnet, eller 656 for ungdomstrinnet og finner dermed et estimert undervisningsårsverksbehov.

Eksempel: En skole trenger 1000 ekstra årstimer til undervisning på 1.-4. trinn for å oppfylle lærernormen. I kalkulatoren beregner vi at skolen har behov for 1,34 ekstra årsverk ($1000/741$).

8.2 Svaralternativene på indeksen støtte hjemmefra og støtte fra lærer i elevundersøkelsen

Indeksen: Støtte fra lærere

Indeksen består av følgende fem indikatorer:

- 1) Opplever du at lærerne dine bryr seg om deg?
- 2) Opplever du at lærerne dine har tro på at du kan gjøre det bra på skolen?
- 3) Opplever du at lærerne behandler deg med respekt?
- 4) Når jeg har problemer med å forstå arbeidsoppgaver på skolen, får jeg god hjelp av lærerne

Svaralternativ: Alle – De fleste – Noen få – Bare en – Ingen

- 5) Lærerne hjelper meg slik at jeg forstår det jeg skal lære

Svaralternativ: Helt uenig – Litt uenig – Verken enig eller uenig – Litt enig – Helt enig

Indeksen: støtte hjemmefra

Indeksen består av følgende tre indikatorer:

- 1) Hjemme viser de interesse for det jeg gjør på skolen

Svaralternativ: Svært ofte eller alltid – Ofte – Av og til – Sjelden - Aldri

- 2) Jeg får god hjelp til leksene mine hjemme
- 3) Hjemme oppmuntrer de voksne meg i skolearbeidet

Svaralternativ: Alltid – Ofte – Noen ganger – Sjelden - Aldri