

Effekter av satsing på økt lærertetthet

Lars J. Kirkebøen, Andreas Kotsadam
og Oddbjørn Raam

Høsten 2012 bevilget Stortinget 1,5 milliarder kroner over fire år til 600 nye lærerstillinger på ungdomstrinnet. Satsingen førte til lavere gjennomsnittlig gruppestørrelse i ordinær undervisning, men vi finner så langt ingen tegn til at elevenes læringsutbytte økte. Resultatene er foreløpige. Videre analyser og data for nye elevkull vil gi sikrere konklusjoner og en mer uttømmende beskrivelse av hva satsningen har ført til.

Innledning

Hensikten med satsingen var å øke lærertettheten i ordinær undervisning. Dette skulle gjøre skolene bedre stilt til å kunne gi tilpasset opplæring og å styrke de grunnleggende ferdighetene. Målet var at elevene skal oppleve økt mestring og bedre læringsutbytte, og at midlene skulle bidra til bedre læringsmiljø og minke behovet for spesialundervisning. Ressursene gikk til en klart definert gruppe på 166 skoler, som hadde relativt lav lærertetthet og relativt svake gjennomsnittlige karakterer. Tiltaket er beskrevet i mer detalj i Boks 1.

I denne artikkelen viser vi de første resultatene fra en pågående evaluering¹ av de ekstra lærerressursene. Målet med evalueringen er å finne ut hvorvidt eller på hvilken måte dette har påvirket skolenes ressursbruk, elevenes læringsutbytte og læringsmiljø. I denne artikkelen forsøker vi å svare på i hvilken grad lærertetthet i ordinær undervisning er økt, og i hvilken grad dette har hatt betydning for elevenes læringsutbytte.

Det finnes en omfattende litteratur som studerer sammenhengen mellom lærerressurser og elevenes resultater. Det er imidlertid liten enighet om effekter, det er f.eks. slående at Hanushek (2003) og Krueger (2003) konkluderer svært ulikt på grunnlag av den samme litteraturen, og det finnes flere studier som analyser STAR² som kommer til ulike konklusjoner. Ettersom ekstra lærere er en kostbar satsning bør effekten være betydelig. Mye forskning tyder imidlertid på at effektene av ekstra lærere ofte er beskjeden, også

norsk forskning (Leuven m.fl., 2008). Men det finnes også forskning, blant annet fra Sverige, som tyder på at klassestørrelse kan ha en betydelig effekt, også på lang sikt, og være kostnadseffektiv (Fredriksson m.fl., 2013).

Utdanningsdirektoratet (2016) har vist at skolene som fikk økt tilskudd, sammenliknet med 2011/12, (i) hadde ansatt omtrent så mange flere lærere som tilskuddet tilsier, (ii) reduserte gruppestørrelsen, (iii) hadde redusert andel elever med spesialundervisning og (iv) hadde økt antall grunnskolepoeng. Øvrige ungdomsskoler/ungdomstrinn, som ikke fikk tilskudd, har til sammenligning hatt et omtrent stabilt antall lærere, hatt en marginalt svakere nedgang i spesialundervisning, og en noe svakere økning i gjennomsnittlige grunnskolepoeng.

Evalueringsdesignet

Formålet med effektanalysen er å studere samlede kausale effekter av tiltaket, dvs. hvordan faktiske resultater er forskjellig fra hvordan de ville ha vært i fravær av økt tilskudd (kontrafaktiske resultater).³ I denne artikkelen er fokus på (i) skolenes ressursbruk, målt ved gruppestørrelse i ordinær undervisning, og (ii) elevenes læringsutbytte, i hovedsak målt ved skriftlig eksamenskarakter i 10. klasse.

Vi er først og fremst interessert i om tiltaket har ført til høyere læringsutbytte, som målt ved skriftlig eksamenskarakter. Tiltaket består imidlertid i å gi ressurser til skolene, via skoleeierne. For at elevenes læringsutbytte skal påvirkes må det skje en atferdsendring på skolene. Fra Utdanningsdirektoratet (2016) vet vi at skolene som mottok tilskuddet har ansatt flere lærere, men årsverksveksten må også sees i lys av elevtallsutviklingen. Utdanningsdirektoratets oppfølging

Lars J. Kirkebøen er seniorforsker ved Gruppe for offentlig økonomi og befolkningsmodeller (kir@ssb.no)

Andreas Kotsadam er seniorforsker ved Frischsenteret (andreas.kotsadam@frisch.uio.no)

Oddbjørn Raam er seniorforsker ved Frischsenteret (oddbjorn.raam@frisch.uio.no)

¹ Evalueringen er utlyst og finansiert av Utdanningsdirektoratet, og utføres av SSB, Frischsenteret og Fafo.

² Et omfattende eksperiment med mindre klasser, gjennomført i 1985-1989 i Tennessee, USA.

³ Dette er en forståelse av effekter i tråd med hva om gjerne omtales som «potential outcomes» (se Angrist og Pischke (2008) for en god beskrivelse): Vi tenker oss at det for hver skole og elev er to forskjellige mulige utfall hhv. med og uten tiltaket, og ønsker å finne forskjellen på disse. I denne samlede effekten inngår respons fra andre aktører på økt tilskudd, herunder skoleeiere, enkeltskoler, foreldre (jf. Todd og Wolpin, 2003).

Boks 1: Om ekstra lærerstillinger i ungdomsskolen

Høsten 2012 ble det vedtatt å bevilge 1,5 milliarder kroner over en fireårsperiode til 600 nye lærerstillinger i ungdomsskolen. Ordningen med økt lærertetthet på ungdomstrinnet startet skoleåret 2013/14 og løper ut skoleåret 2016/17. Målgruppen er elever på ungdomstrinnet ved skoler som skoleåret 2011-2012 hadde en gjennomsnittlig gruppestørrelse på over 20 elever per lærer i ordinær undervisning på ungdomstrinnet, og grunnskolepoeng under snittet for landet (se boks 2). Til sammen ble 166 skoler tildelt øremerkede midler til 573 årsverk. Skolene er fordelt på 98 kommuner, og alle fylkene er representert.

De ekstra ressursene skal gå til å øke antall lærere med inntil fem årsverk for hver skole (ett årsverk for skoler med 0-99 elever, to for 100-199 elever osv.) ut over utgangsnivået for skoleåret 2011/2012. Årsverkene skal benyttes til ordinær undervisning, og til lærere med godkjent utdanning. Midlene kan ikke brukes til spesialundervisning og særskilt norsk. Ut over dette er det opp til kommunene og skolene selv å avgjøre hvordan undervisningen skal organiseres.

Som en oppfølging av tiltaket har Utdanningsdirektoratet jevnlig rapportert til Kunnskapsdepartementet, se f.eks. Utdanningsdirektoratet (2016).

har vært mindre klar på i hvilken grad gruppestørrelse i ordinær undervisning er endret, dette vil være fokus for analysene av ressurser i denne artikkelen. Det er også relevant å se nærmere på om skolene/skoleeierne på grunn av ekstratilskuddet har gjort andre endringer i ressursbruken, sammenlignet med hva skolene ellers ville gjort, f.eks. flere eller mer kompetente lærere eller endring i omfang eller bruk av annet personale (assistenter). Slike analyser vil vi komme tilbake til i framtidig rapportering fra evalueringen.

Når det gjelder læringsutbytte, er skriftlig eksamenskarakter et relevant mål på dette: Denne måler elevenes prestasjoner ved slutten av ungdomsskolen opp mot læreplanen. Skriftlig eksamenskarakter er imidlertid ikke noe uttømmende mål på læringsutbytte og vil ikke fange opp alle mulige relevante effekter. Den store fordelen med skriftlig eksamen er at denne vurderes anonymt av lærere utenfor skolene, slik at vi unngår problemer med at tiltaket påvirker setting av standpunktkarakterer. Dette kan f.eks. skje dersom det ansettes nye lærere med en annen karakterpraksis. Forskjeller i karakterpraksis kan være betydelige, i hvert fall i forhold til forskjeller i bidrag til læring mellom skoler og til effektene det er rimelig å vente av tiltaket (jf. Galloway m.fl., 2011; Hægeland m.fl., 2011).

Vi kan ikke direkte observere den kontrafaktiske ressursbruken eller det kontrafaktiske læringsutbyttet, dvs. slik det ville vært i fravær av tiltaket. Dette er vi nødt til å anslå ved hjelp av en sammenligningsgruppe, fortrinnsvis "like skoler" som ikke fikk tilskudd. Skolene som ikke fikk ekstra ressurser er generelt ikke en relevant sammenligningsgruppe, særlig ettersom

Boks 2: Datagrunnlag og viktige variable

Vi bruker tre datakilder i denne artikkelen:

Grunnskolens informasjonssystem (GSI) inneholder data bl.a. om elevtall og ressurser i norske grunnskoler. Utfra disse kan vi lage flere mål på lærerressurser. Et sentralt mål er gruppestørrelse 1:

$$\frac{\text{(elevtimer totalt)}}{\text{(lærertimer totalt)}}$$

Gruppestørrelse 1 uttrykker hvor mange elever det i snitt vil være per lærer i undervisningen. En del undervisning, bl.a. spesialundervisning, vil ofte foregå i små grupper. Dette vil bidra til å trekke gjennomsnittlig gruppestørrelse ned. Gruppestørrelse 2 (GR2) tar hensyn til dette ved å se bort fra elev- og lærertimer som er knyttet til individuelle rettigheter og som typisk gis i små grupper, og gir et bedre uttrykk for antall elever per lærer i ordinær undervisning:

$$\frac{\text{(elevtimer totalt - (timetall spesialundervisning + timetall særskilt norsk))}}{\text{(ordinære undervisningstimer + timer til oppdeling til samiske språkalternativer)}}$$

Gjennomsnittlig gruppestørrelse 2 på ungdomstrinnet i skoleåret 2011/2012 var omtrent 19,9.

Vi har to datakilder for elevenes læringsutbytte: Nasjonale prøver og karakterer ved avsluttet grunnskole. Elevene gjennomfører nasjonale prøver i lesing, regning og engelsk på 5. og 8. trinn, samt i lesing og regning på 9. trinn. Ved fullført grunnskole registreres alle standpunkt og eksamenskarakterer. Grunnskolepoeng (GP) er et samlemål for disse karakterene, som danner grunnlaget for opptak til videregående utdanning. Grunnskolepoeng beregnes ved å legge sammen alle avsluttende karakterer på vitnemålet – standpunkt og eksamen – og dele på antall karakterer. Deretter multipliseres gjennomsnittet med 10. Elever som mangler karakter i mer enn halvparten av fagene får ikke beregnet grunnskolepoeng. Gjennomsnittlig grunnskolepoeng i skoleåret 2011/2012 var 40,0.

et kriterium for ekstra ressurser var svake resultater i utgangspunktet. En måte å ta hensyn til dette er, som i Utdanningsdirektoratet (2016), å sammenligne med forskjellene som eksisterte før tiltaket. Men dette forutsetter at tiltaksskolene i fravær av tiltaket ville hatt en tilsvarende utvikling som øvrige skoler, og vi vet ikke om det ville vært tilfelle.

Vi har imidlertid et godt grunnlag for å finne relevante sammenligningsskoler ettersom tilskuddet ble gitt etter klare objektive kriterier til skoler som i 2011/12 hadde (i) gruppestørrelse i ordinær undervisning (heretter gruppestørrelse 2 eller GR2, se boks 2) over 20 elever per lærer på ungdomstrinnet, og (ii) grunnskolepoeng (GP) under snittet for landet og dessuten (iii) fra et gitt tidspunkt.

Vi kan dermed sammenligne skoler som (så vidt) kvalifiserte for deltagelse med skoler som (så vidt) ikke kvalifiserte, i et såkalt «regression discontinuity» (RD) design. Dette er en tilnærming som ofte kan gi et

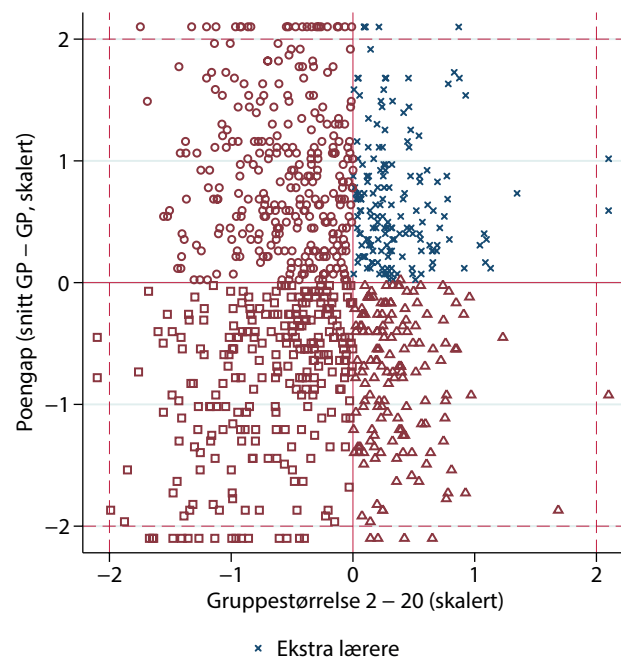
godt grunnlag for troverdige effektmål (se f.eks. Lee og Lemieux, 2010), og som har blitt brukt bl.a. i mange studier av utdanningspolitiske tiltak. I dette konkrete tilfellet bygger en slik sammenligning at skoler som på et gitt tidspunkt har nesten like gjennomsnittlige grunnskolepoeng – eller nesten lik gruppestørrelse – også på andre områder kan forventes å være (nesten) like. Disse i utgangspunktet like skolene hhv. får og ikke får ekstra ressurser. Etter at tilskuddet har fått virke en tid, vil det kunne oppstå en forskjell i resultater mellom skoler som er like over/under grensen. Siden vi har god grunn til å anta at skolene i utgangspunktet var like, og ville forblitt like i fravær av de ekstra ressursene, vil vi tolke en evt. forskjell som en effekt av tilskuddet.

Ettersom skolene måtte oppfylle to kriterier for støtte har vi to grenser der vi kan sammenligne (nesten) like skoler. I Figur 1 viser vi hvordan skolene fordelte seg med hensyn til gruppestørrelse 2 og gjennomsnittlige grunnskolepoeng. Den horisontale akse viser avviket fra gjennomsnittlig gruppestørrelse 2. Skoler til høyre i figuren (med positive verdier) har gruppestørrelse høyere enn 20. Den vertikale akse viser avviket mellom landsgjennomsnittet og skolens gjennomsnittlige grunnskolepoeng, dvs., verdier over null betyr at skolen hadde gjennomsnittlige grunnskolepoeng under landsgjennomsnittet i 2011/12. Det er lite intuitivt at høye avvik svarer til lave poengsnitt. Fordelen med denne måten å måle på er at skolene som deltar i tiltaket har positive verdier både på avvik gruppestørrelse og avvik grunnskolepoeng – det forenkler framstillingen av figurene vi bruker når vi analyserer effekter i fortsettelsen av denne artikkelen. Begge målene er dessuten standardisert (uttrykt som antall standardavvik fra landsgjennomsnittet), slik at de får tilsvarende skala.⁴

Figuren viser også hvorvidt hver enkelt skole fikk ekstra ressurser, skolene som fikk dette er markert med blå kryss. Skolene som hadde grunnskolepoeng under snittet og gruppestørrelse 2 over 20 ligger nordøst i figuren. Vi ser at nesten alle skolene i dette området fikk ekstra ressurser (med unntak for en skole med antall grunnskolepoeng så vidt under snittet), og at ingen skoler med grunnskolepoeng over snittet eller gruppestørrelse 2 under 20 fikk ekstra ressurser.

Figuren illustrerer også sammenligningsgruppene vi vil bruke i effektevalueringen. Rett til venstre for skolene som fikk ekstra lærere (med negative verdier på gruppestørrelseaksen, horisontal akse) finner vi skoler som også har svake karakterer, men som har gruppestørrelse 2 litt under 20 (dvs. skolene markert med sirkler, nordvest i figuren) og dermed ikke fikk støtte. Nedover fra tiltaksskolene (med negative verdier på

Figur 1. Ekstra lærere, skolens gruppestørrelse 2 og poenggap 2011/12 (skalert)



poengaksen, vertikal akse) finner vi skoler som også har store grupper, men som har skoleprestasjoner over snittet (dvs. skolene markert med trekant sørøst i figuren) og som derfor heller ikke fikk støtte. Til slutt har vi noen skoler som hadde relativt små grupper og gode resultater (skolene markert med firkanter sørvest i figuren), og som dermed både måtte hatt større grupper og svakere resultater for å være aktuelle for ekstra ressurser.

Resultater

I denne korte artikkelen presenterer vi foreløpige resultater fra effektevalueringen, og vi ser først og fremst på to utfallsmål knyttet til ressursinnsats og læringsutbytte i ordinær undervisning: (i) Gruppestørrelse 2 som mål på lærertetthet (ressurser) og (ii) gjennomsnittlig karakterer ved avsluttende skriftlig eksamen i 10. klasse.

Effekter på lærertetthet målt ved gruppestørrelse i ordinær undervisning

I figurene 2a og 2b ser vi hvordan gruppestørrelse 2 i skoleårene 2013/14-2015/16 varierer med skolens gruppestørrelse 2 og gjennomsnittlige grunnskolepoeng fra 2011/12. Vi vil diskutere effekter på læringsutbytte i lys av tilsvarende figurer, og forklarer derfor figur 2a i litt ekstra detalj.

Hvert symbol i figur 2a representerer et snitt av flere skoler med omtrent samme gjennomsnittlige grunnskolepoeng i 2011/12. Det blå krysset rett til høyre for den vertikale linjen (som markerer 0, dvs. grunnskolepoeng lik snittet) er dermed skoler som hadde grunnskolepoeng så vidt under snittet i 2011/12. Symbolene svarer til symbolene i figur 1, slik at de blå kryssene markerer skoler som har fått ekstra ressurser. For å si noe om effekter av satsingen ønsker vi å sammenligne

⁴ Standardavvikene er hhv. omtrent 2,1 grunnskolepoeng og 6,5 elever/lærer. Fra figuren ser vi dermed at det er få skoler som har gruppestørrelse 2 høyere enn ca. 26 (ett standardavvik over snittet), eller grunnskolepoeng lavere enn ca. 36 (to standardavvik under snittet). Vi har også, for begge mål, satt verdier større 2,1 eller mindre -2,1 lik disse verdiene, for å unngå at noen svært få ekstreme verdier forstyrrer bildet i figuren.

disse skolene med i utgangspunktet like skoler som ikke fikk ekstra ressurser. Den røde trekanten rett til venstre for den vertikale linjen er skoler som hadde grunnskolepoeng så vidt over snittet i 2011/12, og dermed så vidt ikke fikk ekstra ressurser. Dette tilsier at disse er en god sammenligningsgruppe for å vurdere effekten på skolene som så vidt fikk. En utfordring med denne sammenligningen er imidlertid at det er få skoler rett på hver side av den vertikale linjen, dermed blir gjennomsnittene upresise og sammenligningene usikre.

Vi løser dette ved å bruke også skoler litt lengre unna den vertikale linjen til å beregne gjennomsnitt. Da blir skolene mer ulike i utgangspunktet. Vi beregner derfor glidende gjennomsnitt for alle verdier for alle verdier av grunnskolepoeng, som tar hensyn til dette. Disse gjennomsnittene vises som kurver i figuren. I tillegg til symbolene som markerer gjennomsnitt har vi tegnet inn kurver som viser lokale glidende gjennomsnitt.⁵ Hvert punkt på disse oppsummerer gjennomsnittlig gruppestørrelse 2 i et begrenset område rundt punktet, med størst vekt på de nærmeste skolene.

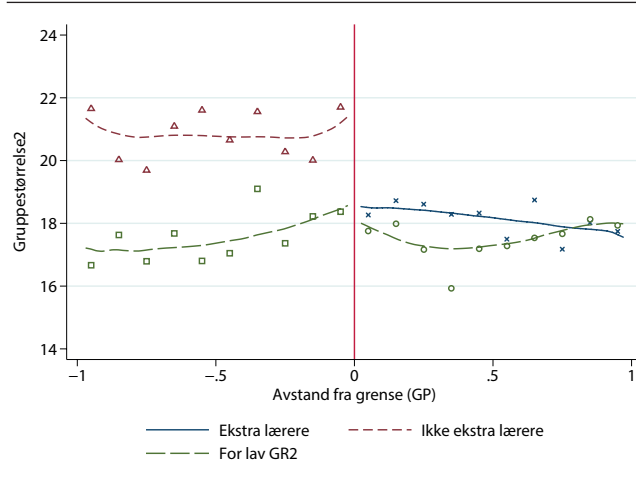
Den heltrukne blå kurven viser gjennomsnittlig gruppestørrelse 2 for skoler som fikk ekstra lærere. Vi ser at gruppestørrelse 2 i de tre etterfølgende år ligger på omtrent 18-19 elever/lærer, og at det ikke er noen sterk sammenheng mellom grunnskolepoeng i 2011/12 og senere gruppestørrelse.

Den stiplede røde kurven viser gjennomsnittlig gruppestørrelse 2 for skolene som hadde høy gruppestørrelse 2 i 2011/12, men som hadde for høyt karaktersnitt til å få ekstra lærere har gruppestørrelse 2 på omtrent 21-22 i senere år. Her er det heller ikke noen sterk sammenheng mellom grunnskolepoeng i 2011/12 og senere gruppestørrelse 2. Kurven krummer riktignok oppover i hver ende, men her er de beregnede gjennomsnittene mer usikre, fordi vi baserer oss på få skoler. Generelt er kurvene i figuren basert på verdiene i forholdsvis små intervaller. Dette gjør de godt egnet til å få fram nyanser i sammenhengen mellom tidligere kjennetegn og senere resultater (her gruppestørrelse 2), men også at de er utsatt for tilfeldigheter eller verdiene til enkelt-skoler. Forskjellen langs den røde stiplede kurven er heller ikke spesielt stor sammenlignet med spredningen til gjennomsnittene rundt kurven (de røde trekantene) eller til forskjellen mellom røde og den blå kurven.

Når vi sammenligner kurvenes endepunkter på hver side av den vertikale linjen kan vi sammenligne

⁵ Kurvene viser en lokal lineær tilpasning, dvs. et gjennomsnitt som tar hensyn til at variabelen vi regner gjennomsnitt for (her gruppestørrelse 2) kan endre seg med endret verdi langs den horisontale akse. I beregningen for hvert punkt inngår observasjoner innen en avstand på 0,5, og observasjoner nær punktet vi beregner gjennomsnitt for gis høyest vekt (triangular vektning). I figurene 2a og 2b er skolene vektet med antall elever på ungdomstrinnet. Merk at figur 2a bare omfatter skoler som hadde gruppestørrelse 2 nær grense for deltagelse (innen ett standardavvik), det er dermed en del av skolene i figur 1 som ikke inngår i analysene i figur 2.

Figur 2a. Effekter på gruppestørrelse i ordinær undervisning: Avstand fra grense grunnskolepoeng



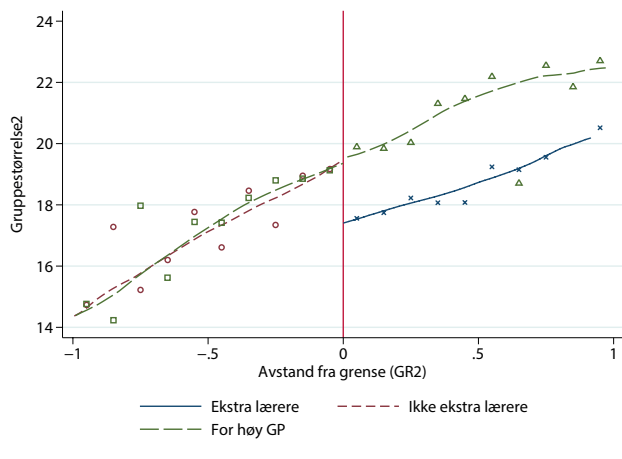
beregnete gjennomsnitt for skoler som i utgangspunktet hadde grunnskolepoeng like under og like over grensen for deltakelse. Vi ser at det er en klar forskjell på omtrent 2-3 elever/lærer.⁶ Ettersom dette vi sammenligner skoler som i utgangspunktet var nesten like, forventer vi at de ville ha vært like i fravær av tiltaket, og at forskjellen vi ser er et resultat av tiltaket. Det vil si, tiltaket reduserte gruppestørrelse 2 med to til tre elever/lærer.

Når vi sammenligner to grupper av skoler, her skoler som i 2011/12 har grunnskolepoeng like over/under grensen for ekstra ressurser, er det en mulighet for at det er finnes andre, relevante forskjeller som vi ikke klarer å ta hensyn til, og som vi dermed sammenblender med tiltaket. En sjekk på hvorvidt dette er tilfelle kan være å sammenligne andre grupper av skoler rundt samme grense, der ingen deltar i noe tiltak, for å se om det er forskjeller mellom disse. De to grønne gruppene av skoler (lang-stiplet) hadde begge for lav gruppestørrelse 2 i 2011/12 til å få støtte. Også her er det bare svake sammenhenger mellom grunnskolepoeng i 2011/12 og senere gruppestørrelse innen hver gruppe. De grønne gruppene viser hvordan senere gruppestørrelse 2 endrer seg rundt gjennomsnittlig grunnskolepoeng for skoler der det ikke er noe tiltak. For disse skolene er gruppestørrelse 2 i snitt omtrent lik 18-19 uavhengig av tidligere grunnskolepoeng ved skolen, både for grunnskolepoeng over og under landsgjennomsnittet. Det er dermed ingen ting som tyder på at det er andre forhold enn tiltaket som er årsak til forskjellen i gruppestørrelse 2 rundt grensen.

I figur 2b viser vi hvordan gruppestørrelse 2 etter tiltaket varierer med gruppestørrelse 2 fra 2011/12. Her ser vi en jevnt stigende sammenheng for alle grupper av skoler. Det er som ventet, og viser bare at skoler som har store grupper i 2011/12 også har det i de tre etterfølgende årene. For skolene med grunnskolepoeng under landsgjennomsnittet er det imidlertid et klart brudd rundt 0 i figuren, det vil si ved tiltaksgrensen ved

⁶ Forskjellen er også klart statistisk signifikant.

Figur 2b. Effekter på gruppestørrelse i ordinær undervisning: Avstand fra grense gruppestørrelse



gruppestørrelse 2 lik 20. Skoler som så vidt var over grensen i 2011/12, og dermed deltar i tiltaket, har i senere skoleår en gruppestørrelse 2 omtrent 2 elever/lærer lavere enn skoler som så vidt ikke deltar. Som i figur 2a er det ingen tilsvarende forskjell rundt grensen blant skoler som uansett ikke ville fått ekstra lærere, her fordi de hadde for høye gjennomsnittlige resultater i 2011/12.

Oppsummert ser vi, både når vi sammenligner skoler som i 2011/12 hadde høy gruppestørrelse 2 og grunnskolepoeng rett under/over snittet og når vi når vi sammenligner skoler som i 2011/12 hadde lave gjennomsnittlige grunnskolepoeng og gruppestørrelse 2 rett under/over 20, at skolene som så vidt fikk ekstra lærere har klart lavere gruppestørrelse 2 senere. Fra begge sammenligningene kan vi konkludere med de ekstra ressursene reduserte gruppestørrelse 2 med omtrent 2-3 elever/lærer, eller omtrent 10 prosent.⁷

Fra figur 2a ser vi at skolene som får ekstra lærere med tiltaket har gruppestørrelse 2 omtrent lik skolene som i utgangspunktet hadde lave gruppestørrelser. Det er også verdt å merke seg at kurvene både i figur 2a og 2b er omtrent parallelle. Dette tyder på at effekten på gruppestørrelse 2 er en reduksjon på omkring 2 elever/lærer, uavhengig av grunnskolepoeng eller hvor store gruppene var i utgangspunktet. Denne konklusjonen holder også for andre typer analyser, med sterkere antagelser om sammenhengene mellom gruppestørrelse 2 og tidligere gruppestørrelse og grunnskolepoeng. Dersom vi ser på alle skoler med støtte og sammenlikner gruppestørrelse med alle andre skoler finner vi tilsvarende effekter som i figur 2, gruppestørrelse 2 er redusert med omkring 2 elever/lærer som følge av ekstra lærerårsverk.

⁷ Når vi ser på skoler rundt grensen for gruppestørrelse 2 går gruppestørrelsen nødvendigvis ned fra dette nivået, skolene så vidt denne grensen har i utgangspunktet gruppestørrelse 2 så vidt over 20. Skolene nær grensen for grunnskolepoeng som fikk ekstra ressurser har i utgangspunktet gruppestørrelse 2 større eller lik 20, men ikke nødvendigvis så vidt over 20, slik at gjennomsnittlig gruppestørrelse 2 for disse skolene i utgangspunktet vil være høyere enn 20.

Læringsutbytte

Vi måler læringsutbyttet i første rekke ved eksamenskarakterer i 10. klasse. Ettersom støtten ble innført høsten 2013, er avgangskullet sommeren 2016 det eneste kullet der elevene på tiltaksskolene opplevde økte lærerressurser gjennom hele ungdomstrinnet. I figurene 3a og 3b analyserer vi eksamenskarakterer slik vi analyserte gruppestørrelse 2 i figurene 2a og 2b. Rundt poenggrensa (figur 3a) finner vi ikke noe hopp for de skolene med store grupper som hadde grunnskolepoeng like under snittet og dermed fikk støtte. Dersom det var effekter av redusert gruppestørrelse ville vi vente at karakterene ved skolene med ekstra lærere (blå heltrukken kurve) nær grensa var klart høyere enn ved skolene uten ekstra lærere (rød stiple kurve) rett på den andre sida av grensa. Som i figur 2a ser vi ingen tegn til forskjeller i resultat rundt grensa for skolene med gruppestørrelse 2 under 20, dermed er det ingenting som tyder på at andre forhold endrer seg ved denne grensa.

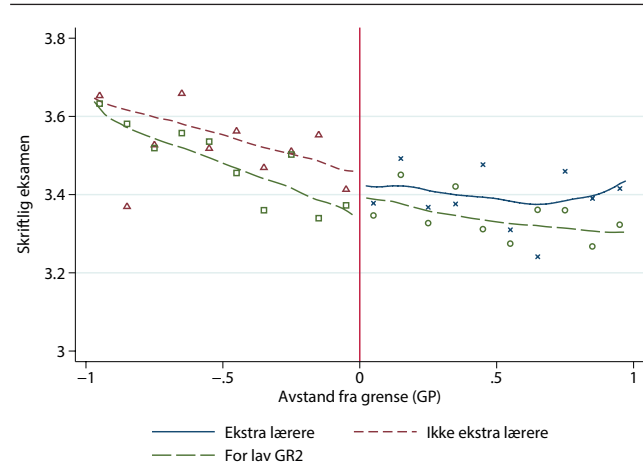
Rundt gruppestørrelsesgrensen i figur 3b er gjennomsnittlig eksamenskarakterer høyere like over grensen for gruppestørrelse, blant skoler som fikk ekstra lærere, enn blant nesten like skoler like under grensen, som ikke fikk ekstra ressurser. Dette resultatet er imidlertid skjørt, forskjellige beregningsmåter for det glidende gjennomsnittet gir forskjellige resultater. Det er også stor statistisk usikkerhet, forskjellen mellom skoler over/under gruppestørrelsesgrensen er ikke statistisk signifikant.

Dersom vi ser på resultater for flere kull samtidig får vi mer presise resultater. Ulempen er at de tre kullene i snitt vil ha opplevd mindre gruppestørrelse i kun 1,5 skoleår, det tilsier at vi kan forvente en mindre effekt. I figurene 4a og 4b viser vi effekter rundt de to grensene for tre avgangskull (2014-2016). Mønsteret likner på det vi fant for 2016-kullet. Når vi ser på skoler med store grupper rundt poenggrensen (figur 4a) er det ingen tegn til bedre resultater blant skolene som så vidt får støtte grunnet litt svakere grunnskolepoeng enn de som så vidt ikke får.⁸ Blant skolene med lave grunnskolepoeng er det imidlertid tegn til bedre eksamenskarakterer blant skolene som så vidt fikk ekstra lærere, men forskjellen er såpass liten at vi ikke kan konkludere med at den ikke skyldes tilfeldigheter.

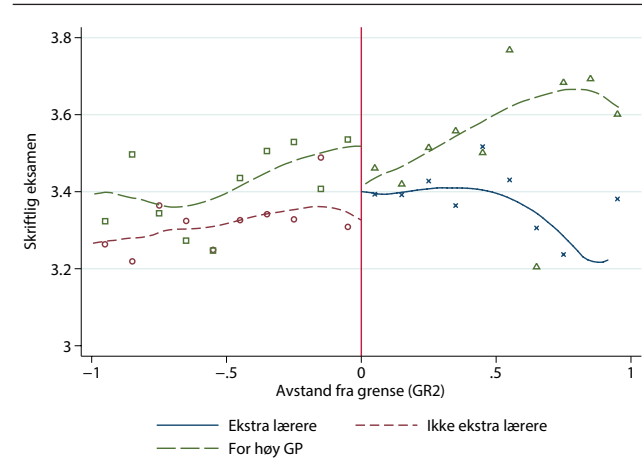
Vi har basert analysene så langt på skoler nær de forskjellige grensene for deltagelse i tiltaket. Ettersom disse i utgangspunktet kan forventes å være nesten like kan vi rimelig sikkert si at evt. systematiske forskjeller skyldes tiltaket. En ulempe med disse sammenligningene av skoler nær grensen er imidlertid at vi har få skoler

⁸ Om noe er det tegn til dårligere resultater i skolene med støtte, men dette resultatet er følsomt for hvordan vi beregner glidende gjennomsnitt og ikke statistisk signifikant. Vi kommer tilbake til presisjon og signifikans i mer detalj ifm. med omtale av figur 5.

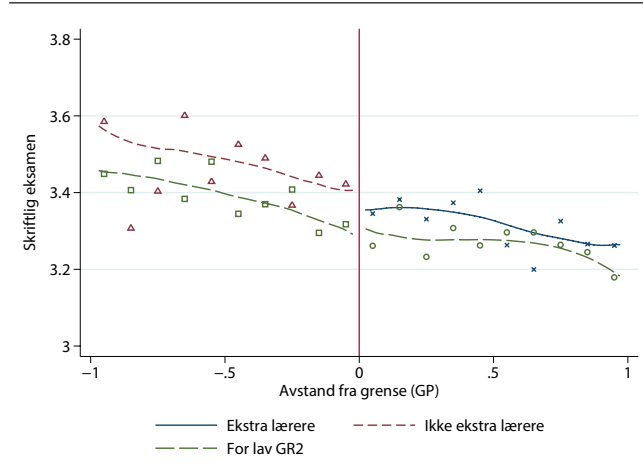
Figur 3a. Effekter på eksamenskarakter 2016: Avstand fra grense grunnskolepoeng



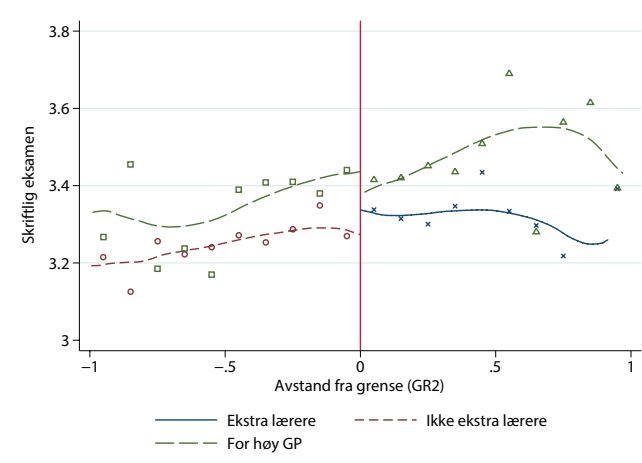
Figur 3b. Effekter på eksamenskarakter 2016: Avstand fra grense gruppestørrelse



Figur 4a. Effekter på eksamenskarakter 2014-2016: Avstand fra grense grunnskolepoeng



Figur 4b. Effekter på eksamenskarakter 2014-2016: Avstand fra grense gruppestørrelse



og relativt få elever å bruke i sammenligningene,⁹ det gir upresise resultater. Vi har derfor også gjort supplerende analyser, der vi gjør sterkere antagelser om hvordan vi kan beskrive skolens resultater, men til gjengjeld kan bruke et større datamateriale og få mer presise estimater.

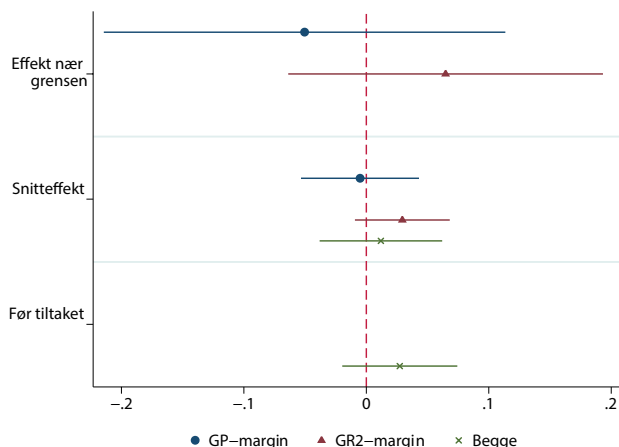
I figur 5a viser vi forskjellige anslag på effekter av tiltaket på karakterer for de tre kullene 2013-2016, med tilhørende usikkerhet. Skalaen er karakterpoeng. Mens midtpunktet på hver linje angir effektanslaget, viser bredden på linjen det såkalte konfidensintervallet (95% nivå). Når dette ikke dekker null kan vi si at effekten er statistisk signifikant og vi kan med rimelig sikkerhet konkludere med at det er en effekt av tiltaket.

⁹ Vi analyserer karakterer på elevnivå. Antall skoler er imidlertid en større begrensning enn antall elever. Ettersom elever i samme skole kan være utsatt for de samme uobserverte påvirkningene justerer vi presisjonen i estimatene for å ta høyde for dette (ved såkalt clustering, eller flernivå-analyse), det har samme betydning for presisjonen i estimatene som en betydelig reduksjon i antall elevobservasjoner.

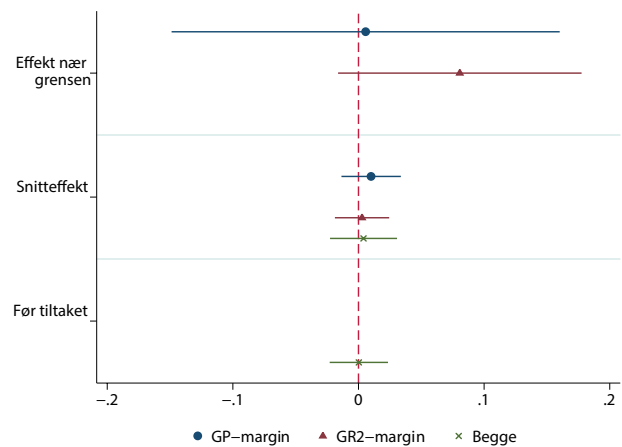
De to øverste effektestimaterne i figur 5a, merket «effekt nær grensen», svarer til estimatene fra figurene 4a og 4b. I figur 4a, der vi sammenligner skoler med grunnskolepoeng rett over/under grensen (vi kaller dette GP-marginen) er det tegn til svakere resultater i skolene som får ekstra lærere. I figur 4b derimot, der vi sammenligner skoler med gruppestørrelse 2 rett over/under grensen (vi kaller dette GR2-marginen) er det tegn til bedre resultater i skolene som får ekstra lærere. Vi ser imidlertid at ingen av disse estimatene gir grunnlag med å konkludere med at det er noen effekt på læringsutbytte, verken positiv eller negativ. Men begge anslagene er upresise. Det betyr at selv om vi ikke finner noen klare tegn til effekter, kan vi heller ikke utelukke betydelige effekter, større enn 0,1 karakterpoeng.

For å oppnå mer presise mål på effekter sammenligner vi også karakterene i alle skolene som fikk støtte med alle andre skoler, også skolene langt fra grensene for deltagelse. For å finne anslag på effekter sammenligner vi ikke lenger skoler som er nesten like. At skolene i utgangspunktet er forskjellige stiller sterkere krav til at vi klarer å ta hensyn til sammenhengen mellom eksamenskarakterer og tidligere resultater og

Figur 5a. Oppsummering av beregnede effekter på eksamenskarakter 2014-2016



Figur 5b. Oppsummering av beregnede effekter på nasjonale prøver 2014-2015



gruppestørrelse.¹⁰ For ytterligere å øke presisjonen i anslagene tar vi også hensyn til elevenes tidligere resultater fra nasjonale prøver på 8. trinn.¹¹ De to første estimatene merket «snitteffekt» i figur 5a er fra denne typen sammenligninger, hhv. av skoler over/under GP-grensen (GP-marginen) og GR2-grensen (GR2-marginen). Vi ser at disse estimatene er mer presise. Estimatene er imidlertid også nærmere null, slik at det fortsatt ikke er tegn til effekter på læringsutbytte.

Estimatene som ikke baserer seg på i utgangspunktet nesten like skoler er mindre troverdige, ettersom det kan være andre forhold vi ikke klarer å ta hensyn til. Vi kan imidlertid bruke det at vi har flere kontrollgrupper til å få troverdige og presise resultater. Dersom vi ikke klarer å fullt ut ta hensyn til betydningen av tidligere resultater, kan det føre til at vi forventer for høye resultater av skoler med svake tidligere resultater. Men dette burde i så fall være tilfellet både for skoler med høy og lav tidligere gruppestørrelse 2, og tilsvarende dersom vi ikke klarer å ta hensyn til betydningen av tidligere gruppestørrelse 2. Det betyr at vi kan finne effekten av tiltaket som resultatforskjellene knyttet til at skoler har både lave tidligere resultater og høy tidligere gruppestørrelse utover forskjellene knyttet til lave tidligere resultater eller høy tidligere gruppestørrelse.

Denne forskjeller-i-forskjell metoden bygger på at tiltaksskolene ville hatt resultater som en «kombinasjon» av skoler med store grupper og svake resultater (i 2011), dersom de ikke hadde fått ansatt nye lærere. Kombinasjonen av et større datamateriale og resultatene fra nasjonale prøver gir svært mye mer presise estimater enn vi fikk i figurene 4a og 4b, jf. estimatet

¹⁰ Estimatene som baserer seg på alle skoler tar hensyn til avstand fra grensene fra hver av grensene, samt kvadratet av hver av disse og produktet av avstandene. Disse analysene vektet også alle skoler likt, uavhengig av avstand til grensene.

¹¹ Dette har liten betydning for de faktiske estimatene. Det fanger imidlertid mange forskjeller mellom elevene som eksisterte forut for tiltaket, og reduserer dermed mengden uforklart variasjon dramatisk. Dette gjør at estimatene blir mer presise.

merket «snitteffekt», «begge» (marginen) i figur 5a. Vi finner at anslaget på denne gjennomsnittseffekten er ganske presist null, vi kan utelukke effekter på eksamenskarakter større enn omtrent 0,06 karakterpoeng.

Denne framgangsmåten baserer seg på å sammenligne skoler i et gitt år. Vi kan gjøre en ytterligere vurdering av om framgangsmåten er rimelig ved å teste om vi finner «effekter» ved å late som om støtte ble gitt til de samme skolene noen år før. Tiltaket kan åpenbart ikke hatt noen effekt før det ble innført, så evt. effekt-estimer før innføring tyder på at det er andre forskjeller mellom skolene som vi ikke lykkes i å ta hensyn til. Her finner vi ingen grunn til bekymring ettersom det ikke er noen «effekt» for tidligere kull. Vi ser at estimatet for tidligere år er nær null, og dessuten svært likt estimatet for årene med ekstra ressurser.

I Figur 5b ser vi resultater for nasjonale prøver i 9. trinn (gjennomsnitt for regning og lesing, målt i enheter av standardavvik¹²), der vi heller ikke finner effekter av ekstra lærere. Her er effektene svært presist null, blant annet fordi vi fjerner svært mye annen variasjon ved å kontrollere for resultater ved tilsvarende nasjonale prøver resultater året før. Vi er dermed i stand til å utelukke effekter større enn 0,03 standardavvik.

Konklusjon

Våre anslag på effekter av satsing på økt lærertetthet er relativt entydige: Skolene som får ekstra ressurser reduserer gruppestørrelsen i ordinær undervisning, men det er få tegn til at elevenes læringsutbytte påvirkes positivt (eller negativt) i disse skolene. Vi finner at gjennomsnittlig gruppestørrelse i ordinær undervisning reduseres omtrent fra 21 til 19 i tiltaksskolene, eller med omtrent 10 prosent. Effekten på skriftlig eksamenskarakter er finner vi å være ganske presist lik null. Vi kan med rimelig stor sikkerhet si at dersom tiltaket har hatt en effekt på gjennomsnittlig eksamenskarakterer i

¹² Et standardavvik svarer til 10 skalapoeng etter omleggingen av nasjonale prøver.

tiltaksskolene er denne mindre enn 0,06 karakterpoeng, dvs. at en av 17 elever går opp en karakter. For nasjonale prøver kan vi avvise gjennomsnittseffekter større enn omkring 0,03 standardavvik (dette tilsvarer 0,3 skala-poeng etter omleggingen av nasjonale prøver).

Resultatene baserer seg på eksamenskarakterer til og med 2016 og nasjonale prøver til og med 2015. Når vi får data fra eksamen 2017 og nasjonale prøver 2016 vil vi kunne få enda litt mer presise resultater. Dette kan være særlig interessant for eksamen, ettersom vi da vil ha to elevkull som har hatt tre skoleår med ekstra lærere. Vi har imidlertid ikke funnet tegn til større effekter for senere årskull så langt.

Det viktigste forbeholdet knyttet til våre resultater er at en endret elevsammensetning ville kunne skape skjevhet og skjule effekter i våre analyser. En målsetning med tiltaket har vært å redusere omfanget av spesialundervisning. Dette kan i prinsippet føre til at flere elever med svake forutsetninger deltar mer i ordinær undervisning og i større grad avlegger nasjonale prøver og/eller skriftlig eksamen. Mer generelt, om flere lærere fører til at flere faglig svake elever blir registrert med eksamenskarakterer eller nasjonale prøver, vil gjennomsnittlige resultatene kunne svekkes og på den måten maskere evt. positive effekter av mindre grupper. Så langt har vi imidlertid ikke funnet tegn til slike sammensetningseffekter, f.eks. har vi ikke funnet at det er færre elever som mangler eksamenskarakterer eller er fritatt fra nasjonale prøver.

Vi vil gjøre videre studier av skolens ressursbruk, for å få et klarere bilde av hvordan skolene prioriterer de ekstra ressursene. Det kan f.eks. hende at skolene i stor grad prioriterer ressursbruk på elever som ikke deltar i nasjonale prøver eller eksamen, og som dermed ikke bidrar til vårt mål på læringsutbytte. Vi finner imidlertid en tydelig effekt på gruppestørrelse i ordinær undervisning, men kan likevel langt på vei utelukke selv ganske små effekter på gjennomsnittlig læringsutbytte.

Om effekten på læringsutbytte i gjennomsnitt er null, kan det likevel være grupper som nyter godt av flere lærere og mindre grupper. I det videre evalueringsarbeidet vil vi blant annet se om vi finner forskjellig effekter for ulike grupper av elever (jenter/gutter, faglig sterke/svak elever mm.). Det kan også være noen skoler eller lærere som i større grad klarer å nyttiggjøre seg de ekstra ressursene enn andre. Ettersom vi allerede langt på vei kan utelukke selv små effekter for elevene samlet kan vi imidlertid ikke vente å finne annet enn evt. små effekter på læringsutbytte og/eller effekter bare for små elevgrupper. Vi vil også studere andre utfall, som læringsmiljø, der vi foreløpig ikke har noe grunnlag for å konkludere.

Effektevalueringen vil suppleres av en kvalitativ implementeringsstudie. Denne vil forsøke å kartlegge i mer detalj hva som har skjedd og hvordan de ekstra ressursene har blitt brukt i utvalgte tiltaksskoler. Vi

håper at analysene til sammen vil kunne tegne et forholdsvis komplett bilde av effektene av tiltaket. Det pågår for tiden flere beslektede forskningsprosjekter. Forskningsrådets program LÆREEFFEKT finansierer studier av ekstra midler til kommunene til økt lærertetthet på 1.-4. trinn samt av ekstra lærerressurser til bestemte tiltak knyttet til matematikk-, lese- og skriveopplæring. Det gjennomføres også flere studier av andre tiltak rundt overgangen mellom ungdomstrinn og videregående, bl.a. effektevalueringen av Ny GIV (Huitfeldt m.fl., 2016) samt flere innenfor Kunnskapsdepartementets satsing Program for bedre gjennomføring. Til sammen vil disse studiene forhåpentligvis gi et grunnlag for å vurdere effekter av forskjellige typer tiltak på forskjellige trinn i grunnopplæringen, og dermed gi et grunnlag for en bedre informert skolepolitikk.

Referanser

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Fredriksson, P., Öckert, B., & Oosterbeek, H. (2013). Long-Term Effects of Class Size. *The Quarterly Journal of Economics*, 128(1), 249-285.

Galloway, T., L.J. Kirkebøen og M. Rønning. Karakterpraksis i grunnskoler (2011). Sammenheng mellom standpunkt- og eksamenskarakterer. Rapport 2011/04. Statistisk sentralbyrå.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *The Economic Journal*, 113(485), 64-98.

Huitfeldt, I., Lars J. Kirkebøen og M. Rønning (2016) Effektevaluering av intensivopplæringen i Overgangsprosjektet, Ny Giv: Andre delrapport, Rapport 2016/08. Statistisk sentralbyrå.

Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal*, 113(485), 34-63.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2), 281-355.

Leuven, E., Oosterbeek, H., & Rønning, M. (2008). Quasi-experimental estimates of the effect of class size on achievement in Norway. *The Scandinavian Journal of Economics*, 110(4), 663-693.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485), 3-33.

Utdanningsdirektoratet (2016). Satsingen på økt lærertetthet på ungdomstrinnet 2016. Rapportering om ordningen fra Utdanningsdirektoratet til Kunnskapsdepartementet.