

Rapporter

Reports

2016/08 ●

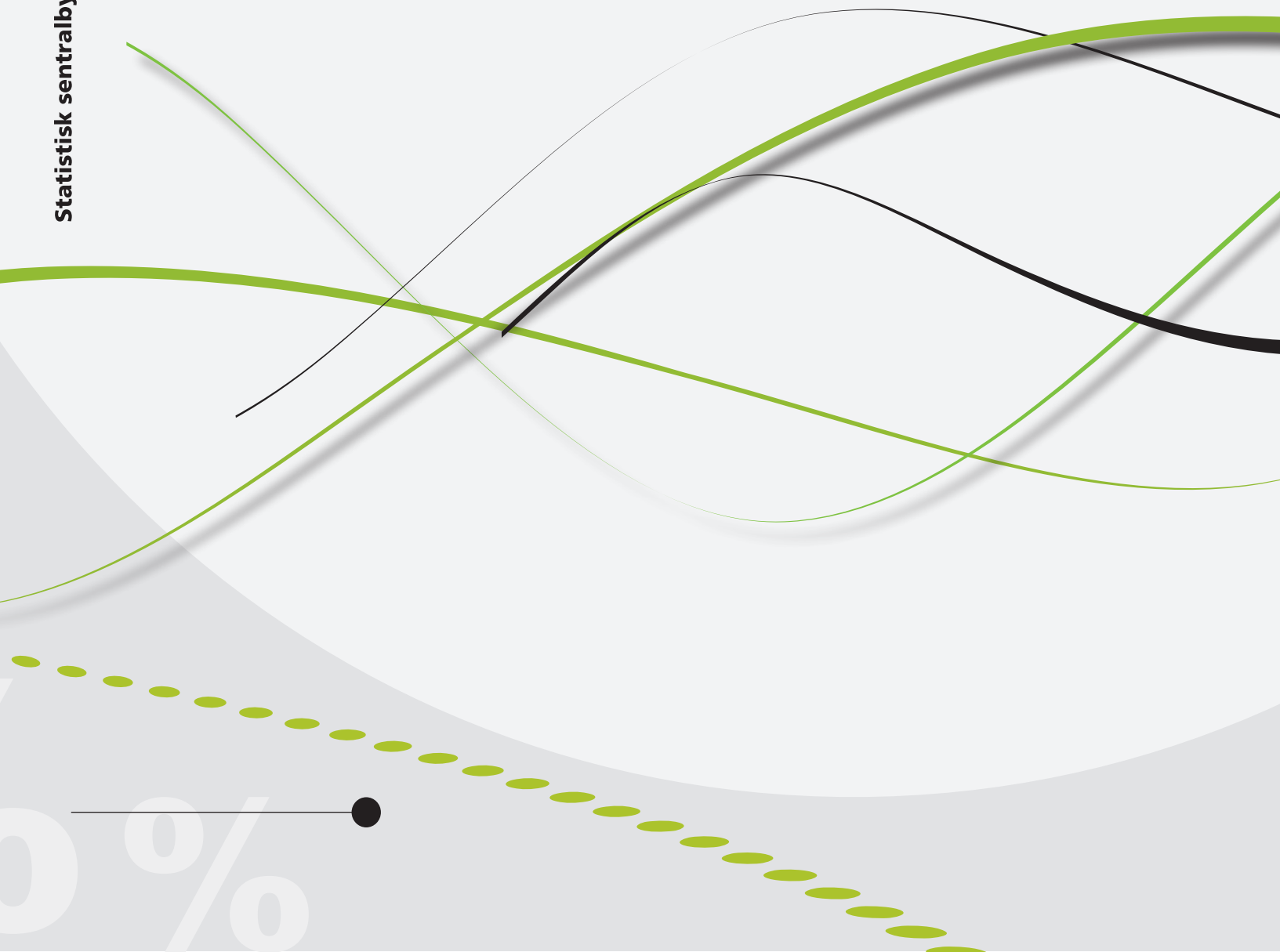
Statistics Norway



Statistisk sentralbyrå

*Ingrid Huitfeldt, Lars Kirkebøen og Marte Rønning*

## **Effektevaluering av intensivopplæringen i Overgangsprosjektet, Ny Giv: Andre delrapport**





*Ingrid Huitfeldt, Lars Kirkebøen og Marte Rønning*  
**Effektevaluering av intensivopplæringen i  
Overgangsprosjektet, Ny Giv: Andre delrapport**

*Rapporter* I denne serien publiseres analyser og kommenterte statistiske resultater fra ulike undersøkelser. Undersøkelser inkluderer både utvalgsundersøkelser, tellinger og registerbaserte undersøkelser.

	<b>Standardtegn i tabeller</b>	<b>Symbol</b>
© Statistisk sentralbyrå	Tall kan ikke forekomme	.
Ved bruk av materiale fra denne publikasjonen skal	Oppgave mangler	..
Statistisk sentralbyrå oppgis som kilde.	Oppgave mangler foreløpig	...
Publisert mars 2016	Tall kan ikke offentliggjøres	:
	Null	-
ISBN 978-82-537-9308-5 (trykt)	Mindre enn 0,5 av den brukte enheten	0
ISBN 978-82-537-9309-2 (elektronisk)	Mindre enn 0,05 av den brukte enheten	0,0
ISSN 0806-2056	Foreløpig tall	*
Emne: Utdanning	Brudd i den loddrette serien	—
	Brudd i den vannrette serien	
Trykk: Statistisk sentralbyrå	Desimaltegn	,

## Forord

Frafall i videregående er ansett å være et betydelig problem, både for de det gjelder og for samfunnet. Høsten 2010 satte Kunnskapsdepartementet (KD) i gang en rekke tiltak for å redusere frafallet. Denne satsingen fikk navnet Ny GIV. Et sentralt tiltak er det såkalte Overgangsprosjektet, der elever med svake faglige prestasjoner midtveis i 10. trinn får tilbud om intensivopplæring i grunnleggende ferdigheter i regning og lesing/skriving. KD har lagt opp til en bred evaluering av Overgangsprosjektet. Statistisk sentralbyrå, i samarbeid med forskere fra Økonomisk institutt, Universitetet i Oslo og Frischsenteret, gjennomfører en effektevaluering, der vi studerer hvordan intensivopplæringen i Overgangsprosjektet har påvirket forskjellige forhold knyttet til fullføring av videregående opplæring. Denne rapporten er andre delrapport fra denne effektevalueringen. Første rapport kom i 2013 (Eielsen m.fl, 2013), og det er planlagt en sluttrapport i 2017. Gaute Eielsen, Edwin Leuven og Oddbjørn Raaum deltok i arbeidet med den første rapporten. De har også gitt bidrag til tidligere utforming av analyseopplegg og tilrettelegging av data samt kommentarer til utkast til denne rapporten.

Overgangsprosjektet og effektevalueringen er finansiert av Kunnskapsdepartementet. KD har bidratt med kommentarer til et tidligere rapportutkast, samt underveis i arbeidet med analysene.

Statistisk sentralbyrå, 5. februar 2016

Christine Meyer

## Sammendrag

Gjennom Overgangsprosjektet i Ny GIV har elever med svake faglige prestasjoner midtveis i 10. trinn fått tilbud om tilpasset intensivopplæring i vårsemestrene 2011-2013. Hensikten har vært å øke gjennomstrømningen i og fullføringen av videregående opplæring (VGO). Denne rapporten er andre underveisrapport fra den pågående effektevalueringen, og presenterer foreløpige resultater.

Hensikten med effektevalueringen er å finne kausale effekter av intensivopplæringen, først og fremst om intensivopplæringen har ført til at flere fullfører og består VGO. Ettersom intensivopplæringen har vært rettet mot elever som fullførte grunnskolen i årene 2011-2013 har vi foreløpig lite data om fullføring av VGO, og kan si lite om eventuelle effekter. Vi studerer derfor effekter på andre resultatmål, som tidligere er vist å henge sammen med fullføring av VGO, først og fremst karakterer fra grunnskolen og tidlige mål på gjennomstrømning i VGO (fullføring av Vg1 og overgang til Vg2).

Med effekten av intensivopplæringen mener vi forskjellen mellom de faktiske oppnådde resultatene og resultatene elevene ville ha fått i fravær av intensivopplæringen. Informasjon om slike effekter gir et grunnlag for å vurdere virkning og prioritere mellom forskjellige tiltak. For å lære om effekter må vi isolere disse fra andre forhold som har betydning for resultatene. Vi bruker to tilnærminger for å isolere effekter: Vi sammenligner utviklingen i skoler som innfører intensivopplæring på forskjellige tidspunkt og vi sammenligner resultatene innad i Stavanger-skoler.

Den første sammenligningen lar oss studere generelle effekter av å innføre intensivopplæring. Vi har et stort datamateriale, bestående av flere fullstendige årskull av elever. Ettersom skoler innfører intensivopplæring på forskjellige tidspunkt kan vi se om resultatutviklingen ved innføring av intensivopplæring er forskjellig fra samtidig resultatutvikling i andre skoler. Vi finner ikke tegn til slike forskjeller, og dermed ikke tegn til effekter av å innføre intensivopplæring, verken på karakterer eller tidlig gjennomstrømning. Det er en viss usikkerhet i estimatene, men det store datamaterialet gjør at denne er beskjeden. Vi kan dermed langt på vei utelukke gjennomsnittlige effekter på tidlig gjennomstrømning større enn omtrent 1 prosentpoeng. Omtrent 10 prosent av elevene deltar i intensivopplæringen, men intensivopplæringen kan ha betydning også for elever som ikke deltar. Dette gjør det krevende å konkludere om effekten på deltakerne. Resultatene tyder ikke på noen effekt for deltakerne, og under bestemte antagelser kan vi utelukke effekter på disse større enn omtrent 5 prosentpoeng.

Den andre sammenligningen baserer seg på en skarp avgrensning av målgruppen i Stavanger-skoler, som lar oss sammenligne i utgangspunktet nesten like elever som deltar / ikke deltar i intensivopplæringen. Denne sammenligningen lar oss studere effekten av å endre antall elever som deltar. Tidlig gjennomstrømning blant Stavanger-elever i målgruppen er omtrent 9 prosentpoeng høyere enn blant nesten like elever som ikke er i målgruppen. En sammenligning med tilsvarende forskjell i Bergen kan tyde på at dette er en effekt av intensivopplæringen. Ettersom målgruppen utgjør 10 prosent av elevene svarer dette til en effekt på knappe 1 prosentpoeng på gjennomsnittlig tidlig fullføring blant alle elever.

Når vi finner tegn til en effekt i Stavanger, men ikke generelt, kan dette skyldes at intensivopplæringen er organisert og gjennomført på en annen måte i Stavanger enn ellers i landet. En økning i tidlig gjennomstrømning på 1 prosentpoeng, fra et nivå på 80-85 prosent, er mindre enn målsetningen for Ny GIV, men likevel stor nok til at en eventuell tilsvarende effekt på fullføring av VGO kan være av betydning. Vi finner foreløpig ikke tegn til effekter på karakterer eller senere gjennomstrømning. For senere gjennomstrømning og fullføring av VGO er datamaterialet foreløpig for knapt til at vi kan konkludere, men eventuelle effekter på dette vil være viktige for den endelige vurderingen av intensivopplæringen.

## Abstract

As part of the Ny GIV initiative, initiated by the Norwegian Ministry of Education and Research, students with weak academic performance midway through the 10th grade were offered intensive training in basic numeracy and literacy in the spring semesters 2011-2013. The intention has been to increase the progress through and completion of secondary education (VGO). This report is the second report from the ongoing impact evaluation, and presents preliminary results.

The purpose of the impact evaluation is to find causal effects of the intensive training, primarily whether it has increased completion of VGO. As the intensive training has been provided to students who completed primary school in the years 2011-2013 we have little data on completion of VGO and can say little about any effects on this outcome. We thus study effects on earlier outcomes, which previously have been shown to correlate with completion of the VGO, primarily grades from compulsory schooling and early measure of progress in VGO (completion of the first year, Vg1, and transition to the second year, Vg2).

With the effect of intensive training we mean the difference between the actual results obtained and the results the students would have received in the absence of intensive training. Information about such effects provides a basis for evaluating the efficiency and prioritize between different policy measures. To learn about effects we must isolate these from other factors that have affected the results. We use two approaches to isolate effects: We compare the development in schools introducing intensive training at different times and we compare the results within Stavanger schools.

The first comparison allows us to study the general effects of introducing intensive training. We have a large data set, consisting of several complete cohorts of students. As schools are introducing intensive training at different times, we can see whether changes in outcomes in schools introducing intensive training are different from changes in other schools. We find no evidence of such differences, thus no signs of effects of introducing intensive training, neither on marks nor early progress through VGO. There is some uncertainty in the estimates, but the vast data material makes this modest. We can thus rule out average effects on early measures of progress greater than about 1 percentage point. Approximately 10 percent of students participate in intensive training, but intensive training can have an effect also for students who do not participate. This makes it difficult to conclude about the effects on participants. The results do not indicate any effect for participants, and, under certain assumptions, we can rule out effects on these larger than about 5 percentage points.

The second comparison is based on a sharp delineation of the target group in Stavanger schools, which allows us to compare initially almost identical students who are participating / not participating in intensive training. This comparison allows us to study the effect of changing the number of pupils participating. Early progress among Stavanger students in the target group is about 9 percentage points higher than among similar students not eligible to participate. A comparison with the corresponding difference in Bergen may indicate that this is an effect of intensive training. Since the target group represents 10 percent of students, this corresponds to an effect of almost 1 percentage point on average early progress through VGO among all students.

When we find evidence of an effect in Stavanger, but not in general, it may be because intensive training is organized and conducted differently in Stavanger than elsewhere in the country. An increase in early throughput of 1 percentage point, from a level of 80-85 percent, is less than the target for Ny GIV, but still large enough that a similar effect on completion of VGO can be economically significant. We find no sign of any effects on marks or progress through higher levels of upper secondary. For progress at higher levels and completion of VGO the data is yet very scarce, not allowing us to draw clear conclusions. Any effects on this would be important in the final assessment of the intensive training.

## Innhold

<b>Forord</b> .....	<b>3</b>
<b>Sammendrag</b> .....	<b>4</b>
<b>Abstract</b> .....	<b>5</b>
<b>Innhold</b> .....	<b>6</b>
<b>1. Innledning</b> .....	<b>7</b>
<b>2. Datagrunnlag</b> .....	<b>10</b>
<b>3. Effektevaluering – empirisk strategi og resultater</b> .....	<b>15</b>
3.1. Forskjeller mellom alle faser og årskull («difference-in-differences», DiD).....	15
3.2. Forskjeller innen Stavanger-skoler .....	24
<b>4. Oppsummering</b> .....	<b>38</b>
<b>Referanser</b> .....	<b>43</b>
<b>Vedlegg A: Supplerende resultater fra analyser av forskjeller mellom faser og årskull («difference-in-differences-modeller»)</b> .....	<b>45</b>
<b>Vedlegg B: Supplerende resultater fra analyser av forskjeller innen Stavanger-skoler («Regression Discontinuity»-modeller)</b> .....	<b>53</b>
<b>Figurregister</b> .....	<b>58</b>
<b>Tabellregister</b> .....	<b>59</b>



## 1. Innledning

Omtrent 30 prosent av elevene som starter videregående opplæring fullfører ikke i løpet av 5 år (Utdanningsdirektoratet, 2015). Dette er høyt sammenliknet med andre OECD land (OECD, 2014). De individuelle og samfunnsøkonomiske kostnadene ved frafall er betydelig (Lillejord m.fl., 2015; Falch m.fl., 2009). Å øke gjennomstrømningen i videregående opplæring (heretter VGO) står derfor høyt på den politiske dagsorden. Fullføring av VGO er vesentlig dårligere blant elever med svake faglige resultater fra grunnskolen (Falch m. fl., 2014; Utdanningsdirektoratet, 2015). Dette er en viktig motivasjon for utformingen av Overgangsprosjektet innen Ny GIV. Den overordnede målsetningen for Ny GIV er å øke andelen som fullfører og består VGO med 6 prosentpoeng, fra 69 prosent for 2004-kullet til 75 prosent for 2010-kullet. I tillegg er det målsetninger for gjennomstrømning underveis, bl.a. at overgang til Vg2 skal øke med 2 prosentpoeng fra 2010 til 2013.

Gjennom Overgangsprosjektet i Ny GIV får utvalgte elever intensivopplæring i grunnleggende lese-, skrive- og regneferdigheter. Målgruppen er de 10 prosent av elevene i hver kommune med svakest resultater ved første termin i 10. klasse. Elevene kan tas ut av den ordinære undervisningen i inntil 7,5 kløkketimer per uke, og intensivopplæringen skal starte tidlig i andre termin og vare ut skoleåret, ifølge Kunnskapsdepartementets invitasjonsbrev til kommunene. I forbindelse med intensivopplæringen er det også noe kursing av lærere, men utover denne kursingen får skolene ikke ytterligere ressurser. Hovedformålet med intensivopplæringen og de øvrige tiltakene i Ny GIV er å øke andelen i målgruppen som gjennomfører videregående. Innføringen av intensivopplæringen foregikk gradvis i tre faser. De første kommunene ble med skoleåret 2010/2011. Neste fase innførte intensivopplæringen skoleåret 2011/2012, mens det var forventet full deltakelse skoleåret 2012/2013.

Ansvar for å implementere ble gitt til fylkeskommunene. Selv om det i invitasjonsbrevet fra KD ble lagt en god del føringer på hvordan implementeringen skulle foregå, ble disse tolket ulikt av de ulike fylkene. Forankringsgraden av Ny GIV varierer derfor på tvers av fylker, skoleeiere, skoleledere og lærere. Dette førte til stor variasjon på tvers av fylkene når det gjelder fordeling av kommuner og skoler på ulike faser, hvordan elevene som fikk tilbud om intensivopplæring faktisk ble plukket ut, samt innhold og gjennomføring av intensivopplæringen. Intensivopplæringen er beskrevet i mer detalj i kartlegginger gjennomført av NOVA (Sletten m.fl., 2011) og NIFU (Holen og Lødding, 2012, 2013). I tillegg har Uni Rokkansenteret studert organisatoriske konsekvenser for skolene (Helgøy og Homme, 2013) og Nordlandsforskning gjennomført klasseromsstudier (Rønning m.fl., 2013). Denne rapporten er andre delrapport fra effektevalueringen av intensivopplæringen. Første delrapport (Eielsen m.fl., 2013) diskuterer tilordning av elever, samt presenterer de første effektestimaterne, men var ikke i stand til å konkludere klart om effekter.

Det finnes en del forskning som studerer effekten av intensivopplæring på elevprestasjoner, både på ungdomsskole- og videregående nivå. De fleste av disse finner at intensivopplæring har en effekt. Lavy og Schlosser (2005) studerer effekten av å gi ekstra undervisning til israelske elever på videregående skoler som presterer dårlig, og finner at dette øker skolens gjennomsnittlige uteksamineringsrate med omtrent 3,3 prosentpoeng. Cortes m.fl. (2014) ser på om en dobling av timetallet i matematikk i det første året av videregående i Chicago har en effekt på fullføring. Studien finner at blant elevene som fikk ekstraundervisning i matematikk var det omtrent 10 prosentpoeng flere som fullførte enn blant andre tilsvarende elever (fullføring økte fra ca. 60 til 70 prosent). Ved å benytte et randomisert kontrollert eksperiment der en veileder gir intensiv matematikkopplæring, finner Cook m.fl. (2015) at deltakelse i veiledningen har en sterk effekt på skoleprestasjoner. Matematikkresultatene økte med 0,2-0,3 standardavvik (i en

norsk karakterfordeling svarer det til omtrent like mange karakterpoeng), og andelen som strøk i matematikk ble halvert. De Haan (2012) benytter data for ungdomsskoleelever i Nederland, og ser på om det har en effekt å gi skolene en vesentlig ekstra finansiering per svaktpresterende elev på yrkesfaglige studieretninger.<sup>1</sup> Dette tiltaket viser seg å øke sannsynligheten for at disse elevene består eksamen med omtrent 4 prosentpoeng eller mer; prestasjoner i matte og lesing bedres også.

Blant de studiene som ser på effekten av intensivopplæring på fullføring av høyere utdanning, er resultatene sprikende. For eksempel finner Bettinger og Long (2009) at sannsynligheten for å fullføre en bachelorgrad innen fire år er høyere for studenter som mottar intensivopplæring i Ohio. På den annen side er det ingen effekt av intensivopplæring på fullføring blant college-studenter i Texas (Martorell og McFarlin, 2011). Vi er ikke kjent med tidligere norsk studier som identifiserer kausale effekter av å gi ekstra ressurser til de svakest presterende elevene på norske elevers læringsutbytte.

I denne rapporten presenterer vi foreløpige effektestimater for hvordan intensivopplæringen har påvirket karakterer og gjennomføring. Rapporten skiller seg fra første delrapport (se Eielsen m.fl., 2013) ved at vi nå har data for alle fasene. Dette gir en del ytterligere analysemuligheter sammenlignet med første delrapport, der vi bare hadde tidlige resultater for elever i fase en. Vi har nå mulighet til å studere resultater på lengre sikt, herunder fullføring av Vg2 og overgang til Vg3. Vi har også tre årganger med deltakende elever fordelt på tre faser. Dette betyr at vi kan studere resultatene til totalt omtrent 12000 elever som deltar i intensivopplæringen, mot bare 2000 i Eielsen m.fl. (2013).

Som i Eielsen m.fl. (2013) har vi to analysestrategier. Vår hovedanalyse baserer seg på den gradvise innfasingen av Ny GIV og har til hensikt å finne den kausale effekten av å innføre intensivopplæringen ved hjelp av en såkalt «difference-in-difference» analyse, heretter DiD analyse (Angrist og Pischke, 2009). Kort forklart innebærer denne analysen at vi kan studere eventuelle forskjeller i utvikling av elevresultatene mellom skoler i forskjellige faser. Under visse forutsetninger kan man tolke disse forskjellene som kausale effekter. I tillegg vil vi studere forskjeller innen Stavanger-skoler. Stavanger ser ut til å være den eneste kommunen som i stor grad har tilordnet elever til intensivopplæring på grunnlag av karakterprestasjoner under tiende persentil. Dette er ikke endret siden Eielsen m.fl. (2013). For Stavanger kommune er det derfor mulig å benytte en såkalt «regression discontinuity» analyse i tillegg (heretter RD analyse), der vi sammenlikner deltakende elever rett under karaktergrensen ved tiende persentil, med ikke-deltakende elever rett over karaktergrensen ved tiende persentil. På grunn av at tilordningen er litt uklart rundt karaktergrensen velger vi å utelate observasjoner i et lite intervall rundt grensen (såkalt «donut hole» i litteraturen) samt supplere RD analysen med en DiD analyse, heretter DiRD (forklares senere), noe som er vanlig å gjøre i analyser med «donut holes» (Leuven m.fl., 2007; Carneiro m. fl., 2014).

Til tross for det større datagrunnlaget finner vi ingen klare effekter i DiD-analysene der vi sammenligner skoler. Vi kan utelukke gjennomsnittlige effekter større enn 1 prosentpoeng på overgang til Vg2. Dette er en klar forbedring av presisjonen sammenlignet med Eielsen m.fl. (2013); vi er nå i stand til å finne effekter av en størrelse på omtrent 30-50 prosent av minste mulige effekt fra 2013-rapporten. Når det gjelder analysene innen Stavanger finner vi tegn til at intensivopplæringen kan ha en effekt på fullføring av Vg1 og overgang til Vg2 på omtrent 9 prosentpoeng. Som for DiD-analysene har vi en forbedret presisjon i analysene innen Stavanger, sammenlignet med tilsvarende analyser Eielsen m.fl. (2013). Ettersom det stadig

---

<sup>1</sup> Yrkesfaglige skoler får 4000 euro per svaktpresterende elev (definert ut fra et sett av kriterier), omtrent en av fire yrkesfagelever er omfattet av ordningen. Denne finansieringen kommer i tillegg til grunnfinansieringen på omtrent 7100 euro. Skolene har stor frihet i hvordan de bruker de ekstra ressursene, mange skoler velger å undervise de svaktpresterende elevene i egne, små grupper.

bare er Stavanger som har en klar grense, er antall deltagende elever omtrent tredoblet. Sammen med endringer i estimeringsmetode gir det høyere presisjon, slik at vi er i stand til å finne effekter omtrent halvparten så store som i Eielsen (2013).

Den overordnede målsetningen med Ny GIV er knyttet til fullføring av VGO. Vi finner ingen effekter på senere gjennomstrømning (fullføring av Vg2, overgang til og fullføring av Vg3, samt oppnådd kompetanse), heller ikke i Stavanger, men for disse resultatmålene kan vi foreløpig ikke trekke noen konklusjoner pga. et begrenset datamateriale. Et større datamateriale blir gradvis tilgjengelig ettersom elevenes gang gjennom VGO registreres. I sluttrapporten i 2017 er det et hovedmål å studere om det er tegn til effekter på fullføring generelt, og spesielt om effekten vi ser på tidlig gjennomstrømning i Stavanger gjenfinnes senere i elevenes utdanningsløp.

## 2. Datagrunnlag

Analysene baserer seg på et omfattende datagrunnlag, i hovedsak registerdata forvaltet av Statistisk sentralbyrå og Utdanningsdirektoratet. Vi bruker data fra følgende registerbaserte kilder:

- Norsk utdanningsdatabase (NUDB)
  - Fullførte og igangværende utdanninger
  - Karakterstatistikk for grunn- og videregående skole (fra VIGO)
  - Nasjonale prøver (NP)
- Skoledata fra Grunnskolens informasjonssystem (GSI)

Vi vil studere effekter på karakterer fra grunnskolen, samt på forskjellige mål for gjennomstrømning. Data kommer fra henholdsvis VIGO (karakterer) og NUDB (gjennomstrømning).

Vi har også sett på karakterer fra VGO, uten å finne noen effekter. Vi rapporterer ikke resultater fra disse analysene. Det er krevende å studere effekter på karakterer fra VGO. Når elever begynner i VGO velger de studieretning og fag innad i studieretning. Det er dermed svært forskjellig hvilke fag og karakterer elever har, og for mange fag er det få elever. Fagene er generelt ikke direkte sammenlignbare og kan også ha forskjellig karakternivå. Dessuten kan intensivopplæringen tenkes å ha en effekt ikke bare på prestasjoner i et gitt fag, men på valg av studieretning og fag. Dersom intensivopplæringen har en effekt på gjennomstrømning kan dette føre til at flere svaktpresterende elever oppnår karakterer i enkelte fag, slik at karakter-snittene i disse fagene trekkes ned. Dette gjør at det kan være effekter på faglige prestasjoner som er vanskelige å finne.

I motsetning til Eielsen m.fl. (2013) viser denne rapporten ikke estimerte effekter på mål for motivasjon fra Elevundersøkelsen (EU). Fram til og med 2012 ble EU gjennomført om våren. Fra og med 2013 er den obligatoriske gjennomføringen flyttet til høsten, dvs. før gjennomføringen av intensivopplæringen. Vi kan derfor ikke studere effekter på elevenes motivasjon for elevkull etter 2012-kullet.<sup>2</sup> Vi har undersøkt besvarelsene fra 2011 og 2012 uten å finne noen effekter av intensivopplæringen på elevenes motivasjon eller vurdering av i hvilken grad læreren er motiverende, men rapporterer ikke disse resultatene i mer detalj.

Vi har tilgang til forskjellige resultatmål fram til og med 2014. Det betyr at vi er i stand til å studere resultater fra fullført grunnskole samt tidlige resultater for gjennomstrømning i VGO for tre kull med intensivelever, som fullførte grunnskolen fra våren 2011 til våren 2013. Totalt studerer vi i denne rapporten resultatene til omtrent 12000 elever som deltar i intensivopplæring, fordelt på tre årskull og på skoler i fase en, to og tre. Dette har stor betydning for analysene vi er i stand til å gjøre, og presisjonen i disse. Vi får et datamateriale med flere elever, og viktigere, med flere elever (i skoler) som deltar i Overgangsprosjektet, enn Eielsen m.fl. (2013), som hadde data til og med 2012 og studerte omtrent 2000 elever i fase en-skoler som deltok i intensivopplæringen våren 2011.

I tillegg til at flere elevkull har deltatt i intensivopplæringen har elevene hatt anledning til å komme lengre gjennom VGO. Dette gjør at vi får flere mål på gjennomføring, f.eks. fullføring av Vg2 og overgang til Vg3. Vi har imidlertid fremdeles bare i veldig beskjeden grad mulighet til å studere faktisk fullføring av VGO; foreløpig kan vi bare studere fullføring innen tre år for elevene som deltok i intensivopplæringen våren 2011. I tabell 1 gir vi en oversikt og beskrivelse over utfallsvariablene vi benytter og hvilke avgangskull de er tilgjengelige for.

---

<sup>2</sup> Det er frivillig for skolene å gjennomføre EU om våren, et begrenset antall skoler gjør dette. Elevene svarer også på EU på Vg1. Ettersom vi bare kan koble resultater fra EU på skolenivå, ikke med de enkelte elevene, kan vi ikke bruke disse besvarelsene til å undersøke motivasjonen til elevene som deltok i intensivopplæringen.

Tabell 1. Utfallsvariable for kull 2009/2010 – 2012/2013

	Kull 2010/11 snitt/ant obs	Kull 2011/12 snitt/ant obs	Kull 2012/13 snitt/ant obs
Deltatt i intensivopplæring	0,031 (63 623)	0,075 (64 657)	0,093 (63 764)
Snitt standpunkt 10. trinn	4,029 (62 828)	4,032 (63 867)	4,044 (62 948)
Snitt skriftlig eksamen 10. trinn	3,456 (59 987)	3,406 (58 194)	3,393 (60 080)
Matematikk stp 10. trinn	3,548 (61 145)	3,513 (62 293)	3,505 (61 326)
Norsk hovedmål stp 10.	3,848 (60 921)	3,840 (61 923)	3,828 (61 055)
Fullført Vg1 t1	0,794 (63 623)	0,800 (64 657)	0,818 (63 764)
Fullført Vg1 t2	0,822 (63 623)	0,828 (64 657)	.
Fullført Vg2 t2	0,714 (63 623)	0,728 (64 657)	(0)
Fullført Vg2 t3	0,761 (63 623)	.	(0)
Fullført Vg3 t3	0,495 (63 623)	.	(0)
Registrert Vg2 t1	0,811 (63 623)	0,812 (64 657)	0,836 (63 764)
Registrert Vg2 t2	0,868 (63 623)	0,869 (64 657)	.
Registrert Vg3 t2	0,734 (63 623)	0,718 (64 657)	(0)
Fullført kompetanse t3	0,483 (63 623)	.	(0)
Antall observasjoner	63 623	64 657	63 764

Første rad i tabellen viser andelen elever registrert som deltagere i intensivopplæringen per kull. Som en del av evalueringen av Overgangsprosjektet har NOVA (Sletten, Bakken og Haakestad, 2011) og NIFU (Holen og Lødding, 2012, 2013) gjennomført kartlegginger av intensivopplæringen. Kartleggingene har blant annet, basert på skolenes registre, samlet data for hvilke elever som har deltatt i intensivopplæringen. Individdata for deltakelse i intensivopplæringen kommer fra disse kartleggingene.<sup>3</sup> Denne øker fra 3,1 prosent av elevene i avgangskullet 2010/11 til nesten 10 prosent for 2012/13 kullet, ettersom flere skoler og kommuner kommer i gang med intensivopplæring. Det er også gjennomført spørreundersøkelser til elever, lærere og skoleledere.

Alle resultatmålene fra fullført grunnskole (snitt standpunkt (stp), standpunkt-karakterene i matematikk og norsk hovedmål samt skriftlig eksamen) er tilgjengelige for alle avgangskullene. Standpunktkarakterene har også stabile snittverdiene over tid, mens snitt skriftlig eksamen er fallende, fra 3,46 for 2010/11 kullet til 3,39 for 2012/13 kullet.<sup>4</sup> Vi har termin én-karakterer for elever fra skoleåret 2006/07 (med noe varierende dekning over tid).<sup>5</sup> For å finne elever i

<sup>3</sup> Vi har bare data for registrert deltagelse. Vi har ingen informasjon om hvilke elever som er tilbudt deltagelse, men som takker nei. Resultatene fra spørreundersøkelsene tyder imidlertid på at dette gjelder et fåtall elever. Gitt at det ikke er feil i registreringen av elever skal det dermed være omtrent det samme om vi studerer elever som er tilbudt deltagelse eller som er registret å ha deltatt. Vi har også vurdert muligheten for å hente inn data om Ny GIV-deltagelse fra VIGO for senere årskull, men det viser seg at det er svært varierende i hvilken grad elever fortsatt er registrert som Ny GIV-deltagere. Sannsynligvis er intensivopplæringen avsluttet og/eller mange i sektoren opplever Overgangsprosjektet som avsluttet. Dette har imidlertid liten betydning for analysemulighetene, ettersom vi i liten grad har anledning til å studere nylige årskulls resultater gjennom VGO.

<sup>4</sup> Snitt skriftlig eksamen er regnet som snittet av skriftlige eksamenskarakterer i norsk hovedmål, matematikk og engelsk. Vi bruker snittkarakter framfor karakterer i enkeltfag for å få et større datamateriale og dermed mer presise resultater. Et stort flertall av elevene har bare skriftlig eksamenskarakter i ett av de tre fagene, slik at snittet svarer til denne karakteren. Et lite mindretall har ingen skriftlig eksamenskarakter, og noen svært få elever har skriftlig eksamenskarakter i to eller tre av fagene.

<sup>5</sup> F.eks. finnes ikke termin én-karakterer i VIGO for Rogaland før skoleåret 2011/12. Vi har fått levert slike fra Stavanger for skoleårene 2009/10 og 2010/11.

målgruppen for intensivopplæringen og vurdere tilordningspraksisen har vi tatt utgangspunkt i registrerte termin én-karakter og deltakere.

Når det gjelder fullføring av henholdsvis Vg1, Vg2 og Vg3 kan vi observere dette på forskjellige tidspunkt. «Fullført Vg1 t1» viser andelen som har fullført Vg1 i løpet av første kalenderår (typisk i vårsemesteret) etter året de avsluttet grunnskolen, dvs. til normert tid.<sup>6</sup> «Fullført Vg1 t2» viser andelen som har fullført Vg1 innen utgangen av andre kalenderår etter avsluttet grunnskole, denne omfatter de som fullførte til normert tid eller ett år senere. For eksempel, blant avgangskullet 2010/11 fullførte 82 prosent Vg1 i løpet av to år. Blant dem fullførte ca. 79 prosent til normert tid mens ca. 3 prosent trengte et ekstra år. Tilsvarende gir «fullført Vg2 t2» og «fullført Vg3 t3» andeler som har fullført Vg2 og Vg3 i løpet av henholdsvis andre og tredje kalenderår etter endt grunnskole, dvs. til normert tid, mens «fullført Vg2 t3» gir andelen som har fullført Vg2 i løpet av tre år etter endt grunnskole, dvs. inntil ett år på etterskudd. 71 prosent av elevene i 2010/11 kullet fullførte Vg2 til normert tid. Ytterligere 5 prosent trengte et ekstra år på å fullføre Vg2, slik at 76 prosent av 2010/11 hadde fullført Vg2 innen to år. Avgangskullet 2010/11 kan vi observere tre år etter fullført grunnskole. Selv for dette kullet kan vi imidlertid bare observere andelen som har fullført studiespesialisering på normert tid. For senere kull har vi færre gjennomstrømningsmål. Vi observerer imidlertid fullføring av Vg1 til normert tid (t1) for alle kullene. Andelen som fullfører Vg1 til normert tid har økt fra 79 prosent for 2010/11 til 82 prosent av elevene i 2012/13 kullet.

I tillegg til fullføring studerer vi også overgang til Vg2 og Vg3. Vi måler her hvorvidt eleven har vært registrert oppmeldt (per oktober) innen et gitt antall kalenderår etter fullført grunnskole. «Registrert Vg2 t1» viser andelen som er eller har vært registrert på Vg2 i det første kalenderåret etter fullført grunnskole. Som et eksempel, en elev som fullførte grunnskolen våren 2013 vil regnes som «fullført Vg1 t1» dersom han/hun har fullført Vg1 våren 2014 (og også «fullført Vg1 t2») og som «registrert Vg2 t1» (og også «registrert Vg2 t2») dersom han/hun er registrert i Vg2 høsten 2014. Derimot vil elever som trengte to år på å fullføre Vg1 og deretter startet i Vg2, være registrert i Vg2 først høsten 2015. Disse vil være «registrert Vg2 t2», men ikke «registrert Vg2 t1». «Registrert Vg3 t2» er andelen oppmeldt til Vg3 to kalenderår etter avsluttet grunnskole, dvs. som starter Vg3 til normert tid. Vi ser at andelen som er oppmeldt i Vg2 til normert tid øker fra 81 prosent for 2010/11 kullet til 84 prosent for 2012/13 kullet.

Den siste utfallsvariabelen vi benytter er andelen med «fullført kompetanse t3», dvs. andelen elever som har oppnådd yrkes- eller studiekompetanse tre kalenderår etter avsluttet grunnskole. I praksis vil dette være elever som har fullført studiespesialisering til normert tid. Som vist i tabellen gjelder dette nesten halvparten av elevene i 2010/11 kullet, for senere kull er det fremdeles for tidlig å studere dette. En stor del av målgruppen for Overgangsprosjektet vil sannsynligvis ikke velge studiespesialisering, slik at det også for 2010/11 kullet er vanskelig å finne effekter på oppnådd kompetanse.

Fra registerdataene henter vi også ut en del kontrollvariable. Dette er variable som kan tenkes å ha en effekt på elevens prestasjoner og gjennomstrømning, og også

---

<sup>6</sup> Merk at vi regner tid fra fullføring av grunnskole. Mål for gjennomstrømning i VGO tar vanligvis utgangspunkt i tidspunkt for påbegynt VGO, ikke fullført grunnskole. Vi ønsker imidlertid å ta utgangspunkt i elever på 10. trinn, da det er her intensivopplæringen gjennomføres. For effektevalueringen er vi heller ikke interessert i elever som ikke har gått i norsk grunnskole eller som har gått i grunnskolen lengre tilbake i tid, ettersom de uansett ikke har vært utsatt for intensivopplæringen. I prinsippet kan intensivopplæringen påvirke overgang fra grunnskole til VGO, i så fall kunne vi gått glipp av interessante effekter ved å ta utgangspunkt i elever i VGO. Ettersom et stort flertall av elevene starter i VGO rett etter fullført grunnskole (mer enn 95 prosent, se f.eks. Eielsen m.fl., 2013) vil dette imidlertid ha mindre praktisk betydning om vi ser på alle elever som fullfører grunnskole eller begrenser oss til de som starter rett i VGO.

være forskjellige mellom elever/skoler/kommuner som henholdsvis deltar / ikke deltar i intensivopplæringen på et gitt tidspunkt. På individnivå:

- Kjønn
- Nasjonale prøver 8. trinn (f.o.m. elevene som fullførte grunnskolen våren 2010)
- Mor og fars utdanning (målt som nivå høyeste fullførte utdanning, grunnskole e.l., påbegynt videregående, lavere og høyere grad fra høyskole eller universitet)
- Hvorvidt eleven er innvandrer eller norskfødt barn av to innvandrere
  
- For skoler har vi (fra GSI):
- Antall elever på 10. trinn
- Antall gutter/jenter med enkeltvedtak
- Andelen ukvalifiserte lærere
- Lærertimer til henholdsvis ordinær undervisning og spesialundervisning per elevtime (det første svarer omtrent til 1/Gruppestørrelse 2)

I tabell 2 presenterer vi snittverdier for relevante kontrollvariable separat for skoler i de ulike fasene. Elever ved skoler i de ulike fasene ser ut til å score omtrent likt på nasjonale prøver i 8. trinn. Andelen jenter er konstant på tvers av faser, og det samme er mors og fars utdanning.<sup>7</sup> Derimot er andelen elever med innvandrerbakgrunn høyere i første fase enn i senere faser. Andelen elever med enkeltvedtak er imidlertid større i fase tre enn i fase en og to. Gruppestørrelsen er størst i fase en og minst i fase tre, også ressursinnsats til spesialundervisning (per elevtime) og andelen ukvalifiserte lærer øker fra fase en til fase to og fra fase to til fase tre. Fra antall elever og skoler ser vi også at skolene i fase en er størst (i snitt omtrent 330 elever), skolene i fase to mindre (omtrent 240 elever) og skolene i fase tre minst (omtrent 160). Det virker som store kommuner og skoler er klart overrepresentert i fase en, og underrepresentert i fase tre. Dette kan ha sammenheng med rask igangsetting av prosjektet, slik at det ble vurdert som praktisk med store skoler i fase en for å få med mange elever for et gitt antall skoler. Forskjeller i f.eks. innvanderandeler og gruppestørrelse følger så naturlig av dette. Eielsen m.fl. (2014) diskuterer forskjellene mellom skoler i mer detalj, og viser bl.a. geografisk fordeling av fase en-kommuner. Eielsen m.fl. studerer også betydningen av skole- og kommunestørrelse. De finner at det er betydelige forskjeller i størrelse, og at disse forskjellene i stor grad forklarer øvrige forskjeller.

**Tabell 2. Beskrivende statistikk for avgangselever 2006/07-2009/10**

	(1) Fase 1 Snitt	(2) Fase 2 Snitt	(3) Fase 3 Snitt
<i>Elevnivå</i>			
Snitt Nasjonale prøver 8. trinn (standardisert) ...	-0,030	-0,003	-0,013
Jente .....	0,485	0,486	0,485
Mors utdanning .....	4,161	4,180	4,142
Fars utdanning .....	4,211	4,180	4,130
Norskfødte med innvandrerforeldre .....	0,066	0,024	0,018
Innvandrere .....	0,062	0,048	0,041
Norskfødte med to norskfødte foreldre .....	0,872	0,928	0,940
<i>Skolenivå</i>			
Andel elever med enkeltvedtak .....	0,106	0,113	0,136
Ordinære lærertimer/elevtimer .....	0,053	0,062	0,077
Lærertimer spesialundervisning/elevtimer .....	0,017	0,019	0,030
Andel ukvalifiserte lærere .....	0,031	0,040	0,051
Antall skolekommuner .....	50	183	211
Antall skoler .....	206	392	418
Observasjoner .....	68 625	95 862	68 408

<sup>7</sup>Mors og fars utdanning er målt som gjennomsnittet av en variabel definert i 9 kategorier, der laveste kategori er ingen skolegang, og høyeste kategori indikerer doktorgrad. Høyere gjennomsnitt for disse variablene indikerer mer skolegang.

Det er noe usikkerhet knyttet til fordelingen av skoler på fase to og tre. Fase to kan forstås på to måter: Som hvilken fase skolene er tilordnet (i KDs lister), eller som når skolene først har registrert elever med intensivopplæring. For fase to ser det ut til å være et betydelig avvik, en rekke skoler (og kommuner) som er registrert som fase to-skoler er ikke registrert med noen intensivopplæring våren 2012.<sup>8</sup> Dette kan henge sammen med at det i fase to er flere mindre skoler (og kommuner). Disse kan i noen tilfeller ha vurdert det til at de ikke har noen elever i målgruppen for intensivopplæringen i et gitt årskull.<sup>9</sup> I denne rapporten vil vi bruke fase som registrert i KDs lister, framfor når det er registrert elever. Dette kan gi mer troverdige effektestimater dersom det er systematikk i hvilke skoler som ikke starter med intensivopplæring (eller ikke registrerer at elever deltar) til tross for at skolen er valgt ut til tiltaket. Vi har imidlertid undersøkt dette i noe detalj og funnet ut at resultatene fra effektevalueringen er uavhengige av om vi tilordner skolene til fase utfra KDs lister eller utfra når vi første gang observerer at elever mottar intensivopplæring.

Holen og Lødding (2012) skriver at det er uklart hvilke Oslo-skoler som tilbyr intensivopplæring, at Oslo ikke har noe klart skille mellom fase én-skoler og andre, men at alle skoler følger samme prinsipper på et gitt tidspunkt. Eielsen m.fl. (2013) omtaler også dette, samt at det er tegn til noen registreringsfeil i gjennomstrømningsdata fra Østfold. Dette kan forkludre våre sammenligninger av skoler i forskjellige faser, og vi gjør derfor robusthetsanalyser der vi utelater Østfold- og Oslo-skoler i sammenligninger mellom skoler og årskull.

Eielsen m.fl. (2013) diskuterer utvelgelsen av elever innen skoler i stor detalj med tanke på konsekvenser for effektevalueringen. Ikke alle de svakest presterende elevene ved tiltaksskolene deltar, og det er mange elever utenfor de 10 prosent svakest presterende som deltar. I sum virker det som elevene til dels er valgt ut på uklart grunnlag, og også på forskjellig grunnlag i forskjellige fylker, kommuner og skoler.<sup>10</sup> Generelt sett er det ikke stor forskjell i andelen deltagere blant elever som så vidt er/ikke er blant de 10 prosent svakestpresterende. Eielsen m.fl. utvikler en søkerutine for finne skoler og kommuner med en klar avgrensning av deltakerne basert på elevenes resultater fra første termin, men finner bare dette for skoler i Stavanger, og fokuserer derfor på disse skolene i analysene rundt grensen for deltagelse.

Også for 2012- og 2013-kullene er det sterkt varierende og til dels lite transparent praksis når det gjelder utvelgelse av elever, til tross for at KD presiserte i løpet av 2012 at det er svaktpresterende elever som skal delta i intensivopplæringen.<sup>11</sup> Vi har arbeidet videre med søkerutinen fra Eielsen m.fl. (2013), men ikke vært i stand til å finne andre kommuner med klar tilordningspraksis når vi har studert 2012- og 2013-kullene. Vi vil derfor fokusere på Stavanger i analysene innen skoler også i denne rapporten.

---

<sup>8</sup> Andre avvik finnes (fase én-skoler som ikke har deltagere våren 2011 eller skoler som har registrerte deltagere før de skal ha startet), men i et ubetydelig omfang. Se også Eielsen m.fl. (2013).

<sup>9</sup> Dette er også omtalt i kartleggingene, se Holen og Lødding (2012, 2013).

<sup>10</sup> Utvelgelsen studeres også i kartleggingene (Sletten, Bakken og Haakestad, 2011; Holen og Lødding, 2012, 2013). En forklaring på manglende sammenheng mellom resultat fra første termin er at bl.a. elever med individuell opplæringsplan (IOP) ofte unntas fra deltagelse. Vi har i analysen av deltagelse tatt registreringen av deltagere for gitt, det er imidlertid også mulig at det er feil i denne registreringen. Dette er i så fall vanskelig å ta hensyn til.

<sup>11</sup> Se Holen og Lødding (2013).



### 3. Effektevaluering – empirisk strategi og resultater

For å kunne vite i hvilken grad et tiltak har virket etter hensikten, er vi nødt til å beregne den kausale effekten av tiltaket. Den kausale effekten er forskjellen mellom det faktiske og kontrafaktiske resultatet - det som ville skjedd i fravær av et tiltak. Vi kan aldri observere det kontrafaktiske resultatet. Dette må derfor estimeres, noe som kan være utfordrende. Utfordringen består i å finne en god kontrollgruppe: En gruppe (her elever) som ikke deltar i tiltaket, men som ellers er like de som deltar. Dersom gruppene faktisk er like på denne måten, vil resultatet til kontrollgruppen svare til det kontrafaktiske resultatet til tiltaksgruppen, og vi kan finne den kausale effekten ved å sammenligne resultatene i gruppene. I praksis er det ofte vanskelig å finne gode kontrollgrupper. Grunnen til at noen elever deltar i tiltak og andre ikke vil ofte være nettopp at de er forskjellige.

Et *randomisert kontrollert eksperiment* (randomized controlled trial, RCT) er gullstandarden for å sikre en gyldig kontrollgruppe. I RCT fordeles deltakerne til tiltak og kontroll ved loddtrekning eller lignende. Vi sikrer oss dermed at det ikke er noen systematiske forskjeller mellom tiltaks- og kontrollgruppen i utgangspunktet. Enhver forskjell etter tiltaket vil derfor være en effekt av tiltaket.

I mange tilfeller er det vanskelig å gjennomføre et eksperiment, for eksempel fordi tiltaket allerede er innført, men uten en eksperimentell utforming. Da er vi avhengige av å bruke andre metoder for å finne effekter. Hvilke metoder som er relevante og i hvilken grad resultatene gir troverdige effektestimater vil avhenge av hvordan tiltaket er utformet. Grunntanken i alle disse metodene er imidlertid å finne en sammenligningsgruppe som er mest mulig lik tiltaksgruppen, eller der det er mulig ta hensyn til hvordan gruppene er forskjellige.

Målet med vår effektevaluering er å fastslå om intensivopplæring til de svakest presterende elevene har en kausal effekt på gjennomstrømning i videregående og på læringsutbytte. Vi vil studere dette ved å sammenligne forskjeller mellom faser og årskull og ved å studere forskjeller innad i Stavanger-skoler. Nedenfor diskuterer vi de to ulike estimeringsstrategiene i mer detalj. Vi har valgt å plassere mer teknisk i omtale i «bokser» utenom hovedteksten. Det skal være fullt mulig å forstå rapporten uten å lese disse. Disse er imidlertid mer matematiske og mer presist formulert.

#### 3.1. Forskjeller mellom alle faser og årskull («difference-in-differences», DiD)

Vi presenterer først vår hovedstrategi for å finne den generelle effekten av tiltaket, den gjennomsnittlige effekten av å innføre intensivopplæringen i skoler over hele landet. Denne analysen baserer seg på sammenligning av utvikling over tid mellom skoler i forskjellige faser, dvs. forskjeller mellom faser i forskjellige årskull. I internasjonal faglitteratur omtales dette gjerne som «difference-in-differences», vi vil i fortsettelsen kalle disse analysene for DiD-analyser. En viktig fordel med DiD-analysene er at vi kan bruke hele datamaterialet; både alle deltagende elever og skoler samt elever og skoler som ikke deltar, i tillegg til årganger fra før Overgangsprosjektet ble innført. Dette gir et omfattende datamateriale, og dermed et grunnlag for presise analyser.

En annen viktig fordel er at vi får et direkte mål på den gjennomsnittlige effekten av å innføre intensivopplæring på en skole, i betydningen at en skole er tilordnet til intensivopplæring. Dette er et tiltak overordnede myndigheter (f.eks. Kunnskapsdepartementet kan gjennomføre). Effekten vil fange hvordan tiltaket faktisk er innført og gjennomført av prosjektledelse, skoleeier og skoleledelse. Dersom disse ikke faktisk bidrar til at det blir endringer i tilbudet til elevene, kan vi ikke vente å finne noen effekt. Dersom implementeringen varierer vil vi finne en

gjennomsnittlig effekt, som vil kunne være mindre enn effekten i de kommunene og skolene som i størst grad lykkes med gjennomføringen.

Det er også to viktige begrensninger i DiD-analysene. For det første baserer analysene seg på en antagelse om kontrafaktisk utvikling, at resultatene i tiltaks-skolene ville ha utviklet seg videre på samme måte som i øvrige skoler dersom tiltaket ikke ble innført. Vi mener denne antagelse er rimelig, utfra hvordan historisk utvikling har vært og ettersom vi ikke kjenner til samtidige relevante endringer, men antagelsen er vanskelig å teste.

For det andre sammenligner vi resultater på skolenivå, dvs. snittresultater til skoler. Det vil si at vi mister mange detaljer knyttet til resultatforskjeller innen skoler. Dessuten vil eventuelle effekter av tiltaket kunne «fortynnes». Ettersom bare omtrent 10 prosent av elevene deltar i intensivopplæringen vil en betydelig effekt som er konsentrert til deltakerne føre til en effekt på snittresultatene til skolen som er mye mindre.

I det følgende går vi først gjennom ideen bak DiD-analysene, antagelser, tolkningsmuligheter, styrker og svakheter, og illustrerer med en uformell grafisk analyse av effekter på standpunkt karakterer fra grunnskolen. Deretter viser vi estimerte effekter på en rekke forskjellige resultatmål, samt diskuterer forskjellige mulige former for heterogenitet i effektene.

### Analysesstrategi og grafisk analyse

Ved hjelp av den gradvise innfasingen av intensivopplæringen kan vi sammenlikne forskjellige årskull innen samme skole/fase og skoler i forskjellige faser i et gitt år. En DiD-analyse utnytter slik tids- og tverrsnittvariasjonen i implementeringen. En intuitiv måte å starte på er å ta for seg alle skolene som implementerte intensivopplæring på tidspunkt  $t$  og sammenlikne resultatene  $Y$  til elevene som gikk i 10. trinn ved skole  $s$  før ( $t-1$ ) og etter ( $t$ ) implementeringstidspunktet:

$$Y_{s,t}^{tiltak} - Y_{s,t-1}^{tiltak}$$

Dersom vi finner at  $Y_{s,t}^{tiltak} > Y_{s,t-1}^{tiltak}$ , kan det gjenspeile endringer i skolens resultater i forbindelse med innføringen av intensivopplæringen. Men det kan også skyldes andre forhold, for eksempel at eksamen i år  $t$  var lettere enn eksamen i år  $t-1$ , lengre og mer generelle trender knyttet til lærernes arbeidsmarked, at de som jobber ved skolen får mer erfaring, endringer i bolig mønstre og dermed elevsammensetning eller endringer i skoleleder eller skoleeiers vektlegging og prioritering.

En måte å ta hensyn til at andre forhold kan påvirke resultatet er å benytte skoler som ennå ikke har implementert intensivopplæring på tidspunkt  $t$  som sammenligningsgruppe. Vi kan da sammenligne resultatutviklingen i tiltaksgruppen rundt implementeringen av tiltaket med resultatutviklingen i sammenligningsgruppen for den samme tidsperioden:

$$(1) \quad (Y_{s,t}^{tiltak} - Y_{s,t-1}^{tiltak}) - (Y_{s,t}^{kontroll} - Y_{s,t-1}^{kontroll})$$

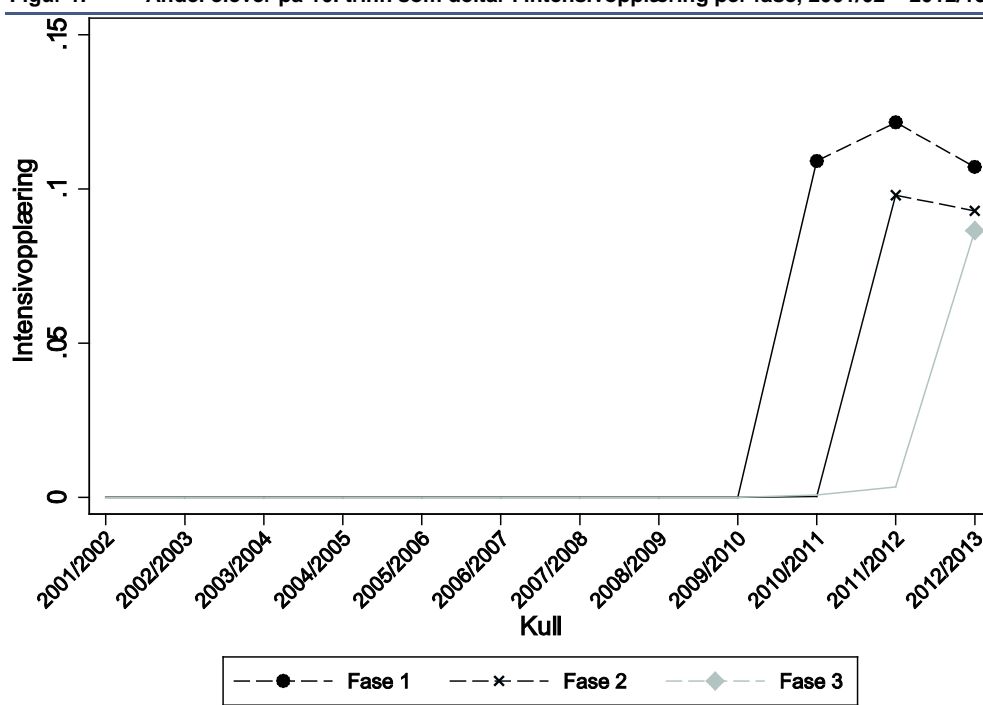
Vi viser ligningene vi estimerer i tekstboksen. På denne måten kan vi finne den kausale effekten av intensivopplæringen, såfremt utviklingen i fravær av tiltaket ville vært tilsvarende i tiltaks- og sammenligningsgruppen. Dette er noe vi må anta, ettersom vi ikke observerer resultatutvikling i fravær av tiltaket. Vi kan imidlertid argumentere for at dette er en rimelig antagelse, f.eks. dersom det ikke skjer andre endringer på samme tidspunkt som påvirker tiltaks- og sammenligningsskolene forskjellig og som ikke henger sammen med intensivopplæringen. Vi kjenner ikke til andre slike forhold rundt innføringen av intensivopplæringen. Vi kan også

vurdere om tiltaks- og sammenligningsgruppene har utviklet seg likt tidligere, før noen hadde satt i gang intensivopplæring. Dersom dette er tilfellet, og vi ikke kjenner til andre endringer som sammenfaller i tid og påvirker tiltaks- og sammenligningsgruppen forskjellig, er det en rimelig antagelse at tiltaks- og sammenligningsgruppen fortsatt ville hatt tilsvarende utvikling i fravær av tiltaket.

Med andre ord er det viktig at trendene i tiltaks- og sammenligningsgruppen har fulgt hverandre parallelt i forkant av implementeringstidspunktet. Dersom skolene var tilfeldig fordelt til tiltak og kontroll ville både utvikling og resultatnivå før tiltaket være likt i tiltak- og kontrollgruppen. Vi er imidlertid ikke avhengig av at skolene er tilfeldig fordelt til deltakelse eller av at resultatnivået i utgangspunktet var likt for å gjøre en effektevaluering, bare av at forskjellen som ville ha vært mellom dem i fravær av intensivopplæringen svarer til historiske forskjeller, som vi kan ta hensyn til.

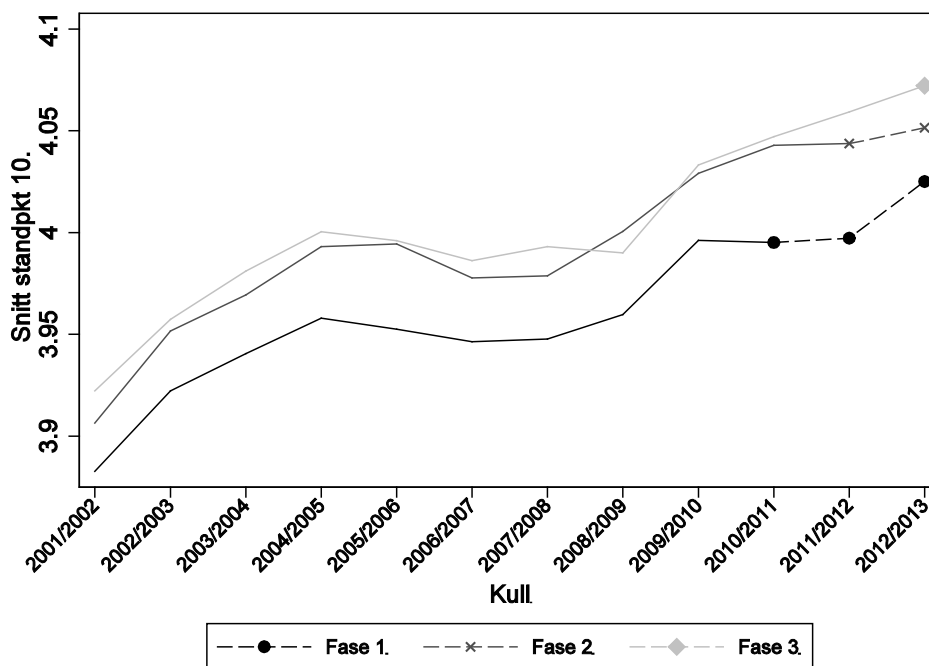
Figur 1 viser utviklingen i andelen av elever som deltar i intensivopplæringen. Som ventet ser vi at andelen går fra 0 til omtrent 10 prosent for skolene i fase 1 i skoleåret 2010/11, og deretter holder seg stabil. For fase 2 ser vi samme utvikling fra skoleåret 2011/12, og for fase 3 fra skoleåret 2012/13. Grunnet innfasingen over tid, vil fase 1 skoler dermed ha innført intensivopplæring i tre kull (2010/2011, 2011/2012, 2012/2013). Fase 2 skoler vil ha to kull (2011/2012, 2012/2013) og fase 3 skoler vil ha kun ett kull (2012/2013). En effekt av intensivopplæringen skulle dermed føre til tilsvarende endringer, først i fase 1 fra og med 2010/2011-kullet, deretter i fase 2 og til slutt i fase 3 fra 2012/2013-kullet.

Figur 1. Andel elever på 10. trinn som deltar i intensivopplæring per fase, 2001/02 – 2012/13



I Figur 2 viser vi utviklingen til ett utfall, standpunkt karakterer for 10. klasse (i vedlegg A.1 viser vi tilsvarende figurer for alle utfallene vi studerer). Som i Figur 1 viser vi utviklingen separat for hver fase. Kull som er markert med punkter har implementert intensivopplæringen, mens kull som er bundet sammen av sammenhengende linjer, er i skoler som ikke har startet ennå. Vi ser at selv om standpunkt karakterene i snitt er bedre etter oppstart av intensivopplæringen enn før, er dette en generell utvikling. Det er ikke noen klar tendens til at resultatene bedres ved oppstart av intensivopplæringen.

Figur 2. Snitt standpunktkarakter 10. klasse per fase, 2001/02 – 2012/13



Siden det er vel så store endringer på andre tidspunkter, er det åpenbart andre forhold som også påvirker standpunktkarakterer. Vi ser imidlertid at snittresultatene i fase en, to og tre følger hverandre tett før Overgangsprosjektet, og en statistisk test bekrefter at det ikke er klare forskjeller. Dette tyder på at forutsetningene for DiD-analyser, tilsvarende utvikling i fravær av tiltak, er oppfylt slik at vi kan bruke denne analysemetoden.<sup>12</sup> Vi kan dermed sammenligne endringen fra 2009/10 til 2010/11 i fase en med den samtidige endringen i fase to og tre, og gjøre tilsvarende sammenligninger av endringene fra 2010/11 til 2011/12 og fra 2011/12 til 2012/13 (når henholdsvis fase to og fase tre starter tiltaket).<sup>13</sup> Vi finner at det ikke er påfallende endringer i gruppen av skoler som starter tiltak, sammenlignet med samtidige endringer i andre skoler, altså finner vi ingen tegn til effekter av intensivopplæringen på standpunktkarakterer.

En måte å overføre grafene i figuren til ligningen ovenfor er å tenke seg at fase 1 skolene som innfører intensivopplæring skoleåret 2010/11 utgjør tiltaksgruppa, mens fase 2 skolene som innfører intensivopplæring i 2011/12, altså ett år senere, utgjør sammenligningsgruppa. Den første differansen i ligningen finner vi ved å ta differansen i elevprestasjoner mellom 2010/11 og 2009/10 for tiltaksgruppa, mens den andre differansen utledes av å gjøre det samme for sammenligningsgruppa for tilsvarende skoleår. Tilsvarende kan vi sammenlikne fase 2 skoler med fase 3 skoler og ta differansen i elevprestasjoner mellom skoleårene 2011/12 og 2010/11. I dette tilfellet er det fase 2 skolene som utgjør tiltaksgruppen, mens fase 3 skolene utgjør sammenligningsgruppen.

<sup>12</sup> Vi ser imidlertid at det er forskjeller i resultatnivået i fase en, to og tre før overgangsprosjektet. Men siden disse forskjellene er stabile kan vi enkelt ta hensyn til de i analysen. Eielsen m.fl. (2013) sammenligner også skolene i fase 1 med matchede, lignende skoler i fase 2 og 3. Det er imidlertid lite å vinne på dette når fasene uansett har en tilsvarende utvikling før oppstart av intensivopplæringen.

<sup>13</sup> Det ville derimot ikke vært informativt å se på utviklingen fra 2012/13 til 2013/14, fordi ingen skoler starter med intensivopplæring. Dermed er vi heller ikke i stand til å skille mellom effekter av tiltaket eller andre samtidige endringer. Vi kan altså bare studere effekter ved hjelp av DiD-analyser på tidspunkt da noen skoler begynner (eller slutter) med tiltaket. I praksis vet vi ikke om registrering av deltakelse er tilstrekkelig systematisk til å se effekter av at skoler eventuelt slutter med tiltaket fra 2012/13 til 2013/14. Vi vet heller ikke om det evt. er spesielle kjennetegn ved skolene som slutter, f.eks. at disse har dårligere erfaring med intensivopplæringen enn andre skoler.

I DiD-analysene studerer vi effekter på snittresultatene til alle elever (i en gitt fase og et gitt årskull), både elever som deltar i intensivopplæring og de som ikke gjør det. Intensivopplæringen kan både ha en effekt på deltakende elever (direkte effekt) og ikke-deltakende elever på skoler som tilbyr intensivopplæring (ringvirkninger). Ringvirkninger kan oppstå dersom intensivopplæringen førte til endringer i gruppesammensetning på tiltaksskoler. Dersom de svakeste elevene tas ut av undervisningen vil øvrige elever motta undervisning i mer homogene og bedre-presterende grupper. I tillegg kan intensivopplæringen ha betydning for hvilke lærere som underviser hvilke elevgrupper. Dersom de «gode» lærerne systematisk blir plassert til å gi intensivopplæring, vil det potensielt kunne ha en negativ effekt på ikke-deltakende elever. Det kan dermed være både positive og negative ringvirkninger til øvrige elever.

Dersom vi presist kunne forutsagt hvilke elever som deltok i intensivopplæring kunne vi sammenlignet med tilsvarende elever i skoler uten intensivopplæring, og slik skilt direkte effekter og ringvirkninger. Dette er ikke mulig, ettersom vi ikke finner noen form for systematikk i tilordningen av elever til intensivopplæring på de enkelte skolene. Det vil si at den effekten vi estimerer er totaleffekten av å innføre intensivopplæring, bestående av både den direkte effekten og ringvirkningene. Dersom det ikke er ringvirkninger vil vi få et fortynnet estimat av den direkte effekten. Effekten på deltakerne vil i så fall være omtrent ti ganger større enn effekten på snittet for alle elever, ettersom omtrent ti prosent av elevene deltar.

**Boks 1**

Mer presist estimerer vi følgende DiD-likning:

$$Y_{ijt} = \beta \times \text{tiltak}_{it} + \text{kull}_t + \text{fase}_j + \varepsilon_{jit}, \quad (1.1)$$

der  $\text{kull}_t$  og  $\text{fase}_j$  er faste effekter for henholdsvis kull og fase. Tiltak er lik 1 dersom skole  $s$ , kull  $t$  har registrert deltakelse i Ny Giv, ellers 0. Standardfeil kløstres på kommune. Det vil si at vi i beregningen av standardfeil tar hensyn til at det kan være forhold som er felles for elever i samme kommune, som fører til at restleddene i (1.1) er korrelert innen kommuner. Punkttestimat og robuste standardfeil for  $\beta$  vises i Tabell 3.

I en alternativ spesifisering inkluderes skolefaste effekter samt karakterer fra nasjonal prøve i åttende trinn. Dette har liten betydning for resultatene, som finnes i vedlegget. Vi har også forsøkt kontroller for familiebakgrunn, som heller ikke påvirker resultatene særlig.

I Tabell 4 har vi estimert effekten av intensivopplæring for hver desil i karakterfordelingen fra første termin:

$$Y_{ijdt} = \beta(\text{tiltak}_{it} \times \text{desil}_d) + \text{desil}_d + (\text{kull}_t + \text{fase}_j) \times \text{desil}_d + \omega_{ijdt}, \quad (1.2)$$

der  $\text{desil}_d$  er avgrenset til fem kategorier: 1. desil, 2. desil, 3. desil, 4. desil og 5.-10. desil (kategori 5). Alle koeffisienter, inkludert konstantledd, er estimert separat for hver kategori.

Samme modell brukes til å studere heterogene effekter for kjønn (binær), innvandringskategori (3 kategorier), foreldres utdanningsnivå (4 kategorier der lavere nivå tilsvarer høyere utdanning) og fylke (19 kategorier). Se vedlegg for disse resultatene.

**Resultater: Beregnede effekter av innføringen av intensivopplæringen**

Tabell 3 rapporterer resultatene fra vår hovedanalyse. Kolonnene viser de ulike utfallsmålene. Variabelen Ny GIV indikerer om en elev er registrert å delta, resten av utfallsmålene er omtalt i kapittel 2. Andre rad viser hvordan de forskjellige utfallene i snitt endres ved en skole når denne innfører intensivopplæring, relativt til samtidige endringer i skoler som ikke innfører intensivopplæring. Under forutsetningene diskutert over kan vi tolke disse forskjellene som effekter av å innføre intensivopplæring på de ulike utfallsmålene. I tillegg presenterer vi antall observasjoner, og gjennomsnittsverdien for utfallet.

Pre-trend gir p-verdier fra en test for hvorvidt utviklingen før innføringen er parallell i de forskjellige fasene. Forutsetningen om like tidstrender må holde for at vi kan stole på estimatene fra denne modellen. P-verdien sier om det er tydelige forskjeller i utviklingen før noen skoler innfører intensivopplæring. Verdier under 0,05 tyder på at det har vært systematiske forskjeller i utvikling. Oppsummert er det lite tegn til systematiske forskjeller, noe som tyder på at analysestrategien gir relevante estimater på kausale effekter. Det er likevel tegn til tidligere systematiske forskjeller for fullføring av Vg2. Vi har ingen grunn til at dette resultatmålet skulle skille seg ut. Ettersom vi gjør et stort antall tester vil vi imidlertid forvente noen tester som tyder på systematiske forskjeller, selv i en situasjon der det ikke er slike forskjeller.

Første kolonne viser, som forventet, at andelen elever ved skoler og kull som har innført intensivopplæring øker med omtrent 10 prosentpoeng. Ingen andre forskjeller er statistisk signifikante; ikke for noe resultatmål er det forskjeller større enn de vi venter utfra observert variasjon. Vi finner dermed ikke tegn til noen effekter av intensivopplæringen. I tillegg til resultatmålene i Tabell 2 har vi også gjort separate analyser av standpunktkarakterer i matematikk og norsk hovedmål, uten å finne effekter på disse heller.<sup>14</sup>

For å få en idé om hvor store effektene er, og hvor store effekter vi kan utelukke, kan lager vi et 95 prosent-konfidensintervall for estimatene. Dette er en måte å oppsummere usikkerheten i resultatene. Hvis vi gjør mange analyser vil vi forvente at et slikt intervall i 19 av 20 tilfeller (95 prosent) inneholder den faktiske sammenhengen. Jo mindre konfidensintervallet er, jo mer presise er estimatene (dvs. jo større er de sammenlignet med variasjon vi ikke kan forklare), og jo mindre usikre vil vi være på sann effekt. Effektestimatet for snitt standpunkt (som er det mest presise estimatet av effekter på karakterer)<sup>15</sup> er -0,013, altså en negativ effekt på standpunktkarakterene. Med en beregnet standardfeil på 0,008, gir dette oss et 95 prosent konfidensintervall på (-0,029, 0,003). Det vil si at det er mer sannsynlig at innføring av intensivopplæringen har redusert standpunktkarakterene enn at det har økt disse, men vi har ikke noe grunnlag for å hevde at noen av delene har skjedd. Intensivopplæringen kan ha redusert standpunktkarakterene med inntil 0,029 karakterpoeng, men den kan også få ført til en marginal økning.<sup>16</sup>

---

<sup>14</sup> Eksamenskarakterene i matematikk og norsk gir vesentlig mindre datamateriale, som gjøre at estimatene blir mer upresise. Som omtalt i kapittel 2 har vi også undersøkt karakterer fra VGO samt resultater fra Elevundersøkelsen. Begge disse datakildene gir særskilte utfordringer for effektevalueringen, vi finner heller ikke noen effekter på disse.

<sup>15</sup> Det er argumenter mot å se på effekter på standpunkt også. Standpunktkarakterer settes av elevenes lærere, og lærerens oppfatning av elevens ferdigheter kan bli påvirket av andre forhold enn de faktiske prestasjonene, se f.eks. Galloway m.fl. (2011). Vi kan til og med tenke oss at det at eleven deltar i intensivopplæringen påvirker lærerens vurdering, uavhengig av effekt på ferdigheter. Dette problemet kan vi unngå ved å studere skriftlige eksamenskarakterer, ettersom disse rettes anonymt. Standpunktkarakterer predikerer imidlertid fullføring av VGO bedre enn eksamenskarakterer (jf. Eielsen m.fl., 2013). Eksamenskarakterer baserer seg imidlertid på mye mindre informasjon, og er mer utsatt for tilfeldig variasjon. Dette gjør også at effektestimatene blir mindre presise. Konfidensintervallet for effekten på skriftlig eksamen blir (-0,025, 0,027), svarende til en direkte effekt på omtrent (-0,25, 0,27) i fravær av ringvirkninger. Når vi er interessert i å se hvor presise estimatene kan være er det derfor interessant å se på standpunktkarakterer.

<sup>16</sup> Konfidensintervallet =  $[-0,013 - 1,96 \cdot 0,008, -0,013 + 1,96 \cdot 0,008]$ . Intervallet er senteret rundt punkttestimatet -0,013. Videre ser vi at konfidensintervallet omfatter 0, det er et uttrykk for at estimatet ikke er statistisk signifikant på 5 prosent-nivå.

Tabell 3. Effekten på elevprestasjoner av intensivopplæring, difference-in-differences estimater

	Ny Giv	Snitt standpkt	Snitt skriftlig eks 10.	Fullført Vg1 t1	Fullført Vg1 t2	Fullført Vg2 t2	Fullført Vg2 t3	Fullført Vg3 t3	Fullført kompetanse t3	Registrert Vg2 t1	Registrert Vg2 t2	Registrert Vg3 t1
Tiltak .....	0,104*** (0,0086)	-0,013 (0,0080)	0,001 (0,0132)	-0,002 (0,0035)	-0,001 (0,0034)	0,003 (0,0039)	0,004 (0,0059)	-0,006 (0,0071)	-0,005 (0,0071)	0,005 (0,0034)	0,004 (0,0030)	-0,002 (0,0038)
Observasjoner	681 636	676 715	642 743	681 636	622 135	622 135	561 717	561 717	561 717	681 636	622 135	622 135
Snitt utfall ...	0,0181	3,989	3,452	0,793	0,822	0,699	0,745	0,486	0,475	0,800	0,861	0,716
Pre-trend ...	.	0,995	0,545	0,665	0,798	0,0311	0,00545	0,693	0,434	0,958	0,255	0,763
Kommuner	405	405	405	405	405	405	405	405	405	405	405	405

Merknad: Robuste standardfeil kløstret på kommunenivå i parentes. Stjerner angir statistisk signifikans: \* p<0,05; \*\* p<0,01; \*\*\* p<0,001.

Disse tallene gir oss summen av effekt på deltagende skoler for elever som deltar (direkte effekt) og elever som ikke deltar (ringvirkninger) i intensivopplæring. En evt. negativ effekt kan skyldes negativ direkte effekt eller ringvirkninger, men som diskutert over er vi ikke i stand til skille mellom disse. Innføring av intensivopplæring fører til at rundt 10 prosent av alle elever ved skolen deltar. Dersom vi tillegger hele effekten til de elevene som faktisk deltok (dvs. vi antar at det ikke finnes ringvirkninger), vil et 95 prosent-konfidensintervall for den sanne direkte effekten være (-0,29, 0,03).<sup>17</sup>

Dersom vi ser på overgang til Vg2 innen to år (som er det målet på gjennomstrømning som gir det mest presise estimatet), ser vi at punkttestimatet er 0,004, det vil si en økning på 0,4 prosentpoeng i andelen som er registrert i Vg2 innen to år. Et 95 prosent-konfidensintervall blir her (-0,002, 0,010), det vil si at intensivopplæringen kan ha økt andelen som er registrert i Vg2 med inntil 1 prosentpoeng, men den kan også ha ført til en marginal reduksjon. En av målsetningene til Ny GIV er at overgangen fra Vg1 til Vg2 skulle øke med 2 prosentpoeng fra 2010 til 2013. Vi finner at effekten av intensivopplæringen på overgang fra grunnskole til Vg2 er maksimalt 1 prosentpoeng.<sup>18</sup>

Det kan imidlertid ha vært en sterkere effekt på målgruppen. Dersom vi, som over, antar at det ikke var noen ringvirkninger er et konfidensintervall for effekten på deltagerne (-0,02, 0,10), det vil si inntil en økning på 10 prosentpoeng (men også mulighet for en reduksjon på inntil 2 prosentpoeng). Effektene på fullføring av Vg1 har litt svakere presisjon, men er enda nærmere null. Dermed kan vi utelukke snitt-effekter større enn omtrent 0,6 prosentpoeng. For de senere målene på gjennomstrømning, fullføring av Vg2 og Vg3, oppmeldt i Vg3 og oppnådd kompetanse har vi mindre presise estimater og kan i mindre grad utelukke interessante effekter.

Oppsummert finner vi ingen klare tegn til effekter av intensivopplæringen. Når vi studerer snittresultater for alle elever kan vi utelukke forholdsvis små effekter, det er f.eks. usannsynlig at overgang fra grunnskole til Vg2 innen to år har økt med mer enn maksimalt 1 prosentpoeng. Dette er en beskjeden endring sammenlignet med målsetningen for Ny GIV. Ettersom det bare er 10 prosent av elevene som deltar kan vi ikke utelukke en betydelig effekt på deltakerne i intensivopplæringen, inntil omtrent 10 prosentpoeng på overgang til Vg2.<sup>19</sup>

I Tabell 4 presenterer vi effekttestimatene for grupper av elever, fordelt ut fra plassering i karakterfordelingen fra første termin i 10. klasse. Desil 1 (de svakestpresterende 10 prosent) til 4 studeres separat, mens vi slår sammen desil

<sup>17</sup> Dette kommer frem ved å multiplisere verdiene i konfidensintervallet med 10, fordi vi antar at hele effekten kommer fra 10 prosent av elevene.

<sup>18</sup> Derimot ser det, fra Tabell 1, ut til at andelen som er registrert i Vg2 ett år etter fullført grunnskole øker med 2 prosentpoeng. Dette ser imidlertid ikke ut til å være et resultat av intensivopplæringen, ettersom vi heller ikke finner noen klar effekt på overgang til Vg2 til normert tid i Tabell 3.

<sup>19</sup> Dette er i fravær av ringvirkninger, dersom intensivopplæringen har ført til negative ringvirkninger kan den ha hatt en stor direkte effekt, uten at vi er i stand til å oppdage dette.



5-10 (de høyestpresterende 60 prosent) til en gruppe.<sup>20</sup> Bortsett fra at vi for hvert resultatmål viser fem effektestimater, et for hver gruppe, har tabellen samme oppsett som Tabell 1. Første kolonne viser andelen i hver gruppe som deltar i intensivopplæring der denne er innført. Dersom skoler og kommuner hadde fulgt KDs anbefalinger om å tilby intensivopplæring til de 10 prosent svakeste elevene skulle punkttestimatet i første kolonne første rad vært lik 1, og de øvrige punkttestimatene i første kolonne vært null. I så fall kunne vi funnet direkte effekter ved å se på estimatene for desil 1, og ringvirkninger for forskjellige grupper fra de øvrige estimatene.

Faktisk andel deltagere er 35 prosent blant elever som befinner seg i 1. desil (altså de 10 prosent svakeste) og 31 prosent for elevene i 2. desil. Også elever i høyere desiler er registrert med intensivopplæring, henholdsvis 14, 6 og 1 prosent i 3., 4. og 5.-10 desil. Med andre ord, som illustrert i Eielsen m. fl. (2013), er det ingen skarp avgrensning i tilordningen av elever til intensivopplæring. Dermed er det vanskelig å skille mellom direkte effekter og ringvirkninger. Vi kunne likevel forvente at effekten er størst i de laveste desilene, ettersom andelen elever som mottar intensivopplæring er høyest i disse desilene. Vi finner imidlertid ikke signifikante resultater av intensivopplæringen her heller. Et flertall av punkttestimatene i Tabell 4 er negative, men svært få er signifikante. Ved mange estimeringer vil vi vente at enkelte estimater blir signifikante ved tilfeldighet. Andelen signifikante estimater i Tabell 4 er ikke vesentlig høyere enn hva vi venter utfra slike tilfeldigheter.

Estimatene for de enkelte desilene i Tabell 4 er mindre presise enn snitteffektene i Tabell 3, dette har betydning for hvor store effekter vi kan utelukke for de enkelte desilene. Punkttestimatet for snitt standpunktkarakter i 10. klasse er -0,033 for elever i første desil, altså effekten for elever blant de 10 % svakeste til første termin i 10. klasse. Et 95-prosent konfidensintervallet blir (-0,072, 0,007). For desil 2 er punkttestimatet -0,034, og konfidensintervallet (-0,064, -0,004). Dette tilsier at det har vært en negativ effekt på standpunktkarakterene til denne gruppen på inntil omtrent 0,06 karakterpoeng, men ettersom dette er et av få signifikante av mange estimater bør det tolkes med varsomhet. Vi kan også her regne ut hvilken direkte effekt dette svarer til i fravær av ringvirkninger. For desil 1 blir konfidensintervallet (-0,21, 0,02) og for desil 2 (-0,21, -0,01) når vi justerer for at henholdsvis 35 og 31 prosent deltar i intensivopplæringen. Vi kan ikke konkludere med at intensivopplæringen har hatt en effekt på standpunktkarakterene til målgruppen, men det er mulig at den har hatt en ganske betydelig negativ effekt.<sup>21</sup>

Tabell 4 viser også effekter på mål for gjennomstrømning. Som for Tabell 3 velger vi å fokusere på andelen som er oppmeldt i Vg2 innen to år. Estimater for desil 1 er her 0,009, en økning på 0,9 prosentpoeng. Konfidensintervallet blir (-0,011, 0,029), det vil si fra 1 prosentpoengs reduksjon til 3 prosentpoeng økning. For desil to er estimatet -0,005 og konfidensintervallet (-0,012, 0,022). Dersom vi antar at det ikke er ringvirkninger blir de direkte effektene henholdsvis (-0,03, 0,08) og (-0,07, 0,04).

<sup>20</sup> Elever uten resultater fra første termin er utelatt fra denne analysen. Effektestimater for denne gruppen er særlig upresise, og i en del tilfeller er det tegn til systematiske forskjeller i utviklingen før innføringen av intensivopplæringen. Det er dermed vanskelig å tolke disse resultatene, og vi har derfor valgt å ikke presentere disse.

<sup>21</sup> Interessant nok er den estimerte effekten på skriftlig eksamen positiv, men mindre presis og langt mindre signifikant. Konfidensintervallene for både desil 1 og 2 er (-0,04, 0,05), som gir en direkte effekt på omtrent (-0,14, 0,16) i fravær av ringvirkninger. Som diskutert i fotnote 15 er det utfordringer knyttet til standpunktkarakterer og lærerens vurdering. Her kan vi imidlertid ikke trekke klare konklusjoner verken for effekter på standpunkt eller eksamen.

Tabell 4. Effekten på elevprestasjoner av intensivopplæring for ulike desiler, difference-in-differences estimater

	Ny Giv	Snitt standpkt	Snitt skriftlig eks 10.	Fullført Vg1 t1	Fullført Vg1 t2	Fullført Vg2 t2	Fullført Vg2 t3	Fullført Vg3 t3	Fullført kompetanse t3	Registrert Vg2 t1	Registrert Vg2 t2	Registrert Vg3 t1
1. desil	0,352*** (0,0146)	-0,033 (0,0203)	0,004 (0,0249)	-0,002 (0,0110)	0,001 (0,0119)	-0,004 (0,0118)	-0,023 (0,0193)	-0,013 (0,0087)	-0,011 (0,0080)	0,012 (0,0109)	0,009 (0,0100)	-0,011 (0,0098)
2. desil	0,313*** (0,0229)	-0,034* (0,0151)	0,004 (0,0244)	-0,021 (0,0112)	-0,012 (0,0115)	-0,024 (0,0123)	0,003 (0,0149)	-0,004 (0,0167)	-0,007 (0,0159)	-0,018 (0,0109)	-0,005 (0,0088)	-0,024 (0,0143)
3. desil	0,144*** (0,0198)	-0,009 (0,0143)	0,052* (0,0256)	-0,007 (0,0092)	-0,007 (0,0090)	0,003 (0,0118)	0,003 (0,0164)	-0,003 (0,0158)	0,005 (0,0151)	0,011 (0,0087)	0,005 (0,0076)	0,010 (0,0118)
4. desil	0,064*** (0,0120)	-0,016 (0,0131)	0,035 (0,0248)	-0,009 (0,0082)	-0,011 (0,0082)	-0,001 (0,0103)	0,013 (0,0150)	-0,041* (0,0207)	-0,040* (0,0198)	0,012 (0,0096)	0,009 (0,0080)	0,001 (0,0112)
5.-10. desil	0,010*** (0,0028)	-0,009 (0,0086)	0,004 (0,0142)	0,001 (0,0034)	-0,001 (0,0036)	0,002 (0,0047)	0,005 (0,0067)	-0,006 (0,0075)	-0,005 (0,0075)	0,004 (0,0032)	-0,000 (0,0029)	-0,004 (0,0034)
Observasjoner	373 745	373 141	354 814	373 745	315 467	315 467	256 266	256 266	256 266	373 745	315 467	315 467
Snitt utfall	0,0327	4,024	3,471	0,816	0,839	0,724	0,763	0,507	0,496	0,830	0,884	0,742
Pre-trend	.	0,143	0,109	0,554	0,536	0,654	0,710	0,0309	0,00968	0,437	0,0205	0,313
Kløstere	405	405	405	405	405	405	386	386	386	405	405	405

Merknad: Stjerner angir statistisk signifikans: \* p<0,05; \*\* p<0,01; \*\*\* p<0,001.

Oppsummert er det heller ingen klare tegn til effekter når vi deler opp elevene etter resultater til første termin. Sammenhengen mellom resultat fra første termin og deltagelse er for svak til at vi presist kan skille mellom ringvirkninger og direkte effekter. De laveste desilene har imidlertid den høyeste andelen deltagere, 30-35 prosent. Vi finner ingen klare effekter for disse gruppene, men kan utelukke effekter på overgang til Vg2 større enn omtrent 2-3 prosentpoeng. Dette svarer til en direkte effekt på omtrent 4-8 prosentpoeng i fravær av ringvirkninger.

Vi har også undersøkt om det er ulike effekter for elever med ulik bakgrunn. Eksempler på bakgrunnsvariabel kan være kjønn, foreldres utdanningsnivå, og om eleven selv er innvandrer, eller har foreldre med innvandrerforeldre. Vi finner heller ikke for disse undergruppene konsistente forskjeller mellom elever ved skoler med og uten intensivopplæring. Resultatene er presentert i Vedlegg A.2.1-A2.3

Overgangsprosjektet har hatt en klar regional dimensjon, med fylkesvise prosjektledere. Det er derfor interessant å se om noen fylker har lyktes bedre enn andre. For å studere dette har vi beregnet separate effekter for de ulike fylkene. Resultatene er presentert i Vedlegg A.2.4. Oslo og Østfold peker seg ut som de fylkene som har signifikante effekter, men som nevnt i kapittel 2 er det noen uregelmessigheter for disse to fylkene. Effektestimatene for disse fylkene er derfor ikke tolkbare. Når det gjelder resterende fylker er det lite systematikk i estimatene, som i de aller fleste tilfellene heller ikke er signifikante.

To fylker skiller seg i noen grad ut. I Aust-Agder ser det ut til å være positive signifikante effekter på skriftlig eksamen og fullføring av Vg1, mens effekten på standpunkt og fullføring av Vg1 i Troms ser ut til å være signifikant negativ. Imidlertid, som allerede nevnt, vil noen signifikante effekter kunne oppstå ved tilfeldighet når man estimerer mange utfall. Det er viktig å huske på det når man tolker resultatene i disse tabellene.

I tillegg beregner vi difference-in-differences effekter hvor vi utelater Østfold og Oslo (jf. diskusjonen ovenfor). Disse resultatene, som er rapportert i Vedlegg A.4 ligner i stor grad på resultatene rapportert i Tabell 3. Det samme gjør resultatene rapportert i Vedlegg A.3 som viser difference-in-differences effekter når vi inkluderer ytterligere kontrollvariable som nasjonale prøver 8. trinn og skolefaste effekter.

### 3.2. Forskjeller innen Stavanger-skoler

Den uttalte målgruppen for intensivopplæringen er de 10 prosent elevene i hver kommune med svakest resultater ved første termin i 10. klasse. Som nevnt har Stavanger kommune, som eneste kommune, i stor grad tilordnet elever etter denne

regelen. For Stavanger er det derfor mulig å gjøre en ytterligere effektevaluering ved utnytte denne karaktergrensen. Denne analysen gir ytterligere i kunnskap om hvilken effekt intensivopplæringen har hatt i Stavanger. Vi kan imidlertid ikke overføre disse resultatene til øvrige fylker, kommuner eller skoler. På samme måte som utvelgelsen av elever tydeligvis har skjedd på en annen måte i Stavanger enn i andre kommuner kan også organiseringen og gjennomføringen av intensivopplæringen hat vært forskjellig, slik at eventuelle effekter av intensivopplæringen også er forskjellige. Nedenfor gir vi en detaljert beskrivelse av analysestrategiene som brukes for Stavanger.

Vi ønsker å sammenligne de elevene som befinner seg *like under* karaktergrensen med de elevene som befinner seg *like over* karaktergrensen. Innenfor gruppen av elever nær grensen for deltakelse kan vi tenke på det som tilfeldig om en elev kommer så vidt over eller under grensen. En tilfeldighet i denne sammenhengen kan være at karakter i fag som kroppsøving, mat og helse og musikk som kanskje i mindre grad reflekterer elevenes ferdigheter i basisfagene, er avgjørende for om en elev havner på den ene eller den andre siden av karaktergrensen. Elevene like under tilbys intensivopplæring, mens elevene over tilbys ikke. Bortsett fra deltakelse i intensivopplæringen er elevene på begge sider av denne karaktergrensen forventet å være like. De elevene som befinner seg like over karaktergrensen (og ikke får tilbud om intensivopplæring) fungerer derfor som en sammenligningsgruppe for de elevene som befinner seg like under karaktergrensen (og som får tilbud om intensivopplæring). Elevene over karaktergrensen gir oss et godt anslag på resultatene elevene under grensen ville oppnådd i fravær av intensivopplæring.

Dette er et eksempel på hva som i den internasjonale litteraturen er omtalt som en (regresjons) diskontinuitet, eller RD. Den største fordelen med en slik RD-tilnærming er at den ligner et randomisert eksperiment (her lokalt rundt den 10. persentilen). Forskjellen i de to gruppenes resultater gir et estimat av den kausale effekten av intensivopplæringen. Et eksempel på en lignende RD-tilnærming er Leuven mfl. (2007), som evaluerer betydningen av subsidier til personell og datamaskiner for grunnskolene i Nederland. Forutsetninger for å gi RD-estimatene en kausal tolkning, er at tilordning av elever på hver side av karaktergrensen er (lokalt) tilfeldig, og at alle kjennetegn ved elevene (samt resultat i fravær av intensivopplæring) endrer seg jevnt og ikke i sprang (dvs. er kontinuerlig) over karaktergrensen. Dette kan ikke bekrefte direkte, men vi kan få indikasjoner på om det stemmer ved å sammenlikne kontrollvariable på hver side av karaktergrensen.

Vi vil i det følgende forsøke å forklare intuisjonen i analysemetoden, og vurderingene vi gjør i forbindelse med den praktiske gjennomføringen. I «Boks 2» gir vi en mer formell presentasjon av ligningene vi estimerer. Vi lar  $y_i^{Stav}(1)$  være utfallet for elev  $i$  som scorer like under den spesifikke karaktergrensen, og dermed får tilbud om intensivopplæring, mens  $y_j^{Stav}(0)$  er utfallet for en annen elev  $j$  som scorer like over karaktergrensen, og dermed ikke får tilbud om intensivopplæring. (Snitt)effekten av intensivopplæring på elevutfall i et RD oppsett får vi derfor ved å ta differansen mellom disse utfallene for alle elevene som befinner seg i et lite intervall rundt den spesifikke karaktergrensen, altså

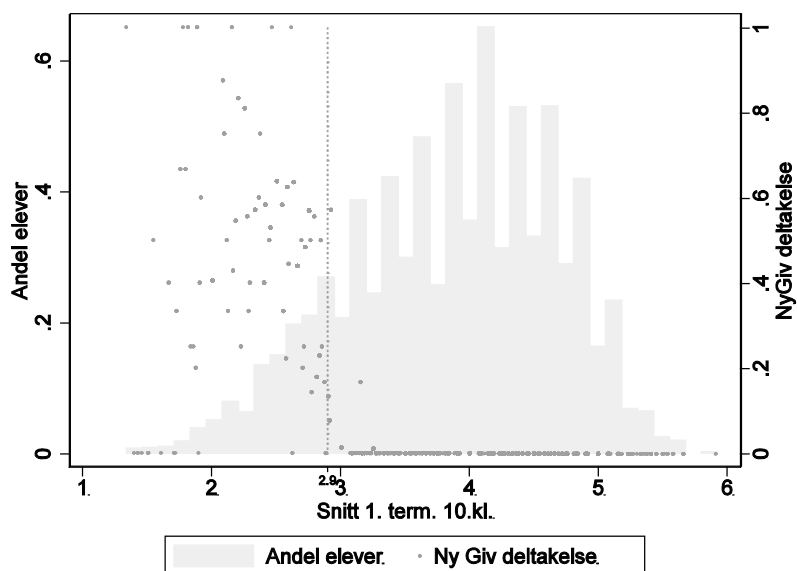
$$(2) \quad (Y^{Stav}(1) - Y^{Stav}(0))$$

En positiv differanse tilsier at det er en effekt av intensivopplæring. Merk at dette RD-estimatet gir snitteffekten av å være like under karaktergrensen, framfor like over. Dette svarer til effekten en gitt elev i umiddelbar nærhet til denne grensen opplever av å tilbys intensivopplæring. To forhold er verdt å merke seg: For det første er dette estimatet lokalt. Elever med vesentlig lavere eller høyere resultater kan ha en annen effekt av intensivopplæringen. For det andre uttrykker estimatet en forskjell rundt grensen for deltagelse, og kan dermed være påvirket av

snittresultatene til elever på begge sider. Det å *innføre* intensivopplæring ved en skole kan ha en effekt på elevene som (så vidt) ikke deltar. F.eks. kan vi tenke oss en situasjon der spesielt dyktige lærere settes til intensivopplæring, og der dette bedrer resultatene til deltakerne, men svekker resultatene til de som ikke deltar (og som nå i mindre grad får undervisning av disse lærerne). I så fall vil et positiv RD-estimat vise den positive effekten av å flytte en elev fra vanlig undervisning til intensivopplæring. Dette vil være sammensatt av en positiv effekt på deltagerne (som er mindre enn RD-estimatet) og en negativ effekt på de som så vidt ikke deltar, sammenlignet med hva som ville ha vært tilfelle i fravær av intensivopplæring.

Figur 3 viser tilordningen av elever til intensivopplæring i Stavanger. Elevprestasjoner (snittkarakter første termin på 10. trinn) er målt langs X-aksen, mens andelen elever totalt og til Ny GIV deltakelse er målt langs Y-aksene. Legg merke til at vi benytter karakterer i stedet for persentiler. For Stavanger er det ca. karakteren 2,9 som korresponderer til 10. persentil. Denne grensen er markert med en stiplet linje i figuren. Ved en streng tolking av tilordningsregelen burde intensivopplæring blitt tilbudt alle elever til venstre for linjen, altså de som scorer lavere enn 2,9, og utelukkende disse. Dette ser ikke ut til å være tilfellet. Selv om det er tydelige tegn til en tilordning av elever til intensivopplæring som svarer til regelen, er det også betydelige avvik.<sup>22</sup>

Figur 3. Karakterfordeling og andel elever i intensivopplæring.



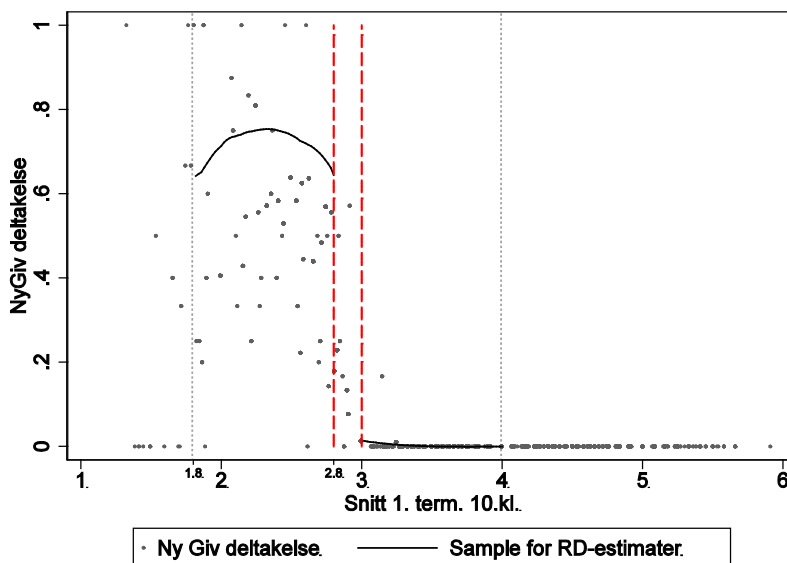
Merknad: Stiplet linje viser 10. karakterdesil.

En uklar tilordning byr på problemer når vi ønsker å sammenligne elever like over og under grensen for deltakelse. Argumentet for RD-analysen baserer seg på at elever med nesten like resultater til første termin har klart forskjellig sannsynlighet for deltakelse i intensivopplæring. Dersom dette ikke er tilfellet kan vi heller ikke knytte eventuelle resultatforskjeller til deltakelse. For å lage et skarpere skille velger vi å utelate observasjoner i intervallet (2,8-3,0), dette omtales som en «smultring» («donut hole») i faglitteraturen. I Figur 4 har vi markert «smultringen» med stiplede linjer. En annen fordel med å skape et skarpere skille i deltakelse

<sup>22</sup> Når vi undersøker tilordningen av elever til intensivopplæring i Stavanger finner vi at denne er mindre klar i skoleårene 2011/12 og 2012/13 enn i 2010/2011, dette er også i tråd med funnene i Lødding og Holen (2013). Stavanger er imidlertid klart tydeligere enn de aller fleste kommuner i alle tre årene. Vi finner noen kommuner med like klar eller klarere tilordningspraksis, men dette er stort sett små kommuner og/eller enkeltår med klar praksis.

rundt karaktergrensen er at det kan redusere sannsynligheten for at vi bruker en feil grense basert på feilregistrering av deltakelse for enkeltelever.<sup>23</sup>

Figur 4. Andel deltakere i intensivopplæring per karakterpoeng



Merknad: Røde stiplede linjer markerer smultringen, der observasjoner innenfor disse utelates. Området mellom grå og rød linjer på hver side av smultringen, markerer størrelsen på båndbredden, her satt til ett karakterpoeng. Figuren viser kvadratiske trender.

Ulempen er at elevene på begge sider av smultringen blir mindre sammenlignbare ettersom vi nå sammenligner elever som har snittkarakter over 3 med elever som har snittkarakter under 2,8. Det kan dermed hende at elevene som nå befinner seg like over karaktergrensen ikke utgjør en gyldig sammenligningsgruppe til de elevene som befinner seg like under. Dette er et generelt problem ved RD-analyser: I fravær av tilstrekkelig observasjoner akkurat rundt diskontinuiteten, tar man i bruk data i et større intervall på hver side av diskontinuiteten for å få et mål på forventet resultat nær grensen. I litteraturen omtales dette som båndbredde. Ved valg av båndbredde står vi overfor en avveining mellom presisjon og skjevhet. Større båndbredde gir mer presise estimater fordi et større antall observasjoner kan brukes i estimeringen. Dette gir oss høyere presisjon. Samtidig vil større båndbredde gjøre estimatet mindre troverdig, fordi vi sammenlikner observasjoner som er lenger unna terskelverdien for karaktergrensen, og dermed mer ulike (større skjevhet). Vi har valgt en båndbredde på ett karakterpoeng, markert som områdene mellom grå og rød vertikale linjer på hver side av smultringen i Figur 4. I RD-analysene er det vanskelig å få estimater som er presise nok til å trekke klare konklusjoner, dette trekker i retning av en stor båndbredde. Samtidig er det lite å vinne på å øke båndbredden ytterligere; fra Figur 3 ser vi at nesten ingen elever har snittkarakter til første termin under 1,8.<sup>24</sup>

Tabell 5 viser beskrivende statistikk for utvalget som inngår i analysene for Stavanger. Med andre ord inkluderes elever med snittkarakter innen intervallene (1,8-2,8) og (3,0-4,0). Kolonne (1) og (2) sammenlikner elever ved skoler og kull

<sup>23</sup> Smultringen kan også redusere sannsynligheten for at noen elever er plassert strategisk over/under grensen. I noen sammenhenger er dette en viktig problemstilling for gyldigheten av RD-analyser. I dette konkrete tilfellet har vi liten grunn til å tro at karakterene til første termin er satt med tanke på deltakelse i intensivopplæring. Dette illustrerer imidlertid fortrinnet med RD-analysene. Elevene som deltar kan være valgt ut delvis basert på (for oss) uobserverbare kriterier som gjør det vanskelig å sammenligne dem med elevene som ikke deltar, f.eks. fordi de er mer motiverte. Med RD-analysene unngår vi dette problemet ved å basere sammenligningen på et mål (her karakterer til første termin) som er bestemt på en mer gjennomsiktig måte.

<sup>24</sup> Ett karakterpoeng var foretrukket båndbredde i RD-analysene i Eielsen m.fl. (2013). Vi har også i denne runden undersøkt flere båndbredder; dette påvirker ikke resultatet nevneverdig.

som deltar i intensivopplæring, men som befinner seg på hver sin side av karaktergrensen. Differansen mellom disse kolonnene, samt t-test for om forskjellene er signifikante, kan ses i kolonne (3). Elever som befinner seg i intervallet (1,8-2,8) presterer som forventet i snitt dårligere enn elever som befinner seg i intervallet (3,0-4,0). I tillegg til å prestere dårligere på karakteren ved første termin, scorer de også dårligere på nasjonal prøve i åttende trinn. Videre ser vi at elevene nedenfor karaktergrensen er i større grad innvandrere, eller har foreldre som er innvandrere. Gutter og elever med foreldre med lavere utdanning er også overrepresentert til venstre for karaktergrensen. Kolonne (1), (2) og (3) i Tabell 5 bekrefter dermed at elever på oversiden av karaktergrensen ikke nødvendigvis representerer et godt kontrafaktisk resultat for elevene på undersiden av karaktergrensen.

En måte å ta hensyn til at elever i intervallet (1,8-2,8) scorer lavere enn elever i intervallet (3,0-4,0), er å betinge på snittkarakteren ved første termin. Da tar vi hensyn til at elever i intervallet (1,8-2,8) scorer lavere enn elever i intervallet (3,0-4,0). Mer presist estimerer vi kvadratiske sammenhenger mellom karakter i første termin og forskjellig resultatmål separat for elever på over- og undersiden av smultringen.<sup>25</sup> En detaljert beskrivelse av likningen som blir estimert er gitt i «Boks 2». RD-estimatet kan tolkes som effekten av en liten utvidelse av omfanget av intensivopplæringen lokalt rundt tilordningsgrensen. Gitt eksistensen av ringvirkninger kan vi tenke på effekten som forskjellen mellom den direkte effekten på den «siste» eleven som deltar og ringvirkningene på den «første» som ikke deltar. I motsetning til DiD-analysen over, der eventuelle positive (gjennomsnittlige) ringvirkninger bidrar til å øke den estimerte effekten vil eventuelle positive ringvirkninger (på elever nær grensen) i RD-analysen bidra til redusere den estimerte effekten.

En annen løsning på utfordringene som oppstår når vi i RD-tilnærmingen risikerer å sammenlikne elever som kan være forskjellige, er å kombinere RD med DiD (se f.eks. Leuven og Oosterbeek, 2004; Carneiro m. fl., 2014). For å kunne gjøre dette trenger vi en ny sammenligningsgruppe med svaktpresterende elever som ikke deltar i intensivopplæring. Vi har vurdert to slike sammenligningsgrupper: Avgangselever i Stavanger i 2010 og avgangselever fra 2010 og 2011 fra Bergenskoler som ikke deltar i Ny GIV på det aktuelle tidspunktet.

Valget av sammenligningsgrupper er i noen grad tilfeldig. Tidligere årskull i samme kommune er et naturlig valg, ettersom disse (i stor grad) vil ha tilsvarende elevbakgrunn, skolestruktur, politikk fra skoleeier mm. Det er imidlertid mangler i datagrunnlaget for karakterer fra første termin for Rogaland; vi har ikke karakterer fra første termin for Stavanger-elever før avgangskullet 2010 (for øvrige kommuner i Rogaland har vi første termin f.o.m. avgangskullet 2012). Dette betyr at vi bare har ett årskull med sammenligningselever, noe som gir mindre presise resultater. Det er også tegn til at sammenhengen mellom resultater til første termin og nasjonale prøver fra 8. trinn er annerledes for dette kullet, men dette er usikkert grunnet den begrensede presisjonen. Bergen framstår som et naturlig alternativ, ettersom både Bergen og Stavanger er storbyer på Vestlandet.<sup>26</sup> Vi vil derfor i

<sup>25</sup> I RD-estimeringene våre utelater vi datapunkter innenfor en smultring (2,8-3,0). Basert på observasjoner i intervallet (1,8-2,8) beregner vi en kvadratisk trend for venstre side av karaktergrensa (vi har også forsøkt lineære sammenhenger, men de viser seg imidlertid å være for lite fleksible til å gi en god tilpasning og forlengelse fram mot grensen for deltagelse). Denne ekstrapoleres helt frem til midtpunktet for smultringen, som er 2,9. Tilsvarende beregner vi en kvadratisk trend for intervallet (3-4), men ekstrapolerer trenden frem til 2,9. Dette gjør at vi kan finne forventet resultatforskjell for elever med karakter rett over og under 2,9, som svarer til RD-estimatet vårt. Merk at vi baserer oss på en forskjell i forventet resultat i et område der vi (bevisst) ikke observerer faktiske resultater. Dette er likevel ikke nødvendigvis noen stor forskjell sammenlignet med ordinære RD-analyser uten smultring, ettersom praktisk talt alle slike baserer seg på modellering av resultater noe unna en grense.

<sup>26</sup> Fra kartleggingen vet vi at antall timer intensivopplæring er relativ lik i Rogaland og Hordaland (6,7 og 6,6 timer per uke i, sammenlignet med et snitt på 5,8 for hele landet).

hovedsak basere oss på Bergen som sammenligningsgruppe, men også presentere resultater med 2010-kullet fra Stavanger som sammenligningsgruppe. For Bergen bruker vi bare elever fra 2010 og 2011-kullene, og bare i den grad skolene deres ikke deltar i Ny GIV (dvs. vi bruker ikke 2011-kullet fra fase 1-skoler; et stort flertall av Bergen-elevene er i fase 2-skoler).

I Stavanger vil det for avgangskullene 2011-2013 være en forskjell i deltakelse i intensivopplæring rundt karaktergrensen. Det vil det ikke være i sammenligningsgruppene, ettersom ingen elever herfra fikk tilbud om intensivopplæring. Vi kan likevel utfra snittresultatene til første termin finne de av elevene i sammenligningsgruppen som (sannsynligvis) ville ha deltatt i intensivopplæringen, dersom de var elever ved en Stavanger-skole som deltok i Ny GIV. Vi kan dermed sammenligne resultatforskjellene mellom elever over og under karaktergrensen i Stavanger-skoler som deltar i Ny GIV (der dette svarer til en forskjell i intensivopplæring) med tilsvarende resultatforskjeller i sammenligningsgruppen (der det ikke er noen slik forskjell).

Vi definerer  $y_i^{Bergen}(1)$  til å være utfallet for elev  $i$  fra sammenligningsgruppen som scorer under karaktergrensen 2,8. For enkelthets skyld kaller vi sammenligningsgruppen Bergen. Samme resonnement gjelder dersom man benytter pre-Ny GIV kohorter i Stavanger som sammenligningsgruppe. Tanken er at resultatene til disse elevene gir et godt anslag på det kontrafaktiske resultatet til de elevene som faktisk får tilbud om intensivopplæring i forbindelse med Ny GIV, altså  $y_i^{Stav}(1)$  som definert i ligning (2). Tilsvarende definerer vi  $y_j^{Bergen}(0)$  til å være utfallet for elev  $j$  fra sammenligningsgruppen som scorer over karakteren 2,8. Disse representerer det kontrafaktiske resultatet til elevene som ikke får tilbud om intensivopplæring, altså  $y_j^{Stav}(0)$  i ligning (2). Den kausale effekten av å tilby intensivopplæring til avgangselever i Stavanger kan dermed avledes ved å ta differansen mellom gjennomsnittresultatene  $Y^{Stav}(1)$  og  $Y^{Stav}(0)$  og sammenlikne den med differansene mellom  $Y^{Bergen}(1)$  og  $Y^{Bergen}(0)$ , beskrevet ved følgende DiD ligning:

$$(3) \quad (Y^{Stav}(1) - Y^{Stav}(0)) - (Y^{Bergen}(1) - Y^{Bergen}(0))$$

Dersom den første differansen er større enn den andre differansen kan det tyde på at det er en kausal effekt på elevresultat av å tilby intensivopplæring. Intuisjonen bak denne analysestrategien er tilsvarende DiD-ligningen for hele landet (ligning (1), dette blir utdypet noe mer nedenfor). For at vi ikke skal sammenblende DiD analysen for hele landet (ligning 1) med DiD analysen for Stavanger (ligning 3) vil vi følgende omtale den sistnevnte som DiDS, hvor S står for Stavanger.

Den første differansen er den samme som i ligning (2). Den andre differansen viser forskjell i gjennomsnittlig utfall for elever under og over (samme) karaktergrense for elever fra avgangskullene 2010 og 2011 i Bergen-skoler som ikke har intensivopplæring i de aktuelle årene. Differanser for utfall og bakgrunnsvariable finner vi i kolonne (4) i Tabell 5 (i vedlegg B viser vi en tilsvarende tabell med avgangselever fra Stavanger fra 2010 som sammenligningsgruppe). Vi finner også for Bergen forskjeller mellom elever med høyt og lavt karaktersnitt tilsvarende kolonne (3). Derimot, når vi sammenlikner kolonne (3) med kolonne (4) er differansene små, og stort sett ikke signifikant forskjellig; se kolonne (5). Ved å ta i bruk data fra et område eller en periode da Ny GIV ikke var innført, klarer vi dermed i betydelig grad å kontrollere for at elever over og under karaktergrensen er ulike. Dette er en indikasjon på at elever i Bergen fungerer som sammenligningsgruppe for elever i Stavanger. DiDS-estimatet i ligning (3), snittforskjellen mellom elever i tiltaks- og sammenligningsgruppen over og under karaktergrensen innenfor

satt båndbredde, gir oss dermed et estimat på kausale effekter. «Boks 2» gir en detaljert beskrivelse av modellen som ligger til grunn for resultatene.<sup>27</sup>

DiDS-analysen med Bergen som kontroll har likheter både med DiD-analysene i del 3.1 og RD-analysene i denne delen. Det er imidlertid verdt å merke seg forskjellene: DiD-analysene i del 3.1 baserer seg på å sammenligne skoler. Dette skyldes at vi ikke er i stand til å finne gyldig sammenligningsgruppe på elevnivå på grunn av uklar tilordning av elever innen skoler. Innen Stavanger-skolene er utvelgelsen av elever mer gjennomsluttet. Det gjør at vi kan sammenligne resultatene til elever med forskjellig første termin karakter innen samme skole med rimelig grad av visshet om at det ikke er andre (systematiske) forskjeller. Dermed unngår vi utfordringer knyttet til at andre forhold varierer mellom skoler. Når vi sammeligner med RD-analysene er den viktige forskjellen at DiDS-analysene ikke studerer effekter akkurat rundt grensen for deltagelse, men heller snitteffekter innenfor båndbredden.

DiDS-analysen baserer seg på å sammenligne elever i Stavanger og Bergen med svake resultater fra termin 1. Vi kan gjøre denne mer fleksible ved å tilpasse fire separate trender henholdsvis for elever på over- og undersiden av smultringen i både tiltaks- og sammenligningsgruppen. Dette gir oss to en sammenligning av to RD-estimer (ligning (2), se også fotnote 25), for henholdsvis Stavanger og Bergen. Resultatene fra denne spesifikasjonen omtaler vi som DiRD-estimer.<sup>28</sup>

Til slutt er det verdt å oppsummere hva som kreves for at de forskjellige estimatene kan gis en kausal tolkning. For RD-estimatene kreves det at vi er i stand til å tilnærme resultatene like over og like under grensen for deltagelse, og at deltagelse i intensivopplæring er det eneste forholdet som endrer seg brått (ikke kontinuerlig) rundt denne grensen. For DiRD-estimatene gjelder en liknende betingelse. Vi kan imidlertid finne gyldige kausale estimater også dersom andre forhold også endrer seg brått rundt grensen, såfremt det er tilsvarende endringer i sammenligningsgruppen.

For DiDS-estimatene er vi avhengige av at sammenligningsgruppen gir et godt bilde av hvordan sammenhengen mellom karakter til første termin og resultatmålene ville vært i fravær av intensivopplæringen. I så fall gir nivåforskjellen over karaktergrensen et godt estimat på den kontrafaktiske nivåforskjellen under karaktergrensen i fravær av intensivopplæringen. Dette er en sterkere antagelse enn hva vi må gjøre for å finne DiRD-estimatene. Vi vil imidlertid argumentere for at dette kan være en rimelig antagelse, og til gjengjeld får vi mer presise estimater.

---

<sup>27</sup> Også i DiDS-analysen kontrollerer vi for karakterer ved første termin ved å estimere kvadratiske sammenhenger mellom karakter i første termin og forskjellig resultatmål. Her har det ikke særlig betydning hvorvidt vi har en kvadratisk tilnærming til sammenhengen mellom resultat og første termin, eller om denne er forskjellig på hver side av karaktergrensen. DiDS-analysen baserer seg på forskjeller mellom Stavanger/kontrollgruppen og over/under karaktergrensen. Såfremt karakterfordelingen er tilsvarende i Stavanger og kontrollgruppen har kontroll for karakterer mindre betydning. Dette er i kontrast til RD-analysen, der vi er avhengig av en tilstrekkelig fleksibel funksjonsform når vi skal ekstrapolere forventet resultat til grensen.

<sup>28</sup> DiRD-analysen baserer seg også på en ekstrapolering, som RD-analysen, jf. fotnote 25. Den ser likevel ut til å være robust for valg av funksjonsform, fordi kontrollgruppen gjør oss i stand til å korrigere for eventuelle feil i funksjonsform – dette er nettopp hensikten med en kontrollgruppe i denne analysen.



**Tabell 5 Beskrivende statistikk. Stavanger-skoler med intensivopplæring og Bergen-skoler uten, båndbredde 1 karakterpoeng.**

	(1) Karakter (1,8-2,8), Stavanger	(2) Karakter (3-4), Stavanger	(3) Differanse kolonne (1) – (2)	(4) Diffe- ranse tilsv.(3) for sammen- ligningskoler	(5) Differanse (DiDS): (3) – (4)
Andel med intensivopplæring .....	0,697	0,002	0,695*** (0,0110)	0,000 (0,0000)	0,695*** (0,0188)
Matte 1. termin .....	1,944	2,988	-1,044** (0,0408)	-1,040 (0,0660)	-0,004 (0,0804)
Nasjonal prøve 8. trinn .....	-0,945	-0,121	-0,824** (0,0377)	-0,702 (0,0682)	-0,121 (0,0771)
Jente .....	0,343	0,418	-0,075 (0,0248)	-0,113 (0,0438)	0,038 (0,0502)
Mors utdanning .....	3,291	4,336	-1,045** (0,0876)	-0,903 (0,1527)	-0,142 (0,1751)
Fars utdanning .....	3,442	4,438	-0,996** (0,0869)	-0,848 (0,1553)	-0,148 (0,1762)
Innvandrere .....	0,177	0,070	0,107 (0,0145)	0,086 (0,0239)	0,020 (0,0289)
Norskfødt barn av innvandrerforeldre .	0,112	0,065	0,047 (0,0134)	0,012 (0,0134)	0,035 (0,0242)
<b>Antall elever .....</b>	<b>498</b>	<b>1802</b>	<b>2300</b>	<b>939</b>	<b>3239</b>

Merknad: I kolonne (1) til (3) ser vi på avgangselever fra Stavanger-skoler fra årene 2011-2013. Kolonne (4) gir en differanse mellom elever med lave karakterer (1,8-2,8) og høyere karakterer (3,0-4,0) for Bergen-skoler i fase 1 (avgangselever 2010) og fase 2 (avgangselever 2010/2011), dvs. ved skoler som på det aktuelle tidspunktet ikke hadde innført intensivopplæring. Stjerner angir statistisk signifikans: \* p<0,05; \*\* p<0,01; \*\*\* p<0,001.

### Resultater innen Stavanger-skoler

I Tabell 6 presenterer vi RD-estimerer, samt DiDS-resultater og DiRD-resultater med Bergen som sammenligningsgruppe. I vedlegg B, tabell B.2 presenterer vi tilsvarende resultater med 2010-kullet i Stavanger som sammenligningsgruppe. Vi bruker, som i Figur 4, en båndbredde på ett karakterpoeng, det vil si at vi inkluderer observasjoner i karakterintervallene 1,8-2,8 og 3,0-4,0.

Kolonne 1 Tabell 6 presenteres rene RD-estimerer tilsvarende ligning (2.1). Disse resultatene baserer seg derfor kun på Stavanger-elever og kull som er omfattet av Ny GIV. Kolonne 2 rapporterer DiDS-estimatene, mens kolonne 3 viser vi DiRD-estimatene. I det følgende går vi gjennom resultatene.

#### Boks 2

RD-spesifikasjonen kan uttrykkes slik:

$$Y_i = \beta(\text{lowgrade}_i) + \alpha_1 \text{slopeLHS} + \alpha_2 \text{slopeRHS} + v_i, \tag{2.1}$$

Der  $\text{lowgrade}_i$  er en indikator lik 1 dersom eleven har snittkarakter fra første termin 10. trinn under 2,8, ellers lik 0.  $\text{slopeLHS}$  og  $\text{slopeRHS}$  er lineære eller kvadratiske trender på hver side av karaktergrensen for deltakelse. Ved båndbredde 1, inkluderes kun elever med karakter ved termin 1 innen intervallene (1,8-2,8) og (3-4). Effekten av intensivopplæring på utfall  $Y_{ij}$  er gitt ved  $\beta$ .

DiDS-likningen vi estimerer er beskrevet slik:

$$Y_{ij} = \beta(\text{lowgrade}_i \times \text{Stavanger}_j) + X' \beta + \text{Stavanger}_j + \text{lowgrade}_i + u_{ij}, \tag{2.2}$$

der  $\text{Stavanger}$  er en indikator som tar verdien 1 for elever i Stavanger (som er omfattet av intensivopplæring) og 0 dersom elevene er fra Bergen (kontrollgruppe). I vektor  $X'$  inkluderes lineære eller kvadratiske trender. Øvrig notasjon følger likning (2.1).

**Tabell 6. Effektestimater innen Stavanger-skoler. RD-, DiDS og DiRD-estimer med Bergen som sammenligningsgruppe. Avgangselever Stavanger 2011-2013 og Bergen 2010-2011**

	(1) RD	(2) DiDS	(3) DiRD
Ny Giv	0,565*** (0,090)	0,692*** (0,023)	0,565*** (0,090)
Index, forventet fullføring Vg1 t1	0,006 (0,023)	-0,023* (0,009)	-0,017 (0,035)
Matematikk 1. termin 10. kl.	-0,104 (0,141)	-0,045 (0,050)	-0,123 (0,194)
Norsk hovedmål 1. termin 10. kl.	-0,142 (0,131)	-0,017 (0,050)	0,136 (0,197)
Snitt standpunkt 10. kl.	0,098 (0,064)	-0,032 (0,025)	0,093 (0,104)
Matematikk standpunkt 10. kl.	0,259 (0,153)	0,019 (0,052)	0,278 (0,205)
Norsk hovedmål standpunkt 10. kl.	0,163 (0,138)	-0,010 (0,055)	0,188 (0,216)
Snitt skriftlig eksamen 10. kl.	0,002 (0,192)	-0,067 (0,076)	-0,033 (0,294)
Fullført VG1 t1	0,103 (0,105)	0,096* (0,041)	0,072 (0,158)
Fullført VG1 t2	0,139 (0,120)	0,088* (0,044)	0,148 (0,168)
Registrert VG1 t2	0,174 (0,104)	0,087* (0,041)	0,059 (0,158)
Registrert VG2 t2	0,200 (0,111)	0,088* (0,042)	0,138 (0,159)
Fullført VG2 t2	0,094 (0,130)	0,066 (0,044)	-0,098 (0,175)
Fullført VG2 t3	0,015 (0,187)	0,118* (0,056)	-0,127 (0,221)
Registrert VG3 t2	0,013 (0,127)	-0,053 (0,044)	-0,033 (0,173)
Registrert VG3 t3	-0,162 (0,185)	-0,009 (0,055)	-0,200 (0,220)
Fullført kompetanse t3	-0,125 (0,087)	0,003 (0,028)	-0,106 (0,108)

Merknad: t1, t2, t3 indikerer henholdsvis 1, 2. og 3. kalenderår etter endt grunnskole. Merknad: Standardfeil er robuste for forskjellig varians av restledd (heteroskedastisitet). Stjerner angir statistisk signifikans: \* p<0,05; \*\* p<0,01; \*\*\* p<0,001.

Første rad viser effekten på deltagelse i intensivopplæring. I RD-analysen finner vi at elever rett under grensen har omtrent 57 prosentpoeng høyere sannsynlighet for å delta i intensivopplæring, dette kan vi også se fra Figur 4. Som forventet gir DiRD-analysen det samme resultatet ettersom det ikke er endring i deltagelse rundt grensen i sammenligningsgruppen. Effekten i DiDS-analysen er derimot større, omtrent 70 prosentpoeng. Det skyldes at andelen som deltar er høyere blant elever med karakter fra første termin rundt 2,3 enn blant elevene med karakter nærmere 2,8, slik at gjennomsnittlig andel under grensen er høyere enn forventet verdi like under grensen. Den klare effekten på deltagelse i intensivopplæringen betyr at analysestrategien med en smultring gir oss i stand til finne lignende (om enn ikke identiske) grupper med svært stor forskjell i sannsynlighet for intensivopplæring. Dersom analysestrategien vår også lykkes i å isolere forskjeller knyttet til intensivopplæringen kan vi også bruke estimatet til å skalere effekten på deltakerne; vi kommer tilbake til dette under.<sup>29</sup>

Andre rad gir et grunnlag for å vurdere om vi kan tolke estimatene som effekter av intensivopplæring. Vi studerer her forhold som kan være viktige for elevenes resultater, men som er gitt før intensivopplæringen. Dette gir dermed en sjekk på om det er relevante forskjeller rundt karaktergrensen i utgangspunktet. Første variabel, «index» er en samleindeks for flere forhold: Snittresultat nasjonale prøver fra 8. trinn, foreldrenes utdanning, samt elevens kjønn og innvandringsbakgrunn. Indeksen er konstruert til å være mellom 0 og 1, hvor 1 er høyest score.

<sup>29</sup> Dette kan tolkes som et førstesteg i et 2SLS-rammeverk.

<sup>30</sup> Fra tabellen ser vi at det ikke er noen signifikant endring rundt grensen i RD-analysen. Indeksen er 0,6 prosentpoeng høyere for elevene rett under grensen enn for elevene rett over. Dette er en liten forskjell, og med god margin innenfor hva vi kan regne med som følge av tilfeldig variasjon. I DiRD-analysen endres forskjellen til 1,7 prosentpoeng lavere, noe som fortsatt er innenfor hva om kan oppstå som tilfeldig variasjon. I DiDS-analysen er det derimot en større forskjell enn vi venter som et resultat av tilfeldigheter. Vi finner her at elevene som i stor grad deltar i intensivopplæring har en lavere verdi på indeksen, omtrent 2,3 prosentpoeng. Denne forskjellen er imidlertid bare litt større enn forskjellen fra DiRD-analysen. At DiDS-estimatet er klart signifikant mens DiRD-estimatet ikke er det, skyldes at DiDS-estimatet er klart mer presist.

For å få en bedre forståelse av de forskjellige resultatene forbundet med andre rad i Tabell 6 viser vi en grafisk analyse i Figur 5. De heltrukne linjene viser en form for glidende gjennomsnitt av indeksen for elever i henholdsvis Stavanger og Bergen, over og under grensen, og gir en fleksibel representasjon av de underliggende dataene.<sup>31</sup> De stiplede linjene viser kvadratiske tilnærminger av sammenhengen mellom indeksen og karakterer ved første termin. De kvadratiske funksjonene ser ut til å gi rimelig gode tilnærminger til dataene. Tilpasningen er dårligst for svært lave karakterer fra første termin. Dette skyldes hovedsakelig at svært få elever befinner seg i disse karakterintervallene, som vi kan se fra Figur 4.<sup>32</sup> En dårlig tilpasning i dette området har dermed liten betydning både for beregnet forskjell ved grensen (dvs. for RD og DiRD) og for snittforskjeller (dvs. for DiDS-estimatet). Dernest ser vi at den kvadratiske tilpasningen for Stavanger er nærmest kontinuerlig rundt grensen, dette svarer til et RD-estimatet tilnærmet null (se første rad første kolonne i Tabell 6). For sammenligningsgruppen er indeksverdien derimot litt høyere for elever under grensen enn for elever over grensen. I tillegg ser vi at den kvadratiske tilpasningen ser ut til å føye dataene dårligere enn for Stavanger. Den tydeligste forskjellen i Figur 5 er at indeksverdien over grensen er høyere i Stavanger enn i Bergen. Dette kan gi en negativ skjevhet i effektestimaterne; vi overvurderer forutsetningen til intensivelevener i Stavanger når vi antar at deres kontrafaktiske resultat er like mye høyere resultatet til sammenligningselevener i Bergen som forskjellen mellom de bedre presterende i Stavanger og Bergen. I vedlegg B gir vi en tilsvarende grafisk analyse med 2010-kullet i Stavanger som sammenligningsgruppe. Det ser ut til at sammenligningselevener i 2010-kullet har forskjellig utvikling i forventet resultat når vi nærmer oss grensen, og at det er vanskelig å lage en god tilpasning til dette. Det kan gjøre det vanskelig å bruke 2010-kullet som en sammenligningsgruppe i en DiRD-analyse.

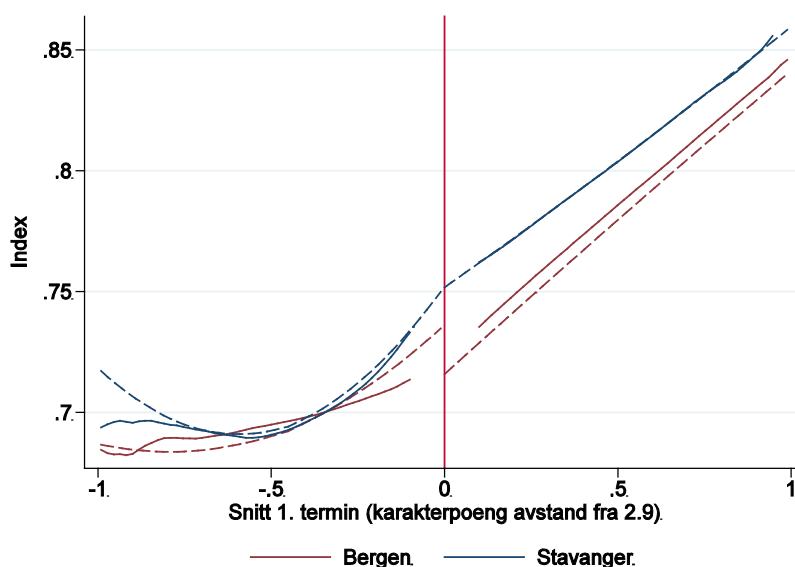
---

<sup>30</sup> Vi har kjørt en lineær regresjon med en indikatorvariabel for fullføring som avhengig variabel, og de omtalte kjennetegnene som kontroller. På bakgrunn av denne har vi predikert fullføring for alle elever. Dette gir en indeks som er et vektet gjennomsnitt av kjennetegnene, der vektene er bestemt av sammenhengene mellom kjennetegn og fullføring. I praksis utgjør resultater fra nasjonale prøver en stor del av indeksen.

<sup>31</sup> Vi beregner gjennomsnitt ved hjelp av lokale lineære regresjoner, gjennomsnittene tar dermed hensyn både til nivå og endring i et lite område (båndbredde 0,2).

<sup>32</sup> Ettersom vi er mest interessert i forskjellen nær grensen gir vi også observasjoner nær denne grensen større vekt, gjennom en triangulær vektfunksjon.

**Figur 5. Sammenheng mellom snitt 1. termin 10. trinn og indeks av nasjonale prøver fra 8. trinn og andre elevkjenntegn (forventet fullføring av Vg1 på normert tid) for avgangselever fra Stavanger 2011-2013 og fra sammenligningsgruppen Bergen 2010-2011**



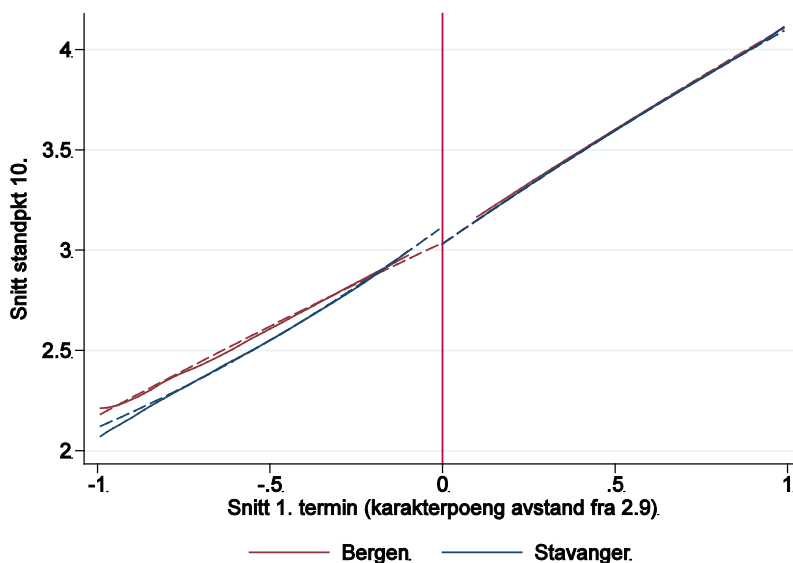
Merknad: Heltrukne linjer viser glidende gjennomsnitt (fra lokal regresjon med båndbredde 0,2). Stiplede linjer viser kvadratiske tilpasninger. Horisontal akse måler snittkarakter som avvik i karakterpoeng fra 2,9.

De påfølgende radene i Tabell 6 viser forskjell i karakterer til første termin i matematikk og norsk hovedmål. Selv om elever med snittkarakter like over og under 2,9 har nesten like snittkarakterer, kan de ha ulike sterke og svake sider. Vi finner ingen klare indikasjoner for disse fagene. Resultatene er imidlertid til dels upresise.

De øvrige radene i Tabell 6 gir effektestimater på forskjellige resultatmål. Vi studerer først karakterer fra fullført grunnskole. Vi ser at det ikke er signifikante effekter for noen karaktermål, verken for snitt standpunkt, standpunktkarakterer i norsk eller matematikk eller for snitt eksamenskarakter. Det er betydelige forskjeller mellom elever over og under grensen både i RD- og DiRD-analysen, særlig for enkeltfagene, men det er ikke tilstrekkelig presisjon til at vi kan trekke klare konklusjoner. For snitt standpunkt, som benytter mest informasjon og gir mest presise estimater, ser vi at vi måtte hatt effekter på 0,15-0,20 karakterpoeng rundt grensen for at RD eller DiRD-estimatene skulle vært signifikante. For DiDS-analysene av snittet til flere elever måtte vi hatt effekter på omtrent 0,05 karakterpoeng.

Figur 6 viser en grafisk analyse for snitt standpunkt. Vi ser at det er en veldig lik sammenheng mellom standpunkt og resultat fra første termin for elever over grensen. Under grensen er bildet mer blandet, men det er ikke veldig store forskjeller. Det kan være utfordringer med å gjøre effektstudier på standpunktkarakterer. Disse settes av elevenes lærere, og det at en elev deltar i intensivopplæring kan ha en effekt på lærerens vurdering og oppfatning av eleven, uavhengig av elevens faktiske faglige prestasjoner. Vi kan unngå dette problemet ved å se på eksamenskarakterer. Disse er til gjengjeld mindre presise, ettersom de er basert på en veldig mye mer begrenset vurdering av eleven. Vi finner ingen tegn til effekter på skriftlige eksamenskarakterer.

Figur 6. Sammenheng snitt 1. termin 10. trinn og standpunkt 10. trinn for avgangselever fra Stavanger 2011-2013 og fra sammenligningsgruppen Bergen 2010-2011



Merknad: Heltrukne linjer viser glidende gjennomsnitt (fra lokal regresjon med båndbredde 0,2). Stiplede linjer viser kvadratiske tilpasninger. Horisontal akse måler snittkarakter som avvik i karakterpoeng fra 2,9.

Derimot er det tegn til effekter på tidlig gjennomføring. I Tabell 6 (rad ni til tolv) ser vi at andelen som har fullført Vg1 eller er oppmeldt i Vg2 innen henholdsvis ett og to år etter fullført grunnskole er 10-20 prosentpoeng høyere blant elevene rett under grensen enn blant elevene rett over (RD-estimatet). Dette er en betydelig forskjell, men presisjonen i estimatene er for lav til at vi kan si sikkert at dette er en reell effekt. Som vi ser fra de estimerte standardfeilene måtte vi hatt effekter på omtrent 20 prosentpoeng for å konkludere med at de var signifikante. I DiRD-analysen er det lignende forskjeller (til dels litt lavere), men fremdeles ikke signifikant. I DiDS-analysen er det en forskjell på 9-10 prosentpoeng, og denne er også statistisk signifikant. Mens effektestimaterne i RD, DiRD og DiDS-analysene er omtrent tilsvarende er presisjonen i disse analysene vesentlig høyere. Det henger sammen med at analysene tar snittet av flere elever, og ikke søker å finne effekten akkurat ved grensen.

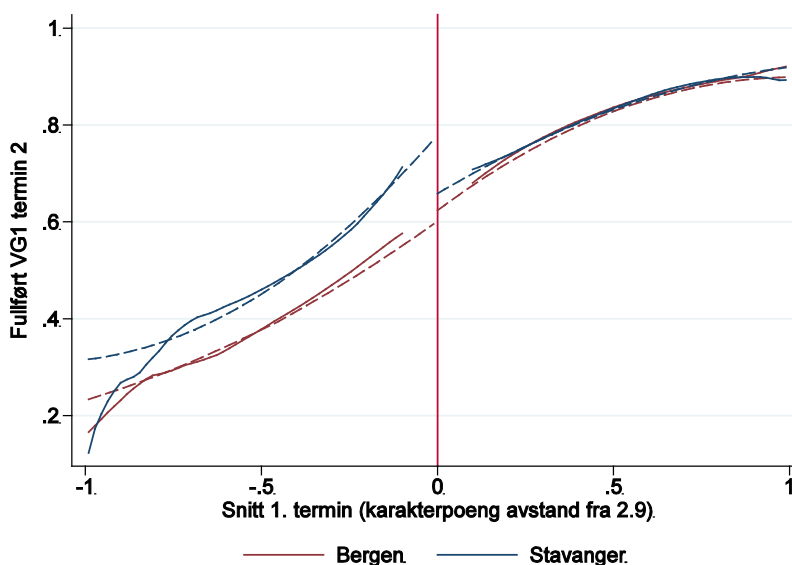
I Figur 7 viser vi en grafisk analyse av andelen som har fullført Vg1 innen to år etter fullført grunnskole, og i Figur 8 andelen oppmeldt i Vg2. For begge resultatmålene ser vi at sammenhengen mellom fullføring og resultat første termin er svært tilsvarende blant elevene over grensen. Særlig for fullført Vg1, men i noen grad også for oppmeldt Vg2 ser vi også at sammenhengene har samme form under grensen, og at sammenhengen mellom første termin og fullføring ser omtrent kontinuerlig ut rundt grensen i sammenligningsgruppen. For Stavanger ser derimot andelen fullført Vg1 og oppmeldt Vg2 til å være forskjøvet mot høyere verdier for elever under grensen, både relativt til elever over grensen og relativt svaktpresterende elever i sammenligningsgruppen. Dette samsvarer med effektene vi finner i rad ni til tolv i Tabell 6. Med andre ord, den grafiske framstillingen støtter inntrykket fra tabellen av at (de potensielt forskjellige) RD, DiRD og DiDS-effektene er av samme størrelse, omtrent 9-10 prosentpoeng, dermed er uttrykk for en ganske tilsvarende effekt på deltakere med forskjellige karakterer fra første termin.

Vi har ikke noe klart grunnlag for å si at Stavanger-elevne over grensen ville ha prestert som Bergen i fravær av intensivopplæring for elevene under grensen; det kan hende de ville prestert bedre eller svakere. Fra figurene ser det ut til at alle analysene fanger en generell økning i tidlig gjennomstrømming på omtrent 9-10 prosentpoeng. Selv om vi ikke kan være sikre på tolkningen, mener vi det er nærliggende å anta at resultatene blant Stavanger- og Bergen-elevne over grensen

ville vært tilsvarende også i fravær av intensivopplæring, det vil si at det ikke er noen ringvirkninger til disse. Sammenligning med 2010-kullet i Stavanger gir et mindre klart, men ganske tilsvarende bilde (se vedlegg B).

Ettersom vi ikke har funnet tegn til andre endringer rundt grensen, og andelen deltagere er omtrent 70 prosentpoeng høyere under enn over, svarer effekten på målgruppen til en effekt på omtrent 14 prosentpoeng på elevene som faktisk deltar.<sup>33</sup> Ettersom målgruppen utgjør 10 prosent svarer det imidlertid bare til en økning på 1 prosentpoeng i gjennomstrømning blant alle elever under ett. Til sammenligning var en av målsetningene i Ny GIV en økning av overgang fra Vg1 til Vg2 på 2 prosentpoeng innen 2013.

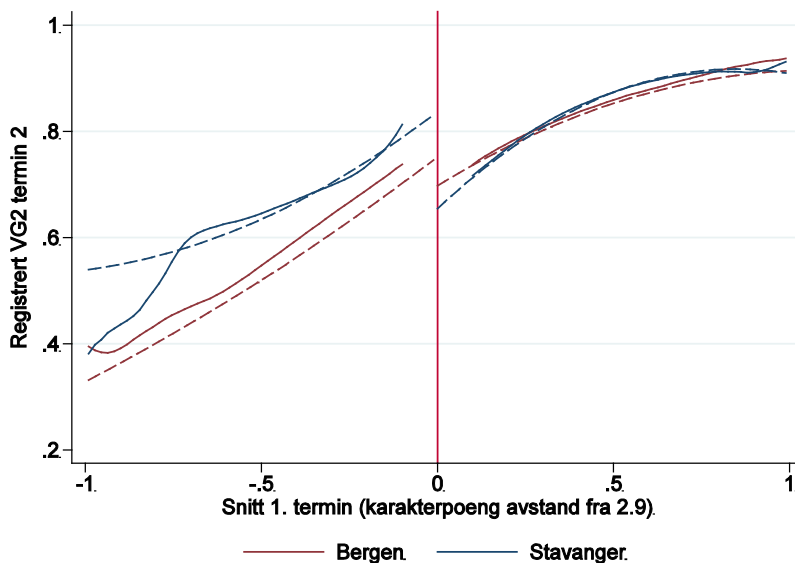
**Figur 7. Sammenheng snitt 1. termin 10. trinn og fullført VG1 t2 for avgangselever fra Stavanger 2011-2013 og fra sammenligningsgruppen Bergen 2010-2011**



Merknad: Heltrukne linjer viser glidende gjennomsnitt (fra lokal regresjon med båndbredde 0,2). Stiplede linjer viser kvadratiske tilpasninger. Horizontal akse måler snittkarakter som avvik i karakterpoeng fra 2,9.

<sup>33</sup> I Tabell 6 er det en signifikant forskjell i samleindeksen. Denne forskjellen går imidlertid i motsatt retning, vi skulle vente lavere fullføring blant elevene som deltar i intensivopplæring. Forskjellen er uansett liten sammenlignet med effektestimaterne.

Figur 8. Sammenheng snitt 1. termin 10. trinn og registrert VG2 t2 for avgangselever fra Stavanger 2011-2013 og fra sammenligningsgruppen Bergen 2010-2011



Merknad: Heltrukne linjer viser glidende gjennomsnitt (fra lokal regresjon med båndbredde 0,2). Stiplede linjer viser kvadratiske tilpasninger. Horisontal akse måler snittkarakter som avvik i karakterpoeng fra 2,9.

Det er ikke klare effekter på fullføring av Vg2, oppmeldt Vg3 eller oppnådd studie- eller yrkeskompetanse. For noen resultatmål finner vi til dels betydelige forskjeller. Disse er imidlertid sjelden signifikante. Resultatene spriker også til dels kraftig mellom forskjellige resultatmål, for forskjellig observasjonstidspunkt (to eller tre år etter fullføring av grunnskole) og mellom analyseteknikker (RD, DiRD og DiD). I grafisk analyse av disse kan det se ut til at vi har for lite data for å lage gode tilpasninger i en del analyser av disse utfallene. Ettersom vi først kan observere disse utfallene to eller flere år etter fullført grunnskole har vi et mer begrenset datagrunnlag enn for fullføring av Vg1 og oppmeldt i Vg2. For utfall tre år etter fullført grunnskole kan vi bare bruke elevene fra 2011-kullet, og når det gjelder oppnådd kompetanse er selv disse ikke forventet å oppnå yrkeskompetanse ennå.

Oppsummert finner vi ingen tegn til effekter på karakterer fra grunnskolen, vi finner tegn til økt fullføring av Vg1 og overgang til Vg2, men ingen effekter på senere utfall. Når vi undersøker mange resultatmål kan vi vente å finne noen signifikante forskjeller, selv om det egentlig ikke skulle være noen effekter. Vi må derfor gjøre en vurdering av om vi i større grad finner signifikante forskjeller enn vi skulle forvente basert på tilfeldigheter. I Vedlegg B.3 gir vi en detaljert beskrivelse av de testene vi utfører for å sjekke om våre signifikante resultater oppstår på grunn av tilfeldigheter.

I sum tolker vi svaret på disse testene som at vi gjennomgående finner noe større og mer positive effekter på tidlig gjennomstrømning enn vi skulle vente. Det er dermed sannsynlig at intensivopplæringen i Stavanger har hatt en effekt. Begrenset presisjon i analysene gjør imidlertid at vi ikke kan konkludere klart. Det er heller ingen klare tegn til effekter på karakterer eller senere gjennomstrømningsmål. Dersom det faktisk er en effekt på tidlig gjennomstrømning ville vi vente konsekvenser også for senere utfall. Effekter på fullføring av Vg2, Vg3 og oppnådd kompetanse er dermed svært relevante både for å støtte konklusjonen om effekt på gjennomstrømning, og for at effekten skal være relevant for politikktutforming – det er sannsynligvis mindre interessant med effekt på fullføring av Vg1 dersom dette ikke henger sammen med senere utfall. Vi har imidlertid foreløpig ikke tilstrekkelig datamateriale for å trekke klare konklusjoner om eventuelle effekter på mer langsiktig gjennomstrømning.

## 4. Oppsummering

Frafall i videregående opplæring (VGO) er ansett å være et betydelig problem, både for de det gjelder og for samfunnet. Høsten 2010 satte Kunnskapsdepartementet (KD) i gang en rekke tiltak for å redusere frafallet. Denne satsingen fikk navnet Ny GIV. Et sentralt tiltak er det såkalte Overgangsprosjektet, der elever med svake faglige prestasjoner midtveis i 10. trinn får tilbud om intensivopplæring i grunnleggende ferdigheter i regning og lesing/skriving. Den overordnede målsetningen for Ny GIV er å øke andelen som fullfører og består VGO med 6 prosentpoeng, fra 69 prosent for 2004-kullet til 75 prosent for 2010-kullet. Det er også målsetninger for overganger underveis, blant annet at overgang til Vg2 skal øke med 2 prosentpoeng fra 2010 til 2013.

Denne rapporten er andre delrapport i effektevalueringen av Overgangsprosjektet. Målet med Overgangsprosjektet er å øke andelen som fullfører videregående opplæring (VGO). Imidlertid er det bare et fåtall av deltakerne som har rukket å komme så langt i utdanningsløpet (avgangselever fra grunnskolen våren 2011, gitt at de startet direkte på studiespesialisering etter fullført grunnskole og har fullført på normert tid). Vi har derfor i hovedsak studert effekter på andre utfallsmål, som karakterer (først og fremst fra fullført grunnskole) og tidlige mål på gjennomstrømning i VGO (først og fremst fullføring av Vg1 og overgang til Vg2). For tidligere elevkull har alle disse målene en positiv sammenheng med fullføring av VGO (Eielsen m.fl., 2013). Dette tilsier at eventuelle effekter på tidlige utfallsmål kan være et forvarsel om framtidige effekter på fullføring.<sup>34</sup>

For å finne kausale effekter har vi hovedsakelig basert oss på to analysestrategier, nemlig forskjeller mellom årskull og skoler i forskjellige faser og forskjeller innen Stavanger-skoler.

Den første analysestrategien, som er en såkalt «difference-in-difference» analyse (DiD), inkluderer alle skoler og sammenligner utviklingen i skolens snittresultat for skoler i forskjellige faser. Vi er interessert i å studere hvorvidt resultatene endrer seg når skolene starter opp intensivopplæringen. Imidlertid kan også snittresultatene endre seg av andre grunner. Vi sammenligner derfor resultatutviklingen i skoler som innfører intensivopplæring (tiltaksskoler) med resultatutviklingen i skoler som ikke gjør det (sammenligningsskoler).<sup>35</sup> Historisk har resultatene til skoler i ulike faser utviklet seg tilsvarende. Dette tilsier at sammenligningsskolene gir et bilde på hvordan resultatene i tiltaksskolene ville ha utviklet seg i fravær av intensivopplæringen. Differansen i resultatutviklingen mellom tiltaks- og sammenligningsskoler er derfor et anslag på den kausale effekten av å innføre intensivopplæring. På grunn av lite systematikk i tilordningen av elever til intensivopplæring innen skoler, greier vi bare å estimere totaleffekten av å innføre intensivopplæring. Totaleffekten består av den direkte effekten på deltakende elever på deltakende skoler og potensielle ringvirkninger på ikke-deltakende elever på deltakende skoler.

Vi finner ingen effekter av intensivopplæringen med DiD-analysene, verken på karakterer eller på gjennomstrømning i VGO. Analysene baserer seg på et omfattende datamateriale, flere fullstendige årskull med grunnskoleelever. Dette gjør at effektestimaterne er ganske presise, slik at vi kan utelukke selv små

<sup>34</sup> Det er derimot ikke sikkert at sammenhengene mellom effekter på forskjellige mål svarer til observerte sammenhenger mellom disse målene. Som et eksempel, dersom først og fremst elevenes motivasjon påvirkes kan dette tenkes å ha en større effekt på fullføring enn f.eks. eksamenskarakterer. Cortes m.fl. (2014) er et eksempel på en studie der effekten er større på fullføring enn på faglige prestasjoner.

<sup>35</sup> Vi er dermed i stand til å ta hensyn til resultatforskjeller mellom skolene som er stabile over tid.



gjennomsnittlige effekter.<sup>36</sup> For standpunkt karakterer ved fullført grunnskole kan vi nesten utelukke positive effekter (karakterene er omtrent stabile, eller kan ha blitt redusert med inntil 0,03 karakterpoeng), mens det for eksamens karakterer kan ha vært positive eller negative effekter på inntil 0,025 karakterpoeng.<sup>37</sup> Et av målene for Ny GIV var at overgang til Vg2 skulle øke med 2 prosentpoeng fra 2010 til 2013. Vi finner ingen effekt av Overgangsprosjektet på overgang til Vg2, men kan langt på vei utelukke effekter større enn 1 prosentpoeng og større enn omtrent 0,5 prosentpoeng på fullføring av Vg1.<sup>38</sup> Den minste effekten vi kan finne for eksamens karakterer fra grunnskolen, svarer til at én av 40 elever går opp en karakter.

Presisjonen i estimatene er vesentlig høyere enn for tilsvarende estimer i Eielsen m.fl. (2013), nær doblet for målene på gjennomstrømning (dvs. vi kan nå finne effekter som er litt over halvparten av minste mulige effekt i Eielsen m.fl.) og økt med omtrent 50 prosent for karakterer (dvs. vi kan finne effekter tilsvarende 2/3 av minste mulige effekt i Eielsen m.fl.). Vi finner heller ikke effekter på fullføring av Vg2, overgang til eller fullføring av Vg2 eller på oppnådd (studie- eller yrkes-) kompetanse. Disse senere gjennomstrømningsmålene har imidlertid mindre presise effektestimater, ettersom vi foreløpig har et mye mer begrenset datagrunnlag. I snitt deltar bare omtrent 10 prosent av elevene ved tiltaksskolene i intensivopplæringen. Det betyr at en liten positiv snitteffekt kan skjule vesentlige større direkte effekter. Som allerede nevnt kan vi ikke skille mellom direkte effekter og ringvirkninger på grunn av uklar implementering. Dersom vi antar at det ikke er ringvirkninger er effektene på deltagerne ti ganger større enn snitteffekten. Under en slik antagelse kan vi utelukke at effekten på deltagerne er mer enn 10 prosentpoeng for overgang til Vg2 og 0,25 karakterpoeng for skriftlig eksamen.

I et forsøk på å skille noe bedre mellom effekter på deltagere og ringvirkninger har vi gjort separate effektestimater for grupper av elever med forskjellige resultater til første termin, til sammen fem ulike grupper. Dersom intensivopplæringen hadde blitt gitt utelukkende til de aller svakeste elevene, ville vi kunne skille mellom effekter på deltagere og ringvirkninger. Selv om dette ikke er tilfelle, i strid med føringene fra KD er det også elever med gode resultater fra første termin som deltar, er det likevel forskjeller i andelen deltagere i de ulike gruppene. De to gruppene med de svakeste resultatene har henholdsvis 35 og 31 prosent deltagere i intensivopplæringen, de øvrige gruppene har færre og i gruppen bestående av de sterkeste elevene er andelen bare 1 prosent. Vi finner ingen klare tegn til effekter av intensivopplæringen for noen av gruppene, og kan utelukke effekter på fullføring av Vg1 og overgang til Vg2 større enn 2-3 prosentpoeng. Hvis vi antar at det ikke er ringvirkninger og justerer resultatene for andel deltakere, svarer det til at vi kan utelukke effekter for deltakerne større enn 4-8 prosentpoeng.

Vi har også undersøkt om det er forskjellige effekter for forskjellige grupper av elever. Vi finner imidlertid ikke tegn til effekter verken når vi studerer effekter separat for kjønn, innvandringsbakgrunn eller foreldres utdanning. Vi har videre estimert effekter for hvert fylke. Overgangsprosjektet har hatt fylkeskommunale prosjektledere, og det er forskjeller fylkene i mellom i gjennomføringen. Det kan dermed tenkes at enkelte fylker har lyktes godt, selv om vi ikke finner noen

---

<sup>36</sup> Med *utelukke* mener vi at effektene ligger utenfor et 95 prosent-konfidensintervall. Det er altså ikke umulig at det er større effekter, men når vi ser på et enkelt estimat er det svært usannsynlig. Når vi ser på mange estimater, er det sannsynlig at noen effekter ligger utenfor sitt konfidensintervall, men det er da like sannsynlig at de er større (mer positive) som at det mindre (mer negative).

<sup>37</sup> Vi har også undersøkt standpunkt karakterer i enkeltfag, samt karakterer fra VGO. Her har vi heller ikke funnet tegn til effekter, men disse estimatene er mindre presise.

<sup>38</sup> Disse tallene er ikke helt sammenlignbare, målet for Ny GIV er knyttet til overgang fra Vg1, mens vi hele tiden studerer resultater en gitt tid etter fullført grunnskole.

generell effekt. Vi finner imidlertid ingen klare tegn til effekter for enkeltfylker heller.<sup>39</sup>

En viktig forskjell i organiseringen av intensivopplæringen er hvordan elevene er valgt ut til å delta. Eielsen m.fl. (2013) viser at Stavanger har en særlig klar utvelgelsespraksis, der nesten utelukkende elever under en gitt karakter fra første termin 10. trinn deltar i intensivopplæring (og der disse i svært stor grad deltar). Eielsen m.fl. bruker dette skarpe skillet i andelen som deltar til å studere effekter innen Stavanger-skoler.

Også i 2012 og 2013 har det vært forskjellig praksis når det gjelder utvelgelse av elever, og generelt ikke en klar avgrensning mellom elever som befinner seg innenfor og utenfor målgruppen. Vi har lett etter flere kommuner med en tilsvarende klar praksis som Stavanger, uten hell. I likhet med Eielsen m. fl. (2013) har vi derfor fortsatt å studere forskjeller innen Stavanger-skoler. Vi benytter hovedsakelig to analysestrategier for å utlede kausale effekter for Stavanger. Det ene er en såkalt «regression-discontinuity» analyse (RD, samt DiRD) som sammenlikner elever i Stavanger som befinner seg like under tilordningsgrensen (tiltaksgruppen) med elever i Stavanger som befinner seg like over tilordningsgrensen (sammenligningsgruppen). Til forskjell fra snitteffektene diskutert over (effekten av å innføre intensivopplæring ved en skole) gir disse sammenligningene marginale effekter. Vi kan tenke på disse som effekten av en liten utvidelse av omfanget på intensivopplæringen. Det vil si at effekten ikke lenger omfatter ringvirkninger på elever som ikke deltar, men utgjør forskjellen mellom den direkte effekten på den siste eleven som deltar og ringvirkningene på den første som ikke deltar.

Den andre analysen er en «difference-in-differences» analyse (DiDS) som sammenlikner snittresultater til elever i Stavanger (tiltaksgruppen) i et bestemt karakterintervall over og under tilordningsgrensen med snittresultater til elever i en sammenligningsgruppe innenfor samme karakterintervall over og under samme tilordningsgrense. Som sammenligningsgruppe bruker vi skoler i Bergen som ennå ikke hadde innført intensivopplæring. En sammenligning av familebakgrunnsvariable og resultater fra nasjonale prøver i tiltaks- og sammenligningsgruppen tyder på at elever i sammenligningsgruppen gir et anslag på hva elever i tiltaksgruppen ville oppnådd i fravær av intensivopplæringen. Dette tilsier at vi kan finne kasuale effekter ved å benytte Bergen som sammenligningsgruppe.

Vi finner ingen klare effekter på karakterer fra grunnskole eller VGO, verken med RD-analysen eller med DiDS-analysen. Derimot viser begge analysestrategiene tegn til økt gjennomstrømning. Vi beregner at blant elevene som har resultater fra første termin rett under grensen for deltagelse i intensivopplæringen, er andelen som fullfører Vg1 omtrent 10 prosentpoeng høyere enn blant elevene rett over grensen, og andelen med overgang til Vg2 omtrent 15-20 prosentpoeng høyere. Disse RD-estimatene er imidlertid upresise, og ikke statistisk signifikante. DiDS-estimatene for fullføring av Vg1 og overgang til Vg2 er derimot statistisk signifikant og kvalitativt tilsvarende RD-estimatene. Andelen elever i Stavanger under tilordningsgrensen med fullført Vg1 og overgang til VG2 er omtrent 9 prosentpoeng høyere enn i Bergen. Ettersom omtrent 70 prosent av de svakt-presterende Stavanger-elevne deltar i intensivopplæring tilsvarer dette en effekt på omtrent 13 prosentpoeng på deltagerne.

Vi finner ingen klare effekter på senere gjennomstrømning, fullføring av Vg2 samt overgang til og fullføring av Vg3. Disse resultatene baserer seg imidlertid på et

---

<sup>39</sup> Oslo og Østfold utmerker seg med store effektestimater. For nettopp disse fylkene er det imidlertid spesielle forhold knyttet til hhv. organiseringen av intensivopplæringen og feilregistreringer i datamaterialet som gjør at vi ikke kan tolke resultatene. Vi finner enkelte signifikante estimat også for andre fylker, men det er lite systematisk i disse, og det er heller ikke flere enn omtrent så mange vi ville vente utfra tilfeldig variasjon.

begrenset datamateriale, og er mindre presise, slik at det er vanskelig å konkludere på bakgrunn av disse.

Ettersom vi gjør mange forskjellige analyser kan vi vente å finne noen signifikante estimater selv om det ikke skulle være noen effekt. Vi finner imidlertid mer signifikante resultater enn vi skulle forvente utfra tilfeldigheter. Dessuten er det en klar systematikk i at resultater for tidlig gjennomstrømning er signifikante, og at disse gjennomgående er positive. Vår tolkning blir dermed at intensivopplæringen i Stavanger sannsynligvis har ført til en økning i tidlig gjennomstrømning på omtrent 9 prosentpoeng.<sup>40</sup> Denne effekten kan i prinsippet også til dels gjenspeile negative ringvirkninger blant elever som ikke deltar, men den grafiske analysen (figur 7 i teksten) tyder ikke på at dette er tilfelle. For å bekrefte funnet og øke relevansen bør det være effekter også på videre gjennomstrømning. Vi har foreløpig ikke data som lar oss studere dette på noen grundig måte.

Vi mener en effekt på målgruppen på 9 prosentpoeng er en potensielt viktig effekt. Ettersom målgruppen bare utgjør 10 prosent svarer det imidlertid bare til en effekt på knappe 1 prosentpoeng for alle elever. Målsettingen i Ny Giv var en økning i gjennomføring av VGO på 6 prosentpoeng over seks år. Effektene vi finner er altså langt under målsetningen for økning i fullføring. Samtidig kan en effekt på fullføring av VGO tilsvarende 1 prosentpoeng være av interesse. Dersom frafallet i Stavanger reduseres med 1 prosentpoeng svarer det til at omtrent 15 ekstra elever fullfører hvert år. Falch m.fl. (2009) finner en samfunnsøkonomisk kostnad av frafall på 900 000 kr per elev, det vil si en samfunnsøkonomisk gevinst på 13,5 millioner per årskull. Dersom frafallet reduseres med 1 prosentpoeng nasjonalt tilsvarer det en samfunnsøkonomisk gevinst på 540 millioner per årskull.<sup>41</sup>

For å oppsummere resultatene finner vi ingen tegn til effekter av å innføre intensivopplæringen når vi studerer alle fylker og kull. Vi kan utelukke at det er effekter på tidlig gjennomføring som er større enn omtrent halvparten av målsetningen i Ny GIV. Det er dermed ikke sannsynlig at intensivopplæringen fører til effekter på fullføring av VGO i nærheten av den overordnede målsetningen. I ettertid framstår målsetningen for Ny GIV som ambisiøs. Dersom denne skulle oppnås gjennom intensivopplæring til de 10 prosent svakestpresterende, måtte enten disse økt sin fullføring med 60 prosentpoeng, eller det måtte vært betydelige positive ringvirkninger til andre elever.<sup>42</sup>

Det er imidlertid betydelig heterogenitet i gjennomføringen av intensivopplæringen. Den klare tilordningspraksisen i Stavanger gjør at vi kan analysere effekter innen Stavanger-skoler. Her finner vi tegn til økt fullføring av Vg1 og overgang til Vg2. Den beregnede effekten er mindre enn målsetningen i Ny GIV, men likevel stor nok til å kunne være av en viss samfunnsøkonomisk betydning, dersom den også finnes på fullføring av VGO.

Det er vanskelig å overføre funnene fra Stavanger til øvrige kommuner. Forskjellen vi ser i tilordningspraksis tyder på at organiseringen i Stavanger er annerledes. Dette kan også ha hatt konsekvenser for gjennomføringen av intensivopplæringen, dermed kan intensivopplæringen ha hatt en større effekt i Stavanger enn andre steder. Dersom Stavanger oppnår tilsvarende resultater på fullføring av VGO som

<sup>40</sup> Gitt at vi stoler på estimatene kan vi utelukke effekter mindre enn 0 og større enn omtrent 17 prosentpoeng.

<sup>41</sup> Anslagene på kostnader ved frafall er naturligvis usikre. Vi kunne ha justert for lønnsvekst, ettersom de baserer seg på historiske inntekter og trygdesatser mm. fra 2008. Vi har ikke gjort det, dette bidrar til at anslagene kan være konservative.

<sup>42</sup> Ny GIV inneholder også andre tiltak, f.eks. Oppfølgingsprosjektet. Det er dermed kanskje ikke relevant å sammenligne effekter av Overgangsprosjektet isolert med målsetninger for Ny GIV sett under ett.

på tidlig gjennomføring vil det imidlertid være interessant å studere nærmere organisering og gjennomføring av intensivopplæringen i Stavanger i mer detalj.

Det kan være flere forklaringer til at vi ikke finner noen generell effekt av intensivopplæringen. For det første er det mulig at den faktiske effekten av intensivopplæringen stort sett er liten eller null. Mulige årsaker til dette kan være at intensivopplæringen ikke er omfattende nok (intensivopplæring i begrenset omfang over kort tid, uten vesentlig tilførsel av ekstra ressurser), og/eller ble gitt for sent (på slutten av grunnskolen), på et tidspunkt hvor det kan være vanskelig å endre elevenes resultater, iallfall på en målbar måte. Vi har på nåværende tidspunkt ikke noe grunnlag for å si at en av disse mulige forklaringene er mer eller mindre sannsynlig eller relevant. Et annet poeng er at når målgruppen er vagt definert, er vi henvist til å studere snitteffekter på store grupper med en beskjeden andel intensivelever (i snitt 10 prosent). Dermed blir det vanskeligere å finne direkte effekter, som også vil sammenblandes med eventuelle ringvirkninger. Selv om vi kan utelukke snitteffekter utenfor et forholdsvis lite intervall kan det være til dels betydelige effekter for deltakerne.

En annen forklaring er i hvilken grad Ny GIV ble forankret blant relevante aktører. Viktigheten av forankring blant lærere illustreres i blant annet Jacob (2015) hvor tilfeldig valgte lærere fra 2. til 5. klasse i sju skoler i Michigan ble kurset i hvordan man på en strukturert måte skulle undervise i lesing. Tiltaket viste seg å ha liten effekt på elevprestasjoner, noe som i stor grad skyldtes at lærerne ikke ble overbevist av kursmaterialet de skulle undervises i (lærerne måtte svare på et spørreskjema i forbindelse med kursingen). Dersom man legger funnene i Lødding og Holen (2013) til grunn, er ikke lav forankringsgrad årsaken til mangelen på positiv effekt av intensivopplæringen. Ved å invitere tilfeldig valgte 234 skoleledere til å delta i en spørreundersøkelse (med en svarprosent på 43,2 prosent) finner de at 99 prosent av lærerne som gjennomfører intensivopplæringen er ganske, eller svært fornøyd med innholdet i skoleringen. Grunn av lav svarprosent, er det likevel usikkert om skolene som besvarte spørreskjemaet representerer et tilfeldig utvalg av alle skolene. Også Helgøy og Homme (2013) gir en detaljert studie av sentrale aktørers erfaring med Ny GIV. Imidlertid kan ikke funnene i den analysen benyttes i den kvantitative effektevalueringen ettersom deres datamateriale kun består av et begrenset antall observasjoner, som ikke nødvendigvis er representative.

Det kan også hende at det er effekter som våre resultatmål ikke fanger opp. For eksempel er intensivopplæringen ment å bedre elevenes grunnleggende ferdigheter. Vårt beste mål for grunnleggende ferdigheter er karakterer i standpunkt og eksamen. I den grad karakterer ikke fanger opp grunnleggende ferdigheter vil vi ikke være i stand til observere effekten. Dersom ferdighetene bedres uten at dette kommer til uttrykk i karakterene, burde vi muligens ha funnet en effekt av intensivopplæring på fullføring av Vg1 (hvor vi observerer alle kullene), noe vi ikke gjør. Dette tyder derfor på at det heller ikke er noen langsiktige effekter, som fullføring av VGO, i vente. Cortes mfl. (2014), som studerer en dobling av matteundervisningen, er likevel et eksempel på at det kan være effekter på fullføring, selv om det er ingen eller bare svake tidlige effekter på faglige prestasjoner.

## Referanser

Angrist, J. D. & Pischke, J. (2008). *Mostly harmless econometrics: An Empiricist's Companion*, Princeton University Press.

Bettinger, E. P., & Long, B. T. (2009). Addressing the Needs of Underprepared Students in Higher Education Does College Remediation Work?. *Journal of Human Resources*, 44(3), 736-771.

Carneiro, P., Løken, K. V., & Salvanes, K. G. (2015). A Flying Start? Maternity Leave Benefits and Long-Run Outcomes of Children. *Journal of Political Economy*, 123(2), 365-412.

Cook, P. J., Dodge, K., Farkas, G., Fryer, J. R. G., Guryan, J., Ludwig, J., Mayer, S. (2015). *Not Too Late: Improving Academic Outcomes for Disadvantaged Youth*. University of Chicago, Northwestern University Working Paper Series WP-15-01

Cortes, K., Goodman, J. & Nomi, T. (2013). *Intensive Math Instruction and Educational Attainment: Long-Run Impacts of Double-Dose Algebra*, working paper, Harvard University

De Haan, M. (2015). The effect of additional funds for low-ability pupils: A nonparametric bounds analysis, *Economic Journal*, forthcoming

Eielsen, G., Kirkebøen, L.J, Leuven, E., Rønning, M. & Raaum, O. (2013): *Effektevaluering av intensivopplæringen i Overgangsprosjektet*, Ny GIV. Første delrapport. Statistisk Sentralbyrå Rapport 54/2013.

Falch, T., Johannessen, A. B. og Strøm, B. (2009). *Kostnader av frafall i videregående oppløring*. SØF-rapport 08/09.

Falch, T., Nyhus, O. H., & Strøm, B. (2014). Performance of young adults: The importance of different skills. *CESifo Economic Studies*, Vol. 60, 2014

Galloway, T. A., Kirkebøen, L. J. og Rønning, M. (2011). *Karakterpraksis i grunnskoler. Sammenheng mellom standpunkt og eksamenskarakterer*. Reports 4/2011 Statistisk Sentralbyrå.

Helgøy, I. & Homme, A. (2013). *Ny GIV Overgangsprosjektet - konsekvenser for skolen. Evaluering av Ny GIV Overgangsprosjektet*. Uni Rokkansenteret. (ISBN 978-82-8095-091-8)

Holen, Solveig og Berit Lødding (2012). *Intensivopplæringen i Ny GIV for 10. trinnslever våren 2012*, Rapport 42/2012, NIFU

Jacob, B. (2015). *When evidence is not enough: Findings from a randomized evaluation of evidenced based literacy instruction (Ebli)*. NBER Working Paper No. 21643

Lavy, V. & Schlosser, A. (2005). Targeted remedial education for underperforming teenagers: Costs and benefits. *Journal of Labor Economics*, 23(4).

Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2007). The effect of extra funding for disadvantaged pupils on achievement. *The Review of Economics and Statistics*, 89(4), 721-736.

Leuven, E. og Oosterbeek H. (2004). Evaluating the effect of tax deduction on training. *Journal of Labor Economics* 22(1), 461-488.

Lillejord, S., Halvorsrud, K., Ruud, E., Morgan, K., Freyr, T., Fischer-Griffiths, P., Eikeland, O. J., Hauge, T. E., Homme, A. D., & Manger, T. (2015). Frafall i videregående opplæring: En systematisk kunnskapsoversikt. Oslo: Kunnskapssenter for utdanning, [www.kunnskapssenter.no](http://www.kunnskapssenter.no)

Lødding, B., & Solveig, H. (2013). Intensivopplæring i eller utenfor klassen? Sluttrapport fra prosjektet Kartlegging av deltakelse, organisering og opplevelse i Overgangsprosjektet innenfor Ny GIV.

Martorell, P., & McFarlin Jr, I. (2011). Help or hindrance? The effects of college remediation on academic and labor market outcomes. *The Review of Economics and Statistics*, 93(2), 436-454.

OECD, (2014). Education at a Glance. *Oecd indicators*, 2014.

Rønning, W., Hodgson, J. og Tomlinson, P. (2013). Å SE OG BLI SETT. Klasseromsobservasjoner av intensivopplæringen i Ny Giv. NF-rapport Nr. 6/2013.

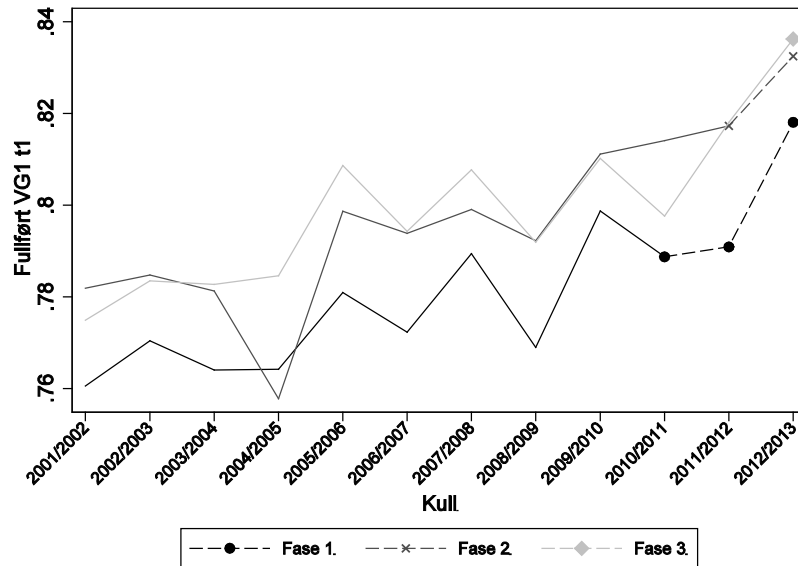
Sletten, M. A., Bakken, A. & Haakestad, H. (2011). Ny start med Ny GIV?, Rapport 23/2011, NOVA

Utdanningsdirektoratet (2015) Utdanningsspeilet 2015

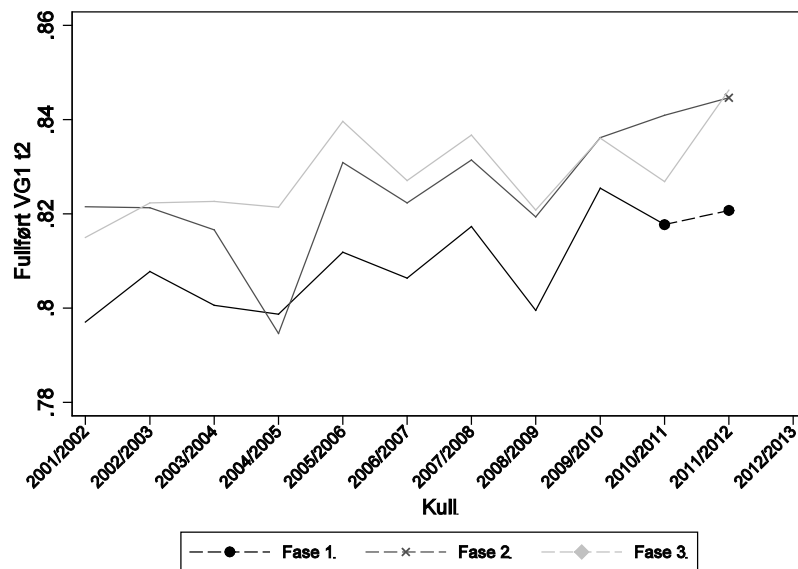
## Vedlegg A: Supplerende resultater fra analyser av forskjeller mellom faser og årskull («difference-in-differences-modeller»)

### A.1 Utvikling over tid i fase 1, 2 og 3

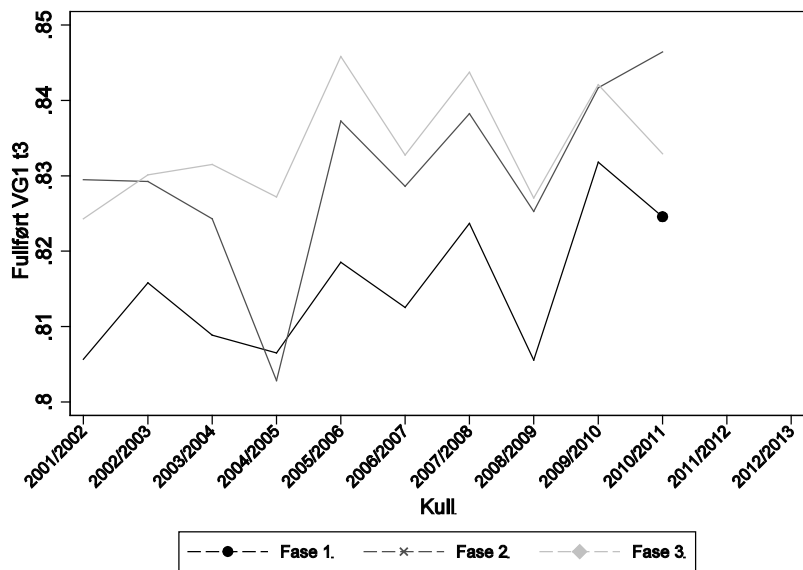
Figur A1.1. Utvikling i fullført VG1 t1 per fase



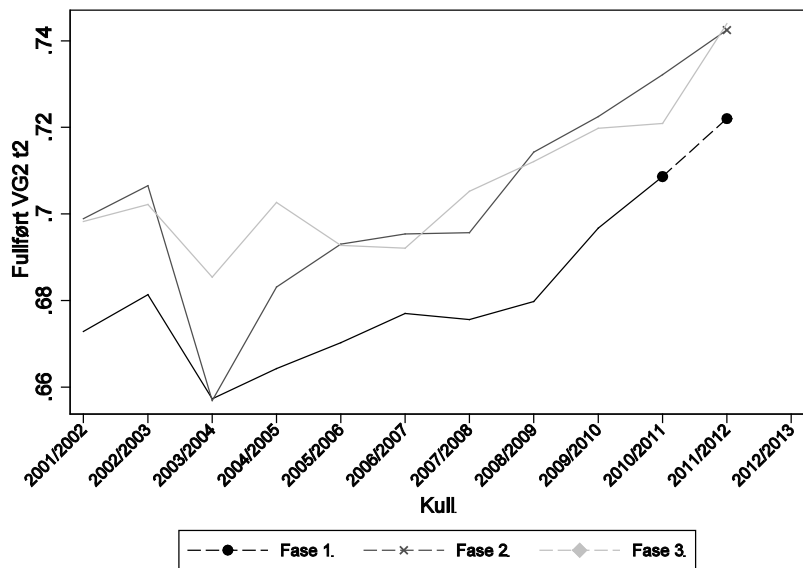
Figur A1.2: Utvikling i fullført VG1 t2 per fase



Figur A1.3. Utvikling i fullført VG1 t3 per fase

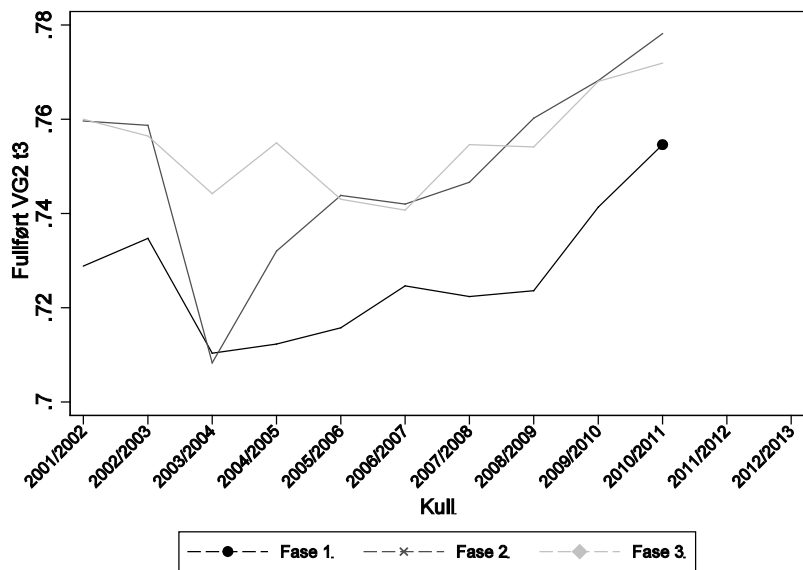


Figur A1.4. Utvikling i fullført VG2 t2 per fase

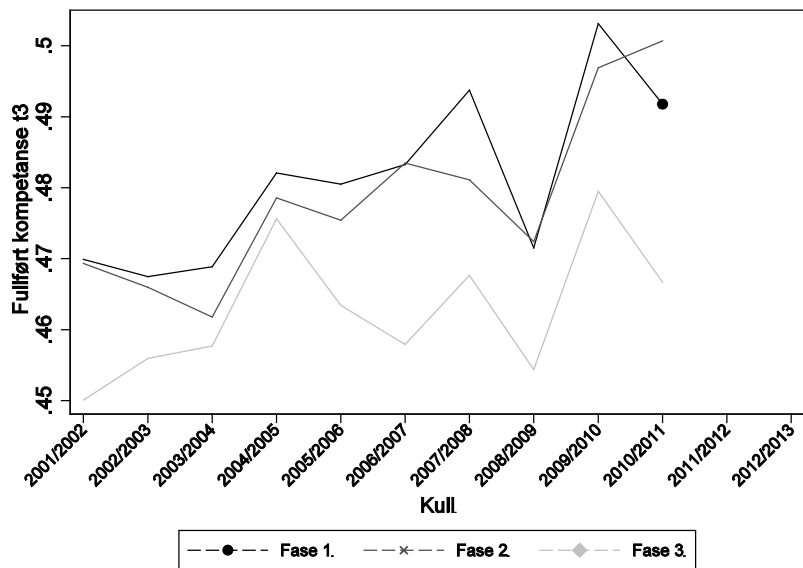




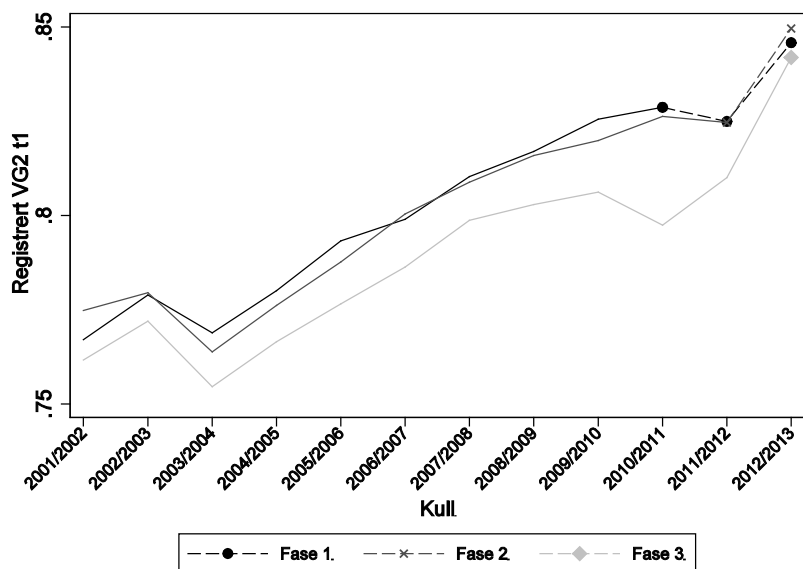
Figur A1.5. Utvikling i fullført VG2 t3 per fase



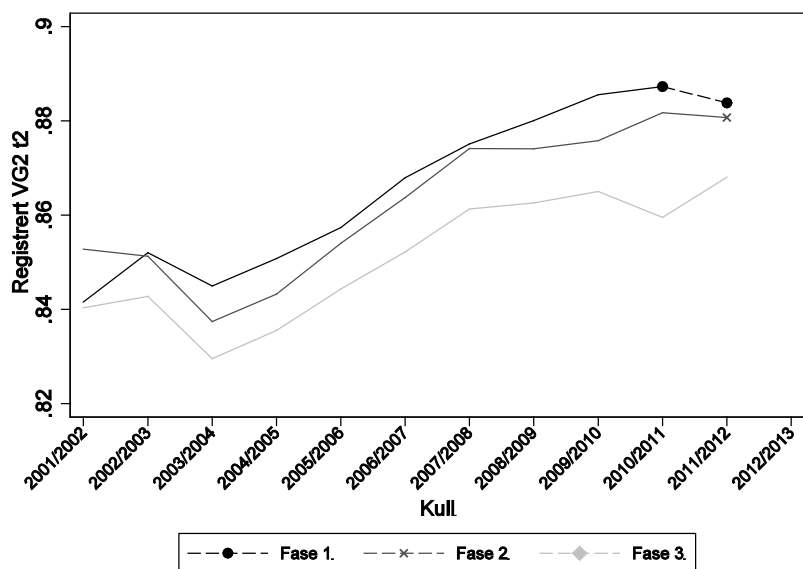
Figur A1.6. Utvikling i fullført kompetanse t3 per fase



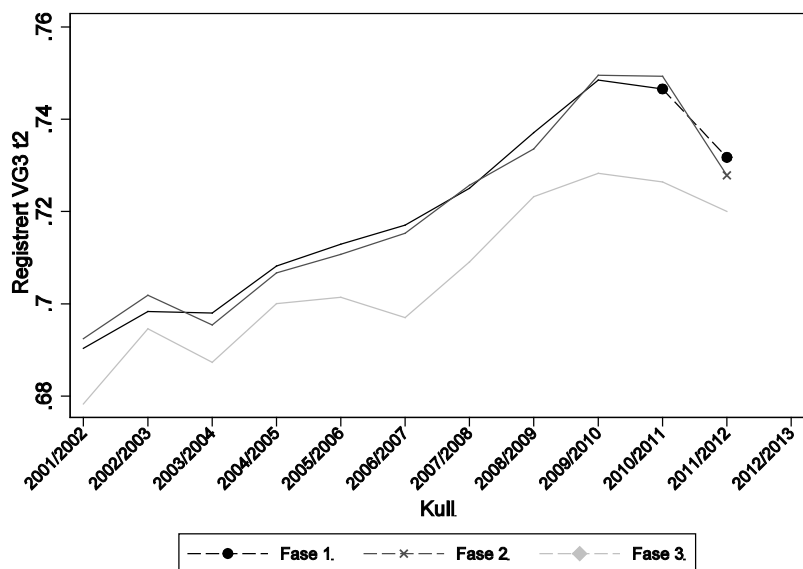
Figur A1.7. Utvikling i registrert VG2 t1 per fase



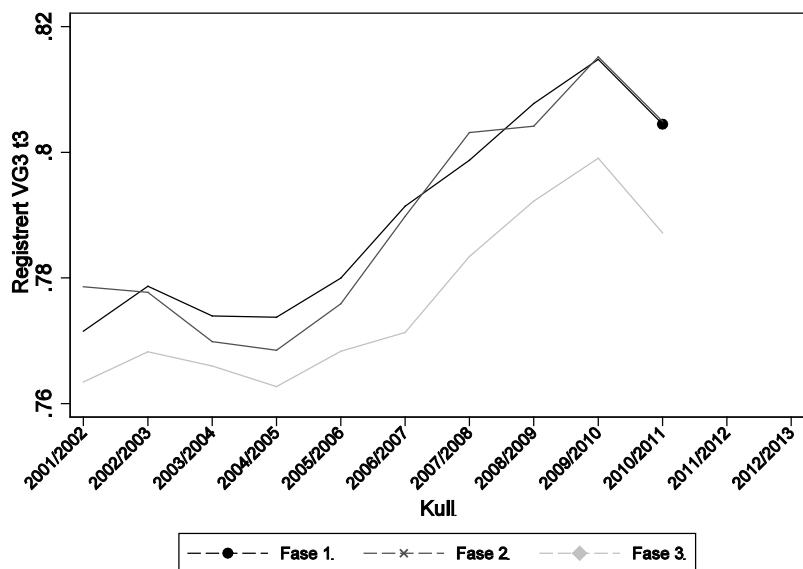
Figur A1.8. Utvikling i registrert VG2 t2 per fase



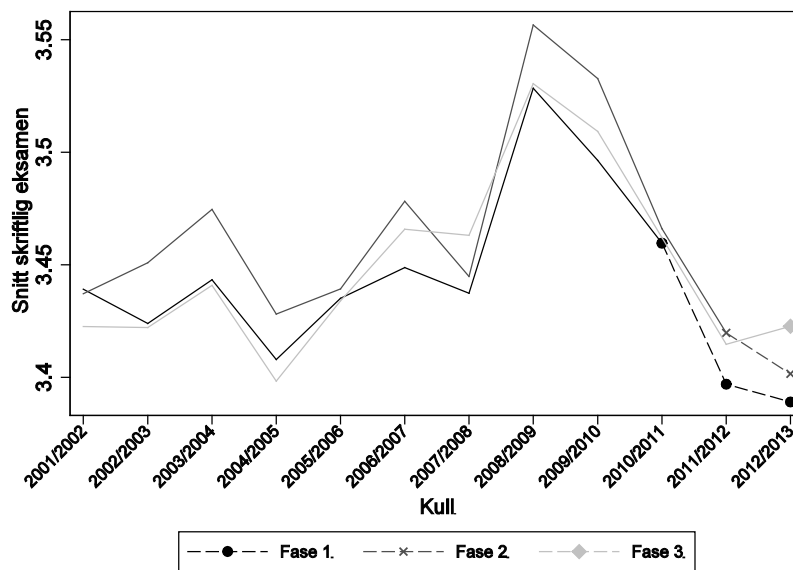
Figur A1.9. Utvikling i registrert VG3 t2 per fase



Figur A1.10. Utvikling i registrert VG3 t3 per fase



Figur A1.11 Utvikling i snitt skriftlig eksamenskarakter



## A.2 Effekten på elevprestasjoner av intensivopplæring, heterogene effekter, difference-in-differences-estimerer.

**Tabell A.2.1. Heterogene effekter, kjønn**

	Ny Giv	Snitt standpkt	Snitt skriftlig eks 10.	Fullført VG1 t1	Fullført VG1 t2	Fullført VG2 t2	Fullført VG2 t3	Fullført VG3 t3	Fullført kompetanse t3	Registrert VG2 t1	Registrert VG2 t2	Registrert VG3 t2
Gutt	0,121*** (0,0060)	-0,004 (0,0125)	0,034 (0,0180)	-0,005 (0,0047)	-0,005 (0,0049)	-0,001 (0,0051)	-0,001 (0,0071)	-0,009 (0,0075)	-0,005 (0,0076)	0,007 (0,0051)	0,000 (0,0033)	-0,004 (0,0057)
Jente	0,084*** (0,0094)	-0,016 (0,0104)	0,001 (0,0153)	0,001 (0,0043)	0,001 (0,0041)	0,002 (0,0059)	0,009 (0,0084)	-0,004 (0,0089)	-0,004 (0,0088)	0,002 (0,0043)	0,001 (0,0034)	-0,003 (0,0049)
Observasjoner	373 674	373 094	354 775	373 674	315 400	315 400	256 203	256 203	256 203	373 674	315 400	315 400
Snitt utfall	0,0327	4,024	3,471	0,817	0,839	0,724	0,763	0,507	0,496	0,830	0,0863	0,742
Pre-trend	.	0,752	0,529	0,402	0,631	0,0276	0,0868	0,666	0,684	0,0785	0,167	0,0625
Kommuner	405	405	405	405	405	405	386	386	386	405	405	405

Merknad: Standardfeil justert for korrelasjoner innen kommuner. Stjerner angir statistisk signifikans: \* p<0,05; \*\* p<0,01; \*\*\* p<0,001.

**Tabell A.2.2. Heterogene effekter, innvandrere**

	Ny Giv	Snitt standpkt	Snitt skriftlig eks 10.	Fullført VG1 t1	Fullført VG1 t2	Fullført VG2 t2	Fullført VG2 t3	Fullført VG3 t3	Fullført kompetanse t3	Registrert VG2 t1	Registrert VG2 t2	Registrert VG3 t2
Høy utd,lang	0,094*** (0,0030)	-0,007 (0,0096)	0,021 (0,0141)	0,001 (0,0036)	0,001 (0,0037)	0,003 (0,0054)	0,003 (0,0046)	0,007 (0,0060)	-0,005 (0,0071)	-0,003 (0,0070)	0,006 (0,0035)	0,003 (0,0031)
VGO	0,167*** (0,0279)	-0,085** (0,0286)	0,027 (0,0422)	-0,029* (0,0141)	-0,031* (0,0153)	-0,003 (0,0179)	-0,048* (0,0194)	-0,029 (0,0254)	-0,039 (0,0228)	-0,041 (0,0214)	-0,009 (0,0123)	0,003 (0,0128)
Grunnskole	0,160*** (0,0271)	0,002 (0,0233)	-0,028 (0,0374)	-0,018 (0,0149)	-0,013 (0,0145)	-0,017 (0,0206)	-0,012 (0,0159)	-0,008 (0,0218)	-0,025 (0,0192)	-0,022 (0,0199)	-0,004 (0,0151)	-0,001 (0,0132)
Observasjoner	373 745	373 141	354 814	373 745	315 467	256 266	315 467	256 266	256 266	256 266	373 745	315 467
Snitt utfall	0,0327	4,024	3,471	0,816	0,839	0,842	0,724	0,763	0,507	0,496	0,830	0,884
Pre-trend	.	0,000728	0,0318	0,0000423	0,000784	0,000412	0,000141	0,0000319	0,166	0,184	0,0445	0,115
Kommuner	405	405	405	405	405	386	405	386	386	386	405	405

Merknad: Standardfeil justert for korrelasjoner innen kommuner. Stjerner angir statistisk signifikans: \* p<0,05; \*\* p<0,01; \*\*\* p<0,001.

**Tabell A.2.3. Heterogene effekter, foreldres utdanningsnivå**

	Ny Giv	Snitt standpkt	Snitt skriftlig eks 10.	Fullført VG1 t1	Fullført VG1 t2	Fullført VG2 t2	Fullført VG2 t3	Fullført VG3 t3	Fullført kompetanse t3	Registrert VG2 t1	Registrert VG2 t2	Registrert VG3 t2
Høy utd,lang	0,027*** (0,0055)	-0,009 (0,0143)	0,035 (0,0227)	-0,003 (0,0074)	-0,002 (0,0072)	-0,006 (0,0079)	0,001 (0,0108)	-0,015 (0,0129)	-0,013 (0,0124)	-0,005 (0,0067)	-0,005 (0,0067)	-0,005 (0,0077)
Høy utd, kort	0,062*** (0,0056)	-0,012 (0,0113)	0,017 (0,0155)	0,004 (0,0046)	0,003 (0,0049)	0,006 (0,0062)	0,006 (0,0077)	-0,008 (0,0102)	-0,007 (0,0105)	0,004 (0,0047)	0,004 (0,0038)	-0,002 (0,0056)
VGO	0,133*** (0,0062)	-0,011 (0,0112)	0,016 (0,0153)	-0,005 (0,0048)	-0,003 (0,0048)	0,001 (0,0059)	0,009 (0,0083)	0,001 (0,0092)	0,003 (0,0087)	0,009 (0,0050)	0,007 (0,0039)	-0,003 (0,0056)
Grunnskole	0,212*** (0,0175)	-0,016 (0,0209)	0,001 (0,0302)	0,001 (0,0132)	-0,006 (0,0129)	-0,003 (0,0138)	-0,017 (0,0156)	-0,031 (0,0162)	-0,028 (0,0158)	0,009 (0,0116)	-0,006 (0,0108)	0,002 (0,0117)
Observasjoner	370 556	370 037	352 236	370 556	312 872	312 872	254 209	254 209	254 209	370 556	312 872	312 872
Snitt utfall	0,0324	4,030	3,477	0,820	0,842	0,727	0,766	0,510	0,498	0,832	0,885	0,745
Pre-trend	.	0,955	0,271	0,0982	0,00511	0,469	0,0905	0,736	0,713	0,360	0,00265	0,643
Kommuner	405	405	405	405	405	405	386	386	386	405	405	405

Merknad: Standardfeil justert for korrelasjoner innen kommuner. Stjerner angir statistisk signifikans: \* p<0,05; \*\* p<0,01; \*\*\* p<0,001.

Tabell A.2.4. Heterogene effekter, fylke

	Ny Giv	Snitt standpkt	Snitt skriftlig eks 10.	Fullført VG1 t1	Fullført VG1 t2	Fullført VG2 t2	Fullført VG2 t3	Fullført VG3 t3	Fullført kompetanse t3	Registrert VG2 t1	Registrert VG2 t2	Registrert VG3 t2
Østfold	0,084*** (0,0020)	-0,128*** (0,0199)	-0,129*** (0,0253)	-0,030** (0,0108)	-0,026* (0,0113)	-0,081*** (0,0090)	-0,064*** (0,0072)	-0,046*** (0,0129)	-0,048** (0,0152)	-0,036*** (0,0070)	-0,017*** (0,0040)	-0,051*** (0,0048)
Akershus	0,073*** (0,0049)	-0,023 (0,0276)	0,030 (0,0359)	-0,008 (0,0085)	-0,002 (0,0085)	-0,010 (0,0116)	0,001 (0,0118)	-0,023 (0,0164)	-0,025 (0,0169)	-0,014 (0,0076)	-0,007 (0,0079)	-0,027** (0,0089)
Oslo	0,177*** (0,0000)	-0,028*** (0,0000)	0,015*** (0,0000)	-0,004*** (0,0000)	-0,005*** (0,0000)	0,003*** (0,0000)	0,002*** (0,0000)	0,019*** (0,0000)	0,020*** (0,0000)	0,011*** (0,0000)	0,009*** (0,0000)	0,007*** (0,0000)
Hedemark	0,092*** (0,0060)	0,046 (0,0532)	0,046 (0,0723)	-0,003 (0,0163)	-0,012 (0,0162)	0,008 (0,0203)	0,004 (0,0185)	0,019 (0,0191)	0,026 (0,0198)	-0,011 (0,0234)	-0,015 (0,0173)	0,012 (0,0195)
Oppland	0,113*** (0,0069)	0,020 (0,0416)	0,063 (0,0525)	-0,003 (0,0137)	-0,009 (0,0133)	0,005 (0,0225)	0,022 (0,0256)	0,005 (0,0256)	0,007 (0,0245)	-0,021 (0,0210)	-0,015 (0,0201)	-0,032 (0,0226)
Buskerud	0,101*** (0,0028)	0,021 (0,0205)	0,011 (0,0546)	-0,016 (0,0170)	-0,021 (0,0187)	0,002 (0,0136)	0,005 (0,0218)	-0,001 (0,0236)	0,004 (0,0242)	0,033** (0,0116)	0,015 (0,0134)	0,001 (0,0106)
Vestfold	0,102*** (0,0095)	-0,070 (0,0511)	0,034 (0,0589)	-0,006 (0,0127)	-0,014 (0,0132)	0,017 (0,0130)	-0,006 (0,0223)	-0,012 (0,0268)	-0,012 (0,0272)	0,019 (0,0130)	0,003 (0,0091)	0,009 (0,0123)
Telemark	0,085*** (0,0056)	-0,022 (0,0328)	-0,087 (0,0527)	0,008 (0,0118)	0,016 (0,0090)	0,013 (0,0135)	0,009 (0,0211)	-0,060*** (0,0179)	-0,055** (0,0189)	0,018 (0,0115)	0,025*** (0,0063)	-0,005 (0,0199)
Aust Agder	0,119*** (0,0058)	0,031 (0,0384)	0,130** (0,0427)	0,025 (0,0186)	0,039* (0,0177)	-0,029 (0,0164)	0,006 (0,0170)	0,046*** (0,0105)	0,046*** (0,0108)	-0,020 (0,0117)	-0,006 (0,0131)	-0,016 (0,0114)
Vest Agder	0,103*** (0,0054)	-0,033 (0,0394)	-0,133* (0,0608)	-0,012 (0,0061)	-0,009 (0,0090)	-0,023 (0,0128)	-0,040*** (0,0093)	-0,076*** (0,0215)	-0,076** (0,0245)	0,002 (0,0097)	0,002 (0,0104)	-0,015 (0,0139)
Rogaland	0,102*** (0,0060)	-0,032 (0,0406)	-0,005 (0,0570)	-0,023 (0,0160)	-0,192 (0,3991)	-0,102 (0,2850)	0,096 (0,2498)	-0,142 (0,2988)	-0,081 (0,2744)	0,024 (0,0213)	-0,299 (0,1639)	-0,064 (0,3074)
Hordaland	0,084*** (0,0034)	0,007 (0,0282)	-0,106*** (0,0311)	0,009 (0,0078)	0,008 (0,0080)	0,011 (0,0109)	0,005 (0,0084)	0,008 (0,0141)	0,011 (0,0129)	0,013 (0,0079)	0,009 (0,0061)	0,006 (0,0148)
Sogn og Fjordane	0,109*** (0,0076)	-0,034 (0,0370)	0,046 (0,0659)	0,016 (0,0199)	-0,002 (0,0162)	0,008 (0,0209)	-0,017 (0,0149)	0,014 (0,0161)	0,014 (0,0154)	0,014 (0,0150)	-0,008 (0,0124)	0,009 (0,0208)
Møre og Romsdal	0,096*** (0,0061)	-0,059* (0,0241)	-0,064 (0,0389)	-0,012 (0,0105)	-0,015 (0,0078)	0,002 (0,0130)	0,029* (0,0126)	-0,011 (0,0203)	-0,019 (0,0227)	0,017* (0,0083)	0,015* (0,0076)	0,001 (0,0101)
Sør-Trøndelag	0,104*** (0,0041)	-0,026 (0,0313)	0,065 (0,0405)	-0,005 (0,0089)	-0,003 (0,0093)	0,008 (0,0089)	0,042*** (0,0080)	0,017** (0,0060)	0,029*** (0,0053)	0,000 (0,0094)	0,004 (0,0103)	0,003 (0,0082)
Nord-Trøndelag	0,102*** (0,0068)	-0,020 (0,0496)	-0,078 (0,0621)	-0,027 (0,0281)	-0,024 (0,0270)	-0,011 (0,0201)	-0,017 (0,0188)	-0,037 (0,0251)	-0,030 (0,0223)	0,027 (0,0211)	0,028 (0,0225)	0,001 (0,0215)
Nordland	0,090*** (0,0111)	0,045 (0,0289)	0,091 (0,0631)	0,019 (0,0156)	0,019 (0,0190)	0,037 (0,0214)	0,071*** (0,0200)	0,027 (0,0174)	0,027 (0,0171)	0,004 (0,0147)	0,001 (0,0121)	-0,006 (0,0176)
Troms	0,133*** (0,0032)	-0,106** (0,0338)	-0,014 (0,0645)	-0,043** (0,0147)	-0,036* (0,0154)	0,032 (0,0202)	0,029 (0,0297)	0,052** (0,0185)	0,052** (0,0168)	-0,013 (0,0090)	-0,027** (0,0087)	-0,003 (0,0191)
Finmark	0,125*** (0,0231)	-0,003 (0,0460)	0,224* (0,1101)	0,021 (0,0131)	0,017 (0,0446)	-0,030 (0,0589)				-0,046 (0,0343)	0,041 (0,0319)	0,029 (0,0357)
Observasjoner	373 745	373 141	354 814	373 745	315 467	315 467	256 266	256 266	256 266	373 745	315 467	315 467
Snitt utfall	0,0327	4,024	3,471	0,816	0,839	0,724	0,763	0,507	0,496	0,830	0,884	0,742
Kommuner	405	405	405	405	405	405	386	386	386	405	405	405

Merknad: Standardfeil justert for korrelasjoner innen kommuner. Stjerner angir statistisk signifikans: \* p<0,05; \*\* p<0,01; \*\*\* p<0,001.

Tabell A.3. Effekten på elevprestasjoner av intensivopplæring når vi kontrollerer ytterligere for nasjonale prøver 8. trinn og skolefaste effekter, difference-in-differences-estimer

	Ny Giv	Snitt standpkt	Snitt skriftlig eks 10.	Fullført VG1 t1	Fullført VG1 t2	Fullført VG2 t2	Fullført VG2 t3	Fullført VG3 t3	Fullført kompetanse t3	Registrert VG2 t1	Registrert VG2 t2	Registrert VG3 t2
Tiltak .....	0,095*** (0,0027)	-0,010 (0,0089)	0,006 (0,0145)	-0,001 (0,0039)	-0,000 (0,0039)	0,005 (0,0044)	0,011 (0,0065)	-0,005 (0,0083)	-0,002 (0,0082)	0,005 (0,0039)	0,003 (0,0033)	-0,001 (0,0042)
Observasjoner	590 538	586 352	556 685	590 538	539 151	539 151	486 930	486 930	486 930	590 538	539 151	539 151
Snitt utfall .....	0,0162	3,987	3,447	0,799	0,828	0,702	0,750	0,478	0,467	0,794	0,856	0,709
Pre-trend .....		0,952	0,168	0,730	0,738	0,325	0,0482	0,755	0,484	0,977	0,684	0,764
Kommuner .....	389	389	389	389	389	389	389	389	389	389	389	389

Merknad: Standardfeil justert for korrelasjoner innen kommuner. Stjerner angir statistisk signifikans: \* p<0,05; \*\* p<0,01; \*\*\* p<0,001.

Tabell A.4. Effekten på elevprestasjoner av intensivopplæring når vi utelater Oslo og Østfold, difference-in-differences estimer

	Ny Giv	Snitt standpkt	Snitt skriftlig eks 10.	Fullført VG1 t1	Fullført VG1 t2	Fullført VG2 t2	Fullført VG2 t3	Fullført VG3 t3	Fullført kompetanse t3	Registrert VG2 t1	Registrert VG2 t2	Registrert VG3 t2
Tiltak .....	0,098*** (0,0045)	-0,010 (0,0088)	0,015 (0,0120)	-0,002 (0,0032)	-0,003 (0,0039)	0,002 (0,0043)	0,007 (0,0064)	-0,009 (0,0070)	-0,007 (0,0067)	0,005 (0,0032)	-0,000 (0,0031)	-0,006 (0,0041)
Observasjoner	226 928	226 226	216 162	226 928	170 179	170 179	112 656	112 656	112 656	226 928	170 179	170 179
Snitt utfall ...	0,0518	4,063	3,469	0,835	0,853	0,745	0,783	0,518	0,506	0,843	0,890	0,756
Kommuner .	405	405	405	405	405	405	405	405	405	405	405	405

Merknad: Standardfeil justert for korrelasjoner innen kommuner. Stjerner angir statistisk signifikans: \* p<0,05; \*\* p<0,01; \*\*\* p<0,001.

## Vedlegg B: Supplerende resultater fra analyser av forskjeller innen Stavanger-skoler («Regression Discontinuity»-modeller)

### B.1 Tabeller

Tabell B.1.1. Deskriptiv statistikk. Stavanger-skoler med intensivopplæring (avgangselever 2011-2013) og uten (avgangselever 2010), båndbredde 1 karakterpoeng

	(1) Karakter (1,8-2,8), kun Ny GIV- skoler/kull	(2) Karakter (3-4), kun Ny GIV-skoler/kull.	(3) Differanse: Kolonne (1) – (2)	(4) Differanse tilsv. (3) for sammenlignings- skoler	(5) Differanse (DiD): (3) – (4)
Andel med intensivopplæring .....	0,721	0,002	0,719***	0,000	0,719***
Matte 1.termin .....	1,924	2,974	-1,051*** (0,0109)	-1,039*** (0,0000)	-0,011 (0,0184)
Nasjonal prøve 8. trinn	-0,945	-0,121	-0,824*** (0,0379)	-0,873*** (0,0622)	0,049 (0,0742)
Jente .....	0,343	0,418	-0,075** (0,0252)	-0,146*** (0,0422)	0,071 (0,0492)
Mors utdanning .....	3,294	4,331	-1,037*** (0,0879)	-1,055*** (0,1486)	0,017 (0,1743)
Fars utdanning .....	3,441	4,434	-0,993*** (0,0873)	-0,881*** (0,1475)	-0,112 (0,1718)
Innvandrer .....	0,154	0,062	0,092*** (0,0140)	0,138*** (0,0244)	-0,046 (0,0276)
Norskfødt barn av innvandrerforeldre .....	0,116	0,067	0,049*** (0,0137)	0,016 (0,0201)	0,033 (0,0260)
Antall elever .....	481	1 759	2 240	783	3 023

Merknad: I kolonne (1) til (3) ser vi på avgangselever fra Stavanger-skoler fra årene 2011-2013. Kolonne (4) gir en differanse mellom elever med lave karakterer (1,8-2,8) og høyere karakterer (3-4) for avgangselever 2010, dvs. ved skoler som på det aktuelle tidspunktet ikke hadde innført intensivopplæring. Stjerner angir statistisk signifikans: \* p<0,05; \*\* p<0,01; \*\*\* p<0,001.

**Tabell B.1.2. Effektestimater. Effektestimater innen Stavanger-skoler: RD, samt DiRD- og DiD-estimater med 2010-elever som sammenligningsgruppe. Avgangselever Stavanger 2010-2013.**

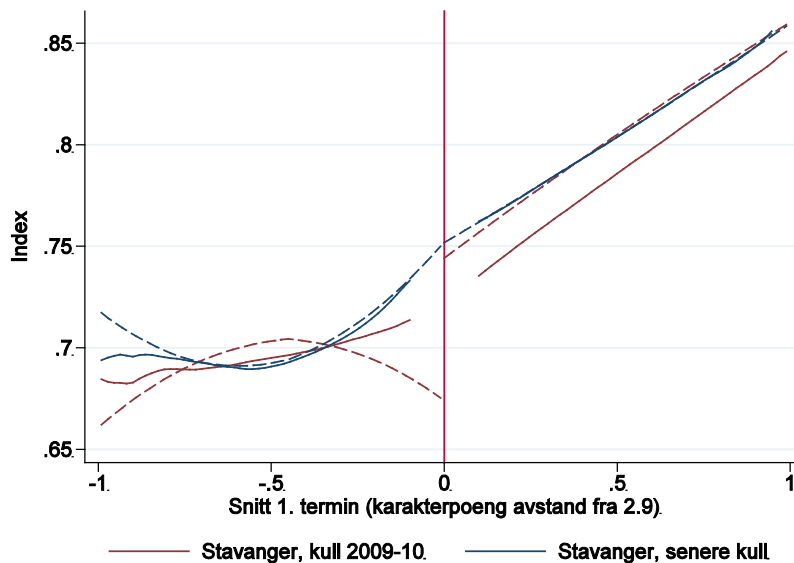
	DiRD	DiD
Ny Giv	0,565*** (0,090)	0,692*** (0,023)
Index	0,081 (0,042)	0,004 (0,011)
Matematikk 1. termin 10. kl.	-0,411 (0,261)	0,005 (0,067)
Norsk hovedmål 1. termin 10. kl.	0,049 (0,250)	-0,050 (0,064)
Snitt standpunkt 10. kl.	0,076 (0,150)	-0,010 (0,039)
Matematikk standpunkt 10. kl.	-0,135 (0,294)	0,179* (0,076)
Norsk hovedmål standpunkt 10. kl.	0,233 (0,253)	-0,004 (0,067)
Snitt skriftlig eksamen 10. kl.	0,346 (0,376)	-0,005 (0,099)
Fullført VG1 t1	0,065 (0,216)	0,062 (0,055)
Fullført VG1 t2	0,072 (0,221)	0,084 (0,056)
Registrert VG1 t2	0,089 (0,216)	-0,025 (0,054)
Registrert VG2 t2	0,047 (0,209)	0,003 (0,052)
Fullført VG2 t2	0,080 (0,232)	-0,032 (0,059)
Fullført VG2 t3	-0,149 (0,267)	0,009 (0,068)
Registrert VG3 t2	-0,103 (0,230)	-0,042 (0,058)
Registrert VG3 t3	-0,375 (0,265)	0,029 (0,066)
Fullført kompetanse t3	-0,132 (0,113)	0,008 (0,034)

Merknad: Standardfeil er robuste for forskjellig varians av restledd (heteroskedastisitet). Stjerner angir statistisk signifikans: \* p<0,05; \*\* p<0,01; \*\*\* p<0,001.



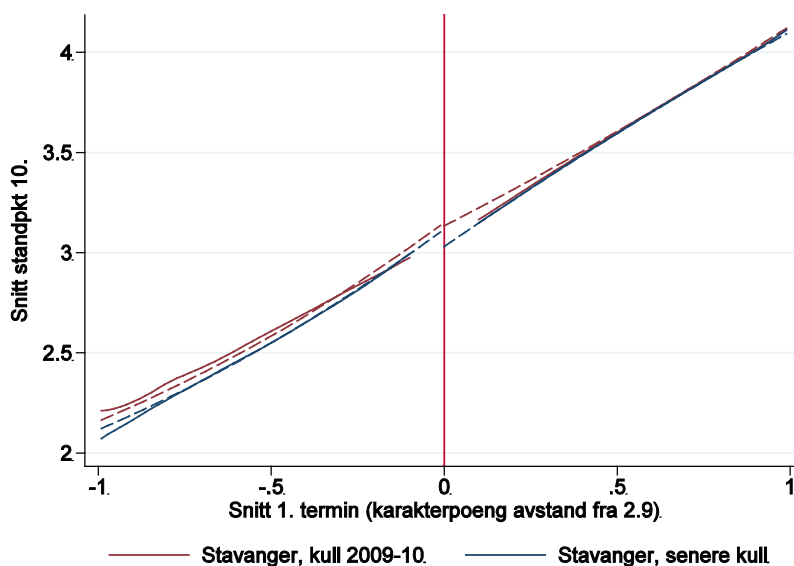
## B.2 Figurer

Figur B.2.1. Sammenheng snitt 1. termin 10. trinn og indeks fra Stavanger 2011-2013 og fra sammenligningsgruppen Stavanger 2009/2010.



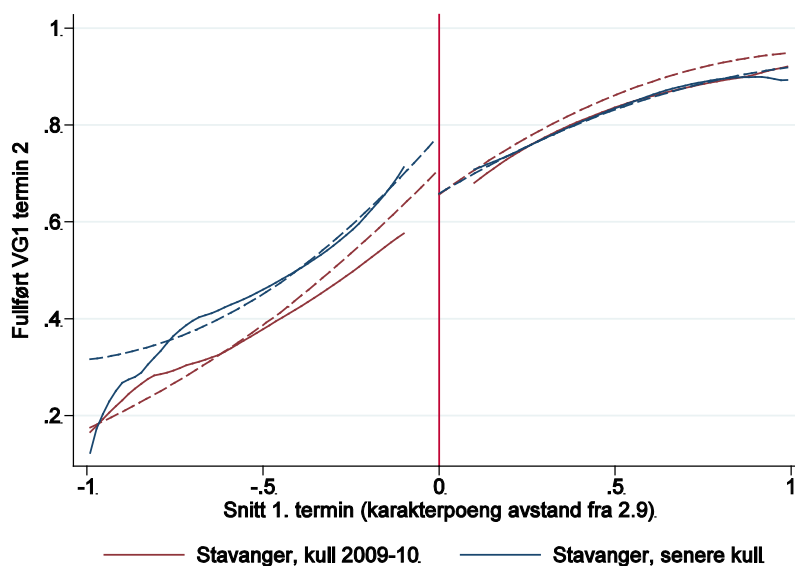
Merknad: Heltrukne linjer viser glidende gjennomsnitt (fra lokal regresjon med båndbredde 0,2). Stiplede linjer viser kvadratiske tilpasninger. Horisontal akse måler snittkarakter som avvik i karakterpoeng fra 2,9.

Figur B.2.2. Sammenheng snitt 1. termin 10. trinn og snitt standpunkt 10. trinn fra Stavanger 2011-2013 og fra sammenligningsgruppen Stavanger 2009/2010.



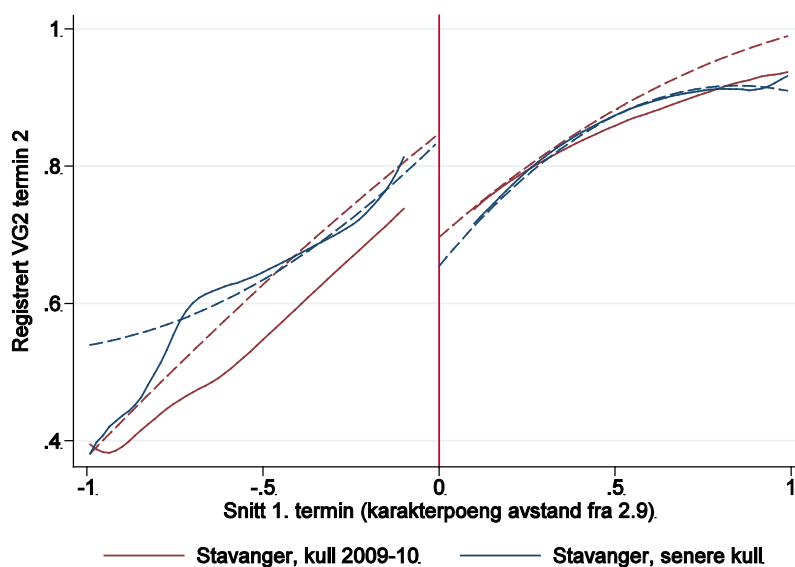
Merknad: Heltrukne linjer viser glidende gjennomsnitt (fra lokal regresjon med båndbredde 0,2). Stiplede linjer viser kvadratiske tilpasninger. Horisontal akse måler snittkarakter som avvik i karakterpoeng fra 2,9.

**Figur B.2.3. Sammenheng snitt 1. termin 10. trinn og fullført VG1 t2 fra Stavanger 2011-2013 og fra sammenligningsgruppen Stavanger 2009/2010.**



Merknad: Heltrukne linjer viser glidende gjennomsnitt (fra lokal regresjon med båndbredde 0,2). Stiplede linjer viser kvadratiske tilpasninger. Horisontal akse måler snittkarakter som avvik i karakterpoeng fra 2,9.

**Figur B.2.4. Sammenheng snitt 1. termin 10. trinn og registrert VG2 t2 fra Stavanger 2011-2013 og fra sammenligningsgruppen Stavanger 2009/2010.**



Merknad: Heltrukne linjer viser glidende gjennomsnitt (fra lokal regresjon med båndbredde 0,2). Stiplede linjer viser kvadratiske tilpasninger. Horisontal akse måler snittkarakter som avvik i karakterpoeng fra 2,9.

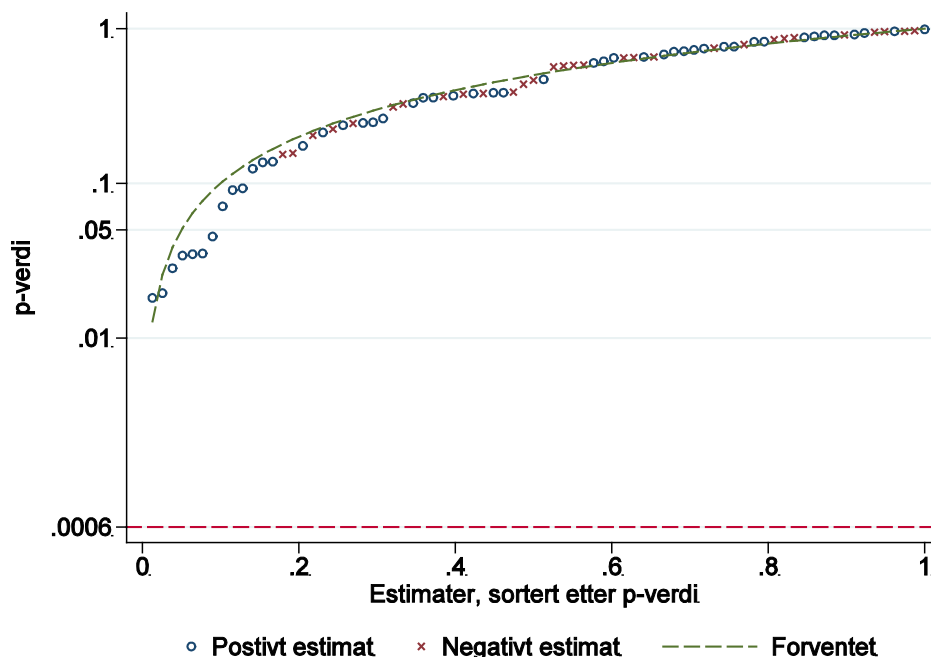
### B.3 Sammenligning med forventet andel signifikante funn

Når vi undersøker mange resultatmål kan vi vente å finne noen signifikante forskjeller, selv om det egentlig ikke skulle være noen effekter. Vi må derfor gjøre en vurdering av om vi i større grad finner signifikante forskjeller enn vi skulle forvente basert på tilfeldigheter.

Slike vurderinger tar gjerne utgangspunkt i p-verdiene til estimatene. P-verdiene oppsummerer sannsynligheten for at vi finner estimater av tilsvarende størrelse, gitt at det ikke er noen effekter. Ved testing av en enkelt hypotese er det vanlig å sammenligne den beregnede p-verdien med en kritisk verdi på 0,05, dvs. at vi bare i ett av 20 tilfeller skal få tilsvarende estimater som resultat av tilfeldighet. Ved testing av flere hypoteser kan denne kritiske verdien reduseres. En justering er såkalt Bonferroni-korreksjon, der kritisk verdi deles på antall hypoteser som testes. Når vi vurderer alle effektestimater i Tabell 6 og B.1.2 gir dette en kritisk verdi på 0,0008. I Figur B.3.1 viser vi p-verdiene fra alle effektestimater (ikke estimert effekt på deltagelse eller balanseringstester), sammenlignet blant annet med denne verdien. Vi ser at ingen av verdiene er så lave, og vi kan dermed ikke konkludere klart med at det er signifikante effekter på noe utfall. (For at DiD-estimatene, som er mest presise, for fullføring av Vg1 skulle vært Bonferroni-signifikante måtte vi hatt effekter på 13-14 prosentpoeng.)

Bonferroni-korreksjon er imidlertid et strengt krav som gir en begrenset styrke, og vi kan lett overse reelle effekter med et såpass strengt krav. Dette gjelder særlig når estimatene er korrelerte, som her. Vi viser derfor også forventet fordeling av p-verdier i Figur B.3.1.<sup>43</sup> For høye p-verdier ser vi at fordelingen av beregnede p-verdier svarer til den forventede. For lavere p-verdier ser det imidlertid ut til at faktiske p-verdier systematisk er noe lavere enn forventet. Figur B.3.1 viser også fortegnet til estimatene, om det er positive eller negative effektestimater. For høye p-verdier er det en blanding, men for lave verdier er det systematisk positive effekter. De signifikante, positive effektene er konsentrert blant tidlige gjennomstrømningsmål, fullføring av Vg1 og overgang til Vg2.

Figur B.3.1. Sammenligning av forventet og faktisk andel signifikante estimater, analyser innen Stavanger



<sup>43</sup> Korrelasjonen mellom estimatene gjør også sammenligningen med forventede p-verdier krevende, de forventede verdiene er beregnet utfra uavhengige estimater.

## Figurregister

Figur 1.	Andel elever på 10. trinn som deltar i intensivopplæring per fase, 2001/02 – 2012/13.....	17
Figur 2.	Snitt standpunktkarakter 10. klasse per fase, 2001/02 – 2012/13 .....	18
Figur 3.	Karakterfordeling og andel elever i intensivopplæring. ....	26
Figur 4.	Andel deltakere i intensivopplæring per karakterpoeng .....	27
Figur 5.	Sammenheng mellom snitt 1. termin 10. trinn og indeks av nasjonale prøver fra 8. trinn og andre elevkjenntegn (forventet fullføring av Vg1 på normert tid) for avgangselever fra Stavanger 2011-2013 og fra sammenligningsgruppen Bergen 2010-2011 .....	34
Figur 6.	Sammenheng snitt 1. termin 10. trinn og standpunkt 10. trinn for avgangselever fra Stavanger 2011-2013 og fra sammenligningsgruppen Bergen 2010-2011.....	35
Figur 7.	Sammenheng snitt 1. termin 10. trinn og fullført VG1 t2 for avgangselever fra Stavanger 2011-2013 og fra sammenligningsgruppen Bergen 2010-2011.....	36
Figur 8.	Sammenheng snitt 1. termin 10. trinn og registrert VG2 t2 for avgangselever fra Stavanger 2011-2013 og fra sammenligningsgruppen Bergen 2010-2011 .....	37
Figur A1.1.	Utvikling i fullført VG1 t1 per fase.....	45
Figur A1.2.	Utvikling i fullført VG1 t2 per fase.....	45
Figur A1.3.	Utvikling i fullført VG1 t3 per fase.....	46
Figur A1.4.	Utvikling i fullført VG2 t2 per fase.....	46
Figur A1.5.	Utvikling i fullført VG2 t3 per fase.....	47
Figur A1.6.	Utvikling i fullført kompetanse t3 per fase .....	47
Figur A1.7.	Utvikling i registrert VG2 t1 per fase .....	48
Figur A1.8.	Utvikling i registrert VG2 t2 per fase .....	48
Figur A1.9.	Utvikling i registrert VG3 t2 per fase .....	49
Figur A1.10.	Utvikling i registrert VG3 t3 per fase .....	49
Figur A1.11.	Utvikling i snitt skriftlig eksamenskarakter.....	50
Figur B.2.1.	Sammenheng snitt 1. termin 10. trinn og indeks fra Stavanger 2011-2013 og fra sammenligningsgruppen Stavanger 2009/2010.....	55
Figur B.2.2.	Sammenheng snitt 1. termin 10. trinn og snitt standpunkt 10. trinn fra Stavanger 2011-2013 og fra sammenligningsgruppen Stavanger 2009/2010.....	55
Figur B.2.3.	Sammenheng snitt 1. termin 10. trinn og fullført VG1 t2 fra Stavanger 2011-2013 og fra sammenligningsgruppen Stavanger 2009/2010.....	56
Figur B.2.4.	Sammenheng snitt 1. termin 10. trinn og registrert VG2 t2 fra Stavanger 2011-2013 og fra sammenligningsgruppen Stavanger 2009/2010. ....	56
Figur B.3.1.	Sammenligning av forventet og faktisk andel signifikante estimater, analyser innen Stavanger .....	57

## Tabellregister

Tabell 1.	Utfallsvariable for kull 2009/2010 – 2012/2013 .....	11
Tabell 2.	Beskrivende statistikk for avgangselever 2006/07-2009/10 .....	13
Tabell 3.	Effekten på elevprestasjoner av intensivopplæring, difference-in-differences estimater .....	22
Tabell 4.	Effekten på elevprestasjoner av intensivopplæring for ulike desiler, difference-in-differences estimater .....	24
Tabell 5.	Beskrivende statistikk. Stavanger-skoler med intensivopplæring og Bergen-skoler uten, båndbredde 1 karakterpoeng .....	31
Tabell 6.	Effektestimater innen Stavanger-skoler. RD-, DiDS og DiRD-estimater med Bergen som sammenligningsgruppe. Avgangselever Stavanger 2011-2013 og Bergen 2010-2011 .....	32
Tabell A.2.1.	Heterogene effekter, kjønn .....	51
Tabell A.2.2.	Heterogene effekter, innvandrere .....	51
Tabell A.2.3.	Heterogene effekter, foreldres utdanningsnivå .....	51
Tabell A.2.4.	Heterogene effekter, fylke .....	52
Tabell A.3.	Effekten på elevprestasjoner av intensivopplæring når vi kontrollerer ytterligere for nasjonale prøver 8. trinn og skolefaste effekter, difference-in-differences-estimater .....	52
Tabell A.4.	Effekten på elevprestasjoner av intensivopplæring når vi utelater Oslo og Østfold, difference-in-differences estimater .....	52
Tabell B.1.1.	Deskriptiv statistikk. Stavanger-skoler med intensivopplæring (avgangselever 2011-2013) og uten (avgangselever 2010), båndbredde 1 karakterpoeng .....	53
Tabell B.1.2.	Effektestimater. Effektestimater innen Stavanger-skoler: RD, samt DiRD- og DiD-estimater med 2010-elever som sammenligningsgruppe. Avgangselever Stavanger 2010-2013 .....	54

## Statistisk sentralbyrå

Postadresse:  
Postboks 8131 Dep  
NO-0033 Oslo

Besøksadresse:  
Akersveien 26, Oslo  
Oterveien 23, Kongsvinger

E-post: [ssb@ssb.no](mailto:ssb@ssb.no)  
Internett: [www.ssb.no](http://www.ssb.no)  
Telefon: 62 88 50 00

ISBN 978-82-537-9308-5 (trykt)  
ISBN 978-82-537-9309-2 (elektronisk)  
ISSN 0806-2056

ISBN 978-82-357-9308-5



9 788235 793089



**Statistisk sentralbyrå**  
Statistics Norway