

*Gaute Eielsen, Lars Johannessen Kirkebøen,
Edwin Leuven, Marte Rønning og Oddbjørn Raaum*

**Effektevaluering av intensivopplæringen i
Overgangsprosjektet, Ny GIV**

Første delrapport

	Standardtegn i tabeller	Symbol
© Statistisk sentralbyrå	Tall kan ikke forekomme	.
Ved bruk av materiale fra denne publikasjonen skal Statistisk sentralbyrå oppgis som kilde.	Oppgave mangler	...
Publisert november 2013	Oppgave mangler foreløpig	...
	Tall kan ikke offentliggjøres	:
	Null	-
ISBN 978-82-537-8805-0 (trykt)	Mindre enn 0,5 av den brukte enheten	0
ISBN 978-82-537-8806-7 (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 effekt-evaluering, der vi studerer hvordan intensivopplæringen i Overgangsprosjektet har påvirket forskjellige forhold knyttet til fullføring av videregående opplæring. Denne rapporten er første delrapport fra denne effektevalueringen. Det er planlagt ytterligere rapporter i 2015 og 2017.

Overgangsprosjektet og effektevalueringen av dette er finansiert av Kunnskapsdepartementet.

Statistisk sentralbyrå, 15. november 2013.

Torbjørn Hægeland

Sammendrag

Gjennom Overgangsprosjektet i Ny GIV har elever med faglige svake prestasjoner midtveis i 10. trinn siden våren 2011 fått tilbud om tilpasset intensivopplæring. Hensikten har vært å øke gjennomstrømning og fullføring av videregående utdanning. I denne rapporten presenterer vi de første resultatene fra den pågående effektevalueringen av intensivopplæringen.

Hensikten med denne effektevaluering er å finne kausale effekter av intensivopplæringen, det vil si i hvilken grad relevante elevresultater er annerledes enn de ville vært dersom intensivopplæringen ikke ble innført. Resultatmålene vi studerer er gjennomstrømning i videregående opplæring (fullføring av Vg1 og overgang til Vg2 på normert tid), skoleprestasjoner i grunn- og videregående skole samt motivasjon. Vi undersøker hvorvidt disse resultatene er endret for elever som selv har deltatt i intensivopplæringen eller som går på skoler der noen elever har gjort det, sammenlignet med hvordan vi kan anta at resultatene ville vært i fravær av intensivopplæringen.

Disse effektene kan ikke måles direkte, ettersom vi ikke kan observere hva som ville skjedd i fravær av intensivopplæringen. Implementeringen av intensivopplæringen har imidlertid to kjennetegn som kan gjøre det mulig å evaluere effekten: 1) En eksplisitt avgrensning av hvilke elever innen hver skole (de 10 prosent svakeste ved første termin) som utgjør målgruppen, og 2) en gradvis innfasing av skoler (totalt tre faser). Den klare avgrensningen innen skolene gjør at vi kan finne nesten like elever som henholdsvis deltar / ikke deltar i intensivopplæringen, og sammenligne senere resultater for disse. Den gradvise innfasingen gjør at vi kan studere hvorvidt resultatutviklingen er annerledes i skoler som deltok i den første fasen, sammenliknet med skoler som deltok i senere faser. I denne omgang studerer vi bare effekter av intensivopplæring i fase én, gjennomført våren 2011.

Vi finner ingen klare tegn til at intensivopplæringen har hatt effekter på resultatmålene vi studerer. Dette gjelder både sammenligninger innen fase én-skoler og mellom skoler som innfører intensivopplæringen på forskjellige tidspunkt. Vi finner verken tegn til direkte effekter på deltakerne, eller til ringvirkninger til andre elever ved fase én-skoler, og vi finner heller ikke tegn til effekter på spesielle grupper av elever eller skoler.

For enkelte resultatmål er det noen forskjeller mellom intensivelever/skoler med intensivopplæring og andre elever/skoler som er ulik null i statistisk forstand. Men ettersom vi studerer mange forskjellige resultatmål og bruker forskjellige analyseteknikker, må vi regne med å finne noen tilfeldige forskjeller, selv om det ikke er noen effekter av intensivopplæringen. Hvilke resultatmål vi finner forskjeller varierer med analyseteknikk, det er ingen systematiske forskjeller. Videre, i flere tilfeller der det er forskjeller i resultater mellom intensivelever og andre elever er det også andre forskjeller enn intensivopplæringen.

Ettersom de første intensivelevne fortsatt er i videregående opplæring er det for tidlig å konkludere om effekter på fullføring. Resultatmålene vi studerer samvarierer sterkt med fullføring av videregående skole. Det er likevel mulig at intensivopplæringen vil få en effekt på andelen som fullfører, til tross for at vi så langt ikke finner noen effekter på andre utfall. Dette først kan vi studere i 2016 og utover, når det har gått fem år siden de første deltakerne gikk ut av grunnskolen.

Det er noe usikkerhet knyttet til hvorvidt våre estimater faktisk fanger opp kausale effekter. Dette skyldes hovedsakelig at utvelgelsen av skoler til fase én og elever til intensivopplæring har skjedd på bakgrunn av uklare og forskjellige kriterier. I tillegg er effektestimaterne basert på utvalg der intensivelevne utgjør en beskjeden andel, noe som også øker (den statistiske) usikkerheten. Vi kan derfor ikke på grunnlag av denne analysen utelukke verken positive eller negative effekter av intensivopplæringen.

Abstract

Through Overgangsprosjektet, which is a part of the larger initiative Ny GIV, students with weak academic performance midway through the 10th grade have since spring 2011 been offered customized intensive training. The aim has been to increase progress through and completion of upper secondary education. In this report we present the first results from the ongoing impact evaluation of intensive training.

The purpose of an impact evaluation is to find causal effects, i.e. the extent to which relevant results are different from what they would have been if the project were not implemented. We study whether progress in upper secondary education, academic performance in lower and upper secondary and motivation has changed for students who have participated in the intensive training or who attend schools where some students have done so, compared with what the outcomes would have been in the absence of the initiative.

These effects can not be measured directly, because we can not observe what would have happened in the absence of the initiative. However, the program has two characteristics that contribute to make it possible to evaluate: An explicit definition of the students in each school who make up the target group, and a gradual phasing in of different schools in three waves. Explicit delimitation in schools means that we can find almost identical students that respectively participate / not participate in intensive training, and compare later outcomes for these. The gradual phasing allows us to compare the evolution of the results in schools of different waves, which introduce intensive training at different times.

We find no clear effects of the intensive training. This applies both for comparisons of students within wave one schools and for comparisons of wave one schools with other schools that adopt intensive training later. We find that for some outcome measures there are differences between students/schools with intensive training and other students / schools that are non-zero in a statistical sense. However, as we study many different outcomes and use different analysis techniques, we must expect to find significant differences as a result of random variation, even if there are no effects. Which measures that differ vary with the technique applied. Moreover, in several cases where there are differences in results between students that participate in intensive training and others there are also evidence of other differences than the intensive training.

Since the first students receiving intensive training still are in secondary education it is too early to conclude whether there are effects on completion. The early outcome measures we study are strongly correlated with eventual completion of upper secondary. It is still possible that intensive training will have an effect on the share completing, despite the fact that we so far have not found any effects on early outcomes. We can not study this until 2016 and onwards, when at least four to five years have passed since the participants left lower secondary education.

We show that the selection of schools and pupils is based on widely varying criteria. This greatly complicates our effort to find effects of the intensive training. A majority of the schools selected students in such a way that it is impossible to find students who are attending / not attending, but otherwise comparable. Schools and local authorities in the different waves are systematically different, and it is difficult to fully take into account these differences.

This means that there will be some uncertainty as to whether our estimates capture differences that are independent of intensive training, and thus fail to identify the effect of the intensive training. Furthermore, our estimates are based on small samples or on groups where students receiving intensive training make up a small share. This reduces the statistical precision of our estimates. We can not on the basis of this analysis rule out either positive or negative effects of intensive training.

Innhold

Forord	3
Sammendrag	4
Abstract	5
1. Innledning	7
2. Om effektevaluering - utfordringer og vår strategi	9
2.1. Generelt om effektevaluering	9
2.2. Hvordan evaluere intensivopplæringen.....	10
3. Utvelgelsen av elever og skoler til intensivopplæring i fase én	18
3.1. Utvelgelse av elever innen fase én-skoler	18
3.2. Utvelgelse av kommuner og skoler til faser	26
4. Evaluering	31
4.1. Sammenheng mellom fullføring av videregående opplæring, tidligere resultater og deltakelse i intensivopplæringen	31
4.2. Effekter på elever innen skoler	33
4.3. Effekter på skolenivå.....	39
4.4. Effekter på intensivelever og ringvirkninger, og andre former for heterogene effekter... ..	47
5. Oppsummering	52
Referanser	58
Vedlegg A: Supplerende figurer og tabeller	59
Figurregister	65
Tabellregister	66

1. Innledning

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 inntil 7,5 klokke timer per uke, og intensivopplæringen skal starte tidlig i andre termin og vare ut skoleåret, ifølge Kunnskapsdepartementets invitasjonsbrev til kommunene. 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. Ansvaret for å implementere ble gitt til fylkeskommunene. I invitasjonsbrevet fra KD står det ikke eksplisitt hvordan utvelgelsen av de 10 prosent svakeste elevene skal foregå. Med andre ord, her var det stor lokal frihet, noe som førte til stor variasjon på tvers av fylkene når det gjelder fordeling av kommuner og skoler på ulike faser, og hvordan elevene som fikk tilbud om intensivopplæring faktisk ble plukket ut.

Hovedformålet med intensivopplæringen og de øvrige tiltakene i Ny GIV er å øke andelen i målgruppen som gjennomfører videregående. Denne rapporten er første delrapport fra effektevalueringen av intensivopplæringen, og studerer hvorvidt intensivopplæringen har virket etter hensikten. Foreløpig har vi ikke data for fullføring av videregående opplæring, ettersom de første elevene fikk intensivopplæring gjennom Ny GIV våren 2011. Faktisk gjennomføring av videregående opplæring kan vi dermed ikke studere før etter 2016, gitt at vi ønsker å studere andelen som har fullført innen fem år. I stedet benytter vi data på skoleprestasjoner i 10. klasse, samt skoleprestasjoner og gjennomstrømning tidlig i videregående skole. I kapittel 4 viser vi, med data for tidligere årskull, at disse variablene er høyt korrelert med fullføring av videregående.

Det har lenge vært en tradisjon i Norge, og også i andre land, å gi ekstra eller spesielt tilpasset opplæring til elever med særskilte behov. Vi vet imidlertid lite om i hvilken grad slike tiltak har en kausal effekt på elevers læringsutbytte. Med kausal menes den faktiske effekten av tiltaket, sammenlignet med hva som ville skjedd dersom tiltaket ikke ble innført. Dette er for eksempel i motsetning til forskjeller i framgang mellom elever som skyldes systematisk utvelgelse av elever til tilpasset opplæring. For eksempel, anta at en skole bestemmer at elever som har karakteren tre eller mindre i snitt tilbud om tilpasset opplæring. En enkel sammenlikning av elevene som får / ikke får tilpasset opplæring vil da undervurdere effekten av tiltaket, fordi elevene som får tilbud om tilpasset opplæring har et dårligere utgangspunkt enn de som ikke får slikt tilbud. Resultatet vi sitter igjen med etter en slik sammenlikning er derfor ikke den rene effekten av tiltaket, men en forskjell som også er påvirket av at elevene i tiltaksgruppen i utgangspunktet systematisk presterer svakere enn elevene de sammenliknes med. Selv om tiltaket skulle ha en positiv effekt, vil evalueringen likevel kunne vise ingen effekt, fordi vi ikke på en god nok måte greier å ta hensyn til at forskjellige elever er fordelt til de to gruppene.

De få studiene som studerer slike tiltak på en troverdig måte er fra andre land enn Norge. 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.¹ Cortes mfl. (2013) studerer effekten av ekstra matteundervisning på senere matteprestasjoner og på fullføring av videregående (high school). De finner en beskjeden effekt på mattekarakterer, og ingen klar effekt for det første året, men en

¹ Et utvalg av skoler med svake resultater fikk tilby 100 elever (ved hver skole) ekstra undervisning på kveldstid, i grupper med inntil fem elever og med lærere som underviste.

effekt på 6-10 prosentpoeng på fullføring.² 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.³ 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. 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 kapittel 2 diskuterer vi kort motivasjonen for å gjennomføre effektevalueringer og generelle utfordringer knyttet til dette. Deretter gir vi en noe mer detaljert beskrivelse av hvordan vi planlegger å gjennomføre effektevalueringen av intensivopplæringen i Overgangsprosjektet, og hvilke typer effekter vi kan identifisere. Kort forklart kommer vi til å benytte to analysestrategier. Den første utnytter at målgruppen består av en definert gruppe av elever (i invitasjonsbrevet fra KD står det at målgruppen skal utgjøre de 10 prosent svakeste elevene til første termin). Den andre baserer seg på den gradvise innfasingen av skoler. Vi presenterer også datagrunnlaget for evalueringen. I denne rapporten begrenser vi oss til å studere elevene som gikk ut av grunnskolen våren 2011, fordi dette er det seneste elevkullet som vi kan følge gjennom Vg1.

I kapittel 3 studerer vi hvordan elever er valgt ut til deltakelse i intensivopplæringen innen fase én-skoler, og hvordan skoler og kommuner er fordelt til faser. Vi finner at det er klare systematiske forskjeller, både mellom elever som deltar/ ikke deltar, og mellom skoler og kommuner som innfører intensivopplæringen i forskjellige faser. Dette er en utfordring for effektevalueringen, som diskutert i kapittel 2. For å håndtere dette finner vi egnede analyseutvalg, henholdsvis skoler og kommuner der utvelgelsen av elever ser ut til å basere seg forholdsvis strengt på karakterer til første termin og skoler i fase to og tre som ligner skolene i fase én.

I kapittel 4 presenterer vi resultatene fra effektevalueringen. Ettersom vi foreløpig ikke kan vite i hvilken grad deltakerne i intensivopplæringen faktisk fullfører videregående opplæring diskuterer vi først sammenhengen mellom fullføring og de resultatmålene vi har så langt. Resultatmålene vi studerer på individnivå er karakterer ved fullført grunnskole, samt forskjellige mål på gjennomstrømning og prestasjoner tidlig i videregående skole. I tillegg har vi data på skolenivå fra Elevundersøkelsen. Vi presenterer resultater for hver av de to analysestrategiene, sammenligninger av intensivelever med andre elever i fase én-skoler og sammenligninger av fase én-skoler med skoler i fase to og tre. Kapittel 4 viser en rekke enkeltresultater, og knytter disse til diskusjonene av analysestrategi i kapittel 2 og av utvelgelse av elever og skoler i kapittel 3. Oppsummering av resultatene utsettes i stor grad til kapittel 5.

I kapittel 5 oppsummerer vi helheten i resultatene fra effektevalueringen så langt. Vi finner ingen klare resultater av intensivopplæringen på noen av resultatmålene vi studerer. Vi diskuterer kort noen mulige forklaringer på at vi ikke finner noen resultater, og peker ut noen videre analyser vi vil komme tilbake til i senere rapporter. Kapittel 5 er ment å kunne leses uavhengig av de mer tekniske kapitlene 2-4, men referer til hvor i kapittel 3 og 4 forskjellige resultater presenteres i mer detalj.

² Tiltaket innebærer omtrent en dobling av matteundervisningen, på bekostning av andre fag. Målgruppen for tiltaket er den svakest presterende halvparten av elevene, og effekten er sterkest er konsentrert blant elever som er dårlige til å lese.

³ 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.

2. Om effektevaluering - utfordringer og vår strategi

2.1. Generelt om effektevaluering

Utdanningsmyndighetene har et åpenbart ønske og en klar ambisjon om å oppnå gode resultater i skolen, for eksempel via bedre ferdigheter, bedre prestasjoner på internasjonale tester og økt gjennomføring av videregående utdanning. Derfor vil de ønske å gjennomføre forskjellige tiltak som kan bedre resultatene.

Ressurser er imidlertid knappe. Innen utdanningssektoren gjelder dette budsjetter til lønn, utstyr med mer, men også andre forhold. Selv med kraftig økte budsjetter vil det være en grense for hvor mange kvalifiserte lærere en evner å rekruttere. Videre, lærere pålegges en rekke oppgaver. Dersom de får flere oppgaver eller klarere føringer for hvordan noen oppgaver skal utføres vil dette kunne føre til at andre oppgaver prioriteres ned. Til slutt, mange opplever ”reformtrettehet” i utdanningssektoren. Nye reformer, tiltak eller initiativer kan potensielt undergrave motivasjon hos lærere og tid til å forberede god undervisning. Derfor er det viktig å vite hvilke tiltak som gir en betydelig effekt, slik at innsatsen kan konsentreres om disse.

Dette er formålet med en effektevaluering, å finne den kausale effekten av et tiltak, slik at det blir mulig å sammenligne tiltak og prioritere de mest effektive. En kausal effekt er hvordan tiltaket har endret resultatene, forskjellen mellom et resultat vi faktisk kan observere, og hva som ville vært det tilsvarende resultatet i fravær av tiltaket. Det er verdt å merke hvordan vi definerer effekt. Effekten, for eksempel av intensivopplæringen på karakterer ved avsluttet grunnskole, er noe annet enn utviklingen i karakterer mens tiltaket varer, som vi forholdsvis lett kan observere. I stedet er effekten abstrakt, definert som forskjellen mellom utviklingen som faktisk skjer, og som vi lett kan observere, og den kontrafaktiske utviklingen i fravær av tiltaket, som vi aldri kan observere.

Denne definisjonen av effekt henger naturlig sammen med motivasjonen for effektevaluering. Når vi skal velge hvorvidt vi innfører et tiltak, eller det mest effektive fra en mengde mulige tiltak, ønsker vi nettopp å vite om det gir en gevinst (eller størst mulig gevinst) sammenlignet med å ikke gjennomføre det.

Enhver metode for å finne effekter innen et tankeskjema med kontrafaktiske utfall bygger på antakelser. Ettersom effektene er noe abstrakt, og relativt til noe vi ikke kan observere, er de utfordrende å finne. I praksis finner vi effekten av et tiltak ved å sammenligne resultatene til de som deltar med en sammenligningsgruppe som ikke deltar. For at det skal være mulig å finne faktiske effekter må deltakerne og sammenligningsgruppen være like på alle andre (relevante) måter enn deltakelse i tiltaket, eller vi må være i stand til å ta hensyn til forskjellene.

Gullstandarden for effektevaluering er et eksperimentdesign. I et eksperiment har vi kontroll på gruppene som deltar eller ikke deltar, for eksempel gjennom lodd-trekning. Dette sikrer oss at det ikke vil være noen forskjeller mellom deltakerne og sammenligningsgruppen i utgangspunktet, og enhver forskjell etter tiltaket vil være en effekt av dette. Vi kunne for eksempel fra gruppen bestående av svakt-presterende elever trekke et tilfeldig utvalg som blir tilbudt intensivopplæring (tiltaksgruppen). Gruppen som ikke blir tilbudt intensivopplæring fungerer da som en gyldig sammenligningsgruppe til tiltaksgruppen nettopp fordi utvelgelsen av elever til de to gruppene er tilfeldig, og dermed uavhengig av andre kjennetegn ved eleven eller dens familie. Resultatet i sammenligningsgruppen ville da gi et godt anslag på det kontrafaktiske resultatet for tiltaksgruppen, og differansen i skoleprestasjonene i de to gruppene ville svare til den forventede effekten av å delta for en tilfeldig elev i målgruppen.

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.

2.2. Hvordan evaluere intensivopplæringen

Gjennom Overgangsprosjektet i Ny GIV får enkeltelever intensivopplæring i grunnleggende lese-, skrive- og regneferdigheter. Målgruppen er de 10 prosent elevene i hver kommune med svakest resultater ved første termin i 10. klasse. Intensivopplæringen kan være på åtte til ti timer per uke. I forbindelse med intensivopplæringen er det også noe kursing av lærere, men utover denne kursing får skolene ikke ytterligere ressurser. Hovedformålet med Ny GIV består i å øke andelen av denne gruppen elever som gjennomfører videregående opplæring med studiekompetanse eller fagbrev innen fem år.

Intensivopplæringen kan ha en effekt på elevene som får intensivopplæring, i det følgende vil vi omtale dette som den direkte effekten eller deltakereffekten av intensivopplæringen. Intensivopplæringen er imidlertid også et tiltak på skolenivå. Det er mulig at også andre elever ved fase én-skolene påvirkes av at intensivopplæringen innføres. Dette kan for eksempel være tilfellet dersom innføringen av intensivopplæringen har betydning for hvilke lærere som underviser hvilke elever, dersom intensivopplæringen fører til annen organisering av (eller i det minste en annen sammensetning av elevmassen i) undervisning for de øvrige elevene eller dersom intensivelevne som et resultat av bedre ferdigheter eller motivasjon bidrar på en annen måte i annen undervisning. I det følgende vil vi omtale disse effektene på andre elever som ringvirkninger.

Målet med vår effektevaluering er å fastslå om intensivopplæring til de svakest presterende elevene har en kausal effekt på gjennomstrømming i videregående og på læringsutbytte. Vi vil studere dette ved å sammenligne resultatene til forskjellige grupper elever og skoler.

En enkel sammenlikning av elevprestasjonene i gruppen som får intensivopplæring (tiltaksgruppen) og gruppen som ikke får det (sammenligningsgruppen) vil gi et galt bilde på hva som er effekten av intensivopplæring. Grunnen til det er at det er de svakest presterende elevene som mottar intensivopplæring. Disse vil ved avsluttet grunnskole sannsynligvis fremdeles prestere under snittet av øvrige elever. Dette sier imidlertid lite om effekten av intensivopplæringen, bare at det ikke var tilstrekkelig til å utjevne forskjellen som var der i utgangspunktet.

En mye brukt tilnærming i forskjellige evalueringer er å forsøke å ta hensyn til forskjeller i som er der i utgangspunktet ved å kontrollere for forskjellige elevkjennetegn i regresjonsanalyser. Et naturlig utgangspunkt er karakterer til første termin, som både er et sentralt kriterium for utvelgelsen av elevene, og som sannsynligvis fanger mye av forskjellene i utgangspunkt. Med detaljerte norske registerdata er det også mulig å ta hensyn til flere kjennetegn, slik som resultater ved nasjonale prøver i 8. trinn, foreldres utdanning med mer. Vi kan da, under noen forutsetninger, sammenligne resultatene til elever som er like hva gjelder disse kjennetegnene. Dette gir likevel ikke noen troverdig effektevaluering. Selv om vi kan sammenligne elever som er like langs mange dimensjoner vil det fremdeles kunne være vesentlige forskjeller. Detaljene i hvordan elevene er valgt ut vil da være avgjørende for om vi overvurderer eller undervurderer effekten av intensivopplæringen når vi sammenligner på denne måten. Dersom elevene som deltar er valgt fordi de er motivert for å gjøre en innsats og forbedre seg vil vi sannsynligvis, når vi sammenligner med andre elever med tilsvarende første termin-karakterer, finne en forskjell som er større enn effekten av intensivopplæringen. Dersom

intensivelevne er valgt fordi de i utgangspunktet har en negativ utvikling og trenger et tiltak vil vi kunne finne en forskjell som er mindre enn effekten.

I stedet vil vi basere oss på to kjennetegn ved intensivopplæringen, som bidrar til å gjøre tiltaket evaluerbart: En eksplisitt avgrensning av hvilke elever innen hver skole som utgjør målgruppe, og en gradvis innfasing i forskjellige skoler. Avgrensningen innen skolene gjør at vi kan finne nesten like elever som henholdsvis deltar / ikke deltar i intensivopplæringen, og sammenligne disse for å finne effekter. Den gradvise innfasingen gjør at vi kan studere eventuelle forskjeller i utvikling mellom skoler i forskjellige faser. I det følgende vil vi diskutere disse tilnærmingene i mer detalj.

2.2.1. Forskjeller mellom intensivelever og andre elever ved fase én-skoler

I invitasjonsbrevene fra KD, som ble sendt til alle fylkeskommuner og kommuner i oktober 2010, står det at ”målgruppen for prosjektet er de 10 prosent svakest presterende elevene etter første termin i tiende trinn i hver av de kommunene som fylkeskommunen inviterer til å delta i prosjektet”.

I den grad det finnes en slik klar grense for deltakelse i intensivopplæringen ved den 10. persentilen⁴ kan vi sammenligne de elevene som befinner seg *like under* den 10. persentilen med de elevene som befinner seg *like over* den 10. persentilen. Elevene like under tilbys intensivopplæring, mens elevene over tilbys ikke. Bortsett fra deltakelse i intensivopplæringen vil vi forvente at disse elevene er like. Elevene som akkurat ikke får tilbud om intensivopplæring fungerer da som en gyldig sammenlikningsgruppe for elevene som akkurat får tilbud om intensivopplæring.

Innenfor denne gruppen kan vi tenke på det som tilfeldig om en elev kommer så vidt over eller under grensen for deltakelse. 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.

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-tilnærmingen er at den ligner et randomisert eksperiment (her lokalt rundt den 10. persentilen) som beskrevet i avsnitt 2.1. En annen fordel med RD-tilnærmingen er at den ikke er avhengig av en sammenlikningsgruppe utenfor skolen eller årskullet, dette gjør den egnet til å studere utviklingen i effekten av intensivopplæringen. 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.

Denne tilnærmingen baserer seg på en sammenligning av intensivelever med andre elever ved fase én-skoler. Den estimerte effekten er forskjellen mellom den direkte effekten av intensivopplæringen på elevene som så vidt deltar (deltakelseeffekten) og potensielle ringvirkninger på elevene som så vidt ikke deltar:

$$RD\text{-effekt} = \text{deltakelseeffekt} - \text{ringvirkninger}$$

Bare hvis det ikke er ringvirkninger vil vi slik få et gyldig estimat av deltakelseeffekten. Denne lokale forskjellen rundt grensen for deltakelse, som vi finner med RD, er den effekten vi venter dersom målgruppen for

⁴ En persentil er et mål på verdien som skiller en viss andel av fordelingen fra den øvrige. Eksempelvis er den tiende persentilen det karaktersnittet som 10 prosent av elevene i en kommune akkurat har eller ligger under, og nødvendigvis 90 prosent av elevene ligger over. Med andre ord så har den beste eleven i den tiende persentilen dette karaktersnittet.

intensivopplæringen skulle utvides litt, for eksempel fra 10 til 11 prosent av elevene.

I praksis er vi ikke i stand til å finne presise effektestimater ved å sammenligne noen få elever rundt grensen for deltakelse i intensivopplæringen. I stedet beregner vi en form for gjennomsnitt av resultatene på hver side av den 10. persentilen. Såfremt vi er i stand til å skille elever med nesten tilsvarende resultater, for eksempel ved å se på snittet av karakterene til første termin,⁵ kan vi da se hvordan resultatene endrer seg med små forskjeller i første termin-karakterer. Dersom intensivopplæringen ikke har noen effekt vil vi vente å se en forskjell i resultater rundt grensen for deltakelse som er omtrent lik resultatforskjeller ved like stor forskjell i snitt første termin, men for andre verdier av snitt første termin. Derimot, dersom vi ser en påfallende forskjell i resultat for en liten forskjell første termin akkurat rundt grensen for deltakelse tyder dette på at vi ser en effekt av intensivopplæringen.

Forutsetningen for at RD-estimering skal gi gyldige effektestimater er at sannsynligheten for å delta i intensivopplæringen faller brått rundt tilordningsgrensen, men andre forhold varierer lite (lokalt rundt tilordningsgrensen). Gitt at elever er tilbudt deltakelse på bakgrunn av første termin-karakterer vil vi kunne se at sannsynligheten for deltakelse i intensivopplæringen endrer seg brått.

Vi er ikke avhengige av at alle elever med lavere karakter deltar i intensivopplæringen, eller av at ingen meld lavere karakterer gjør det. Andelen som deltar trenger ikke falle fra 100 til 0 prosent, men i den grad det ikke er et klart skille påvirker dette tolkningen av estimatene våre. Hvis vi forestiller oss et brått fall fra 70 prosent som deltar under grensen til 20 prosent over vil resultatene på begge sider av grensen være påvirket både av direkte effekter og ringvirkninger. Imidlertid, siden det er stor forskjell i andelen som deltar rundt grensen vil vi kunne knytte eventuelle forskjeller i resultater til intensivopplæringen. Vi vil da se på forskjellen både for deltakelse i intensivopplæringen og resultater mellom elevene over og under grensen for deltakelse.⁶ Siden forskjellen i deltakelse er 50 prosentpoeng vil vi konkludere med at forskjellen mellom direkte effekt og ringvirkninger er dobbelt så stor som den resultatforskjellen vi observerer rundt grensen for deltakelse.

At ingen andre relevante forhold endrer seg brått er en grunnleggende antagelse, som vi aldri fullt ut kan bekrefte eller avkrefte. Men vi kan vurdere gyldigheten av en RD-tilnærming ved å studere hvorvidt andre elevkjenninger (for eksempel resultater fra nasjonale prøver eller familiebakgrunn) er like for elever så vidt tilbudt intensivopplæring og de så vidt ikke tilbudt. Dersom slike forhold ikke endrer seg rundt tilordningsgrensen tyder det på at elevene på hver side er like, og at RD-estimering gir gyldige effektestimater.

2.2.2. Sammenlikninger mellom skoler og årskull

I tillegg til å sammenligne elever innen skoler kan vi også se på snittresultatene ved skoler som har innført intensivopplæringen. Fordelen med denne tilnærmingen er at vi unngår problemet med hvorfor den enkelte elev er valgt ut til deltakelse, og hvordan disse elevene skiller seg fra øvrige elever. Vi er også her avhengige av en sammenligningsgruppe, som vil være samme skole på et tidligere tidspunkt eller andre skoler på samme tidspunkt. Valg av sammenligningsgruppe bygger på

⁵ Når vi ser på snittkarakter vil det være små forskjeller på verdiene vi kan se, 1/12 karakterpoeng dersom elevene har karakterer i 12 fag. Dersom vi ser på karakter i et enkeltfag er det derimot vanskeligere å hevde at elevene med to er nesten like elevene med tre i dette faget.

⁶ Merk at vi vil sammenligne samtlige elever like over og under denne grensen. Dersom vi for eksempel begrenset oss til å sammenligne deltagere under grensen med ikke-deltagere over ville vi fått det samme problemet med hvorfor noen elever er valgt ut til deltagelse og hvordan disse skiller seg fra de som ikke deltar som når vi forsøker å betinge på observerbare kjennetegn.

antakelser om hva som skiller skoler, hvilket er nødvendig for å kunne identifisere forskjeller som er uttrykk for effekter av intensivopplæringen.

Den første og mest intuitive tilnærmingen er å se på endringene i resultatene til en skole ved innføringen av intensivopplæringen. Det er rimelig å tenke seg at mange forhold ved en skole, for eksempel gjennomsnittlige elevkjenntegn, lærernes kompetanse og undervisningsmetoder, skoleleders og skoleeiers fokus og prioriteringer, er omtrent stabile over korte tidsrom. Dersom vi ser betydelige endringer i skolens resultater i forbindelse med innføringen av intensivopplæringen vil disse sannsynligvis gjenspeile effekter av intensivopplæringen. Det vil imidlertid ganske sikkert også være andre endringer over tid. Dette kan for eksempel skyldes lengre trender som utskifting av lærere, eller at de som jobber ved skolen får mer erfaring, endringer i boligmonstre og dermed elevsammensetning eller endringer i skoleleder eller skoleeiers vektlegging. Vi kan også tenke oss mer kortsiktige variasjoner, flere spesielt gode elever i et årskull, spesielt god kjemi mellom en lærer og en klasse eller forstyrrende byggearbeider ett enkelt år. Vi er derfor avhengige av å kunne vurdere hva som utgjør en "betydelig" endring. Dette kan vi enkelt gjøre ved å sammenligne med hvor mye resultatene varierer før innføringen av intensivopplæringen.

En annen tilnærming er å sammenligne skoler som på et gitt tidspunkt deltar i intensivopplæringen med skoler som ikke deltar. Dersom skolene var tilfeldig fordelt til deltakelse/ikke ville dette vært et eksperiment, som diskutert tidligere. Skolene ville i utgangspunktet vært like, og vi kunne funnet effekten av intensivopplæringen ved å sammenligne resultatene etter at den ene gruppen hadde innført intensivopplæringen og den andre ikke.

Slik Overgangsprosjektet er utformet har vi ingen grunn til å tro at skolene i fase én og senere faser i utgangspunktet er like. Vi er dermed avhengige av å ta hensyn til forskjellene som er uavhengige av intensivopplæringen. Vår første tilnærming til dette er å kombinere endringer over tid med forskjeller mellom skoler og se på forskjeller i utvikling mellom skoler. I den internasjonale litteraturen omtales dette som en "difference-in-differences"-metode (Angrist og Pischke, 2009).

Under bestemte forutsetninger egner denne metoden seg godt til å estimere den kausale effekten dersom vi har data på skolenivå både før og etter innføringen av intensivopplæringen, og ikke alle skolene innførte intensivopplæringen samtidig. Ideen her er at skolene som ikke innførte intensivopplæringen fungerer som en gyldig sammenlikningsgruppe for skolene som innførte intensivopplæringen.

Kort forklart kan metoden skisseres slik. Dersom elevprestasjonene på en fase én-skole går opp det året intensivopplæringen implementeres sammenliknet med året før kan det tyde på at intensivopplæringen har hatt en effekt på skolerresultatene. Men, som diskutert over kan en slik økning også skyldes en rekke andre forhold, som for eksempel at eksamen i skoleåret 2009/2010 var vanskeligere enn i skoleåret 2010/2011. Slike andre forhold vil imidlertid også påvirke skoler hvor intensivopplæringen ikke ble implementert skoleåret 2010/2011. Vi tar derfor hensyn til slike forhold som er felles for alle skoler ved å sammenlikne utviklingen på skolene som innførte intensivopplæringen med utviklingen på skolene som ikke gjorde det (for samme årskull). Med andre ord, denne strategien baserer seg på sammenligninger endringene mellom forskjellige årskull på tvers av skoler.

Hovedforutsetningen for at dette skal gi et gyldig estimat på effekten av intensivopplæringen er at det ikke skjer andre endringer på samme tidspunkt som ikke henger sammen med intensivopplæringen, men som påvirker fase én-skolene og de øvrige skolene forskjellig. Vi er imidlertid ikke avhengig av at skolene er tilfeldig fordelt til deltakelse, bare at forskjellen som ville ha vært mellom dem i

fravær av intensivopplæringen svarer til historiske forskjeller.⁷ I så fall kan vi ta hensyn til denne, og finne effekten av intensivopplæringen som forskjellen i utvikling.

At det ikke ville vært slike forskjeller er en antagelse om den kontrafaktiske utviklingen, som vi aldri kan teste. Dersom vi har historiske resultater kan vi imidlertid se på utviklingen i disse, og se om det har vært tilsvarende utvikling i skoler i forskjellige faser tidligere, før intensivopplæringen var innført i noen av dem. I så fall er det rimelig at det fortsatt ville ha vært en tilsvarende utvikling i fravær av intensivopplæringen, og at eventuelle avvik fra en slik tilsvarende utvikling er en effekt av intensivopplæringen.

En annen tilnærming til det at skolene i de forskjellige fasene forskjellige er å forsøke å finne noen skoler som er likere og sammenlignbare. Dette kan gjøres ved matching-metoder. For at denne tilnærmingen skal gi gyldige estimater må vi være i stand til å finne skoler i fase to og tre (og som dermed ikke har innført intensivopplæringen) som ligner på fase én-skolene (som har innført intensivopplæringen). Vi er ikke avhengige av at alle skolene, eller en gang av at snittet eller flertallet av skolene i fase to og tre ligner på fase én-skolene. Men dersom ingen skoler i fase to tre ligner fase én-skolene vil vi ikke finne troverdige effektestimater.

Det er flere mulige matching metoder. Vi vil se på en rekke kjennetegn ved fase én-skolene, og så for hver fase én-skole finne den skolen som i størst mulig grad har tilsvarende kjennetegn.⁸ Dette betyr at vi sammenligner med snittresultatet fra fase én-skolene med et vektet snittresultat fra øvrige skoler. Dette vektete snittet vil imidlertid bare basere seg på de skolene i fase to og tre som ligner på en fase én-skole. Til gjengjeld vil fase to og tre-skoler som ligner flere fase én-skoler få ekstra stor vekt.

En utfordring med matching-metoder er å vurdere om skolene vi sammenligner er tilstrekkelig like. For å vurdere dette kan vi kombinere matching med en difference-in-differences-tilnærming. Vi sammenligner da utviklingen ved innføringen av intensivopplæringen i fase én-skolene med utviklingen i samme periode for de matchede skolene. Slik kan vi sammenligne skoler som er like hva gjelder observerbare kjennetegn, mens vi tar hensyn til forskjeller i tidligere resultater. Vi kan også se på forskjeller i utvikling før innføringen av intensivopplæringen, for å se om de har utviklet seg forskjellig. Dersom vi ser en forskjell i utviklingen ved innføring av intensivopplæringen, men ikke tidligere, tyder dette på at intensivopplæringen har hatt en effekt.

En begrensning med disse tilnærmingene er at de bare kan brukes på de første årskullene, dvs. mens det ennå finnes skoler som ikke har introdusert intensivopplæringen som vi kan bruke som sammenligningsgruppe.

Disse tilnærmingene vil finne en annen effekt enn RD-tilnærmingen diskutert i avsnitt 2.2.1. Når vi ser på forskjellene mellom skoler estimerer vi en total gjennomsnittseffekt:

$$DiD\text{-effekt} = \text{deltakelseeffekt} + \text{ringvirkninger}$$

Siden det er både elever som deltar i intensivopplæringen og andre elever på fase én-skolene vil vi estimere en samlet effekt på alle disse. Dette kan være nyttig fordi

⁷ Dette vil være tilfellet dersom tidspunktet Ny GIV innføres i en skole er tilfeldig. Med tilfeldig menes at det ikke skal være en bakenforliggende årsak til at Ny GIV innføres i en bestemt skole på et bestemt tidspunkt. En slik bakenforliggende årsak kan for eksempel være at en skole har mange svake elever i utgangspunktet og ser på Ny GIV som en mulighet til å gi de et tilbud.

⁸ Andre matching-metoder estimerer en sammenheng mellom kjennetegn og deltagelse, og sammenligner skoler som har omtrent tilsvarende sannsynlighet for deltagelse, eller sammenligner hver deltakende skole med et vektet gjennomsnitt av flere andre skoler.

intensivopplæringen er et tiltak på skolenivå, og det potensielt kan være viktige ringvirkninger. Ettersom intensivelevne vil utgjøre omtrent 10 prosent av elevene ved fase én-skolene vil imidlertid disse elevene ha beskjeden vekt i estimeringen av effekter. Dersom det ikke er noen ringvirkninger vil vi dermed vente å finne en *DiD-effekt* som er bare omtrent en tiendedel av *RD-effekten*. Ettersom vi kan observere andelen som deltar i intensivopplæringen vil vi imidlertid, under en antagelse om fravær av ringvirkninger, kunne beregne hvilken effekt på intensivelevne en gitt samlet effekt svarer til.

Dette kan vi bruke til å skille deltakelseeffekt og ringvirkninger. I den grad intensivelevne er valgt ut på bakgrunn av et klart kriterium kan vi beregne samlede effekter separat for intensivelever og for andre elever ved fase én-skolene. Effekten på intensivelevne kan vi finne ved å sammenligne deltakerne med elever med tilsvarende prestasjoner ved andre skoler, som ville deltatt i intensivopplæringen dersom de gikk på en fase én-skole, denne effekten svarer til den direkte effekten. Effekten på øvrige elever kan vi finne ved å sammenligne disse med elever ved andre skoler som presterer bedre enn intensivelevne til første termin, og derfor ikke ville deltatt i intensivopplæringen. Denne effekten svarer til ringvirkningene.

Dersom det ikke er noe klart kriterium for deltakelse i intensivopplæringen vil vi ikke kunne skille like klart mellom direkte effekter og ringvirkninger. Ved å dele opp elevgruppen etter første termin-karakterer vil vi imidlertid sannsynligvis kunne finne grupper med svært forskjellig grad av deltakelse i intensivopplæringen. Grupper med svake termin én-karakterer og høy deltakelse i intensivopplæringen vil være da kunne gi et uttrykk for den direkte effekten, mens en gruppe med sterkere karakterer og lav deltakelse i intensivopplæringen vil gi et uttrykk for ringvirkninger.

2.2.3. 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)
- Karakterstatistikk for grunn- og videregående skole (fra VIGO)
- Nasjonale prøver (NP)
- Skoledata fra Grunnskolens informasjonssystem (GSI)
- Elevundersøkelsen (EU): Elevers svar på en omfattende spørreundersøkelse

Alle resultatene vi studerer kommer fra disse datakildene. I tillegg er det, som en del av evalueringen av Overgangsprosjektet, gjennomført kartlegginger av NOVA (Sletten, Bakken og Haakestad, 2011) og NIFU (Holen og Lødding, 2012, samt en ytterligere kartlegging av intensivopplæringen i 2013). Disse omfatter en registerbasert registrering av alle intensivelever, samt spørreundersøkelser til elever, lærere og skoleledere. I denne rapporten studerer vi resultater for 2011-kullet (vi kommer tilbake til årsakene til dette under), dermed baserer vi oss i hovedsak på registreringen gjort av NOVA i 2011. Vi vil komme tilbake til senere årskull og dataene samlet inn av NIFU i kommende analyser.

Gjennomstrømning i videregående kommer fra NUDB, skoleprestasjoner i grunn- og videregående skole kommer fra de respektive karakterstatistikkenes, og vi bruker mål på motivasjon fra EU. Konkret vil vi studere følgende resultatmål på individnivå:

- Skriftlig eksamenskarakter fra grunnskolen (i enten engelsk skriftlig, matematikk eller norsk hovedmål skriftlig)
- Gjennomsnittlig standpunktkarakter fra grunnskolen (snitt alle fag)
- Gjennomsnittlig standpunktkarakter i fag med skriftlig eksamen (engelsk skriftlig, matematikk, norsk hovedmål og sidemål)

- Gjennomsnittlig standpunktkarakter i fag med muntlig eksamen (engelsk muntlig, RLE, naturfag, samfunnsfag, norsk muntlig)
- Gjennomsnittlig standpunktkarakter i fag uten eksamen (kroppsøving, kunst og håndverk, mat og helse, musikk)
- Videregåendepoeng (snittkarakter fra Vg1, og opptaksgrunnlag for Vg2)
- Fullført og bestått Vg1 på normert tid (ett år etter fullført grunnskole)
- Registrert i Vg2 på normert tid (registrert i oktober året etter fullført grunnskole)
- Registrert i videregående opplæring høsten på normert tid (registrert i oktober året etter fullført grunnskole)

Besvarelsene fra EU kan ikke kobles med enkeltelever, men med skole. Følgende resultatmål, alle basert på besvarelser avgitt av elever på 10. trinn (som på de aktuelle tidspunktene besvarte EU på våren) studerer vi på skolenivå:

- Motivasjon (gjennomsnittlig svar på spørsmålet "Er du interessert i å lære på skolen?")
- Motivasjon i målgruppen for intensivopplæringen
- Elevenes vurdering av lærerne (gjennomsnittlig svar på "Har du lærere som gir deg lyst til å jobbe med fagene?")
- Elevene i målgruppen for intensivopplæringen sin vurdering av lærerne

Vi har termin én-karakterer for elever fra skoleåret 2006-07 (med noe varierende dekning over tid). For å finne elever i målgruppen for intensivopplæringen har vi tatt utgangspunkt i registrerte termin én-karakter og deltakere. Basert på dette har vi beregnet sannsynlighet for deltakelse i intensivopplæringen som en fleksibel funksjon av termin én-karakterer. Vi har så regnet ut en tilsvarende sannsynlighet for deltakelse i intensivopplæringen for hver besvarelse i EU, ut fra karakterene respondentene oppgir. Merk at vi har gjort dette uavhengig av hvorvidt skolen er i fase én og i gang med Ny GIV, *målgruppen* for intensivopplæringen kan vi definere også på andre skoler og i andre år.

Fra registerdataene henter vi også ut en del kontrollvariable. Dette er variable som kan tenkes å ha en effekt på elevens prestasjoner og gjennomstrømming, og også være forskjellige mellom elever/skoler/kommuner som henholdsvis deltar / ikke deltar i intensivopplæringen på et gitt tidspunkt. På individnivå:

- Kjønn
- Nasjonale prøver 8. trinn
- Mor og fars utdanning (målt som år normert varighet)
- Hvorvidt eleven er innvanderer 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 tillegg vil vi regne ut snitt av noen elevkjennetegn (innvandrere/barn av innvandrere, mor og fars utdanning), skriftlig eksamen (snitt over årene 2008-2010) og motivasjon og vurdering av lærere fra EU (også snitt over 2008-2010). I noen sammenhenger vil vi også bruke noe historisk informasjon: Antall elever i 2002, samt skriftlig eksamen i 2002-2004.

Til slutt, på kommunenivå har vi informasjon om:

- Antall elever i kommunen (fra GSI, også data for 2002)
- Sentralitet
- Arbeidsledighet (målt i juni 2010)

I denne rapporten studerer vi effekter på elevene som gikk ut av grunnskolen våren 2011, dvs. det første grunnskolekullet som deltar i intensivopplæringen. Det er tre grunner til at vi har avgrenset analysene på denne måten. For det første, hovedmålsetningen til Ny GIV er å øke fullføringen av videregående opplæring. Vi er derfor spesielt interessert i elever som vi kan se hvordan presterer her. Karakterdata fra videregående blir tilgjengelig omtrent seks måneder etter at elevene avslutter skoleåret, i desember, samtidig blir foreløpige tall for fullføring av foregående skoleår og oppmelding for inneværende skoleår tilgjengelige. Det betyr, at de nyeste dataene tilgjengelig for analysene i denne rapporten har vært karakterer, fullføring og oppmelding fra 2012. Grunnskolekullet 2011 vil, med normert progresjon, ha startet i Vg1 høsten 2011, fullført Vg1 våren 2012 og startet i Vg2 høsten 2012. Alt dette er det tilgjengelige data for. Videre vil 2011-kullet, fremdeles gitt normert progresjon, ha fullført Vg2 og startet på Vg3 i løpet av 2013, men dette kan vi foreløpig ikke observere.

For det andre, det er noe usikkerhet knyttet til fordelingen av skoler på fase to og tre. Når vi skal studere resultatene for grunnskolekullet 2012 vil vi ønske å sammenligne skoler i fase én og to, som har gått i gang med intensivopplæringen, med skoler i fase tre, som startet først skoleåret 2012/2013. Fase to kan forstås på to måter: Ut fra hvilken fase skolene er tilordnet, eller ut fra når skolene først gikk i gang 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 har ingen elever i fase to. Dette kan henge sammen med at det i fase to er flere mindre skoler (og kommuner). Disse kan i noen tilfeller ha vurdert å ikke ha noen elever i målgruppen for intensivopplæringen i et gitt årskull, men dette blir et tema for videre studier. For å analysere effekter på 2011-kullet er vi bare avhengige av å skille mellom fase én og senere faser. Her er det bare ubetydelige avvik mellom de skolene som er registrert som fase én-skoler og de der vi faktisk finner intensivelever.

For det tredje, den uoversiktlige tilordningen av elever og skoler har ført til at det har blitt mer arbeidskrevende enn opprinnelig forventet å gjøre en noenlunde robust effektevaluering. Vi prioriterer derfor i denne rapporten å presentere et forholdsvis detaljert bilde av utvelgelsen av elever og skoler, og, etter vår mening, robuste effektestimater på tidlige utfall for 2011-kullet. De øvrige Ny GIV-kullene vil vi komme tilbake til i senere analyser.

Det ser ut til at det også for 2012-kullet er sterkt varierende og til dels lite gjennomiktig praksis når det gjelder utvelgelse av elever. Foreløpige analyser av grunnskolerresultater, innen eller mellom skoler tyder heller ikke på vesentlig forskjellige resultater fra hva vi finner for 2011-kullet. Det er imidlertid for tidlig å trekke noen klar konklusjon her. KD presiserte i løpet av 2012 at det er svaktpresterende elever som skal delta i intensivopplæringen, det er mulig at dette har ført til at det ligger bedre til rette for sammenligninger innen skoler for 2013-kullet.

Vi gjør i en del analyser noen ytterligere avgrensninger. For sammenligninger innen skoler er vi avhengige av at det er et tydelig skille mellom elever som deltar / ikke deltar i intensivopplæringen, dette diskuterer vi i en del detalj i neste kapittel. Videre har vi funnet tegn til noen registreringsfeil i gjennomstrømningsdata fra Østfold. Av den grunn utelater vi Østfold fra sammenligninger mellom skoler og årskull. Videre har det blitt hevdet at Oslo ikke har noe klart skille mellom fase én-skoler og andre, men at alle skoler følger samme prinsipper på et gitt tidspunkt. Dette forkludrer våre sammenligninger av skoler i forskjellige faser, og vi vil derfor utelate Oslo-skoler i sammenligninger mellom skoler og årskull.

3. Utvelgelsen av elever og skoler til intensivopplæring i fase én

I dette kapitlet beskriver vi implementeringen av intensivopplæringen. I avsnitt 3.1. ser vi på hvordan elevene som får intensivopplæringen skiller seg fra de elevene som ikke får intensivopplæring og hvorvidt det ser ut til å være forskjeller mellom skoler og kommuner i hvordan elever blir valgt. I flertallet av skolene er elevene valgt ut på en måte som gjør RD-estimering vanskelig, men ved hjelp av informasjon fra prosjektledere og en søkeprosedyre finner vi fram til et utvalg av skoler og kommuner der RD-estimering kan være mulig. I avsnitt 3.2. ser vi på fordelingen av kommuner og skoler på faser, og hva som kjennetegner de som er med i fase én relativt til fase to og tre. Forskjellene er i utgangspunktet betydelige. Ved å koble hver skole i fase én med en mest mulig lik skole fra de to andre fasene, basert på observerbare kjennetegn på kommune- og skolenivå, kan vi imidlertid redusere forskjellene betraktelig.

3.1. Utvalgelse av elever innen fase én-skoler

I den empiriske analysen vil vi sammenlikne elever på fase én-skoler som får tilbud om intensivopplæring med elever på fase én-skoler som ikke får tilbud om intensivopplæring. For at dette skal være en gyldig analysestrategi må elever som får / ikke får tilbud i utgangspunktet være sammenlignbare, eller, mer realistisk, vi må være i stand til å ta hensyn til forskjellene.⁹

Tabell 3.1 gir gjennomsnittsverdien med tilhørende standardavvik for en rekke kjennetegn ved elever som deltar i intensivopplæring i kolonne én og elever som ikke får tilbud om intensivopplæring i kolonne to. I kolonne tre rapporterer vi den estimerte forskjellen mellom kolonne én og to med estimerte standardfeil.

Tabell 3.1. Hvem var intensivelevne i fase én? Sammenligning av kjennetegn ved intensivelever og andre elever i fase én-skoler

	(1) Snitt intensivelever (standardavvik)	(2) Snitt andre elever (standardavvik)	(3) Forskjell (standardfeil)
Snitt termin 1	2,863 (0,592)	3,840 (0,957)	-0,977** (0,022)
Mangler termin 1	0,013 (0,112)	0,023 (0,150)	-0,010** (0,003)
Termin 1 matematikk	2,178 (0,658)	3,543 (1,126)	-1,365** (0,026)
Termin 1 norsk hovedmål skriftlig	2,767 (0,712)	3,792 (0,928)	-1,025** (0,022)
Nasjonale prøver 8. trinn	-0,806 (0,677)	0,081 (0,872)	-0,887** (0,021)
Jente	0,404 (0,491)	0,488 (0,500)	-0,084** (0,012)
Mors utdanning (år)	11,287 (4,169)	13,244 (3,844)	-1,957** (0,096)
Fars utdanning (år)	11,209 (3,962)	12,921 (4,079)	-1,713** (0,104)
Innvandrere	0,129 (0,336)	0,070 (0,256)	0,059** (0,006)
Barn av innvandrere	0,123 (0,328)	0,068 (0,251)	0,055** (0,006)
<i>N</i>	1 972	16 112	18 084

Kolonne (1) viser gjennomsnittlige kjennetegn for intensivelever. Kolonne (2) viser gjennomsnittlige kjennetegn ved elever ved fase én-skoler som ikke deltar i intensivopplæring. Kolonne (3) viser forskjellen mellom (1) og (2), og tester om denne er statistisk signifikant. Kolonne (1) og (2) har standardavvik i parentes, kolonne (3) beregnede standardfeil. Stjerner angir signifikansnivå: * $p < 0,10$, ** $p < 0,05$

Elever som får intensivopplæring får lavere snittkarakter til første termin og nasjonale prøver i 8. klasse enn elever som ikke får tilbud om intensivopplæring.

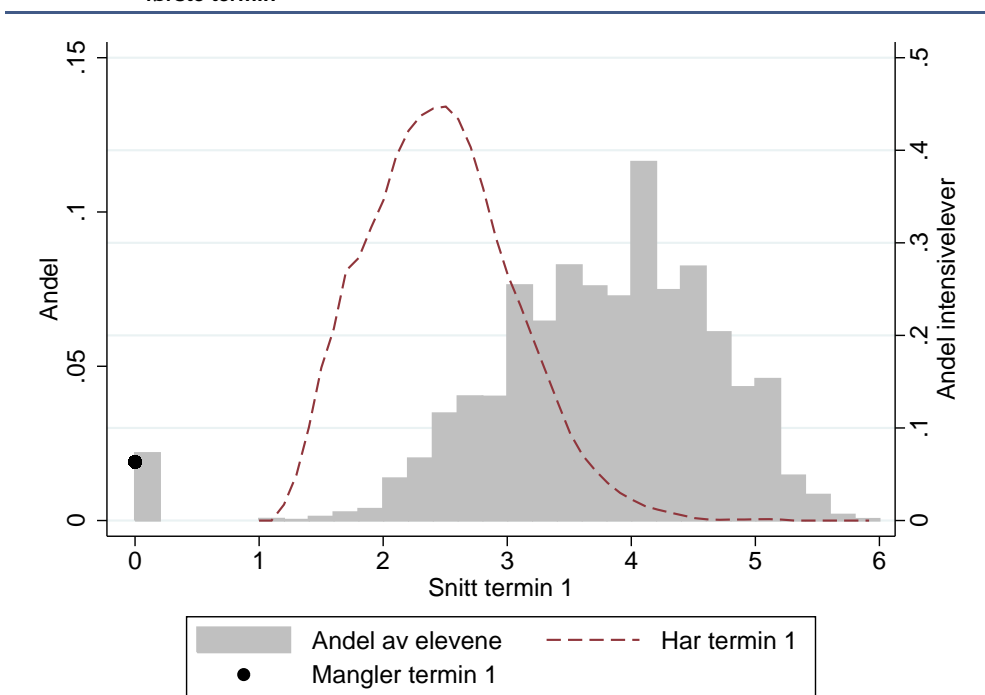
⁹ I prinsippet er det en viktig forskjell på elever som deltar og elever som tilbys deltagelse, men takker nei. Elever som takker nei er sannsynligvis ikke sammenlignbare verken med de som deltar eller med elever som ikke tilbys deltagelse. Ettersom alle rapporter tilsier at det er svært få elever som takker nei vil vi imidlertid ikke skille mellom tilbud om deltagelse og faktisk deltagelse.

Forskjellen i norsk hovedmål og (særlig) matematikk er større enn forskjellen i snitt første termin. Blant de elevene som får intensivopplæring er det også en høyere andeler gutter og elever med innvandrerbakgrunn, og foreldrene har i snitt lavere utdanning. Dette er som ventet, når kriteriet for deltakelse er svake prestasjoner og gutter, elever med innvandrerbakgrunn og elever med lavt utdannede foreldre er overrepresentert blant svaktpresterende elever. Alle forskjellene mellom kolonne én og to er statistisk signifikant til fem prosents nivå. Dette viser tydelig at intensivelever er forskjellige fra øvrige elever, og sannsynligvis ville ha prestert under snittet i fravær av intensivopplæring.

3.1.1. Tilordningspraksis

Til tross for at intensivopplæringen var tiltenkt de 10 prosent svakest presterende elevene i kommunen etter første termin i 10. trinn så viser kurven i Figur 3.1. at blant de aller svakeste elevene ved fase én-skolene er det få som får tilbud om intensivopplæring. Høyre y-akse måler andelen elever som får tilbud om intensivopplæring, mens x-aksen måler snittkarakter til første termin på en skala fra 1 til 6. Blant elevene med snittkarakter til første termin mellom to og tre er det 30-40 prosent av som deltar i intensivopplæringen, både for lavere og bedre karakterer er andelen vesentlig lavere.

Figur 3.1. Elever ved fase én-skoler: Antall elever totalt og andel intensivelever etter snitt første termin



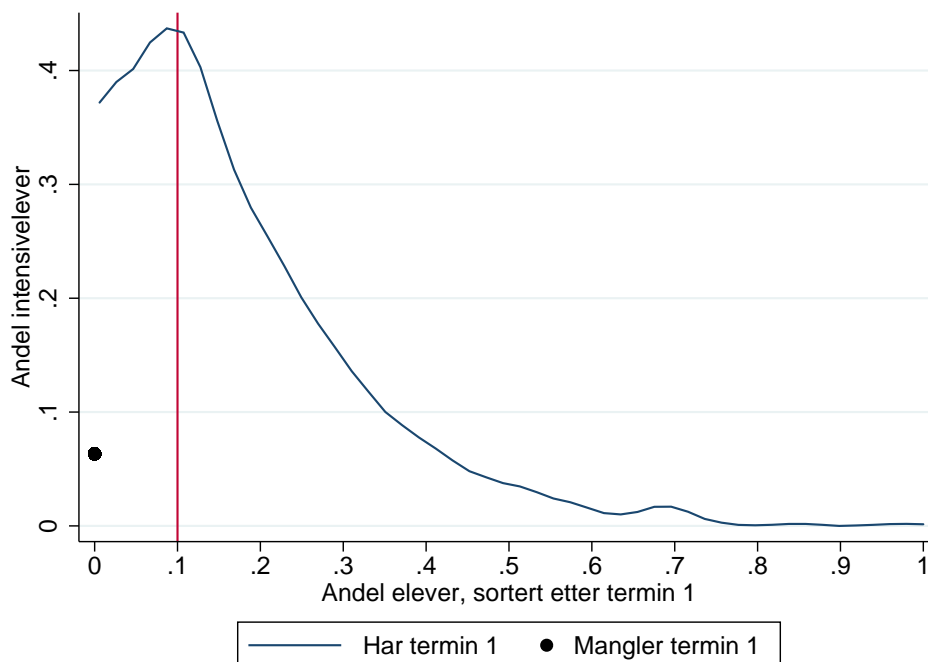
Kurven (høyre y-akse) viser hvordan andelen av elevene ved fase én-skolene som deltar i intensivopplæringen varierer med elevenes snittkarakterer til første termin. Histogrammet (venstre y-akse) viser fordelingen av elever på forskjellige snittkarakterer. Søylen og punktet til venstre oppsummerer antallet elever uten første termin-karakterer og andelen intensivelever blant disse.

Det er imidlertid få elever som har snittkarakter under to, og flere som har snittkarakter over tre enn under tre, jamfør histogrammet i Figur 3.1. Dermed er det litt mer enn halvparten av som har karakterer mellom to og tre, de øvrige stort sett mellom tre og fire.

Hvis vi heller ser på andelen intensivelever blant gruppen elever som presterer svakt, vil gruppen som presterer aller svakest og har svært lav deltakelse i intensivopplæringen utgjør en beskjeden andel av denne. I Figur 3.2 ser vi at andelen intensivelever er høyest blant de elevene som så vidt er blant de 10 prosent svakestpresterende, i overkant av 40 prosent. Dette er på fase én-skoler når intensivopplæringen er innført. Merk at under halvparten av målgruppen, de 10 prosent svakest presterende elevene, faktisk deltar i intensivopplæringen. Andelen

deltakere er så vidt lavere blant de svakeste elevene, og faller ganske bratt med bedre prestasjoner. Det er likevel en betydelig andel intensivelever blant elevene som ligger rundt 20. til 30. persentil, og også noen få deltakere med faglige prestasjoner over medianen.

Figur 3.2. Andelen deltakere i fase én-skoler etter plass i fordelingen av snittkarakterer til første termin



3.1.2. Betydningen av første termin – utvalg til RD-analyse

I praksis er det stor variasjon mellom kommuner når det gjelder utvelgelsen av elever til den første fasen våren 2010. Det dreier seg hovedsakelig om forskjeller i hvilken fagkombinasjon kommunene har brukt til å beregne gjennomsnittskarakter og dermed bestemme hvem som er de svakeste, og i hvilken grad elever med individuell opplæringsplan (IOP) deltar i intensivopplæringen.

På grunn av at lokale prosjektledere har lagt ulike strategier til grunn i utvelgelsen av elever, er det ingenting i dataene som tyder på intensivopplæringen har vært rettet mot de 10 prosent svakeste elevene. Det eneste unntaket er Stavanger, som også har lagt karaktersnittet (av alle karakterer) til grunn i utvelgelsen av elever.

I prosessen med å anslå de faktiske karaktergrensene har vi tatt utgangspunkt vi informasjon fra e-post korrespondanse med de kommunale prosjektlederne (i april 2012), hvor prosjektlederne ble bedt om å beskrive utvelgelsesprosessen i sine ansvarsområder. Videre bruker vi informasjon fra mer spesifikke spørsmål til den samme gruppen fra en spørreundersøkelse sendt i desember samme år.

Informasjonen fra prosjektlederne viser stor variasjon når det gjaldt hvorvidt grense for deltakelse ble satt på kommune- eller skolenivå. En del prosjektledere svarte imidlertid at de hadde hatt en såkalt «streng» tilordning til intensivopplæringen, det vil si at kun karaktersnittet bestemte hvilke elever som deltok. Noen opplyste også både hvilken kombinasjon av første termin-karakterer de brukte og hvilken prosentandel eller faktisk karaktersnitt som hadde bestemt om en elev fikk tilbud om tiltaket eller ikke. Gjennomgående var det imidlertid lite samsvar mellom kriteriene prosjektledere oppga og hvilke elever som var registrert som deltakere.

Vi har derfor søkt etter regler, en viss andel av elevene, sortert etter en snittkarakter i en bestemt fagkombinasjon, som forklarer deltakelse i intensivopplæringen. Vi

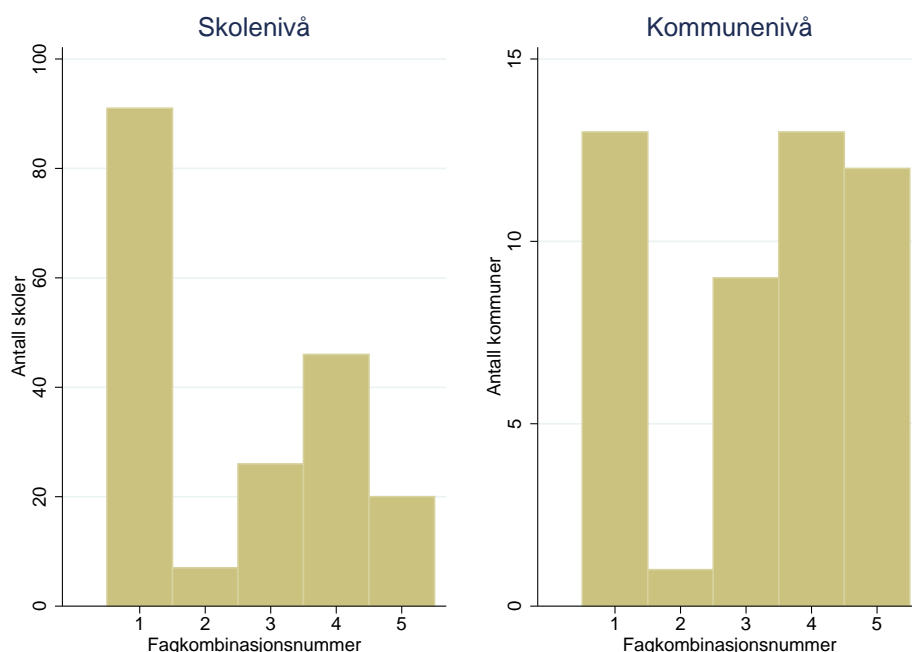
har tatt utgangspunkt i følgende fem fagkombinasjoner, som er de som ble nevnt av prosjektlederne:

1. Alle første termin-karakterer i tiende klasse
2. Alle unntatt faget mat og helse
3. Alle unntatt faget musikk
4. Alle unntatt karakteren i norsk muntlig
5. Alle unntatt karakterene i mat og helse og norsk sidemål

Ettersom vi vet hvilke elever som har fått tiltaket og kjenner karakterene deres, kan vi for gitt fagkombinasjon finne den persentil i karakterfordelingen som er best beskriver faktisk deltakelse for hver skole og kommune. Dette gjøres ved hjelp av en søkealgoritme som til slutt velger den av de fem fagkombinasjonene som gir den totalt høyeste forklaringsgraden. Denne prosedyren gjentar vi for samtlige kommuner og skoler som deltar i den første fasen av intensivopplæringen. Vi får finner dermed kommune- og skolespesifikke karaktergrenser, som varierer med hensyn på fagkombinasjon som inngår i beregningen av gjennomsnittet og prosentandel deltakere. For Stavanger kommune og enkelt skolene i Stavanger var det samsvar mellom fagkombinasjonen som best forklarte utvelgelsen (kombinasjon én, alle karakterer) og hva vi fikk opplyst fra prosjektledelsen i kommunen.

Fra søkeprosedyren får vi den relative hyppigheten kombinasjonene har blitt brukt, vist i Figur 3.3. Det ser ut til at de fleste skoler og kommuner har valgt å se bort fra én eller flere karakterer når de har beregnet elevenes snittkarakterer.

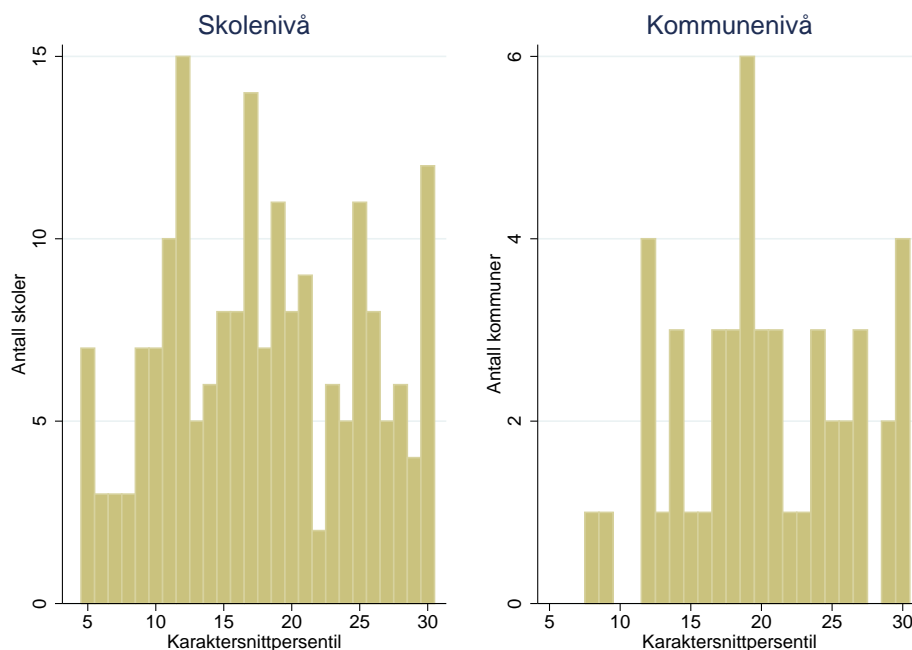
Figur 3.3. Fordeling fagkombinasjoner



Figur 3.4 viser hvilken persentil som best forklarer deltakelse i intensivopplæringen. Figuren viser at det er stor variasjon både på tvers av kommuner og skoler. Dette kan komme av flere grunner, men en viktig forklaring er grad av inkludering av IOP-elever. For eksempel, anta at to kommuner begge har 15 prosent IOP-elever, og den ene i større grad velger å inkludere IOP-elevene i målgruppen for intensivopplæringen. Da vil, alt annet likt, intensivopplærings-elevene på den skolen som i stor grad ekskluderer IOP-elever fra målgruppen ha et

høyere karaktersnitt og plassere seg høyere i karakterfordelingen enn intensivopplæringselevne på den skolen som inkluderer de.¹⁰

Figur 3.4. Tilordningsgrenser på skole- og kommunenivå. Persentilene som best forklarer tilordning av elever til intensivopplæringen på henholdsvis skole- og kommunenivå

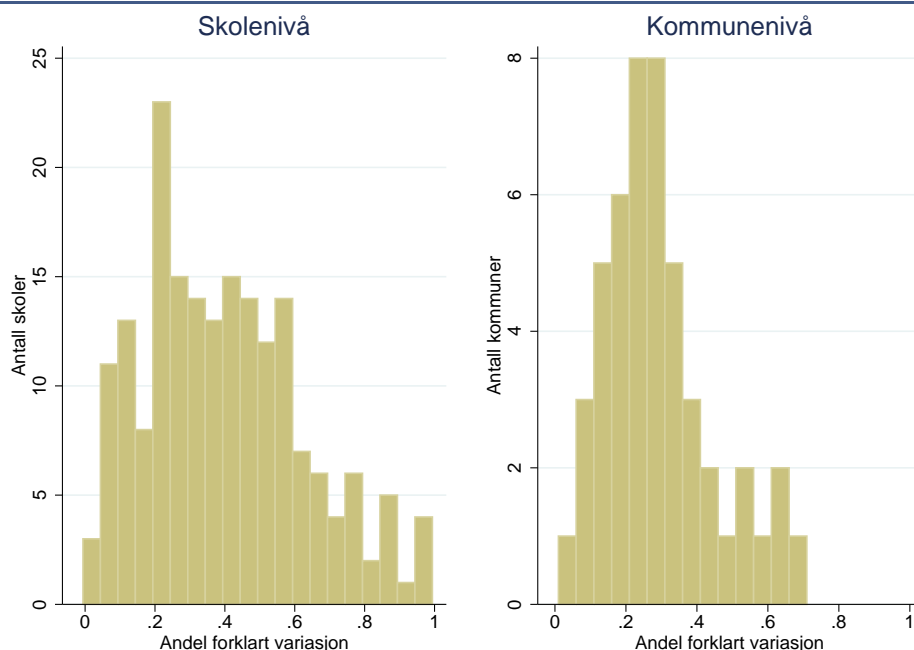


En annen forklaring på den store variasjonen er at skolene i stor grad har brukt skjønn i utvelgelsesprosessen. Dersom det er tilfelle vil metoden vår, som kun belager seg på at utvelgelsen er basert på karakterer gi lav forklaringsgrad. I Stavanger finner vi at en 12 prosentandel forklarer utvelgelsen best, mot 10 prosent som har blitt rapportert fra prosjektledelsen. Da forskjellen dreier som om en veldig liten forskjell i karaktersnitt som ligger til grunn for å få tiltaket kan dette skyldes tilfeldigheter, men det kan også komme av at noen få elever med IOP ikke har blitt ansett for å ha nytte av intensivopplæringen. De fleste med IOP i Stavanger må imidlertid ha fått tiltaket.

Figur 3.5 viser hvor streng tilordning skoler og kommuner har hatt. Med streng tilordning menes at karaktersnittet i stor grad har bestemt utvelgelsen av elever til intensivopplæring og dermed har en forklaringskoeffisient med høy verdi. Forklaringskoeffisienten uttrykker hvor mye av variasjonen vi er i stand til å forklare. Høyeste mulige verdi er én, som svarer til at vi fullt ut kan forklare tilordningen av elever til intensivopplæringen med karaktersnitt, mens null svarer til ingen sammenheng. Kun et fåtall skoler, og enda færre kommuner har hatt en streng tilordning, Stavanger er den ene kommunen der vi er i stand til å forklare over 80 prosent av variasjonen, alle øvrige kommuner ligger under 75 prosent, og et stort flertall under 40 prosent. I det følgende vil vi omtale over 50 prosent som ”strengt” og mellom 25 og 50 prosent som ”delvis strengt”.

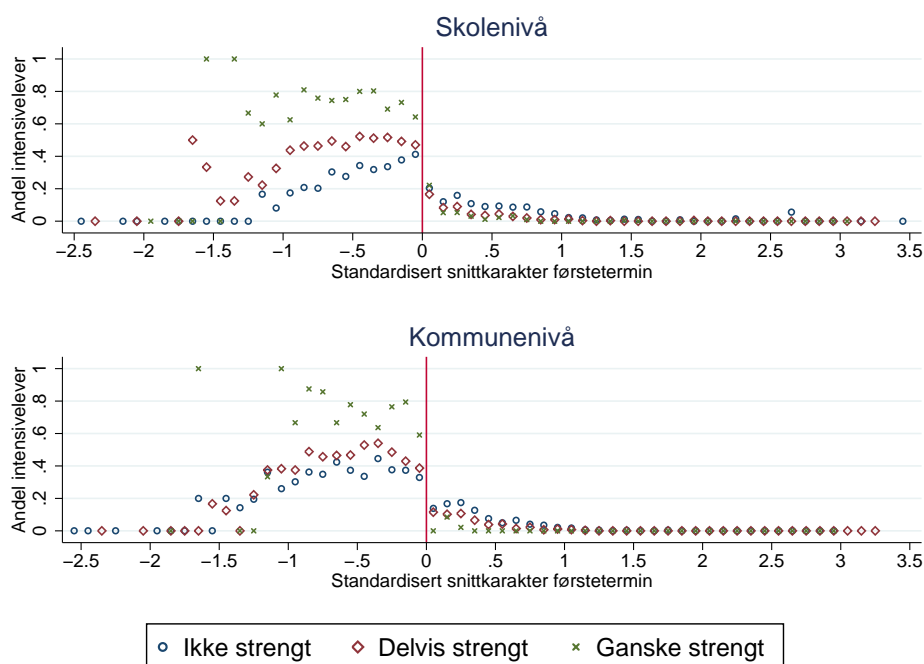
¹⁰ Vektet med elevtall er det persentil 22 som best forklarer utvelgelsen. 10 prosent av elevene skal få intensivopplæringintensivopplæringintensivopplæring, mens det i snitt er 12 prosent av elevene som har IOP. Dette tyder på at IOP-elever i stor grad presterer svakt til første termin, men ikke deltar i Ny GIV. Vi har ikke informasjon om IOP på individnivå, så det er vanskelig å studere dette i detalj (i avsnitt 3.1.2 ser vi på sammenhengen mellom Ny GIV-deltagelse og forhold vi kan observere). Imidlertid, når vi sammenholder skolenes avgrensning av intensivopplæringintensivopplæringintensivopplæringen og andelen elever med IOP finner vi et ganske godt samsvar. Vi finner også at høyere andel IOP-elever går sammen med lavere snittkarakterer til første termin. Dette er ikke noe definitivt bevis, men peker i retning av at IOP er en stor grad av forklaringen på de relativt høye grensene for tilordning til Ny GIV og for at det er mange svaktpresterende elever som ikke deltar.

Figur 3.5. Grad av streng tilordning, skole- og kommunenivå

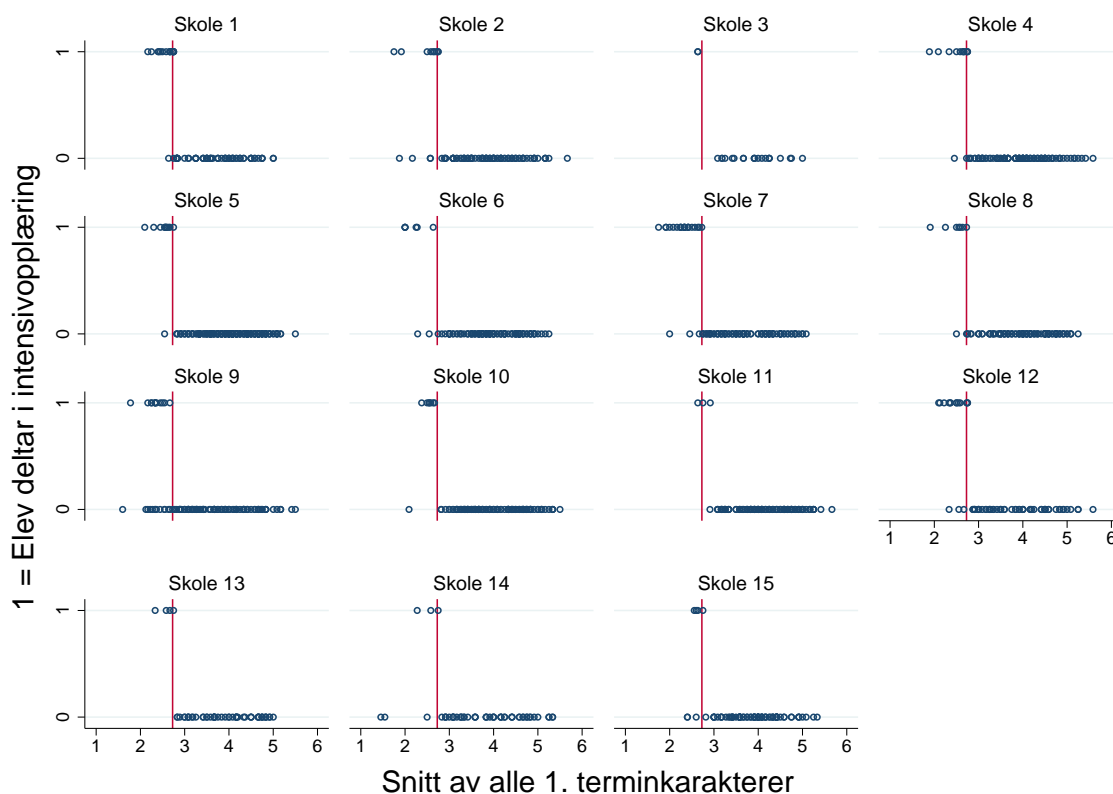


I Figur 3.6 standardiserer vi ved å måle hvor mye den aktuelle snittkarakteren avviker fra karaktergrensen og ser på andelen som deltar i intensivopplæring rundt denne grensen. Det vil si at hvis en kommune har brukt et snitt på karakteren 2,8, mens en annen har brukt 3,2, så vil begge ha en grense på null i figuren. En elev med snittkarakter 3 vil imidlertid være plassert ved 0,2 i den første kommunen, og til høyre for grensen, mens en elev med snittkarakter 3 i den andre kommunen vil være plassert ved -0,2, til venstre for grensen. Videre har vi delt skoler og kommuner i kategorier etter grad av streng tilordning, definert over. Som forventet ser vi at skoler og kommuner med forklaringskoeffisienter større enn 50 prosent i større grad viser et klart fall i andelen som deltar i intensivopplæringen ved karaktergrensen.

Figur 3.6. Sannsynlighet for å delta i intensivopplæringen, fordelt etter kategorier av "strenghet", skole- og kommunenivå



Figur 3.7. Sannsynlighet for å delta intensivopplæringen i skoler i Stavanger etter snittkarakter til første termin, grense satt på kommunenivå



Det er forskjellen i andelen rundt grensen vi vil bruke for å estimere effekter innen skoler. Hvor tydelig diskontinuitet man skal kreve må i dette tilfellet veies mot å få tilstrekkelig antall elever tilgjengelig for å kunne estimere effekter. For videre analyser ser vi på to ulike utvalg som vektlegger de to kriteriene ulikt. I det første utvalget prioriterer vi å få et stort antall elever å estimere på og legger til grunn en forklaringskoeffisient på en halv som tilstrekkelig streng utvelgelsespraksis. Det gir et utvalg av skoler med totalt 4849 elever, hvorav 565 fikk intensivopplæringen.

Det er mulig å tenke seg andre forklaringer for at søkealgoritmen finner diskontinuiteter selv om det reelt sett ikke har blitt praktisert en streng utvelgelse. For eksempel kan andre kriterier, som at eleven bør ha en viss motivasjon for tiltaket, ha blitt lagt til grunn. Vi vil likevel kunne finne et klart fall i sannsynligheten for deltakelse i intensivopplæringen ved en tilfeldig karakterverdi dersom det er en tilfeldig klumping av elever med rette kjennetegn for deltakelse i intensivopplæringen, slik skolen vurderer det, akkurat ved denne karakteren. Vi er ikke i stand til å skille dette fra et fall som faktisk skyldes en streng praksis. For en gitt skole er det usannsynlig at vi skal finne en slik klumping av elever, men når vi undersøker mange skoler og kommuner er det sannsynlig at vi vil kunne finne noen tilfeller av dette. Samplet med et tydelig fall i sannsynligheten for deltakelse i intensivopplæringen vil dermed kunne være sammensatt av enheter der det faktisk er streng praksis, og enheter der det kun ser slik ut.

På grunn av dette vil vi også estimere tilsvarende effekter for skoler i Stavanger, som i særklasse er den kommunen hvor det er klare samsvar mellom beskrivelsen av utvelgelse og det vi observerer i data. Figur 3.7 viser hvordan utvelgelsen i Stavanger har foregått ved at prosjektledelsen har satt en grense for kommunen og skolene så har tilbudt deltakelse i intensivopplæringen til elevene som har et lavere snitt enn denne grensen. Det er noen uregelmessigheter på enkeltskoler, to skoler har en andel forklart variasjon i utvelgelsen på mindre enn en halv. For Stavanger sett under ett er andelen forklart variasjon på hele 0,7 så her har vi en stor forskjell i sannsynlighet ved grensen til å estimere effekter. Ulempen med å kun velge Stavanger er at vi da kun sitter igjen med totalt 1465 elever, hvorav 130 deltok i

intensivopplæringen. Dette fører til at resultatene har begrenset presisjon, og vil ikke være i stand til å finne eventuelle små til moderate effekter.

3.1.3. Andre forhold

I forrige avsnitt så vi at for mange skoler forklarer snittkarakteren fra første termin mye av forskjellen i sannsynlighet for å få intensivopplæringen. Det ble også tydelig at mye av variasjonen må forklares av andre faktorer. Ettersom det er få av de svakeste elevene som deltar, og så en andel som faller mot null for elevene midt i karakterfordelingen, kan sammenhengen være komplisert. Figur A.3 i Vedlegg A viser imidlertid at sammenhengen mellom deltakelse i intensivopplæringen og det karaktersnittet og den grensen vi finner ved den enkelte skole kan tilnærmes godt med en forholdsvis enkel funksjonsform.¹¹

I Tabell 3.2. ser vi på korrelasjonene mellom elevkjennetegn og sannsynligheten for deltakelse i intensivopplæringen for elever i fase én-skoler, kontrollert for den tilnærmede sammenhengen med termin én-karakterer fra Figur A.3. I første kolonne ser vi på sammenhengen mellom deltakelse i intensivopplæringen og elev- og familiekjennetegn. Vi ser at det, gitt termin én-karakterer, ser ut til å være større sannsynlighet for at jenter deltar i intensivopplæringen. Forskjellen er imidlertid ikke veldig stor, omtrent 2 prosentpoeng. Selv om kjønnsforskjellen er signifikant også i kolonne fem i tabellen, hvor alle kjennetegnene er inkludert i modellen, reduseres den til omtrent 1,3 prosentpoeng. Det å mangle nasjonale prøver fra 8. trinn er negativt assosiert med deltakelse, men denne forskjellen er ikke lenger signifikant i kolonne fem. Forskjellen i mors utdanning er også ubetydelig. Det er ikke noe som tyder på at innvandrerbarn eller barn av innvandrere i mindre grad har blitt inkludert i intensivopplæringen, til tross for at dette har blitt oppgitt fra prosjektledelsen i flere kommuner. Vi ser faktisk snarere en svakt større sannsynlighet for at innvandrere deltar.

I kolonne to til fire bruker vi informasjon fra karakterdataene utover snitt første termin. I kolonne to ser vi på antall standpunktkarakterer.¹² Å mangle standpunktskarakterer ser ut til å ha en negativ sammenheng med deltakelse. Det kan komme av at IOP-elever, som ofte mangler enkelte karakterer, i mindre grad får intensivopplæringen. Forskjellen blir imidlertid betydelig mindre når man ser på den rikere modellen i kolonne fem.

I kolonne tre og fire ser vi nærmere på resultater i enkeltfag, både tallkarakterer og andre registreringer.¹³ Det å være fritatt for matte er negativt assosiert med deltakelse, mens elever fritatt for Norsk i større grad deltar. Å mangle standpunktkarakterer i disse fagene av andre grunner enn fritak ser ikke ut til å være assosiert med endret sannsynlighet for deltakelse. Til slutt legger vi spesielt merke til at elever med karakterene én og to i matte har større sannsynlighet for deltakelse. Dette kan kanskje forklares med at enkelte skoler har hatt som kriterium at alle elever med én og to skal være med i intensivopplæringen uavhengig av karaktersnittet for øvrig. Vi ser imidlertid ingen tilsvarende sammenheng for norsk.

Tabell 3.2 rapporterer også i hvor stor grad de forskjellige modellene er i stand til å forklare deltakelse i intensivopplæringen, og forskjellen i sannsynligheten for

¹¹ Figur A.3 viser en fleksibelt modellert sammenheng mellom Ny GIV-deltagelse og karaktersnittet og –grensen vi fant at best forklarer utvelgelsen i forrige avsnitt. Funksjonen som tilnærmer denne sammenhengen er to uavhengige annengrads polynom på hver side av grensen. Analysene tillater også forskjeller i sannsynligheten for Ny GIV-deltagelse mellom skoler for gitt karakter til første termin. Histogrammet viser fordelingen av elever, og vi ser at tilnærmingen er god der det er et betydelig antall elever.

¹² Vi kontrollerer også for om eleven mangler enda flere karakterer. Det er forholdsvis få elever som gjør det slik at disse resultatene blir mindre presise, men kvalitativt sett ligner de på hva vi ser for elevene som mangler fire karakterer.

¹³ Vi rapporterer bare resultatene knyttet til fagene matte og norsk hovedmål, men kontrollerer for tilsvarende resultater også for øvrige fag i estimeringene.

deltakelse mellom elever over og under grensen. Vi ser at begge disse påvirkes noe av hvorvidt vi tar hensyn til om elever har en eller to i enkeltfag. Hovedbildet er imidlertid at ingen av kontrollvariablene vi inkluderer i Tabell 3.2 har spesielt stor innvirkning på forklart variasjon eller forskjell i deltakelse over/under grensen. Til tross for at enkelte variable har en klar sammenheng med deltakelse i intensivopplæringen ligger andelen forklart variasjon i alle tilfeller på 30-35 prosent, og forskjellen i deltakelse i intensivopplæringen over/under grensen på omtrent 35 prosentpoeng. Dette tilsier at vi fanger det vesentligste av tilordningspraksisen bare ved å se på termin én-karakterer, og at det er forholdsvis lite å hente ved å ta hensyn til ytterligere kjennetegn.

Tabell 3.2. Sammenheng mellom utvalgte elevkjennetegn og deltakelse i intensivopplæringen, elever ved fase én-skoler og gitt første termin-karakterer

	(1)	(2)	(3)	(4)	(5)
	Deltar	Deltar	Deltar	Deltar	Deltar
Forventet å delta fra termin én-karakterer	0,359** (0,011)	0,361** (0,011)	0,362** (0,011)	0,344** (0,011)	0,345** (0,011)
Jente	0,020** (0,004)				0,013** (0,004)
Nasjonale prøver 8. trinn	-0,003 (0,003)				-0,007** (0,003)
Mangler nasjonale prøver 8. trinn	-0,101** (0,011)				-0,019 (0,013)
Innvandrere	0,008 (0,010)				0,026** (0,010)
Barn av innvandrere	-0,010 (0,009)				0,006 (0,009)
Mors utdanning (år)	-0,001** (0,001)				-0,001** (0,001)
Fars utdanning (år)	0,000 (0,001)				-0,000 (0,001)
Mangler 4 standpunktkarakter		-0,179** (0,026)			-0,062* (0,035)
Mangler 3 standpunktkarakter		-0,144** (0,017)			-0,061** (0,026)
Mangler 2 standpunktkarakter		-0,073** (0,016)			-0,001 (0,023)
Mangler 1 standpunktkarakter		-0,022** (0,007)			0,037** (0,016)
Mangler standpunktkarakter i matte			-0,010 (0,043)		-0,023 (0,043)
Mangler standpunktkarakter i norsk hovedmål			0,058 (0,038)		0,020 (0,039)
Fritatt i matte			-0,174** (0,047)		-0,127** (0,048)
Fritatt i norsk hovedmål			0,126** (0,048)		0,139** (0,048)
Karakteren 1 i matte				0,115** (0,014)	0,096** (0,014)
Karakteren 1 i norsk hovedmål				0,022 (0,033)	0,037 (0,033)
Karakteren 2 i matte				0,129 (0,006)	0,111** (0,006)
Karakteren 2 i norsk hovedmål				0,027 (0,008)	0,033** (0,008)
N	17 445	17 445	17 445	17445	17 445
R ²	0,306	0,318	0,319	0,329	0,343

Resultater fra estimering av lineære sannsynlighetsmodeller, avhengig variabel er deltakelse i intensivopplæringen. Alle spesifikasjoner kontrollerer for andreordens polynom i positiv/negativ avstand til beregnet grense for deltakelse. Beregnede standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

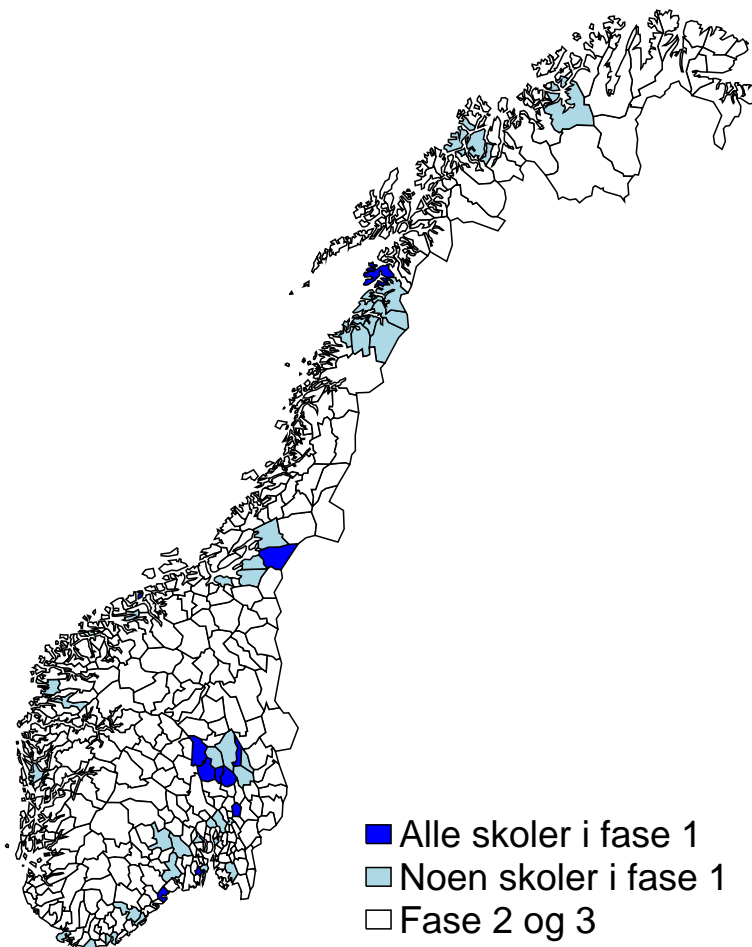
3.2. Utvelgelse av kommuner og skoler til faser

Som allerede nevnt skulle kommunene gradvis innføre intensivopplæringen i tre faser. For å gjøre implementeringen så enkel som mulig foreslo KD at de største kommunene skulle være med i den første fasen. Fylkesadministrasjonene ser ut til å ha fulgt KD sin anbefaling i stor grad.

Figur 3.8 viser den geografiske fordelingen av kommuner til fase én og de to andre fasene. I alle fylker var det kommuner som deltok i fase én. I tillegg ser vi at det er

en stor konsentrasjon av kommuner i fase én rundt store byer som Oslo, Bergen, Stavanger, Tromsø.

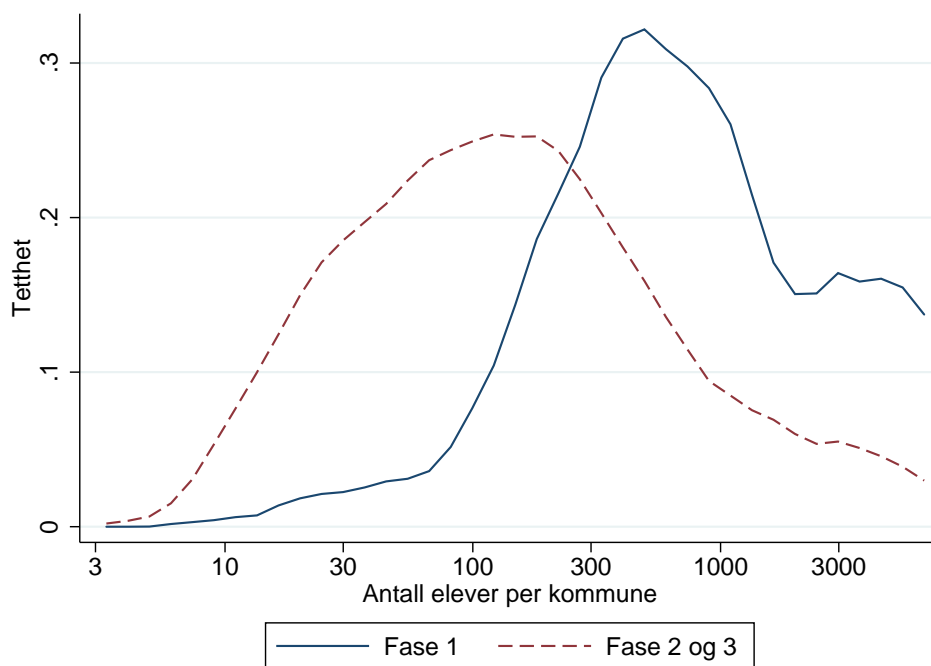
Figur 3.8. Geografisk fordeling av kommuner med første deltakelse i fase én, to og tre



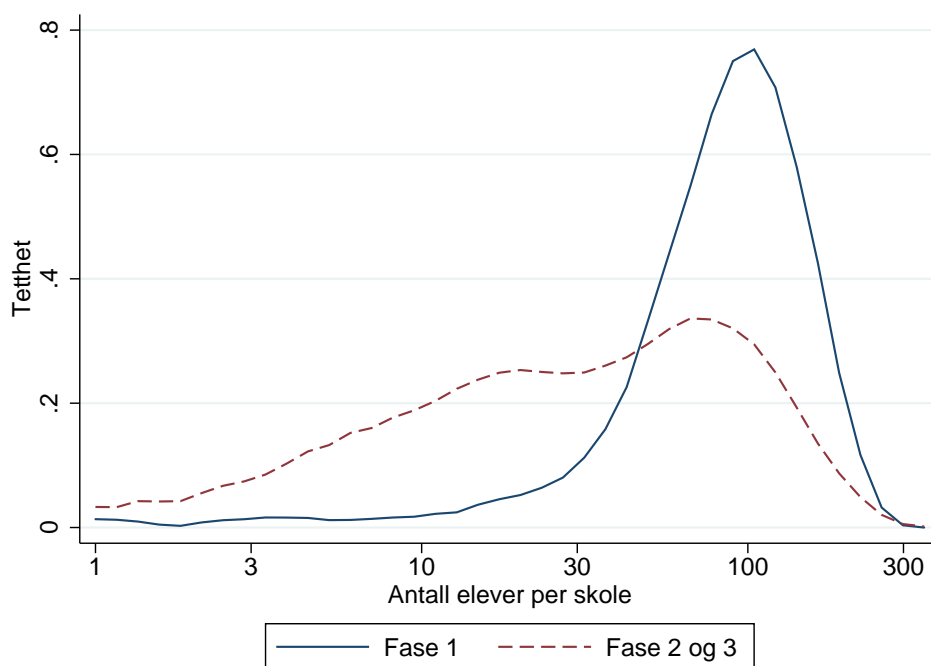
Figur 3.9 viser størrelsesfordelingen blant kommune i de forskjellige fasene, og bekrefter at det er flere store kommuner i fase én enn i fase to og tre. I figuren viser den blå linja, som gir størrelsesfordelingen til kommunene i fase én, at kommunene i den første fasen typisk er større enn kommunene i fase to og tre.¹⁴ Vi ser tilsvarende at skoler i fase én ofte er større enn skoler i fase to og tre i Figur 3.10. Det ser altså ut til at fylkene valgte store kommuner som igjen valgte store skoler innad i kommunen. Det er noe variasjon mellom fylker og mellom kommuner (se Figur A.1 og Figur A.2 i Vedlegg A), men i all hovedsak er bildet det samme.

¹⁴ Merk at vi i Figur 3.9 og Figur 3.10 viser fordelingen av logaritmen til antall elever, ikke antall elever (lineært). Dette er fordi det er noen få kommuner som er svært mye større enn de øvrige. Vi finner også at det er en tydeligere sammenheng mellom log størrelse og Ny GIV-deltagelse enn mellom størrelse og Ny GIV-deltagelse. Dette er rimelig, ettersom det svarer til at effekten av en prosentvis økning, for eksempel en dobling, av størrelsen er konstant, uavhengig av nivået. En sammenheng med (lineær) størrelse svarer til at effekten av en gitt økning i antall elever er konstant, dvs. at en økning fra 10 til 20 elever svarer til en økning fra 1010 til 1020.

Figur 3.9. Fordelingen av kommuner på størrelse etter første deltakelse i intensivopplæringen



Figur 3.10. Fordelingen av skoler på størrelse etter første deltakelse i intensivopplæringen



I Tabell 3.3 viser vi i første kolonne gjennomsnittsverdier og standardavvik for ytterligere kjennetegn ved skoler i fase én. I kolonne to viser vi forskjellen mellom disse skolene og skolene i fase to og tre og estimerte standardfeil som viser hvor presist vi estimerer disse forskjellene. Et positivt fortegn betyr at skolene i fase én har en høyere gjennomsnittsverdi enn skolene i fase to og tre.

Når det gjelder kjennetegn ved kommunene ser vi, ikke overraskende, at de skolene som ble med i fase én ligger i kommuner som i gjennomsnitt er større og mer sentrale enn skoler som ble med i de to neste fasene. I tillegg ser vi at den gjennomsnittlige utdanningen til mor og far samt arbeidsledigheten er høyere i de

kommunene som har skoler som kom med i fase én sammenliknet med kommuner som har skoler som kom med senere.

Når det gjelder kjennetegn ved skoler ser vi også her at det er flere store skoler i fase én enn i fase to og tre. Videre har skolene i fase én færre lærertimer til ordinær undervisning per elevenstime og også færre lærertimer til spesialundervisning, mens andelen med ukvalifiserte lærere er noe lavere i skoler i fase én enn i to og tre. En viktig forskjell mellom fasene ser vi i antall elever med enkeltvedtak i skolene. I fase én er det i snitt 2,7 flere gutter og 1,2 flere jenter enn i fase to og tre. Utdanningsnivået til foreldrene er også tydelig forskjellig, med i snitt omtrent 2 måneder mer registrert utdanning for både far og mor i fase én. Og det er også en betydelig større andel barn av innvandrere på fase én skoler med over dobbelt så stor andel registrert.

Tabell 3.3. Forskjeller mellom fase én og fase to og tre skoler før og etter matching

	(1)	(2)	(3)
	Snitt fase én	Forskjell før matching	Forskjell etter matching
Kommunekjennetegn			
Log elever i kommunen	6,161 (1,012)	1,394** (0,121)	0,213 (0,145)
Lite sentrale kommuner	0,061 (0,240)	-0,058* (0,030)	-0,000 (0,032)
Noe sentrale kommuner	0,341 (0,476)	0,158** (0,039)	-0,008 (0,064)
Mest sentrale kommuner	0,576 (0,496)	0,158** (0,048)	0,008 (0,066)
Arb.ledighet i kommunen (juni 2010)	2,670	0,198**	0,064
kolekjennetegn			
Log elever ved skolen	4,425 (0,687)	0,731** (0,085)	0,018 (0,083)
Elevenes vurdering av lærer i 2008-10	3,244 (0,177)	-0,017 (0,021)	-0,011 (0,022)
Elevenes motivasjon i 2008-10	3,987 (0,160)	0,014 (0,017)	0,008 (0,018)
Ordinære lærertimer/elevtime	0,051 (0,014)	-0,009** (0,002)	0,001 (0,002)
Lærertimer spesialundervisning/elevtime	0,015 (0,008)	-0,006** (0,002)	0,001 (0,001)
Andelen ukvalifiserte lærere	0,023 (0,040)	-0,018** (0,006)	0,007 (0,005)
Gutter med enkeltvedtak	6,962 (4,325)	2,686** (0,362)	0,705 (0,535)
Jenter med enkeltvedtak	3,212 (2,710)	1,150** (0,218)	0,348 (0,324)
Elever med enkeltvedtak	10,174 (6,161)	3,836** (0,510)	1,053 (0,749)
Snitt skriftlig eksamen 2008-10	3,484 (0,247)	0,027 (0,025)	-0,035 (0,031)
Log elever ved skolen i 2002	4,269 (0,730)	0,685** (0,085)	-0,003 (0,089)
Log elever i kommunen i 2002	5,963 (1,001)	1,316** (0,116)	0,190 (0,142)
Snitt skriftlig eksamen 2000-02	3,451 (0,215)	0,047* (0,024)	-0,008 (0,027)
Snitt mors utdanning (år)	13,291 (0,877)	0,159** (0,076)	-0,020 (0,105)
Snitt fars utdanning (år)	12,811 (0,864)	0,153* (0,083)	-0,157 (0,108)
Andel innvandrere	0,058 (0,050)	0,007 (0,005)	0,009 (0,006)
Andel barn av innvandrere	0,033 (0,052)	0,019** (0,003)	0,005 (0,006)
Antall skoler	132	723	226
Test, alle forskjeller lik null (χ^2)		177,9	25,9
p-verdi		0,000	0,211
Pseudo R^2 fra probit-modell		0,259	0,071

Beregnete standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

Oppsummert, det viser seg å være store forskjeller mellom fase én og fase to og tre-skolene. Dette tilsier at det kan være resultatforskjeller også, som vi må ta hensyn til i estimeringen av den kausale effekten. Det gir neppe gyldige estimater å sammenligne resultatene i fase én-skoler med andre skoler, uten å ta hensyn til forskjellene.

En måte å ta hensyn til forskjellene er å koble fase én-skolene med andre, like skoler ved hjelp av matching-teknikker. Kolonne tre viser forskjellene mellom fase én-skolene og dette matchede datasettet.¹⁵ Vi ser at matching-teknikken som vi har benyttet har på langt vei balansert variablene vi bruker. Forskjellene er betydelig redusert. I motsetningen til i det opprinnelige datasettet er ingen av forskjellene individuelt signifikante, og de er heller ikke statistisk signifikante sett under ett. Merk imidlertid at vi mister en rekke skoler fra datasettet. I det opprinnelige datamaterialet har vi 723 skoler (132 fase én-skoler), mens i det koblede datasettet har 226 observasjoner (113 fase én-skoler). Antall unike skoler er noe lavere, ettersom noen av skolene i sammenligningsgruppen er koblet med mer enn en fase én-skole.

¹⁵ Vi har brukt Stata-programmet `psmatch2` (Leuven og Sianesi, 2003) til matchingen. Ved matching konstruerer man et mål for likhet, og kobler så hver av enhetene som skal matches med en eller flere andre enheter. Vi har sammenlignet skoler ved hjelp av såkalt Mahalanobis-avstand, en metrikk som søker å balansere alle kjennetegn som inngår i beregningen. Kjennetegnene vi har latt inngå er alle i Tabell 3.3. Hver fase én-skole er matchet med den skolen i sammenligningsgruppen som er nærmest.

4. Evaluering

I denne seksjonen vil vi presentere resultatene så langt fra effektevalueringen. Først, i avsnitt 4.1, vil vi se på sammenhengen mellom utfallsvariablene vi skal studere og fullføring av videregående og forskjeller i resultater mellom intensivelever og andre elever. I avsnitt 4.2 vil vi benytte oss av beskrivelsen av tilordningspraksis innen skoler fra avsnitt 3.1 til å sammenligne elever ved fase én-skoler som deltar og ikke deltar, mens vi i avsnitt 4.3 vil sammenligne skoler i fase én med skoler i fase to og tre.

4.1. Sammenheng mellom fullføring av videregående opplæring, tidligere resultater og deltakelse i intensivopplæringen

Før vi rapporterer effektanslag er det naturlig å motivere valget av utfallsvariable og analysestrategi ved å se på sammenhengene mellom våre utfallsvariable og fullføring av videregående opplæring og mellom våre utfallsvariable og deltakelse i intensivopplæringen.

Tabell 4.1. Alternative utfallsmål. Sammenhenger mellom fullføring etter 5 år og tidligere resultatmål. Grunnskolekullene 2002-2004

	(1)	(2)	(3)	(4)
	Gjennomsnitt	Bivariat sammenheng	Bivariat R^2	Multivariat sammenheng
Skriftlig eksamen	3,43 (1,10)	0,164** (0,001)	0,152	0,011** (0,001)
Snitt standpunkt	3,95 (0,83)	0,305** (0,001)	0,298	0,059 (0,046)
Standpunkt, skriftlige fag	3,64 (0,94)	0,245** (0,001)	0,247	0,011 (0,014)
Standpunkt, muntlige fag	3,97 (0,97)	0,255** (0,001)	0,282	0,074** (0,018)
Standpunkt, fag uten eksamen	4,25 (0,74)	0,322** (0,001)	0,258	0,066** (0,014)
Fullført Vg1 år 1	0,78 (0,42)	0,563** (0,002)	0,251	0,245** (0,003)
Registrert i Vg2 år 1	0,81 (0,40)	0,516** (0,002)	0,191	0,186** (0,003)
Registrert i VGO år 1	0,92 (0,27)	0,404** (0,004)	0,054	0,027** (0,004)
<i>N</i>	197 408			185 755
R^2				0,378

Resultater fra en lineær sannsynlighetsmodell, med oppnådd kompetanse innen fem år som avhengig variabel. Beregnede standardavvik (kolonne (1)) / standardfeil (kolonne (2) og (4)) i parentes, * $p < 0,10$, ** $p < 0,05$

I Tabell 4.1 ser vi på sammenhengen mellom utfallsmålene vi studerer og det målet som intensivopplæringen er ment å påvirke, fullføring av videregående opplæring (oppnådd yrkes- eller studiekompetanse innen fem år etter oppstart). Ettersom årskullene som deltar i intensivopplæringen kun er omtrent halvveis i videregående baserer Tabell 4.1 seg på grunnskolekullene 2002-2004. Kolonne én viser gjennomsnittet for de forskjellige utfallsmålene med standardavvik i parentes under. Vi ser at 92 prosent av elevene er registrert i videregående opplæring (VGO) etter et drøyt år. 81 prosent er registrert på det andre året på videregående (Vg2), så flertallet av de gjenværende 11 prosent ligger sannsynligvis etter normert progresjon. Ellers er det betydelig variasjon i alle utfallene.

Fra kolonne to ser vi tydelig at alle utfallsmålene også er klart korrelert med fullføring av VGO. Både bedre skriftlig eksamenskarakter, bedre standpunkt-karakter (i alle grupper av fag), fullført Vg1 til normert tid, registrering i Vg2 til normert tid og å være registrert i VGO etter ett år går sammen med høyere grad av fullføring av VGO. Det er imidlertid forskjeller i hvor sterke sammenhengene er. For eksempel er det en sterkere sammenheng mellom standpunkt-karakterer og fullføring enn mellom skriftlig eksamen og fullføring, enten vi måler sammenhengen med regresjonskoeffisienten i kolonne to eller andel forklart variasjon i kolonne tre. Til slutt, i kolonne fire ser vi at også når vi tar hensyn til alle forhold

samtidig er det en sammenheng mellom fullføring og hvert av de tidlige resultatene. Noen av sammenhengene mellom standpunkt karakterene og fullføring er riktignok ikke statistisk signifikante, men dette skyldes at det er vanskelig å skille mellom de forskjellige høyt korrelerte standpunktmålene, ikke at sammenhengene er svake.

Tabell 4.2. Resultater i grunnskole- og Vg1, ubetingede og betingede forskjeller mellom intensivelever og andre elever

	(1) Alle elever	(2) Alle elever	(3) Termin 1 under 3	(4) Termin 1 under 3
Skriftlig eksamen	-1,025** (0,032)	0,018 (0,038)	0,063 (0,045)	0,038 (0,040)
Snitt standpunkt	-1,083 (0,055)	-0,018** (0,009)	0,004 (0,011)	0,003 (0,011)
Standpunkt, skriftlige fag	-1,145 (0,042)	0,026 (0,024)	0,039** (0,020)	0,016 (0,015)
Standpunkt, muntlige fag	-1,266 (0,067)	-0,044** (0,015)	-0,020 (0,014)	-0,023 (0,014)
Standpunkt, fag uten eksamen ...	-0,810 (0,051)	-0,029 (0,020)	0,022 (0,024)	0,038 (0,025)
Fullført Vg1 år 1	-0,257 (0,020)	-0,075** (0,016)	0,002 (0,022)	0,010 (0,022)
Registrert i Vg2 år 1	-0,110 (0,021)	-0,025 (0,017)	0,043** (0,017)	0,050** (0,017)
Registrert i VGO år 1	-0,006 (0,007)	-0,008 (0,007)	0,021** (0,010)	0,023** (0,009)
Videregående poeng	-10,476 (0,532)	-0,798** (0,305)	0,184 (0,485)	0,329 (0,499)
Mangler videregående poeng	0,021 (0,008)	0,002 (0,007)	-0,018 (0,012)	-0,020* (0,011)
Kontrollvariable	Ingen	Snitt termin 1	Snitt termin 1	Termin 1 (fleksibelt), NP8, kjønn, foreldres utdanning

Forskjeller i resultater mellom intensivelever og alle andre elever, kontrollert for variablene angitt nederst i tabellen. Beregnede standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

I Tabell 4.2 ser vi sammenhengene mellom deltakelse i intensivopplæringen og de forskjellige utfallsmålene. Kolonne én sammenligner resultatene til intensivelever og andre elever. I denne sammenligningen vil det åpenbart være andre forhold enn deltakelse i intensivopplæringen som også varierer, jamfør Tabell 3.1. Derfor er det ikke overraskende at intensivelevne jevnt over presterer vesentlig svakere. Dette sier lite om effekten av intensivopplæringen, bare at intensivopplæringen i hvert fall ikke er tilstrekkelig til å utjevne forskjellene som var der. I kolonne to tar vi hensyn til en vesentlig forskjell, snittkarakter til termin én. Dette reduserer forskjellen mellom intensivelever og andre betydelig. Det er imidlertid fremdeles forskjeller, blant annet i snitt standpunkt karakter og fullføring av Vg1 til normert tid. Nå er det også noen forskjeller i favør intensivelevne, om enn ikke statistisk signifikante.

I kolonne to har vi gjort en forholdsvis sterk antakelse om hvordan utfallene våre samvarierer med termin én, ved å anta at en gitt forskjell i termin én går svarer til samme forskjell i utfall, uavhengig av hvilke elever vi sammenligner eller hva nivået på termin én er. Vi kan gjøre denne antakelsen mindre streng ved å sammenligne elever som i utgangspunktet er likere, og i kolonne tre begrenser vi datamaterialet til elever med snitt fra termin én under tre. Vi ser at dette har til dels betydelig påvirkning på forskjellene vi beregner, for eksempel går sammenhengen mellom deltakelse i intensivopplæringen og normert registrering i Vg2 fra (ikke signifikant) negativ til signifikant positiv. Det er imidlertid ingen grunn til å tro at snittet fra termin én oppsummerer alle forskjellene mellom intensivelever og andre svaktpresterende elever. I kolonne fire tar vi hensyn til termin én-karakterer på en fleksibel måte, i tillegg til at vi tar hensyn til nasjonale prøver fra 8. trinn, kjønn, innvandringsbakgrunn og foreldres utdanning. En del av resultatene fra kolonne tre endres lite, men sammenhengen mellom deltakelse i intensivopplæringen og standpunkt karakter i skriftlige fag blir nærmest borte og er ikke lenger signifikant.

Tabell 4.2 illustrerer utfordringen knyttet til å sammenligne resultatene til forskjellige elever: Vi kan aldri vite hvorvidt vi har fanget opp alle relevante forhold.

Hvordan ville resultatene i kolonne fire endret seg dersom vi hadde data for elevenes motivasjon? Eller for utviklingen deres i løpet av høstsemesteret 10. trinn? Dette kan vi umulig svare på, og derfor vil fortsette effektevalueringen med strategiene beskrevet i kapittel to.

4.2. Effekter på elever innen skoler

Våre første effektestimater baserer seg på RD-tilnærmingen beskrevet i avsnitt 2.2.1, og studerer forskjeller mellom elever innen fase én-skoler. For at RD-estimering skal gi gyldige effektestimater er vi avhengig av en relativ streng utvelgelse basert på en form for karaktersnitt, slik at elever over og under grensen er nesten like. Som vi har sett i avsnitt 3.1 er det bare en begrenset andel av skolene der dette ser ut til å være tilfellet. I disse kan vi imidlertid tenke oss at det er tilfeldigheter eller veldig små forskjeller mellom elevene som avgjør om en elev er over eller under dette faktiske karaktersnittet.

Som forklart tidligere er i så tilfelle elevene med karaktersnitt rett under karaktergrensen en god sammenligningsgruppe for elevene rett over. Hvor stort intervall man bruker rundt karaktergrensen kommer an på hvor mange elever som har et snitt like i nærheten av grensen. Det er nødvendig med en viss størrelse for å få presisjon i estimatene og dermed blir det en avveining mot at også forventet gjennomsnittlig forskjell mellom gruppene av elever øker jo lenger man beveger seg fra karaktergrensen. I praksis ser vi på forskjellige intervaller og ser om estimatene er følsomme for valg av intervall.

4.2.1. Utvalget av skoler med relativ streng utvelgelse

Vi ser først på utvalget definert i avsnitt 3.1.2 med grenser funnet på skolenivå og med en andel forklart variasjon på over 0,5. I kolonne to i tabellene 4.3-4 sammenligner vi elevene som er inntil et halvt karakterpoeng under grensen med de som er inntil et halvt poeng over. Dette er i litteraturen omtalt som å bruke en båndbredde på et halvt karakterpoeng. Selv om vi har 4849 elever i utvalget bruker vi da kun 1290 elever i estimeringen av effekten av intensivopplæringen. Blant disse er 531 elever akkurat under grensen, hvorav 169 er intensivelever. Over grensen er det 759 elever, hvor 43 av dem er deltakere. Et så lite intervall gir altså en kraftig reduksjon i antall elever, men med et så stort utvalg er det fortsatt relativt mange elever å estimere på. I kolonnene én, tre og fire bruker vi en båndbredde på henholdsvis et kvart, et helt og halvannet karakterpoeng. Med de større intervallene vil antallet elever vi bruker i estimeringen øke, med den nevnte ulempen at elevene på ytterkantene av intervallene også blir stadig mer forskjellige. Etter en avveining mellom presisjon i estimatene og i hvilken grad estimatene fanger effekter av intensivopplæringen eller andre forhold er vår foretrukne båndbredde et halvt karaktersnitt, som oppgitt i kolonne to og som er det vi bruker i Figurene 4.1-4.4.¹⁶

Den første av forutsetningene nødvendig for RD-estimering, et klart fall i sannsynligheten ved karaktergrensen, er tilfredstilt mekanisk her da det nettopp er kriteriet for å bli inkludert i utvalget. Vi ser i kolonne to i Tabell 4.3, og i øverste graf til venstre graf til venstre i Figur 4.1, at sannsynligheten for deltakelse i intensivopplæringen er knappe 40 prosentpoeng høyere for elevene med snitt rett under grensen enn for gruppen av elever rett over. Dette betyr imidlertid at på begge sider av karaktergrensen er det både intensivelever og andre elever, slik at snittresultatet på begge sider av grensa vil fange både en eventuell direkte effekt og ringvirkninger. Dersom det ikke er noen ringvirkninger vil vi, ettersom fallet i

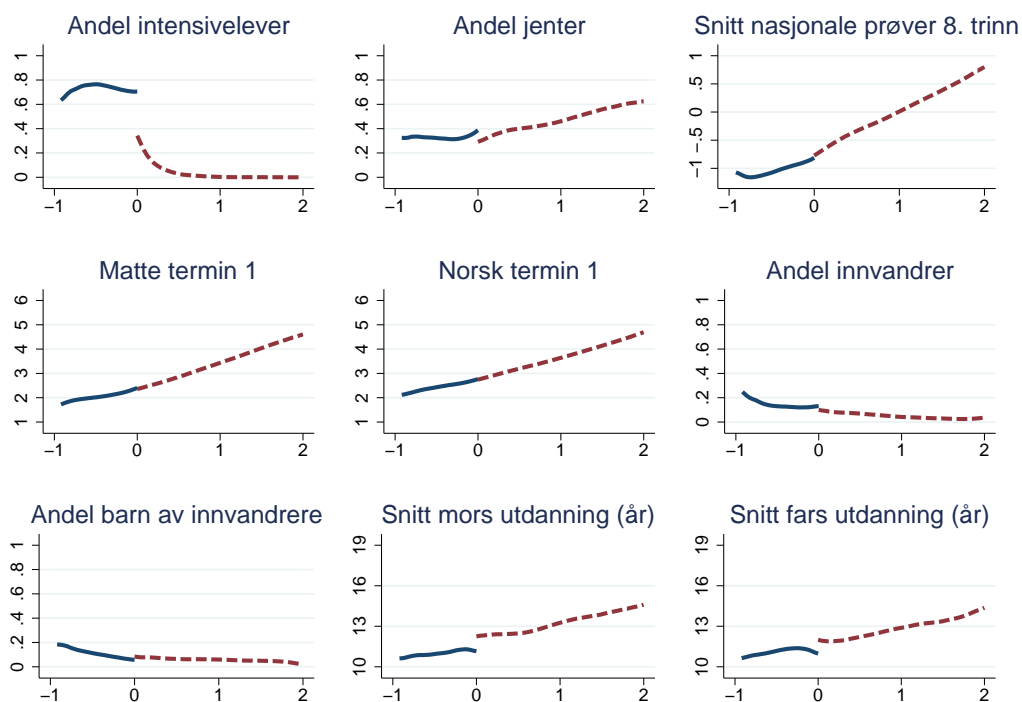
¹⁶ RD-estimeringen gjøres ved Stata-kommandoen `rd` (Nichols, 2011). Metoden beregner en form for glidende gjennomsnitt over og under karaktergrensen. Disse gjennomsnittene, vist som kurvene i Figurene 4.1-4.4, tar hensyn til at det kan være andre forskjeller mellom elever med forskjellige karakterer, og at disse forskjellene kan variere med karakternivået. Dette gjøres ved å gjøre lokale lineære regresjonsanalyser på en rekke forskjellige punkter. Effekten av å være under karaktergrensen finnes ved å sammenligne det glidende gjennomsnittet rett under grensen med det tilsvarende snittet rett over grensen. I tillegg til å basere seg på elever innen en viss båndbredde vil estimatene av å være under karaktergrensen legge større vekt på elever som er nær grensen enn elever som er lengre unna.

andelen deltakere rundt grensen er på 40 prosentpoeng, finne estimer av å være under karaktergrensen som er 40 prosent av effekten av å delta i intensivopplæringen.

Vi kan ikke teste direkte den andre forutsetningen, om at det er tilnærmet tilfeldig på hvilken side av grensen en elev havner. Men dersom elevkjenntegn som kjønn, tidligere faglige prestasjoner på nasjonale prøver og i første termin, foreldres utdanning og elevens og foreldres bakgrunn, vist i Figur 4.1, ikke viser tegn til diskontinuiteter ved grensen, vil dette være en god indikasjon på at denne forutsetningen er realistisk.

Figur 4.1. viser imidlertid at det er noen problemer med antagelsen om at elevene rundt grensen er like. Spesielt ser vi at elevenes snitt fra nasjonale prøver i 8. trinn er noe lavere under karaktergrensen, om enn ikke statistisk signifikant i analysen med båndbredde 0,5 karakterpoeng. Forskjellen øker med større intervaller, og gjennomsnittlig NP8 er signifikant forskjellig på et fem prosentnivå for de to største intervallene. Videre ser elever under grensen til å ha foreldre med noe lavere utdanning. Her er imidlertid stort sett ikke forskjellene signifikant forskjellige fra null. For de andre kjennetegnene er det få tegn til diskontinuiteter (forskjeller) rundt karaktergrensen. Sett under ett er det ikke klare tegn til at elevene er forskjellige når vi bruker en båndbredde på 0,25 eller 0,5 karakterpoeng, en båndbredde på 1 karakterpoeng er på grensen til å gi signifikante forskjeller sett under ett, mens det er klare tegn til forskjeller mellom elevene når vi bruker en båndbredde på 1,5 karakterpoeng.¹⁷

Figur 4.1. Elevsammensetning rundt grensen for deltakelse. Forskjeller i elevkjenntegn etter avstand fra beregnet skolespesifikk grense



Figurene viser hvordan forskjellige kjennetegn varierer med avstand til skolespesifikk karaktergrense for deltakelse i intensivopplæringen. x-aksen viser avstanden (karakterpoeng) til denne grensen, se avsnitt 3.1.2 for mer detaljer om datamaterialet og beregning av avstand. Se fotnote 16 for mer detaljer om estimeringen av gjennomsnittlige kjennetegn for forskjellig avstand til karaktergrensen.

Vi går nå videre til å se på estimerte effekter av intensivopplæringen, ved å se hvordan en rekke resultatmål varierer rundt karaktergrensen. Dersom elevene rett under karaktergrensen, der 40 prosentpoeng flere har deltatt i intensivopplæringen,

¹⁷ I Figur A.4 og Tabell A.1 viser vi forskjeller i skolekjennetegn rundt karaktergrensen.

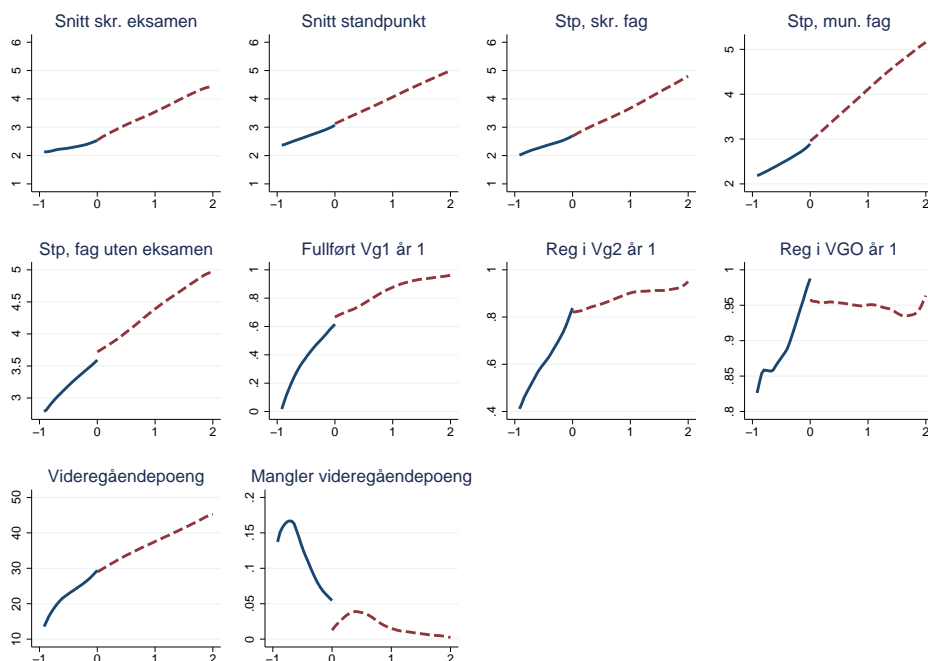
presterer bedre tyder det på at intensivopplæringen har hatt en positiv effekt. Det er imidlertid viktig å ha forskjellene i andre kjennetegn, spesielt tidligere resultater på nasjonale prøver, i bakhodet når vi tolker effektestimaterne i Tabell 4.4 og Figur 4.2. Vi ser at elever rett under karaktergrensen ser ut til å gjøre det noe dårligere enn elevene rett over på skriftlige eksamener og i standpunkt karakter for skriftlige og muntlige fag. Forskjellene er ikke signifikante ved båndbredde 0,5, men med større båndbredde er de signifikante på et 5 prosentnivå.

Tabell 4.3. Elevsammensetning rundt grensen for deltakelse. Forskjeller i elevkjenntegn etter avstand fra beregnet skolespesifikk grense

	(1)	(2)	(3)	(4)
	0,25	0,50	1,00	1,50
Andel intensivopplæring	0,221** (0,083)	0,395** (0,053)	0,497** (0,036)	0,539** (0,031)
Andel jenter	-0,008 (0,082)	0,020 (0,057)	-0,018 (0,041)	-0,014 (0,035)
Snitt nasjonale prøver 8. trinn	-0,127 (0,137)	-0,117 (0,092)	-0,167** (0,066)	-0,180** (0,056)
Matematikk termin1	0,036 (0,140)	-0,054 (0,089)	-0,033 (0,061)	-0,011 (0,052)
Norsk hovedmål termin1	0,107 (0,099)	0,034 (0,071)	-0,013 (0,051)	-0,011 (0,044)
Andel innvandrere	-0,054 (0,057)	-0,000 (0,037)	0,021 (0,026)	0,023 (0,022)
Andel barn av innvandrere	0,015 (0,044)	-0,030 (0,031)	-0,021 (0,023)	-0,014 (0,019)
Snitt mors utdanning (år)	-1,084 (0,579)	-0,579 (0,461)	-0,582* (0,345)	-0,446 (0,302)
Snitt fars utdanning (år)	-0,816 (0,634)	-0,656 (0,467)	-0,478 (0,347)	-0,458 (0,302)
Antall observasjoner	4849	4 849	4 849	4 849
Test om noen forskjeller utenom andel intensivopplæring er signifikante (χ^2)	8,270	6,789	13,273	16,630
p-verdi for test	0,408	0,560	0,103	0,034

Se avsnitt 3.1.2 for mer detaljer om datamaterialet og karaktergrenser, og fotnote 16 for mer detaljer om estimeringsmetode. Beregnede standardfeil i parentes, $p < 0,10$, $** p < 0,05$

Figur 4.2. RD-effekter av intensivopplæringen: Forskjeller i utfall etter avstand fra beregnet skolespesifikk grense



Figurene viser hvordan forskjellige resultatmål varierer med avstand til skolespesifikk karaktergrense for deltakelse i intensivopplæringen. x-aksen viser avstanden (karakterpoeng) til denne grensen, se avsnitt 3.1.2 for mer detaljer om datamaterialet og beregning av avstand. Se fotnote 16 for mer detaljer om estimeringen av gjennomsnittlige resultater for forskjellig avstand til karaktergrensen.

Dette bør likevel ikke tolkes som en negativ effekt av å ha deltatt i intensivopplæringen, ettersom forskjellen i resultater er omtrent tilsvarende som de for-

skjellene vi observerte i tidligere prøveresultater. En plausibel tolkning er heller at det er noe svakere elever som så vidt blir inkludert i intensivopplæringen og at disse forblir noe svakere etter tiltaket. Dette tyder i så fall på at vi ikke har lyktes i å finne et utvalg av skoler der elevene på hver side av karaktergrensen som best forklarer deltakelse i intensivopplæringen er nesten like. I stedet ser det ut til å være systematiske forskjeller, som gjør at vi ikke kan trekke noen klare konklusjoner fra disse analysene.

Tabell 4.4. RD-effekter på ulike utfallsmål. Forskjeller i utfall etter avstand fra beregnet skolespesifikk grense

	(1)	(2)	(3)	(4)
	0,25	0,50	1,00	1,50
Fullført Vg1 år 1	-0,011 (0,080)	-0,023 (0,055)	-0,016 (0,040)	-0,020 (0,035)
Registrert i Vg2 år 1	-0,018 (0,064)	-0,013 (0,045)	-0,022 (0,033)	-0,021 (0,029)
Registrert i VGO år 1	0,023 (0,036)	0,020 (0,024)	0,009 (0,018)	0,008 (0,015)
Snitt skriftlig eksamen	-0,203 (0,160)	-0,161 (0,107)	-0,190** (0,077)	-0,204** (0,067)
Snitt standpunkt	0,007 (0,073)	-0,041 (0,047)	-0,076** (0,033)	-0,075** (0,028)
Standpunkt, skriftlige fag	-0,048 (0,089)	-0,057 (0,059)	-0,102** (0,041)	-0,092** (0,035)
Standpunkt, muntlige fag	-0,053 (0,091)	-0,069 (0,061)	-0,093** (0,043)	-0,099** (0,037)
Standpunkt, fag uten eksamen	0,108 (0,092)	0,003 (0,059)	-0,035 (0,042)	-0,039 (0,037)
Videregående poeng	0,802 (1,801)	-0,124 (1,145)	-0,568 (0,802)	-0,664 (0,687)
Mangler videregående poeng	-0,002 (0,027)	0,020 (0,020)	0,005 (0,016)	0,000 (0,015)
Antall observasjoner	4849	4 849	4 849	4 849
Test om noen forskjeller er signifikante (χ^2) ...	4,352	7,111	24,439	33,627
p-verdi for test	0,930	0,715	0,007	0,000

Se avsnitt 3.1.2 for mer detaljer om datamaterialet og karaktergrenser, og fotnote 16 for mer detaljer om estimeringsmetode. Beregnede standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

4.2.2. Sammenligning innen fase én-skoler i Stavanger

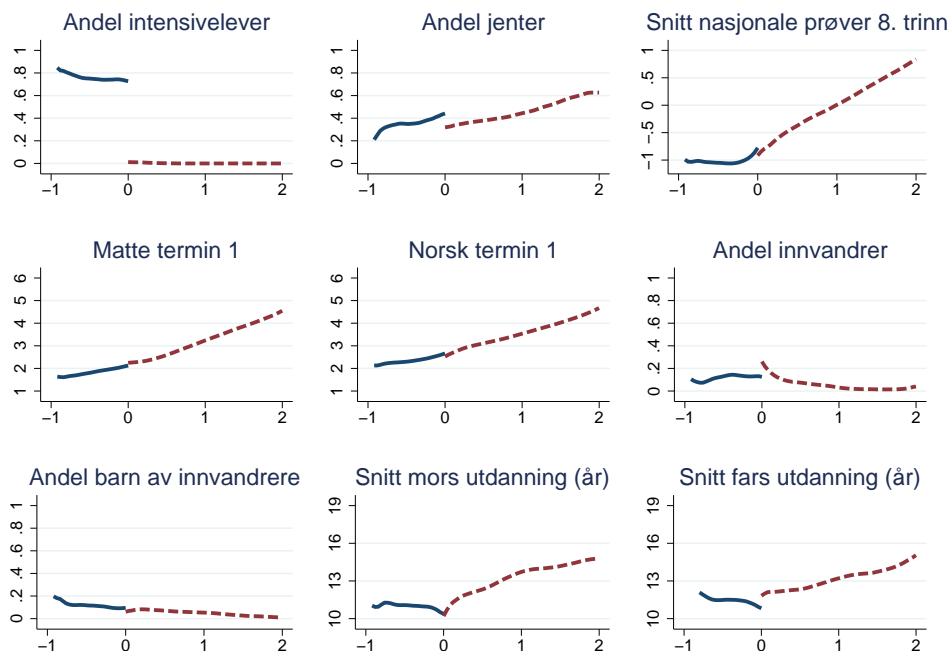
I Stavanger vet vi i større grad hvordan utvelgelsen av intensivelever foregikk. Ettersom denne var strengt basert på karaktersnitt er det sannsynligvis mulig å sammenligne intensivelever med andre elever som ikke er systematisk forskjellige, utover forskjellene i karaktersnitt, som vi kan observere. Dette begrenser imidlertid datamaterialet kraftig. Skolene i Stavanger hadde i skoleåret 2010/2011 1465 elever, hvorav 130 deltok i intensivopplæringen. Hvis vi igjen ser først på elever med et halvt karakterpoeng under og over karaktersnittgrensen som i utgangspunktet bestemte deltakelse i intensivopplæringen ser vi et litt annet bilde enn for utvalget over. Nå er det 99 av 135 elever som deltok i intensivopplæringen under, mens kun 1 av 218 over.

Forskjellen i deltakelse over og under karaktergrensen vitner om en vesentlig strengere praksis i Stavanger, som vi også ser fra første graf i Figur 4.3 og den første raden i Tabell 4.5. Sannsynligheten for deltakelse i intensivopplæringen er rundt 70 prosentpoeng høyere for gruppen under karaktergrensen. For de øvrige grafene er det viktig å merke seg at for båndbredden på et halvt karaktersnitt har vi kun 353 elever til å estimere forskjeller. Dermed vil tilfeldig variasjon (støy) kunne få stor betydning, og se ut som faktiske forskjeller i Figur 4.3. Eksempelvis ser det ut til at det er vesentlig høyere sannsynlighet for at en elev under grensen er jente enn rett over grensen. Denne forskjellen er relativt stabil for de ulike intervallene, men baserer seg på få elever, og på grunn av stor usikkerhet i estimatene kan vi ikke forkaste at forskjellen skyldig tilfeldigheter.

Den store usikkerheten er et generelt problem for utvalget med Stavanger-skoler, og det hadde helt klart vært ønskelig å ha flere kommuner som hadde praktisert tilsvarende streng praksis. Selv om det ser ut til at det kan være problemer rundt karaktergrensen for andelen med innvandrerbakgrunn og for mor og fars ut-

danning, kan dette også være tilfeldigheter. Med unntak av sannsynligheten for deltakelse i intensivopplæringen er det ikke signifikante forskjeller i elevkjenne-tegn mellom gruppene, uavhengig av valg av båndbredde. Det er heller ikke tegn til signifikante forskjeller når vi ser på alle resultatene under ett.

Figur 4.3. Forskjeller i elevkjenne-tegn etter avstand fra beregnet kommunespesifikk grense. Skoler i Stavanger



Figurene viser hvordan forskjellige kjennetegn varierer med avstand til kommunespesifikk karaktergrense for deltakelse i intensivopplæringen. x-aksen viser avstanden (karakterpoeng) til denne grensen, se avsnitt 3.1.2 for mer detaljer om datamaterialet og beregning av avstand. Se fotnote 16 for mer detaljer om estimeringen av gjennomsnittlige kjennetegn for forskjellig avstand til karaktergrensen.

Tabell 4.5. Forskjeller i elevkjenne-tegn etter avstand fra beregnet kommunespesifikk grense. Skoler i Stavanger

	(1)	(2)	(3)	(4)
	0,25	0,50	1,00	1,50
Andel intensivopplæring	0,691**	0,713**	0,720**	0,730**
	(0,092)	(0,073)	(0,059)	(0,054)
Andel jenter	0,119	0,125	0,099	0,110
	(0,163)	(0,113)	(0,083)	(0,072)
Snitt nasjonale prøver 8. trinn	0,002	0,142	-0,051	-0,127
	(0,310)	(0,186)	(0,124)	(0,105)
Matematikk termin1	-0,095	-0,113	0,008	0,070
	(0,196)	(0,135)	(0,100)	(0,087)
Norsk hovedmål termin1	0,129	0,125	-0,039	-0,054
	(0,191)	(0,137)	(0,100)	(0,087)
Andel innvandrere	-0,174	-0,119	-0,026	-0,008
	(0,133)	(0,088)	(0,060)	(0,050)
Andel barn av innvandrere	0,12	0,031	0,007	0,009
	(0,085)	(0,065)	(0,049)	(0,043)
Snitt mors utdanning (år)	-0,126	-0,030	-0,459	-0,522
	(1,910)	(1,219)	(0,826)	(0,715)
Snitt fars utdanning (år)	-1,587	-1,091	-0,910	-0,718
	(1,109)	(0,839)	(0,646)	(0,578)
Antall observasjoner	1450	1 450	1 450	1 450
Test om noen forskjeller utenom andel intensivopplæring er signifikante (χ^2)	7,413	7,085	4,268	6,952
p-verdi for test	0,493	0,527	0,832	0,542

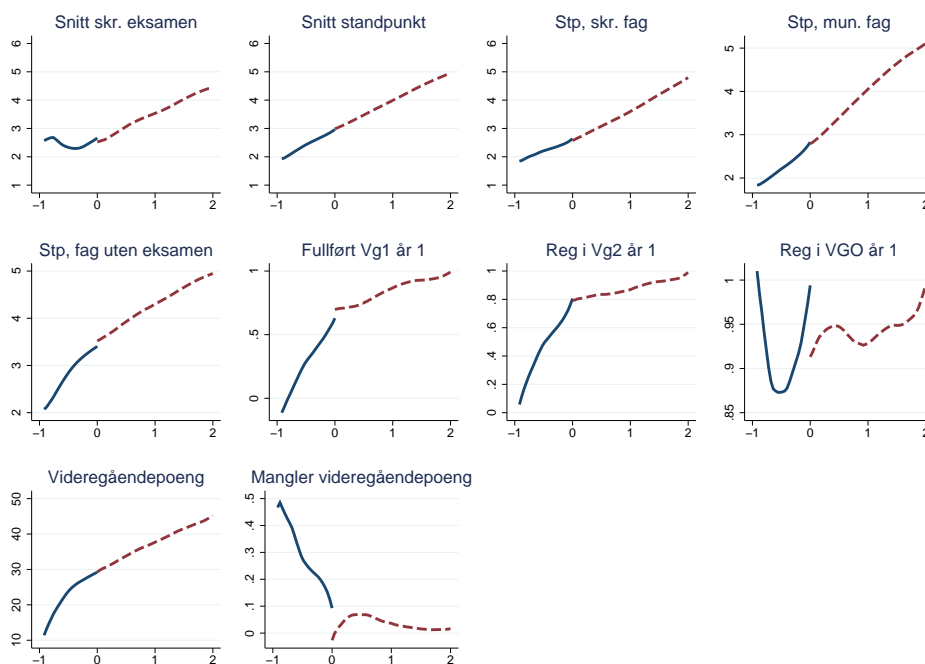
Se avsnitt 3.1.2 for mer detaljer om datamaterialet og karaktergrenser, og fotnote 16 for mer detaljer om estimeringsmetode. Beregnede standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

Vi finner heller ingen effekter av intensivopplæringen på de ulike utfallene vi er studerer. Estimaten varierer for mange av utfallene mye med valg av båndbredde, noe som tilsier at de er følsomme for spesifisering av funksjonsform. Når vi ikke finner effekter er det viktig å presisere at med så store standardfeil, det vil si stor

usikkerhet i estimatene, ville det for flere av utfallene ikke vært mulig å fastslå om det er effekter, med mindre de hadde vært dramatisk store. For eksempel, for det første utfallet i Tabell 4.6, andelen som fullfører det første året av videregående på normert tid, tilsier estimatet med vår foretrukne båndbredde at å være under karaktergrensen for deltakelse i intensivopplæringen førte til en reduksjon i fullføring på snau seks prosentpoeng. Her er imidlertid usikkerheten så stor at med den konvensjonelle graden av sikkerhet på 95 prosent kan vi ikke utelukke at den reelle effekten er alt fra en reduksjon i fullføring på 27 prosentpoeng til en økning på 16 prosentpoeng. Dersom det ikke er ringvirkninger, og alle effekter er på elever som deltar i intensivopplæringen svarer dette til et intervall fra en negativ effekt på 39 prosentpoeng til en positiv effekt på 23 prosentpoeng, ettersom forskjellen i deltakelse i intensivopplæringen er omtrent 70 prosentpoeng.¹⁸ For dette utfallet måtte altså effekten av intensivopplæringen ha vært svært stor for at vi kunne ha funnet den i et utvalg med så få elever.

Usikkerheten er relativt sett mindre for effekten intensivopplæringen har på gjennomsnittlig standpunkt karakter. Her er effekten upresist estimert til en reduksjon i snittkarakteren på 0,017 karakterpoeng, mens estimatet måtte vært en endring i karaktersnittet på mer enn 0,13 for at vi skulle kunne konkludere med en effekt den ene eller andre veien, igjen på et fem prosentnivå.¹⁹ Dette er en endring på omtrent en åttedels standardavvik, som tilsvarer at elevene i gjennomsnitt forbedrer (eller går ned) én karakter i ett til to av tretten forskjellige fagene som følge av intensivopplæringen. Vi kunne altså funnet en plausibel effekt, men det faktiske estimatet er svært nær null.

Figur 4.4. Forskjeller i resultater etter avstand fra beregnet kommunespesifikk grense. Skoler i Stavanger



Figurene viser hvordan forskjellige resultatmål varierer med avstand til kommunespesifikk karaktergrense for deltakelse i intensivopplæringen. x-aksen viser avstanden (karakterpoeng) til denne grensen, se avsnitt 3.1.2 for mer detaljer om datamaterialet og beregning av avstand. Se fotnote 16 for mer detaljer om estimeringen av gjennomsnittlige resultater for forskjellig avstand til karaktergrensen.

For det å være registrert i videregående opplæring det første året er det signifikante forskjeller på et 10 prosentnivå, men dette kan også være spuriøst når vi ser på så

¹⁸ I fravær av ringvirkninger vil effekten på de deltagende elevene være effekten på snittet delt på effekten på andelen som deltar rundt deltagesgrensen. En effekt på 2 prosentpoeng i hele elevmassen svarer da til en effekt på 20 prosentpoeng blant Ny GIV-elevne.

¹⁹ Dette ville svart til en endring på omtrent 0,18 karakterpoeng i fravær av ringvirkninger.

mange som ti utfall. Det er faktisk forventet ut fra tilfeldig variasjon at vi finner én signifikant effekt på et 10 prosents signifikansnivå når vi ser på ti uavhengige utfall. Tilsvarende på et 5 prosents nivå vil vi forvente å finne én signifikant effekt når vi ser på 20 uavhengige utfall.²⁰ Dette er et viktig poeng vi vil vende tilbake til flere ganger når vi diskuterer effektene av intensivopplæringen i dette kapittelet. Andelen som mangler videregående poeng kan se ut til å være noe forskjellig også, men dette er snakk om en så liten gruppe at det heller ikke er noe å legge for mye i. Fra Figur 4.4 ser en eventuell effekt ut til å være drevet av en kraftig redusert andel blant elevene som så vidt ikke deltar i intensivopplæringen.

Tabell 4.6. Forskjeller i resultater etter avstand fra beregnet kommunespesifikk grense. Skoler i Stavanger

	(1)	(2)	(3)	(4)
	0,25	0,50	1,00	1,50
Fullført Vg1 år 1	-0,086 (0,159)	-0,058 (0,111)	-0,049 (0,080)	-0,043 (0,070)
Registrert i Vg2 år 1	0,082 (0,153)	0,029 (0,094)	-0,018 (0,070)	-0,008 (0,061)
Registrert i VGO år 1	0,182 (0,101)	0,087 (0,052)	0,036 (0,035)	0,027 (0,031)
Snitt skriftlig eksamen	-0,039 (0,269)	0,148 (0,189)	0,103 (0,143)	0,002 (0,128)
Snitt standpunkt	-0,011 (0,088)	-0,017 (0,063)	-0,017 (0,047)	-0,016 (0,040)
Standpunkt, skriftlige fag	0,112 (0,148)	0,073 (0,097)	0,017 (0,069)	0,027 (0,059)
Standpunkt, muntlige fag	0,024 (0,129)	0,054 (0,092)	0,030 (0,070)	0,019 (0,061)
Standpunkt, fag uten eksamen	-0,105 (0,151)	-0,110 (0,097)	-0,050 (0,070)	-0,064 (0,061)
Videregående poeng	0,166 (2,680)	-0,048 (1,732)	0,283 (1,276)	0,218 (1,121)
Mangler videregående poeng	0,060 (0,055)	0,098 (0,045)	0,071 (0,045)	0,043 (0,045)
Antall observasjoner	1450	1450	1450	1450
Test om noen forskjeller er signifikante (χ^2) ...	6,160	10,664	5,416	3,727
p-verdi for test	0,802	0,384	0,862	0,959

Se avsnitt 3.1.2 for mer detaljer om datamaterialet og karaktergrenser, og fotnote 16 for mer detaljer om estimeringsmetode. Beregnede standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

4.3. Effekter på skolenivå

Vi går nå over til å se på effekter på skolenivå, ved å sammenligne skoler i forskjellige faser. Som diskutert i avsnitt 2.2, vil vi nå studere et litt annet effektbegrep enn vi gjorde i avsnitt 4.2. Effekten her vil være en gjennomsnittseffekt for intensivelever (elever som selv deltar) og andre elever ved fase én-skoler som er utsatt for eventuelle ringvirkninger. Etersom bare omtrent 10 prosent av elevene faktisk deltar i intensivopplæringen vil intensivelevne bare ha begrenset påvirkning på gjennomsnittet.

Vi har gjort en avgrensning av datamaterialet for analysen av forskjeller mellom skoler. Det er uklart hvordan intensivopplæringen er gjennomført i Oslo, om det faktisk er noen reell forskjell på hvordan det undervises i fase én-, to- og tre-skoler på et gitt tidspunkt. Videre er det tegn til registreringsfeil i gjennomstrømningsdataene for noen årganger for Østfold. Derfor har vi utelatt Oslo og Østfold fra analysene på skolenivå. Vi har undersøkt i hvilken grad resultatene våre ville endres dersom vi inkluderte Oslo og Østfold, og har funnet at dette ville ha beskjeden effekt på de effektestimaterne. Det blir imidlertid mindre klart om forskjellene vi da finner skyldes effekter av intensivopplæringen eller andre forhold.

I alle analysene som følger tar vi hensyn til en del kommune- og skolekjennetegn, oppsummert i Tabell 3.3. Resultatene er omtrent tilsvarende dersom vi ikke gjør dette. Videre kan det hende at elever på forskjellige skoler i samme område påvirkes av de samme hendelsene, som ikke trenger å påvirke elever i andre deler av landet, eller at elever i en kommune eller skole er like over tid. Vi tar derfor

²⁰ Dette er noe mer komplisert når vi som her ser på utfall som i større eller mindre grad samvarierer.

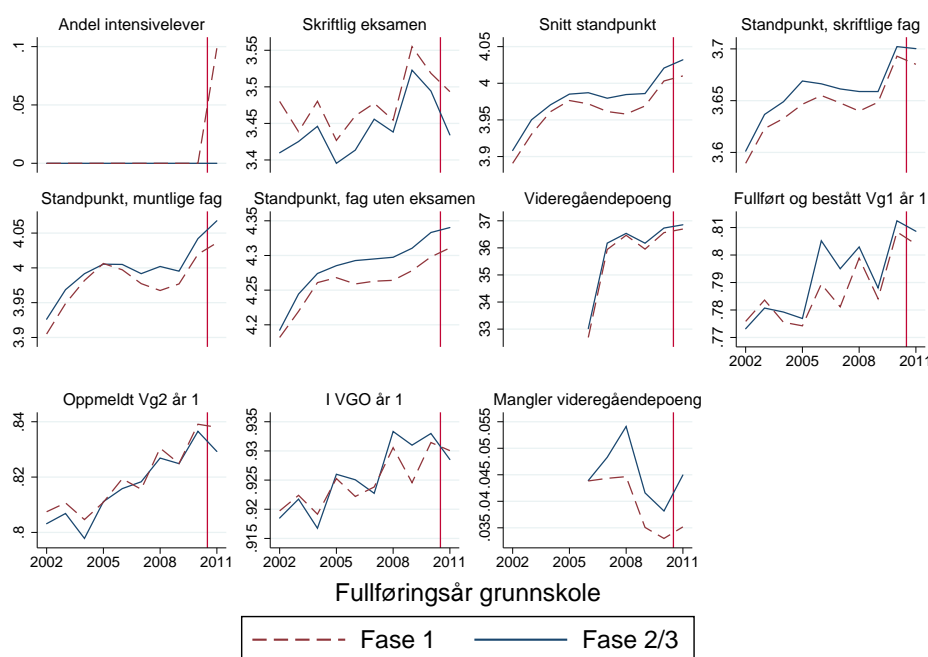
hensyn til at det kan være likheter i uobserverte kjennetegn mellom elever i samme skole, også mellom forskjellige årskull.²¹

Vi begynner, i avsnitt 4.3.1, med å studere forskjeller i utvikling mellom skoler. Deretter, i avsnitt 4.3.2, vil vi sammenligne skolene i fase én med skoler i fase to og tre som er likest mulig, og i avsnitt 4.3.3 også sammenligne utviklingen over tid i disse gruppene. I neste avsnitt, 4.4, vil vi forsøke å bygge en bro til estimatene i 4.2, ved å se på grupper med forskjellig grad av deltakelse i intensivopplæringen.

4.3.1. Forskjeller i utvikling mellom skoler

Vi begynner med å se på forskjellene mellom skoler i utvikling i forskjellige utfallsmål. Figur 4.5 viser dette, med skolene fordelt på fase én og fase to og tre sammen. Figur 4.6 viser tilsvarende utvikling for resultatmålene fra EU. Vi ser at andelen intensivelever øker brått fra 0 til ca. 10 prosent i fase én-skolene fra 2010 til 2011. Dette er hva vi ser etter også i de øvrige utfallsmålene, en brå endring som svarer til innføringen av intensivopplæringen. Vi ser imidlertid ingen tilsvarende klare endringer i noen av resultatmålene. Det er tydelig at det, i motsetning til for deltakelse i intensivopplæringen, er endringer fra 2010 til 2011, men også på tidligere tidspunkt, og dermed uavhengig av intensivopplæringen. Ikke for noen av de øvrige utfallene peker endringene i fase én-skolene fra 2010 til 2011 seg umiddelbart ut, selv om det kan se ut til å være tegn til høyere motivasjon i fase én-skolene. Vi kan dermed ikke trekke noen konklusjoner om effekter av intensivopplæringen bare ved å se på utviklingen innen fase én-skolene.

Figur 4.5. Forskjeller mellom fase én- og fase to og tre-elever over tid



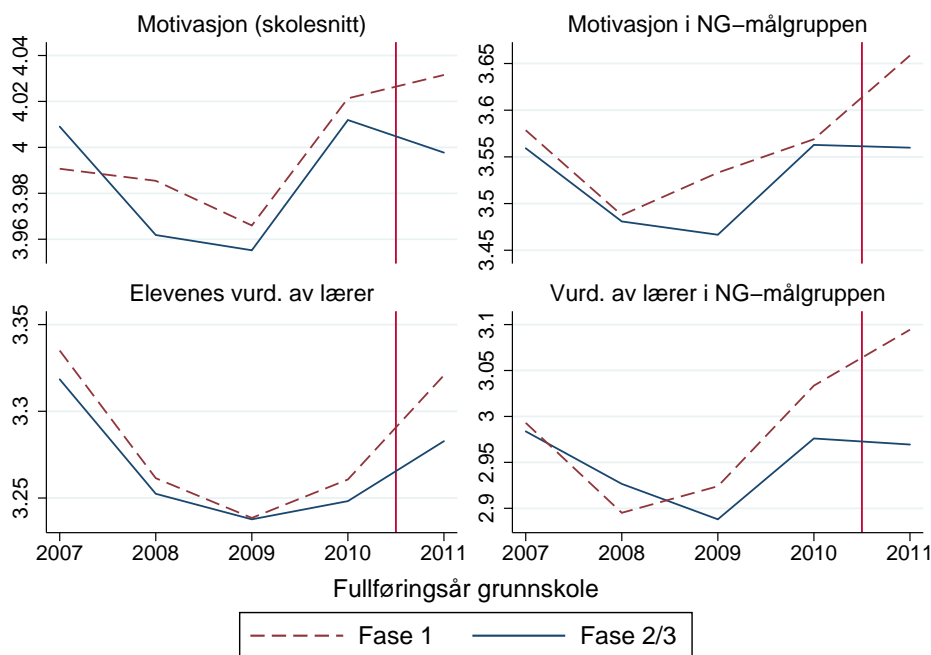
Vi ser imidlertid at utviklingen før 2011 i fase to- og tre-skolene ligner på den vi ser i fase én-skolene. Dette er rimelig dersom endringene i resultatmålene vi studerer i stor grad påvirkes av forhold utenfor de enkelte skolene. Eksempler på slike forhold kan være arbeidsmarkedsforhold, endringer ellers i utdannings-systemet, slik som ferdighetene elevene har fra grunnskolen eller kravene for opptak i videregående eller høyere utdanning, eller generelle trender som for eksempel i retning av høyere utdanningsambisjoner. Det vil også være endringer i de enkelte skolene, men såfremt disse i snitt er omtrent tilsvarende i fase én- og

²¹ Vi gjør dette ved å tillate såkalt *clustering* av residualene innen for hver kommune, over alle årene vi bruker, i alle estimeringene.

fase to- og tre-skolene vil vi kunne observere en lignende utvikling. I det følgende vil vi, med utgangspunkt i denne observasjonen, studere effekter av intensivopplæringen ved å se på forskjeller i utvikling mellom fase én-skoler og fase to- og tre-skoler.

Som beskrevet i avsnitt 2.2 vil vi da sammenligne forskjellene mellom fase én- og fase to- og tre-skoler i 2011 med forskjellene i tidligere år, for å se om denne forskjellen er stabil. Dersom forskjellen endrer seg vil dette, under bestemte forutsetninger, kunne tolkes som en effekt av intensivopplæringen.

Figur 4.6. Forskjeller mellom fase én- og fase 2- og 3-elever over tid – utfall fra Elevundersøkelsen



I et veldeignert eksperiment kunne vi fordelt skoler tilfeldig til fase én, to og tre. Dette ville, gitt et tilstrekkelig antall skoler, mekanisk sørget for at skolene i de tre fasene var like i utgangspunktet, og eventuelle forskjeller i resultater etter oppstarten av intensivopplæringen kunne vi tolket som effekter av intensivopplæringen. Vi har imidlertid i avsnitt 3.2 sett at skolene i fase én er forskjellige fra skolene i fase to og tre på en rekke områder. Derfor vil vi i utgangspunktet vente at de også har forskjellige resultater. Den kritiske forutsetningen er at forskjellen mellom fase én og fase to- og tre-skoler i fravær av intensivopplæringen ville vært stabil. Dersom forskjellene mellom fasene i fravær av intensivopplæringen ville vært den samme i 2010 og 2011 kan en endring i forskjellen fra 2010 til 2011 tolkes som en effekt.

Denne forutsetningen kan vi aldri teste. Vi kan imidlertid, ved å se på hvordan forskjellen har utviklet seg før innføringen av intensivopplæringen, vurdere om den har vært stabil tidligere. Dersom de er tilfellet, og endringen fra 2010 til 2011 er stor sammenlignet med endringene tidligere år, ville det tydet på at intensivopplæringen har en effekt.

I Figur 4.7 viser vi den beregnede forskjellen mellom fase én og fase to- og tre for en rekke forskjellige resultatmål, og for hvert av avgangskullene 2002-2011. Alle forskjellene er relativt til forskjellen vi finner for 2010-kullet. Kullet som gikk ut av ungdomskolen i 2010 blir dermed vår referanse, og vi undersøker i hvilken grad

vi ser den samme forskjellen for øvrige kull.²² Vi ser at for kullene 2002-2009 er det gjennomgående små forskjeller, den beregnede differansen i Figur 4.7 er hele veien nær null. Fra figuren ser vi at forskjellene i de aller fleste tilfeller heller ikke er statistisk signifikante, det er bare helt unntaksvis tegn til at det er noen forskjeller utover tilfeldig variasjon. Estimerte forskjeller, med standardfeil og signifikansnivå, er rapportert i Tabell A.3, Tabell A.4 og Tabell A.5. At det er noen kombinasjoner av år og resultatmål som gir signifikante forskjeller er heller ikke overraskende. Også i fravær av reelle effekter ville vi vente at noen av forskjellene er statistisk signifikante når vi tester såpass mange, som diskutert i avsnitt 4.2.2. Formelle statistiske tester for om det i det hele tatt er noen forskjeller i perioden 2002-2009, som også er vist i Tabell A.3 til Tabell A.5 tilsier bare at det er slike forskjeller for andelen som mangler videregåendeopng.

Det er for de fleste resultatmålene heller ingen klar trend over perioden. Andelen elever i videregående opplæring (VGO) etter ett år er et unntak, det ser ut til at denne andelen blant elever fra fase én-skolene faller over tid, sammenlignet med tilsvarende andel i fase to- og tre-skolene. Forskjellene over tid er imidlertid langt fra statistisk signifikante, og som parallelt med hva vi diskuterte for signifikans av forskjellene over, det er ikke overraskende at vi finner tegn til en trend for ett resultatmål når vi studerer mange.

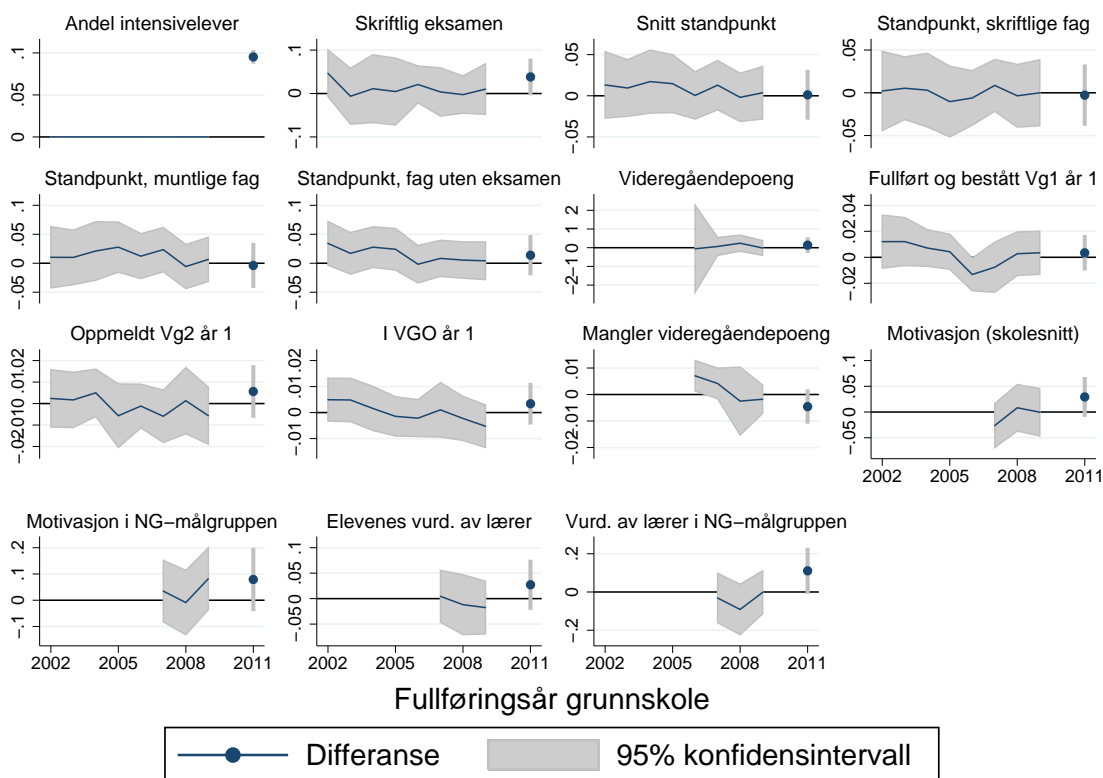
Oppsummert ser det ut til at utviklingen i de forskjellige resultatmålene har vært tilsvarende blant fase én og fase to og tre over perioden 2002-2010. Dersom vi ser klare endringer i 2011 vil dette sannsynligvis gjenspeile effekter av intensivopplæringen. I Figur 4.7, som viser effektestimater med tilhørende konfidensintervall, er det imidlertid lite påfallende endringer fra 2010 til 2011. Forskjellene er gjennomgående små i tallverdi, og langt fra statistisk signifikante. De er heller ikke spesielt store sammenlignet med forskjellene tidligere år. Noen mulig unntak finnes: Det kan være en økning i skriftlig eksamenskarakter, motivasjon og vurdering av lærere i fase én, men disse resultatene er ikke veldig tydelige.

I Tabell 4.7 har vi samlet effektestimaterne. Effektestimaterne fra figuren over er gjengitt i kolonne en, der vi sammenligner forskjellen mellom fase én og fase to og tre i 2011 med tilsvarende forskjell i 2010. Ettersom vi finner at det ikke er noen forskjeller i utvikling innad i perioden 2002-2010 sammenligner vi i kolonne to med hele denne perioden. Det betyr at sammenligningsgrunnlaget blir mindre følsomt for resultater i enkeltår, slik at estimatene kan bli mer presise. Vi finner tilsvarende resultater i begge kolonner. Det er som ventet, nettopp fordi det ikke er tydelige endringer før 2011. I noen tilfeller blir estimatene også betydelig mer presise.

Vi finner en svak og marginalt signifikant positiv effekt (på 10 prosentnivå) på skriftlig eksamen, og litt tydeligere positive effekter på motivasjon og vurdering av lærere. For vurdering av lærere finner vi også en sterkere effekt blant elever som i stor grad er intensivelever, dette er som vi skulle vente dersom intensivopplæringen har hatt en betydelig effekt. Dette er imidlertid ikke tilfellet for motivasjon. Når vi vurderer signifikansen av estimatene må vi også her ta med i betraktningen at vi vil vente å finne noen signifikante forskjeller når vi studerer mange resultatmål. Videre, for resultatene som bygger på data fra Elevundersøkelsen, dette gjelder motivasjon og vurdering av lærere, har vi bare noen få tidligere årganger. Dermed blir det vanskeligere å vurdere om en endring er stor relativt til tidligere endringer.

²² For variablene utenom motivasjon og vurdering av lærere fra Elevundersøkelsen har vi analysert data på elevnivå. For resultatene fra Elevundersøkelsen har vi analysert data på skolenivå, og vektet hver skole med antall besvarelser fra Elevundersøkelsen.

Figur 4.7. Resultater over tid, relativt til avgangskullet 2010 og fase to og tre



Tabell 4.7. Effektestimater fra sammenligning av utvikling over tid, fase én i 2011 relativt til tidligere avgangskull og fase to og tre

	Sammenlignet med	
	2010	2002-2010
Andel intensivelever	0,095** (0,004)	0,095** (0,004)
Fullført Vg1 år 1	0,003 (0,007)	0,001 (0,006)
Registrert i Vg2 år 1	0,006 (0,006)	0,006 (0,005)
Registrert i VGO år 1	0,003 (0,004)	0,003 (0,003)
Snitt skriftlig eksamen	0,038 (0,021)	0,029* (0,015)
Snitt standpunkt	0,001 (0,015)	-0,006 (0,012)
Standpunkt, skriftlige fag	-0,003 (0,018)	-0,003 (0,014)
Standpunkt, muntlige fag	-0,004 (0,020)	-0,016 (0,015)
Standpunkt, fag uten eksamen	0,014 (0,018)	0,001 (0,015)
Videregående poeng	0,142 (0,214)	0,091 (0,276)
Mangler videregående poeng	-0,004 (0,003)	-0,006* (0,003)
Motivasjon (skolesnitt)	0,029 (0,020)	0,034** (0,015)
Motivasjon i NG-målgruppen	0,079 (0,061)	0,050 (0,041)
Elevenes vurdering av lærer	0,027 (0,025)	0,033* (0,019)
Vurd. Av lærer i NG-målgruppen	0,111 (0,061)	0,141** (0,049)

Beregnete standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

4.3.2. Matching

I avsnitt 3.2 så vi at kommuner og skoler i fase én avviker systematisk fra skoler og kommuner i senere faser. Det kan derfor hende at vi kan få en mer relevant sammenligningsgruppe ved å studere de skolene som i størst grad ligner fase én-

skoler. Vi så i avsnitt 3.2 at vi er i stand til å finne slike skoler, som i stor grad ligner skolene i fase én.

I Tabell 4.8 sammenligner vi en rekke resultatmål for 2011-kullet ved fase én-skoler med tilsvarende resultater for fase to og tre. I første kolonne gjør vi en enkel sammenligning, tilsvarende forskjellen i skolekjenntegn i kolonne to i Tabell 3.3. I kolonne to begrenser vi oss til å sammenligne fase én-skolene med de matchede skolene, skolene i fase to og tre som ligner på fase én-skolene. Her ser vi på forskjeller på skolenivå. Effektene vi er opptatt av er først og fremst på elevnivå, det er mindre viktig for at resultatene blir bedre ved mange skoler med få elever enn at de bedres for et stort antall elever. Etersom vi matcher kobler skolene med like skoler, og skolene i fase én gjennomgående er store skoler, vil vi stort sett sammenligne store skoler. Det er likevel relevant å eksplisitt ta hensyn til antall elever. I kolonne tre sammenligner vi skoler, men vektet skolene ut fra deres elevtall. Skoler med mange elever får dermed større vekt, og vi får effektestimater som svarer til den gjennomsnittlige effekten på en elev, framfor på et gjennomsnitt på skolenivå. Til slutt, i kolonne fire kombinerer vi matching og elevvekter. Da sammenligner vi fase én-skoler med like skoler i fase to og tre, og legger mer vekt på resultatene for de store fase én-skolene.

Tabell 4.8. Effekter på flere utfall, estimert ved sammenligning av resultater for 2011-kullet i fase én-skoler og matchede skoler

	(1) Ingen vektning	(2) Matching- vekter	(3) Elevvekter	(4) Maching- elevvekter
Andel intensivelever	0,114** (0,006)	0,102** (0,006)	0,095** (0,004)	0,097** (0,004)
Fullført Vg1 år 1	-0,003 (0,008)	0,002 (0,009)	0,001 (0,008)	0,003 (0,008)
Registrert i Vg2 år 1	0,013 (0,007)	0,013** (0,007)	0,013** (0,006)	0,015** (0,007)
Registrert i VGO år 1	0,008 (0,005)	-0,002 (0,004)	0,000 (0,004)	-0,000 (0,004)
Snitt skriftlig eksamen	0,047 (0,025)	0,050* (0,028)	0,045 (0,027)	0,028 (0,029)
Snitt standpunkt	-0,020 (0,021)	-0,008 (0,026)	-0,017 (0,025)	-0,008 (0,029)
Standpunkt, skriftlige fag	-0,021 (0,026)	0,008 (0,026)	0,002 (0,026)	0,009 (0,029)
Standpunkt, muntlige fag	-0,032 (0,025)	-0,024 (0,029)	-0,040 (0,028)	-0,023 (0,032)
Standpunkt, fag uten eksamen	-0,004 (0,023)	-0,006 (0,032)	-0,008 (0,030)	-0,006 (0,034)
Videregående poeng	-0,167 (0,294)	0,463 (0,332)	0,337 (0,322)	0,453 (0,347)
Mangler videregående poeng	-0,010 (0,005)	-0,007 (0,004)	-0,007* (0,004)	-0,006 (0,004)
Motivasjon (skolesnitt)	0,031 (0,026)	0,017 (0,029)	0,029 (0,023)	0,032 (0,026)
Motivasjon i NG-målgruppen	0,042 (0,051)	0,025 (0,060)	0,049 (0,050)	0,039 (0,057)
Elevenes vurdering av lærer	0,056 (0,026)	0,031 (0,037)	0,041 (0,029)	0,023 (0,034)
Vurd. Av lærer i NG-målgruppen	0,100 (0,056)	0,035 (0,069)	0,055 (0,059)	0,039 (0,070)

Beregnete standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

Vi ser at de forskjellige vektingene bare i beskjeden grad påvirker forskjellene i resultater. Men det er også få signifikante effektestimater. Uavhengig av vektning finner vi en viss økning i andelen som følger normert progresjon til andre året av videregående (Vg2), på omtrent 1,3 prosentpoeng. Vi finner også en svak økning i skriftlig eksamen, på inntil 0,5 karakterpoeng, men denne reduseres og er ikke signifikant når vi bruker både match- og elevvekter. Forskjellen i elevenes vurdering av lærerne svekkes når vi vektet, og verken denne eller forskjellen i motivasjon er signifikant for noe sett av vekter.

4.3.3. Forskjellig i utvikling mellom fase én-skoler og matchede skoler

Vi har argumentert for at matching og å se på forskjeller i utvikling er to forskjellige, men relevante måter å evaluere effekter på skolenivå. Vi vil nå kombinere disse, og se på forskjellene i utvikling i fase én-skolene og de matchede fase to og tre-skolene.

I Tabell 4.9 viser vi hvordan forskjellene i utvikling i andelen som fullfører Vg1 på normert tid endres når vi sammenligner like skoler. Ikke i noen av kolonnene er det signifikante forskjeller mellom fase én-skolene og de matchede skolene, og en statistisk test for om det er en forskjell er ikke i noe tilfelle i stand til å konkludere med at det er slike forskjeller. Vi ser imidlertid at denne testen i enda mindre grad tyder på forskjeller i tidligere resultater når vi vekter. Om vi vekter ved å sammenligne like skoler eller med elevtall ser imidlertid ut til å ha en tilsvarende effekt. I kolonne fem tar vi også hensyn til at det kan være forhold på skolenivå som er konstante over tid. Dette har en beskjeden betydning for forskjellene i de enkelte årene. Når vi ser på forskjellen for 2011-kullet, som er vårt estimat for effekten på fullføring av Vg1 på normert tid, ser vi at den estimerte effekten er svært nær null.

Tabell 4.9. Effekter på fullføring av Vg1 til normert tid, estimert ved å sammenligne utviklingen i fase én-skoler og matchede skoler i fase to/tre

	(1) Ingen vektning	(2) Matching- vekter	(3) Elevvekter	(4) Matching- elevvekter	(5) Som (4) + faste skoleeffekter
2002-kullet	0,016 (0,012)	-0,010 (0,019)	-0,006 (0,016)	-0,006 (0,019)	-0,004 (0,020)
2003-kullet	0,014 (0,012)	0,009 (0,015)	0,003 (0,013)	0,005 (0,015)	0,004 (0,015)
2004-kullet	-0,011 (0,010)	-0,019 (0,012)	-0,008 (0,012)	-0,015 (0,012)	-0,013 (0,012)
2005-kullet	-0,002 (0,010)	0,008 (0,015)	0,001 (0,011)	0,005 (0,013)	0,005 (0,014)
2006-kullet	-0,010 (0,010)	-0,009 (0,013)	-0,016 (0,011)	-0,015 (0,011)	-0,015 (0,012)
2007-kullet	-0,017 (0,012)	-0,013 (0,015)	-0,018 (0,012)	-0,014 (0,013)	-0,011 (0,013)
2008-kullet	0,000 (0,011)	-0,000 (0,012)	0,002 (0,012)	-0,003 (0,012)	-0,004 (0,012)
2009-kullet	-0,003 (0,011)	-0,005 (0,012)	0,009 (0,016)	-0,002 (0,011)	-0,001 (0,011)
2011-kullet	0,002 (0,009)	0,001 (0,012)	0,001 (0,009)	0,001 (0,012)	0,002 (0,012)
N	8222	2 258	2 258	2 258	2 258
F-test: Ingen forskjeller for gsk- kullene 2002-2009	1,446	1,038	1,022	1,147	0,903
p-verdi	0,176	0,413	0,425	0,339	0,518

Beregnete standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

I Tabell 4.10 presenterer vi effektestimater for alle resultatmålene. Denne tabellen baserer seg på en sammenligning av elever før og etter innføringen av intensivopplæringen, og viser de samme forskjellige vektingene som Tabell 4.9. Resultatene ligner på hva vi så i Tabell 4.8: Det er tegn til en beskjeden effekt på andelen som har normert progresjon til Vg2, og det er svake tegn til bedret skriftlig eksamenskarakter, men disse svekkes når vi vekter med elevtall. Estimerte effekter på motivasjon og vurdering av lærere er også gjennomgående positive, men svekkes og blir ikke signifikant forskjellige fra null med enhver vektning.

Forutsetningen for at sammenligningen vi gjør i Tabell 4.10 mellom før og etter innføringen av intensivopplæringen skal være gyldig er at utviklingen i resultatene til fase én-skolene og de matchede fase to- og tre-skolene ikke avviker av andre grunner. Som når vi så på forskjeller i utviklingen for alle skoler i avsnitt 4.3.1, viser vi i Figur 4.8 og Figur 4.9 tilsvarende forskjeller i utviklingen mellom fase én-skolene og de matchede skolene. Figur 4.8 er ikke vektet med elevtall, og svarer til kolonne to i Tabell 4.10, mens Figur 4.9 er vektet med elevtall, og svarer til kolonne fire. Vi ser at forutsetningen om tilsvarende utvikling i fravær av intensivopplæringen virker rimelig. I begge figurene ser vi at forskjellene i tidligere

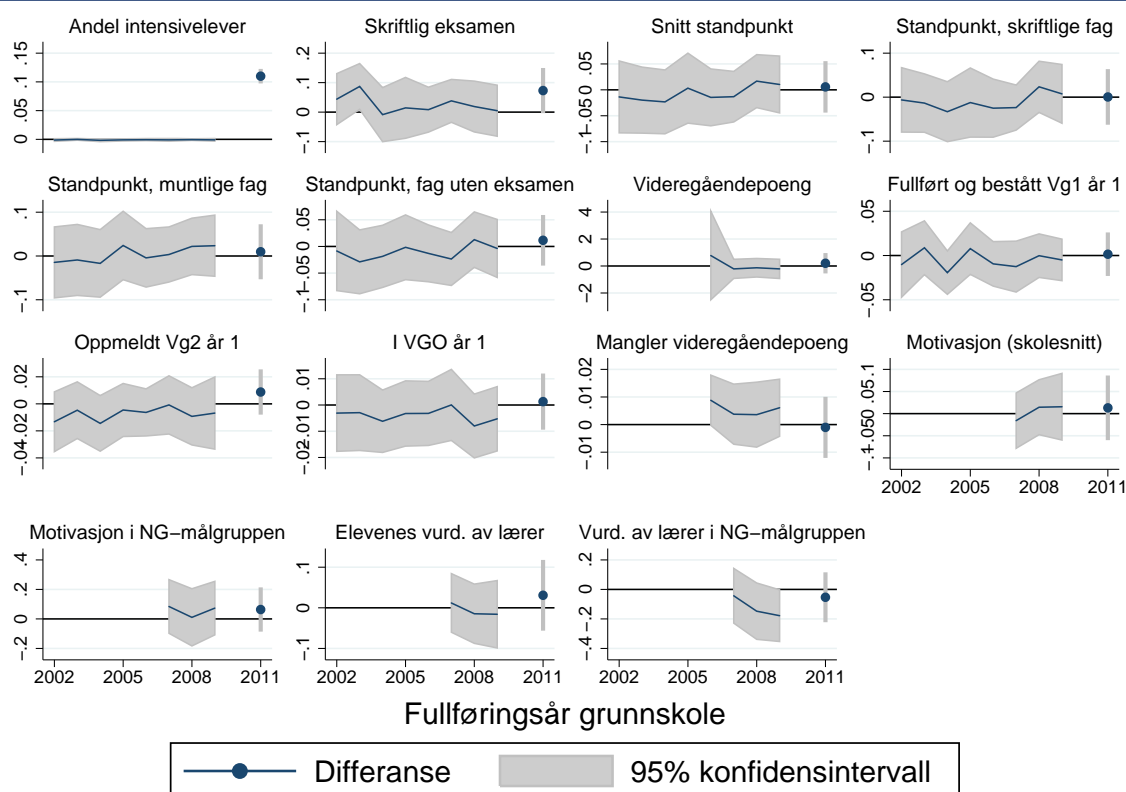
år i all hovedsak er små og ikke signifikante, og at det ikke er noen klar trend i utviklingen. Men, vi ser imidlertid også at det ikke er tegn til effekter. Forskjellene i 2011, som i motsetning til estimatene i Tabell 4.10 er relativt til 2010, ikke hele perioden 2002-2010, er som i tidligere år små og ikke signifikante.

Tabell 4.10. Effekter på flere utfall, estimert ved å sammenligne resultater i fase én og matchede fase to/tre-skoler før og etter innføringen av intensivopplæringen

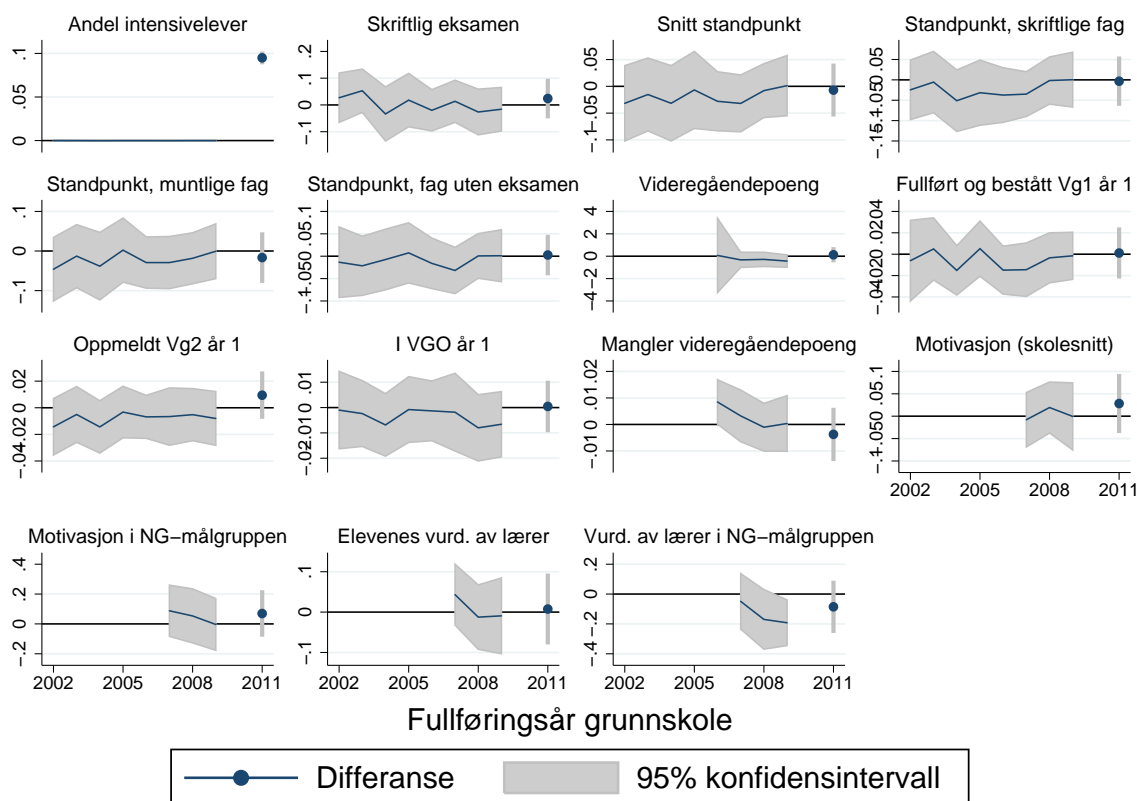
	(1)	(2)	(3)	(4)	(5)
	Ingen vektning	Matching-vekker	Elevvekker	Matching-elevvekker	Som (4) + faste skoleeffekter
Andel intensivelever	0,112** (0,008)	0,111** (0,007)	0,095** (0,004)	0,095** (0,004)	0,095** (0,004)
Fullført Vg1 år 1	0,004 (0,006)	0,006 (0,009)	0,005 (0,008)	0,006 (0,009)	0,006 (0,009)
Registrert i Vg2 år 1	0,007 (0,006)	0,015** (0,007)	0,013* (0,006)	0,017** (0,007)	0,017** (0,007)
Registrert i VGO år 1	0,002 (0,004)	0,005 (0,005)	0,005 (0,005)	0,004 (0,004)	0,004 (0,004)
Snitt skriftlig eksamen	0,033 (0,022)	0,050* (0,026)	0,033 (0,026)	0,022 (0,025)	0,023 (0,026)
Snitt standpunkt	-0,003 (0,019)	0,012 (0,019)	0,004 (0,017)	0,010 (0,018)	0,009 (0,019)
Standpunkt, skriftlige fag	-0,009 (0,027)	0,010 (0,026)	0,009 (0,021)	0,017 (0,021)	0,017 (0,022)
Standpunkt, muntlige fag	-0,008 (0,021)	0,007 (0,023)	-0,011 (0,021)	0,003 (0,024)	0,002 (0,024)
Standpunkt, fag uten eksamen .	0,007 (0,017)	0,021 (0,020)	0,019 (0,019)	0,012 (0,020)	0,011 (0,021)
Videregående poeng	-0,014 (0,363)	0,158 (0,369)	0,466 (0,324)	0,327 (0,335)	0,331 (0,345)
Mangler videregående poeng	-0,008 (0,004)	-0,006 (0,005)	-0,005 (0,004)	-0,006 (0,004)	-0,006 (0,005)
Motivasjon (skolesnitt)	0,016 (0,022)	0,010 (0,030)	0,032 (0,022)	0,026 (0,026)	0,028 (0,029)
Motivasjon i NG-målgruppen	0,050 (0,044)	0,021 (0,058)	0,047 (0,056)	0,035 (0,059)	0,041 (0,067)
Elevenes vurdering av lærer	0,062 (0,026)	0,035 (0,039)	0,028 (0,028)	0,002 (0,036)	0,003 (0,039)
Vurd. av lærer i NG-målgruppen	0,119 (0,056)	0,039 (0,072)	0,045 (0,067)	0,017 (0,073)	0,020 (0,081)

Beregnete standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

Figur 4.8. Forskjeller mellom snittresultat fase én-skoler og snittresultat matchede skoler, relativt til 2010



Figur 4.9. Forskjeller mellom snittresultat fase én-skoler og snittresultat matchede skoler, relativt til 2010, elevvektet



Når vi studerer forskjeller i utvikling mellom skoler finner vi noen tegn til bedre eksamenskarakter, og høyere motivasjon og vurdering av lærerne. Når vi sammenligner fase én-skoler med tilsvarende skoler i andre faser blir imidlertid disse effektene langt på vei borte. I stedet finner vi en effekt på overgang til Vg2, som vi ikke finner når vi ser på alle skoler og elever. En slik effekt kan igjen bare være et resultat av tilfeldig variasjon med så mange resultatmål. Ettersom de signifikante effektestimaterne vi finner er såpass lite robuste, og stort sett bare marginalt signifikante, er det mulig at de ikke er uttrykk for reelle effekter. Oppsummert finner vi ingen klare effekter av intensivopplæringen på skolenivå.

Det kan være verdt å tenke gjennom hvor stor effekter vi er i stand til å finne i våre analyser. Vi ser at presisjonen i estimatene, uttrykt ved den beregnede standardfeilen, er omtrent 0,5-1 prosentpoeng for gjennomstrømningsmålene, og omtrent 0,2-0,3 karakterpoeng for snittkarakterene. Det betyr at det er en usikkerhet i, i positiv og negativ retning, på omtrent 1-2 prosentpoeng og 0,5 karakterpoeng i resultatene våre. En effekt på 2 prosentpoeng eller 0,5 karakterpoeng kan være en ganske betydelig effekt. For eksempel er en målsetningen med Ny GIV at overgang fra Vg1 til Vg2 skal øke med 2 prosentpoeng fra 2010 til 2013.

4.4. Effekter på intensivelever og ringvirkninger, og andre former for heterogene effekter

Så langt har vi studert effekter på skolenivå, uten å skille mellom hvorvidt de enkelte elevene faktisk deltar i intensivopplæringen eller ikke, bare hvorvidt skolen er i fase én. Bakgrunnen for dette er, som tidligere diskutert, at fase én-skoler sannsynligvis i større grad er sammenlignbare med andre skoler enn intensivelever sammenlignet med andre elever. Dette betyr imidlertid at vi finner en gjennomsnittseffekt for alle elever ved fase én-skolene, og ikke er i stand til å skille mellom effekter på elever som faktisk deltar og ringvirkninger til elever på samme skole som ikke deltar. Dette skillet er interessant for å forstå eventuelle effekter av intensivopplæringen. Når vi tidligere i dette avsnittet ikke finner noen klare effekter kan

dette være et resultat av at intensivelevens prestasjoner øker (svækkes), men at endringen motsvares av at øvrige elevers prestasjoner svekkes (øker). Dette kan for eksempel være tilfellet dersom spesielt gode lærere settes til intensivopplæring, og i mindre grad underviser øvrige elever.

Et ytterligere poeng er at, dersom det ikke er vesentlige ringvirkninger, men bare effekter på intensivelevne, vil det bare være beskjedne snitteffekter. Og motsvarende, vi vil bare være i stand til å finne forholdsvis store effekter som er konsentrert til intensivelever. Når vi i forrige avsnitt fant at estimatene våre for gjennomsnittseffekten har en presisjon på 1-2 prosentpoeng, svarer det til en effekt på 10-20 prosentpoeng konsentrert blant intensivelevne. Dette vil åpenbart være en stor effekt.

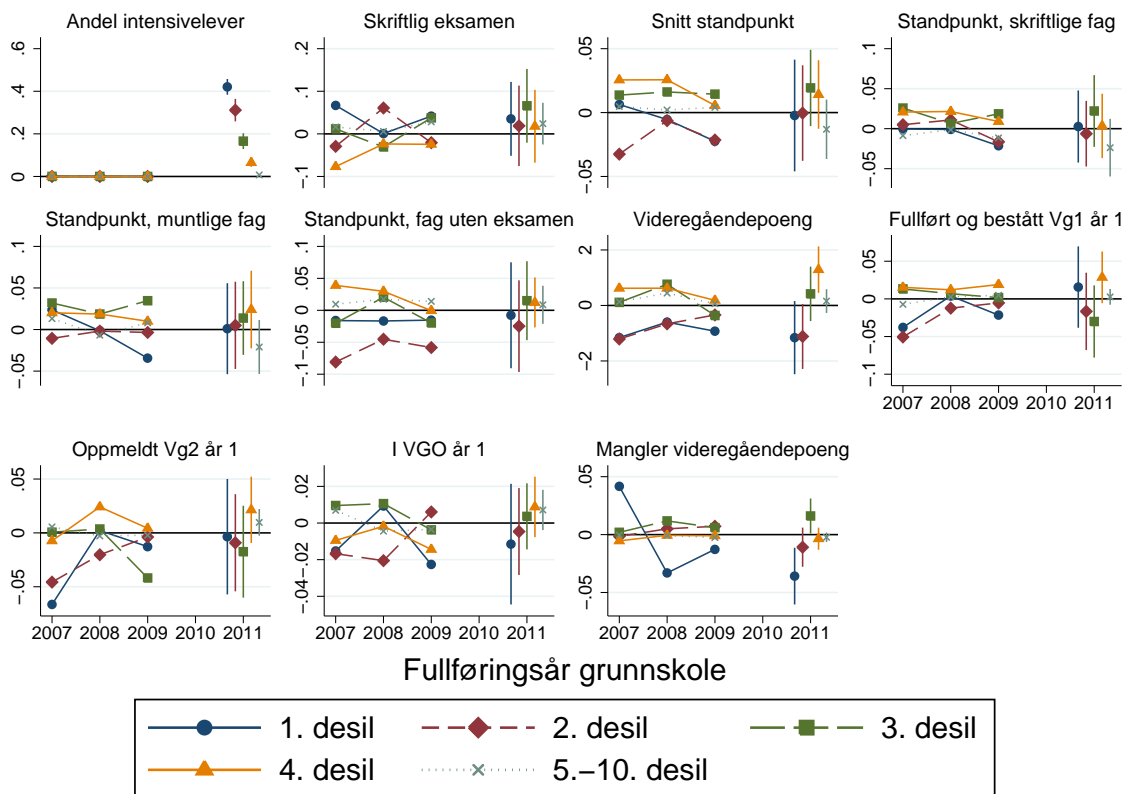
Ved å studere målgruppen for intensivopplæringen og øvrige elever hver for seg kan analyseopplegget vårt brukes til å skille mellom direkte effekter og ringvirkninger. Vi er da avhengig av en måte å definere målgruppen for intensivopplæringen på som gjør oss i stand til å finne de tilsvarende elevene ved andre skoler. Dette gjør vi ved å sammenligne elever med forskjellige karaktersnitt til termin én. Figur 4.10 viser effektestimater tilsvarende estimatene i avsnitt 4.3.1 for elever som henholdsvis er i første desil (de 10 prosent svakeste, andre desil (de neste 10 prosent), tredje desil, fjerde desil og femte eller høyere desil (de høyestpresterende 60 prosent). For hver gruppe viser vi hvordan forskjellene mellom fase én og fase to og tre samlet for grunnskolekullene 2007-2009 og 2011-kullet avviker fra tilsvarende forskjell i 2010.²³

I tråd med hva vi skulle vente ser vi en tydelig forskjell i deltakelse i intensivopplæringen mellom de forskjellige gruppene: For elevene med svakest termin én-karakterer øker deltakelse i intensivopplæringen fra 0 til 40 prosent, for de nest svakeste elevene til om lag 30 prosent, for tredje desil til omtrent 17 prosent, for fjerde desil til 7 prosent, mens det for den sterkestpresterende gruppen bare er en marginal økning. Dette betyr at eventuelle effekter på disse siste gruppene i all hovedsak vil være ringvirkninger, siden det er få deltakere vil det være lite direkte effekter. For de svakest presterende gruppene, desil én til tre, er det imidlertid også et flertall av elevene som ikke deltar i intensivopplæringen, slik at effektestimaterne fanger en blanding av direkte effekter og ringvirkninger. Dette er en følge av den uklare tilordningen av elever til deltakelse i intensivopplæringen, målgruppen for intensivopplæringen er ikke tilstrekkelig klart definert. De direkte effektene vil imidlertid være relativt sett viktigere for de svaktpresterende gruppene.

Når vi ser på effektene for de forskjellige gruppene er det imidlertid ingen klare effekter for noen grupper eller utfallsmål. De vertikale linjene angir usikkerheten i hvert av effektestimaterne, og bare i noen svært få utelukker disse intervallene ingen effekt. Vi finner tegn til svakere videregåenderesultater (videregående poeng) for desil 1 og 2, og økte videregående poeng for desil 4. Dette kan være konsistent med en negativ direkte effekt av intensivopplæringen, men positive ringvirkninger til middels sterke elever. Generelt ser det ut til at effektestimaterne for gruppene med størst andel intensivelever, første og andre desil, er svært nær null. De største effektene ser ut til å være i desil tre og fire, som begge inneholder en beskjeden andel intensivelever. Disse estimatene peker heller ikke i noen klar retning, for eksempel har desil tre den mest positive utviklingen i standpunkt- og eksamenskarakter fra grunnskolen, men den mest negative utviklingen i andelen som fullfører Vg1 og er registrert i Vg2.

²³ De nøyaktige estimatene finnes i Tabell A.6 i vedlegget. Der viser vi også resultater for gruppen uten termin én-karakter. Antall i denne gruppen og disses resultater varierer mer enn for de andre gruppene. Det kan tyde på at gruppen ikke er veldefinert over tid, vi diskuterer dette mer senere i dette avsnittet.

Figur 4.10. Forskjeller mellom snittresultat fase én og fase to og tre, relativt til 2010, etter termin én



For 2011-kullet viser de vertikale linjene et 95 prosent konfidensintervall for forskjellene. Resultatene for 2011-kullet er også forskjøvet litt langs x-aksen for å gjøre det lettere å skille mellom de forskjellige gruppene.

Ettersom estimatene er upresise er det bare i svært få tilfeller at vi kan si at et estimat med noen grad av sikkerhet tilsier at det er en effekt for den aktuelle gruppen. I enda færre tilfeller er det klare tegn til forskjeller mellom grupper. Det er imidlertid ikke slik at vi finner mange store forskjeller, og at manglende presisjon er den eneste grunnen til at vi ikke kan si at det er noen effekter. De fleste estimatene er nær null. Noen av estimatene er av en moderat størrelse, for eksempel finner vi at fullføringen av Vg1 svekkes med tre prosentpoeng.

Vi må også være forsiktige med å tolke de få statistisk signifikante forskjellene som finnes som faktiske effekter, ettersom vi også i Figur 4.10 tester vi en lang rekke effekter (mange utfallsmål for flere grupper). Da forventer vi å finne noen signifikante forskjeller, bare som et resultat av tilfeldig variasjon. Vi kan vurdere betydningen av dette ved å se hvorvidt det tilsvarende forskjeller i årene før intensivopplæringen. Figuren viser at forskjellene i 2011 ikke er spesielt store sammenlignet med forskjellene tidligere år.

I Tabell A.6 viser vi statistiske tester av hvorvidt det er signifikante forskjeller i årene 2007-2009, vi finner at det er dette for alle resultatmålene. Forskjellene er særlig store knyttet til gruppen som mangler termin én-karakterer, denne gruppen er ikke vist i Figur 4.10, for å gjøre figurene mer oversiktlig.²⁴ Andelen uten termin én-karakterer varierer over tid og mellom fylker som tyder på at sammensetningen av elevene i hver gruppe også vil variere. Dette vil ikke bare gjelde gruppen uten termin én-karakterer, når denne gruppen reduseres vil det føre til at nye elevgrupper kommer inn og forskyver forholdene mellom de øvrige gruppen. I Tabell A.6 tester vi også for om det er forskjeller mellom desil én til ti i årene 2007-2009. For flere variable er det ikke slike forskjeller, men for andre er det forskjeller, og tydeligst nettopp for videregående poeng.

²⁴ Vi viser imidlertid forskjellene for denne gruppen i Tabell A.6.

En ytterligere utfordring når vi ønsker å skille direkte effekter og ringvirkninger ved å se på elever med forskjellig termin én-karakterer, og dermed forskjellig sannsynlighet for deltakelse i intensivopplæringen, er at praksis for termin én-karakterer (hvilke elever som får og ikke får, og hvilken karakter en gitt ferdighet svarer til) kan variere mellom skoler. Praksisen, og forskjellene mellom skolene, kan også endre seg over tid. I ytterste konsekvens kan det skje som et resultat av at skolen deltar i intensivopplæringen, for eksempel dersom lærerne gir elevene dårligere (bedre) termin én-karakterer for å sikre (unngå) deltakelse i intensivopplæringen.²⁵

En mulig løsning er å dele datamaterialet etter resultater fra nasjonale prøver, 8. trinn (NP8) i stedet. Det er en sterk korrelasjon mellom NP8 og karakterer ved første termin eller avsluttet grunnskole, slik at elever som presterer svakt på NP8 i større grad også vil prestere svakt ved første termin og delta i intensivopplæringen. Sammenhengen mellom NP8 og deltakelse i intensivopplæringen er imidlertid svakere enn for termin én-karakterer. Fra Tabell A.7 ser vi at 28 prosent av elevene i desil én NP8 deltok i intensivopplæringen, mens 12 prosent av elevene i fjerde desil deltok. Når vi ser på de beregnede effektene fra Tabell A.7 er disse grovt sett konsistente med de tilsvarende i Tabell A.6. Det er generelt lite signifikante estimater, men økte videregående poeng for fjerde desil er et av de få. Det er også en nedgang i videregående poeng for første desil, men denne er ikke signifikant. Vi har NP8 bare for grunnskolekullene 2010 og 2011, dermed har vi ikke mulighet til å si om estimatene for 2011 er store eller små sammenlignet med tidligere forskjeller. Vi har derfor ikke tilstrekkelig grunnlag til å konkludere hvorvidt økningen i videregående poeng for fjerde desil er reell eller spuriøs.

Dermed er resultatene fra denne analysen konsistente med analysene innen skoler, presentert i seksjon 4.1. Der fant vi ingen effekter av intensivopplæringen. Resultatene herfra tyder på at det var fordi det ikke var noen klare effekter på resultatmålene vi studerer, verken direkte effekter på deltakerne eller ringvirkninger på elever som så vidt ikke deltar.²⁶

4.4.1. Øvrig heterogenitet i effekter

Resultatene så langt i dette avsnittet bidrar til å skille mellom direkte effekter på intensivelever og ringvirkninger. Etersom vi har sett på flere grupper med forskjellige termin én-karakter vil vi også kunne fange eventuelle heterogene effekter. Vi kan tenke oss at intensivopplæringen har forskjellige effekter på forskjellige grupper av elever, eller at det er forskjellige ringvirkninger. Fra resultatene vi har vist her ser det ut til at den samlede deltakereffekten og ringvirkningene er nær null for grupper, uavhengig av termin én-karakter.

Det kan imidlertid være annen interessant heterogenitet mellom grupper av elever med forskjellige kjennetegn. Vi kunne for eksempel tenke oss at intensivopplæringen har en sterkere effekt på gutter enn på jenter. Vi har imidlertid undersøkt en rekke forskjellige grupper separat, uten å finne klare effekter for noen av disse. Gruppene vi har undersøkt er henholdsvis gutt, jenter, innvandrere, barn av innvandrere samt elever som har foreldre med forskjellig utdanningslengde. For alle disse gruppene har vi sammenlignet resultater til elever med det aktuelle kjennetegnet som er over/under grensen for deltagelse innen Stavanger-skoler (som

²⁵ En kanskje mer sannsynlig forklaring er at registrering av termin 1 har pågått forholdsvis kort, og at dataene ikke brukes til andre formål (i motsetning til avsluttende karakterer, som har betydning for opptak til videregående). Dermed er det begrenset kvalitetssikring, og det er mulig lærere og skoler tar lett på rapporteringen. Uten at vi kan si noe om årsakene, er det tegn til endringer i praksis når det gjelder termin 1-karakterer.

²⁶ Vi har avgrenset datamaterialet på forskjellige måter, slik at vi ikke uten videre kan sammenligne resultatene fra avsnitt 4.2 med resultatene her. Når vi utelater Oslo og/eller Østfold fra datamaterialet for analysene i avsnitt 4.2.1 får vi imidlertid tilsvarende resultater. Når vi gjentar analysene på skolenivå på datamaterialet vi har brukt i avsnitt 4.2.1 eller 4.2.2 får vi i noen grad problemer med at det er få skoler. I den grad vi får noen klare resultater svarer de til resultatene vi har vist her for det mer omfattende datamaterialet.

i avsnitt 4.2.2), og vi har sammenlignet snittresultater for de forskjellige gruppene ved fase én-skoler med snittresultater for tilsvarende grupper ved andre skoler (som i avsnitt 4.3).

En annen og potensielt interessant form for heterogenitet er forskjeller i effekt mellom skoler som gjennomfører intensivopplæringen på forskjellige måter. Som en del av NOVAs kartlegging ble det gjennomført en spørreundersøkelse til skoleledere ved skoler som gjennomførte intensivopplæring (Sletten, Bakken og Haakestad, 2011). Skolelederne svarte blant annet på forskjellige spørsmål om hvordan intensivelevne ble valgt ut og hvordan intensivopplæringen ble gjennomført. Vi har gruppert skolene på forskjellige måter ut fra skoleledernes svar, og estimert effekter separat for disse gruppene. Som for grupper definert fra elevkjenntegn har vi sett separat på elever i Stavanger-skoler (som i avsnitt 4.2.2). Vi har også sammenlignet forskjellige skoler (som i avsnitt 4.3). Her er det ikke mulig å gruppere skolene som ikke er i fase én, vi har derfor separat sammenlignet grupper av fase én-skoler med forskjellig praksis med alle skoler i fase to og tre, eller med de skolene vi ved matching finner at ligner mest. Vi finner ingen klare effekter, slik vi ikke har funnet tydelige gjennomsnittseffekter eller effekter på noen grupper av elever.

Det er verdt å påpeke at selv om vi hadde funnet heterogenitet i effekter mellom skoler, ville vi ikke dette gitt tilstrekkelig grunnlag for å konkludere at en form for gjennomføring var mer eller mindre effektiv. Dette er fordi det ikke trenger å være tilfeldig hvilke skoler som gjennomfører intensivopplæringen på hvilken måte. Dermed kan vi også fange andre forskjeller mellom skolene, for eksempel hvor motiverte skoleledere eller lærere er for intensivopplæringen.

Ettersom vi studerer grupper av intensivelevne eller fase én-skolene reduseres datamaterialet, og resultatene blir mindre presise. For noen av analysene, som for kjønn, har vi fremdeles ganske store grupper. For andre grupper, som elever med forskjellig innvandringsbakgrunn eller en del grupperinger av skoler etter hvordan intensivopplæringen gjennomføres, blir antallet observasjoner for begrenset til at vi kan trekke noen konklusjoner. Vi vil komme tilbake til analyser av heterogenitet i effekter i kommende rapporter, når vi har flere årganger med intensivelever og – skoler.

5. Oppsummering

I denne rapporten har vi beskrevet elever og skoler som deltar i fase én av intensivopplæringen i Overgangsprosjektet, som er en del av Ny GIV. De første resultatene fra den pågående effektevalueringen presenteres også. Hensikten med effektevalueringen er å undersøke hvorvidt intensivopplæringen har endret resultatene til de elevene som deltar (intensivelever), og om det finnes ringvirkninger for andre elever på skoler som tilbyr intensivopplæring. I dette kapitlet oppsummerer vi resultatene vi har presentert i kapittel 3 og 4. Vi går først gjennom de enkelte resultatene, og refererer til hvor vi har presentert resultatene i mer detalj i figurer og tabeller. Til slutt i dette kapitlet vil vi forsøke å oppsummere og tolke helheten i resultatene.

Rapporten dokumenterer i avsnitt 3.1 at intensivelevne, som ventet, i utgangspunktet er svaktpresterende. Intensivelevne har en snittkarakter til første termin på 10. trinn som er omtrent ett karakterpoeng lavere enn andre elever og forskjellen i termin én-karakterene i matematikk og norsk er enda litt større. De har også hatt svake resultater tidligere. Gjennomsnittlige poeng på nasjonale prøver i 8. trinn er omtrent ett standardavvik lavere blant intensivelevne. (Ett karakterpoeng svarer også omtrent til ett standardavvik. Forskjellene i snittkarakter til termin én i 10. trinn og i nasjonale prøver i 8. trinn er altså omtrent tilsvarende.) Som ventet er forskjellige grupper som i gjennomsnitt har svake prestasjoner overrepresentert blant intensivelevne. Det er en overvekt av gutter blant deltakerne, omtrent 60 prosent av intensivelevne. Innvandrere og barn av innvandrere er overrepresentert (hver av gruppene utgjør 12-13 prosent av intensivelevne og omtrent 7 prosent av øvrige elever), og deltakernes foreldre har i snitt kortere utdanning enn foreldrene til øvrige elever ved de samme skolene (hver av foreldrene har i snitt nesten to år kortere utdanning). Det er imidlertid bare svake sammenhenger mellom de forskjellige kjennetegnene og deltakelse i intensivopplæringen når vi tar hensyn til resultater fra første termin.

Forskjeller vi kan observere mellom elever som deltar og ikke deltar er det mulig å ta hensyn til når vi skal anslå effekter. Utfordringen ligger i ulikheter vi ikke kan observere. Tilbakemeldinger fra lokale prosjektledere tyder på at motivasjon og faglig/sosial utvikling gjennom høsten på 10. trinn kan ha spilt en rolle i utvelgelsen av elever. Disse forholdene vil også påvirke elevenes videre prestasjoner, og er dermed viktig å ta hensyn til når vi ønsker å finne effekter av intensivopplæringen. Ettersom vi ikke kan kontrollere direkte for uobservert ulikhet finner vi i avsnitt 3.1.2 skoler og kommuner med elever som henholdsvis deltar / ikke deltar i intensivopplæringen, men ellers er (nesten) like. Dette er skoler og kommuner der utvelgelsen av elever ser ut til å basere seg bare på termin én-karakterer i 10. trinn, og ikke andre forhold. Det viser seg at Stavanger er den eneste kommunen der dette ser ut til å være tilfelle, men vi finner også noen enkeltskoler i andre kommuner.

I avsnitt 3.2 viser vi hvordan skoler og kommuner i fase én (som tilbyr intensivopplæring fra våren 2011) skiller seg fra skoler og kommuner i fase to og tre (som begynner med intensivopplæring i 2012 eller 2013). Kommunene som deltar i fase én er større enn andre kommuner, og i stor grad konsentrert rundt større byer. Fase én-kommunene har også i snitt høyere arbeidsledighet. Skolene i fase én er, som kommunene de ligger i, større enn øvrige skoler, med høyere utdannede foreldre og flere innvandrerforeldre. Fase én-skolene har flere elever med enkeltvedtak og færre ukvalifiserte lærere, men mindre undervisningsinnsats per elevtime, både til ordinær undervisning og spesialundervisning. Det er en tendens til sterkere eksamensresultater ved fase én-skolene, men forskjellen var tydeligere for ti år siden. Motivasjon, og i hvilken grad lærerne oppleves som motiverende (begge disse motivasjonsmålene kommer fra Elevundersøkelsen) er omtrent tilsvarende.

I kapittel 4 presenterer vi resultater fra effektevalueringen så langt. Målet for intensivopplæringen er å øke andelen som fullfører videregående opplæring (VGO). Deltakerne har imidlertid foreløpig ikke kommet så langt. I stedet studerer vi effekter på karakterer i grunnskole og videregående, tidlige mål for gjennomstrømning (fullføring av Vg1 og overgang til Vg2) samt motivasjon (fra Elevundersøkelsen, som omtalt over). I avsnitt 4.1 viser vi at det er en positiv sammenheng mellom alle disse tidlige resultatmålene og fullføring av videregående. Bedre karakterer fra grunnskolen, bedre resultater fra Vg1, fullføring av Vg1 ett år etter fullført grunnskole og å være oppmeldt på Vg2 året etter fullført grunnskole går alle sammen med høyere sannsynlighet for å fullføre VGO innen fem år. Dette tyder på at disse tidlige resultatmålene kan fange relevante effekter av intensivopplæringen.

I avsnitt 4.1 viser vi også forskjeller i resultater mellom intensivelever og andre elever. Intensivelevne presterer dårligere på et bredt spekter av resultatmål. Dette er som ventet, ettersom intensivelevne i utgangspunktet er svaktpresterende elever. Men også når vi tar hensyn til forskjeller i tidligere resultater er det noen forskjeller mellom intensivelevne og andre elever. Størrelsen og mønstret i disse forskjellene avhenger imidlertid helt av på hvilken måte vi tar hensyn til tidligere resultater. Dette tyder på at det er uobserverbare forskjeller mellom intensivelever og andre elever. Fra et effektevalueringperspektiv betyr dette at forskjeller i resultater *ikke* svarer til effekten av intensivopplæringen, men i betydelig grad kan tilskrives andre forskjeller som eksisterte før intensivopplæringen ble innført. Derfor må vi bruke andre tilnærminger, som diskutert i kapittel 2 og 3, for å finne eventuelle effekter av intensivopplæringen.

I avsnitt 4.2 studerer vi effekter ved å se på forskjeller mellom intensivelever og andre elever i fase én-skoler ved hjelp av en såkalt Regression Discontinuity design. Kort forklart går denne metoden ut på å sammenlikne resultatene til de elevene som så vidt deltar i intensivopplæringen med resultatene til de elevene som så vidt ikke deltar. For å unngå at sammenligningen påvirkes av andre forhold begrenser vi i avsnitt 4.2.1 analysene til skoler der karakterer til første termin i stor grad forklarer utvelgelsen av elever. Når vi sammenligner elever med nesten tilsvarende termin én-karakterer i 10. trinn, men som er henholdsvis rett over og rett under grensen for deltagelse, vil vi forvente at de to gruppene er omtrent like før intensivopplæringen. Det ser imidlertid ut til at også andre forhold enn termin én-karakterer har blitt vektlagt på mange skoler. Elevene som så vidt er under grensen for deltagelse i intensivopplæringen har svakere tidligere prestasjoner på nasjonale prøver og lavere utdannede foreldre enn de som så vidt er over grensen. Vi kan dermed ikke tolke resultatforskjeller mellom disse elevene som effekter av intensivopplæringen.

I avsnitt 4.2.2 begrenser vi oss til å sammenligne elever innad i Stavanger-skoler, som faktisk har hatt en praksis for utvelgelse av elever som baserer seg strengt på termin én-karakterer på 10. trinn. Som ventet finner vi ingen klare forskjeller i kjennetegn rundt grensen for deltagelse i Ny GIV. Men vi finner heller ingen forskjeller i resultater, det vil si, ingen effekter av intensivopplæringen. Ettersom denne analysen er begrenset til Stavanger-skolene baserer resultatene seg på forholdsvis få elever og skoler. Det gjør at alle resultater blir upresise. For eksempel finner vi at elever som har termin én-karakterer så vidt under grensen for deltagelse i intensivopplæringen har fedre som har ett år mindre utdanning enn elevene så vidt over, men dette er en mindre forskjell enn hva vi kan regne med som følge av tilfeldig variasjon. Videre er det blant elever som så vidt er under grensen for deltagelse 6 prosentpoeng færre som fullfører Vg1 enn så vidt over. Men også dette kan skyldes tilfeldigheter. For at vi skulle kunne avvise at en forskjell i andelen som fullfører er tilfeldig måtte den ha vært omtrent 20 prosentpoeng, hvilket er en svært stor effekt. For de forskjellige snittkarakterene er vi i stand til å avvise at mer moderate forskjeller skyldes tilfeldigheter. Effektene på disse

resultatmålene må typisk være fra 0,1-0,3 karakterpoeng for å være statistisk signifikante. Vi finner imidlertid ingen slike effekter.

I avsnitt 4.3 ser vi på forskjeller i skolegjennomsnitt. Vi ser dels på endringer innen skoler ved innføring av intensivopplæring og dels på forskjeller mellom skoler i fase én, som tilbyr intensivopplæring i 2011, og fase to og tre, som starter intensivopplæringen senere. Heller ikke med denne tilnærmingen finner vi klare effekter av intensivopplæringen.

Fordelen med å sammenligne skolegjennomsnitt er at vi unngår problemet med at elever som er utvalgt til intensivopplæring er annerledes enn andre elever på samme skole. Dette gjennomsnittet fanger både direkte effekter på intensivelevne og ringvirkninger på andre elever. Det er krevende å skille mellom direkte effekter og ringvirkninger, ettersom utvelgelsen av intensivelever i så stor grad varierer. Videre, ettersom andelen intensivelever er bare omtrent 10 prosent må effektene på denne gruppen være store for at vi skal finne dem ved å studere gjennomsnittet av alle elever.

I avsnitt 4.3.1 konstaterer vi først at størrelsen på forskjellene i snittresultater mellom skoleår er omtrent tilsvarende over tid, uavhengig av hvorvidt forskjellen fanger opp innføringen av intensivopplæringen. Vi kan dermed ikke konkludere om effekten av intensivopplæringen bare ved å se på hvordan resultatene utvikler seg når den innføres. Vi sammenligner derfor utviklingen i fase én-skolene med utviklingen i skoler i fase to og tre. Eventuelle forskjeller i utvikling kan, under bestemte forutsetninger, tolkes som effekter av intensivopplæringen. For de fleste resultatmålene vi studerer finner vi ingen forskjeller av betydning. For eksempel svarer forskjellene i de forskjellige målene for gjennomstrømming til effekter på 0,1-0,6 prosentpoeng. Det er også få forskjeller som er statistisk signifikante, dvs., som vi med en rimelig grad av sikkerhet kan si at ikke skyldes tilfeldigheter.

Noen forskjeller imidlertid verdt å kommentere nærmere. Vi finner at gjennomsnittlig eksamenskarakter ved innføringen av intensivopplæringen øker med omtrent 0,03 karakterpoeng. Dette er en beskjeden endring, men *dersom* effekten er konsentrert blant intensivelevne (en av ti elever), svarer det til en effekt på omtrent 0,3 karakterpoeng på hver deltaker, eller at omtrent en av tre elever skårer ett karakterpoeng bedre. Denne forskjellen er stor sammenlignet med forskjeller i tidligere år, og omtrent på grensen for statistisk signifikans. Vi finner også tegn til at elevene ved fase én-skolene i 2011 er mer motiverte og i større grad vurderer lærerne som motiverende enn hva de gjorde i tidligere år. Effekten ser også ut til å være sterkere for elever som oppgir at de har svake termin én-karakterer, og som dermed i større grad deltar i intensivopplæringen. På den annen side, det er bare svært svake tegn til effekter når vi sammenligner med 2010-kullet, så det kan hende effektestimaten fanger en utvikling som skjer uavhengig av intensivopplæringen. For resultatene fra Elevundersøkelsen har vi imidlertid bare data for noen få tidligere årganger. Dette gjør det vanskeligere å vurdere om forskjellene vi ser etter innføringen av intensivopplæringen er forskjellige fra hva vi ser i tidligere årganger.

Det er også ytterligere usikkerhet knyttet til disse skolesammenligningene, ettersom avsnitt 3.2 viste at skolene og kommunene i fase én er forskjellige fra skolene og kommunene i fase to og tre. I avsnitt 4.3.2 velger vi ut en sammenligningsgruppe av fase to og tre-skoler som i større grad ligner på skolene i fase én. Når vi sammenligner resultatene til 2011-kullet og utviklingen fram til 2011 med denne nye "kontrollgruppen av skoler" svekkes effektene vi fant tegn til da vi studerte forskjeller i utvikling mellom skoler. Både forskjellen i skriftlig eksamen og i motivasjon og vurdering av lærerne blir mindre. Dessuten øker den statistiske usikkerheten, slik at vi er ganske langt i fra å kunne konkludere med at det er noen effekt av intensivopplæringen.

Når vi sammenligner resultatene til 2011-kullet i fase én-skolene med sammenligningsgruppen og tidligere årganger finner vi imidlertid en økning i andelen som er registrert i Vg2 høsten året etter fullført grunnskole, det vil si på normert tid. Dette er overraskende, ettersom det ikke var tegn til noen slik effekt når vi studerte forskjeller i utvikling i avsnitt 4.3.1. Andelen registrert i Vg2 øker med omtrent 1,7 prosentpoeng (jf. Tabell 4.9). Dette er en såpass stor forskjell at det i utgangspunktet er usannsynlig at den skulle dukket opp som en tilfeldighet. Men når vi i Figur 4.9 sammenligner resultatene i 2011 med sammenligningsgruppen og 2010 (i motsetning til alle årene 2002-2010 samlet) finner vi en mye svakere effekt, som kan skyldes tilfeldigheter. Dette kan tyde på at økningen i andelen som starter i Vg2 fra fase én-skolene startet tidligere og uavhengig av intensivopplæringen.

I analysene innen Stavanger-skoler er det svært begrenset presisjon, som gjør at vi, i det minste for noen resultatmål, bare kan finne signifikante effekter dersom de er svært store. Når vi sammenligner skoler har vi et større datamateriale og mer presise resultater. I analysene som baserer seg på å sammenligne skolegjennomsnitt ville vi være i stand til finne signifikante effekter dersom de var på omtrent 1-2 prosentpoeng for gjennomstrømning eller omtrent 0,03-0,05 karakterpoeng for de forskjellige snittkarakterene. Dette er små effekter. Men ettersom intensivelevene bare utgjør omtrent 10 prosent kan dette likevel svare til store effekter i denne gruppen. I fravær av ringvirkninger vil effekten på intensivelevene være omtrent 10 ganger effekten på hele gruppen. Det betyr at vi, dersom det ikke er ringvirkninger, vanskelig kan finne effekter mindre enn 10-20 prosentpoeng eller 0,3-0,5 karakterpoeng, som er store eller moderate effekter. Det er imidlertid ikke slik at vi finner mange store, men insignifikante effekter. De fleste effektestimaterne er nær null.

Til slutt, i avsnitt 4.4, estimerer vi effekter for forskjellige grupper av elever. Analysene i avsnitt 4.2, som sammenligner elever i fase én-skoler, og analysene i avsnitt 4.3, som sammenligner fase én-skoler med fase to og tre vil i forskjellig grad fange direkte effekter på deltakerne og ringvirkninger. Når vi sammenligner elever rundt grensen for deltagelse i intensivopplæring innen fase én-skolene finner vi forskjellen mellom direkte effekter og ringvirkninger. Når vi sammenligner skoler i fase én med skoler i fase to og tre vil vi finne summen av direkte effekter og ringvirkninger. Ved å estimere effekter for elever med forskjellig termin én-karakterer kan vi, i noen grad, skille mellom effekter på deltakere i intensivopplæringen og ringvirkninger til andre elever. Dette kompliseres imidlertid av at intensivelevene ikke er valgt ut fra noe klart kriterium. Vi finner heller ingen klare effekter av intensivopplæringen, verken på elever med svake termin én-karakterer (som i stor grad deltar i intensivopplæringen) eller på elever med bedre termin én-karakterer (som i liten grad deltar). Eventuelle effekter på gruppene med stor andel intensivelever må være på omtrent 5 prosentpoeng eller 0,05-0,1 karakterpoeng for at vi skal være i stand til å finne dem.

Vi har også sett separat på forskjellige grupper av elever definert ut fra kjennetegn (kjønn, foreldres utdanning og innvandringsbakgrunn) og grupper av skoler definert ut fra skoleledernes beskrivelser av hvordan intensivelevene velges ut og intensivopplæringen gjennomføres. Vi finner ingen klare effekter for noen gruppe av elever eller skoler. En del av disse gruppene blir imidlertid små. Vi vil derfor komme mer tilbake til effekter for forskjellige grupper når vi har data for flere skoler og elever med intensivopplæring.

Oppsummert finner vi få tegn til effekter av intensivopplæringen. I kapittel 4 finner vi i noen tilfeller statistisk signifikante forskjeller mellom elever/skoler som deltar/ikke deltar. Men når vi tester et stort antall forskjeller må vi regne med at noen av disse er signifikante, selv om det ikke er noen reelle effekter. Disse forskjellene vil da være signifikante som en følge av tilfeldig variasjon. De signifikante forskjellene vi finner er få og lite systematiske (for eksempel finner vi i forskjellige analyser signifikante forskjeller for forskjellige resultatmål) at de ikke gir noe klart grunnlag for å hevde at de er uttrykk for effekter av

intensivopplæringen. Dessuten, i noen tilfeller der vi finner signifikante forskjeller i resultater er det tegn til tilsvarende sammensetningseffekter. Dette betyr at vi finner tegn til at det var forskjeller mellom elevene vi studerer allerede før intensivopplæringen, og at disse svarer omtrent til forskjellene i resultater. Et eksempel er analysene innen fase én-skoler hvor vi finner lavere karakterer, men også svakere resultater fra nasjonale prøver i 8. trinn blant elever som så vidt ligger over de beregnede grensene for deltakelse i intensivopplæringen.

Når vi ikke finner noen effekt av intensivopplæringen i grunnskole eller tidlig i videregående har dette flere mulige forklaringer. For det første er det mulig at den faktiske effekten av intensivopplæringen 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, det kan være vanskelig å raskt endre elevenes resultater 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. Det er også mulig at effekten av intensivopplæringen endres over tid, for eksempel at den bedres etter hvert som skolene får mer erfaring med intensivopplæringen.

Selv om vi ikke finner noen effekter på resultater fra slutten av grunnskolen eller tidlig i videregående betyr ikke dette at vi kan utelukke effekter på sikt, for eksempel på faktisk fullføring av VGO. Det kan også hende at det er effekter på kort sikt som våre resultatmål ikke fanger opp. For eksempel er intensivopplæringen ment å bedre elevenes grunnleggende ferdigheter. Vårt beste mål for grunnleggende ferdigheter er karakterer, her har vi ikke funnet noen effekt av intensivopplæringen. Men det er mulig at karakterer ikke fanger grunnleggende ferdigheter og at ferdighetene bedres uten at dette kommer til uttrykk i karakterene. I så fall vil vi ikke være i stand til observere effekten. I tillegg til snittkarakterene vi har presentert resultater for i kapittel 4 har vi imidlertid også sett på karakterer i matematikk og norsk hovedmål. Disse fagene uttrykker sannsynligvis i større grad grunnleggende ferdigheter i henholdsvis regning og lesing/skriving, dette er kanskje særlig tilfellet for matematikk. Når vi ikke finner effekter på intensivopplæringens karakterer i disse fagene kan det tyde på at effekten på fullføring vil vise seg å bli liten. Cortes mfl. (2013), 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.

Det er også mulig at intensivopplæringen faktisk har en effekt på resultatmålene vi studerer, uten at vi har vært i stand til å finne denne på grunn av usikkerhet i estimatene våre. Intensivopplæringen er i utgangspunktet utformet slik at det ligger til rette for evaluering, gjennom eksplisitt avgrensning innen skoler og gradvis innfasing. Vi finner imidlertid at den faktiske implementeringen er mindre oversiktlig, og forvansker evalueringen. Dette skjer på to måter: For det første velges skoler og elever til deltakelse på en måte som ikke er tilfeldig, og kanskje på bakgrunn av variable vi ikke kan observere. Dette er det vanskelig å fullt ut ta hensyn til i evalueringen, og det vil bidra til noe usikkerhet rundt enhver konklusjon.

For det andre blir vi ikke i stand til fullt å benytte de data som er tilgjengelige. Når vi analyserer forskjeller innen skoler er vi henvist til å bruke et begrenset antall skoler med klar avgrensning for deltakelse, og dermed relativt få elever. Når skolene i forskjellige faser er forskjellige mister vi informasjon når vi begrenser oss til å sammenligne like skoler. I begge tilfeller fører tapet av informasjon til at vi får mindre presise resultater enn vi ville fått dersom vi kunne studere alle skoler. For analysene innen fase én-skoler er denne effekten særlig dramatisk. Når vi begrenser oss til å studere skoler i Stavanger kan vi ikke finne effekter på fullføring av Vg1 eller overgang til Vg2 som er mindre enn omtrent 20 prosentpoeng.

Videre, når målgruppen er vagt definert er vi henvist til å studere snitteffekter på store grupper med en beskjeden andel intensivelever. Dermed blir det vanskeligere å finne direkte effekter, som også vil sammenblandes med eventuelle ringvirkninger.

Den tydeligste konklusjonen så langt fra effektevalueringen av intensivopplæringen er at tiltaksutforming er avgjørende for at det skal være mulig å evaluere og dermed lære av et tiltak. Gullstandarden er randomisering, tilfeldig fordeling til deltakelse, som gjør at forskjeller mellom grupper svarer til effekten. Konkret i dette tilfellet kunne skoler vært tilfeldig tilordnet til forskjellige faser, det ville gjort analysene mellom skoler og år både mer presise og mer troverdige. I fravær av randomisering på elevnivå er en klar avgrensing, som i Stavanger, et godt hjelpemiddel for å analysere effekter innen deltakende skoler. Dette gir også bedre muligheter for å skille mellom direkte effekter og ringvirkninger.

Til slutt, intensivopplæringen er forholdsvis vagt definert. Brevene til fylker og kommuner åpner for forskjellig organisering, og det virker også som det har vært forskjellige oppfatninger om hvordan intensivopplæringen bør eller rent praktisk kan innrettes. Dette kan ha bidratt til å svekke effekten av intensivopplæringen, for eksempel dersom uklarhet om hva som skulle gjøres førte til lavere motivasjon blant lærere eller redusert gjennomføring av intensivopplæringen. Derimot er dette ikke noe prinsipielt problem for evalueringen. Vi vil likevel kunne finne en gjennomsnittseffekt av intensivopplæringen, slik den faktisk gjennomføres. Imidlertid, dersom er ønskelig å lære mer presist hva som virker, eventuelt ikke virker, kan det også være relevant med mer presis styring av innholdet i tiltaket. Variasjon i tiltaket øker potensialet for hva en kan lære, ved at tiltaket effektivt består av flere tiltak. Men om dette potensialet skal kunne utnyttes er det viktig å ha kontroll på hvordan for eksempel skoler fordeles til forskjellige deltiltak, eller forskjellig tiltaksutforming.

Referanser

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

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

De Haan, M. (2012), “The effect of additional funds for low-ability pupils: A nonparametric bounds analysis”, upublisert manuskript, Universitetet i Oslo

Hanushek, E., J. Kain og S. Rivkin (2002), “Inferring program effects for special populations: Does special education raise achievement for students with disabilities?” *Review of Economics and Statistics*, 84(4), 584–599.

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

Lavy, V. og A. Schlosser (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 B. Sianesi (2003), “PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing” *Statistical Software Components S432001*, Boston College Department of Economics, <http://ideas.repec.org/c/boc/bocode/s432001.html>, revidert 19.07.2012

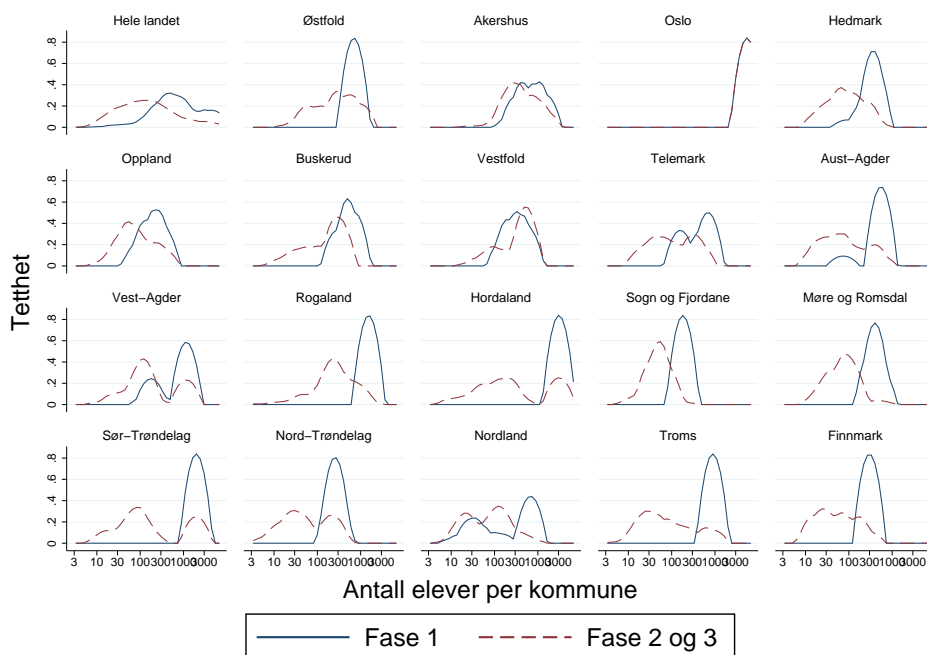
Nichols, A. (2011), “rd 2.0: Revised Stata module for regression discontinuity estimation”. *Statistical Software Components S456888*, Boston College Department of Economics, <http://ideas.repec.org/c/boc/bocode/s456888.html>, revidert 11.06.2011

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

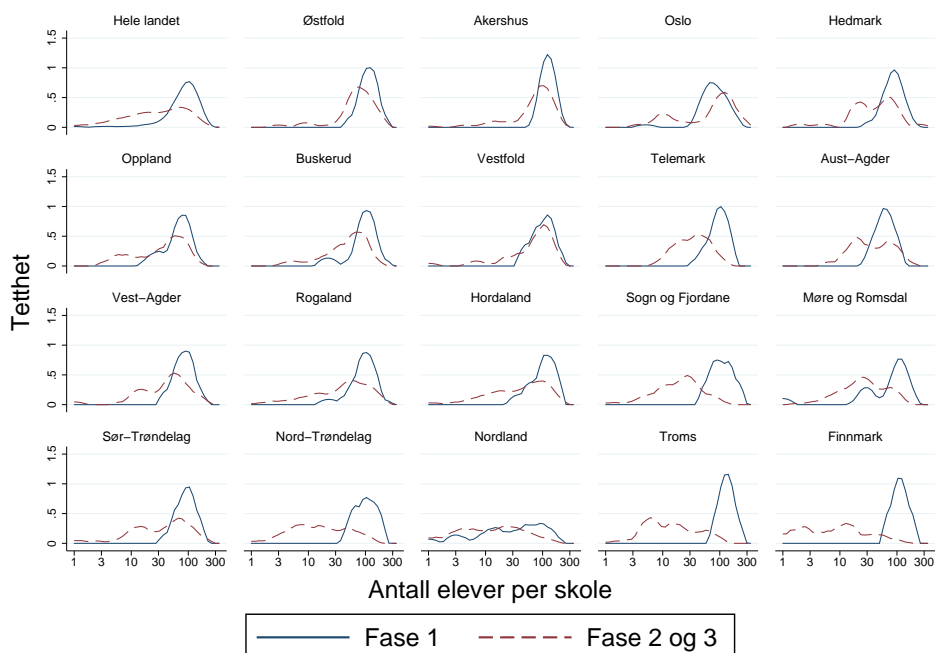
Vedlegg A: Supplerende figurer og tabeller

Fylkesspesifikke fordelinger

Figur A.1. Fordelingen av kommuner på størrelse etter første deltakelse i intensivopplæringen, per fylke

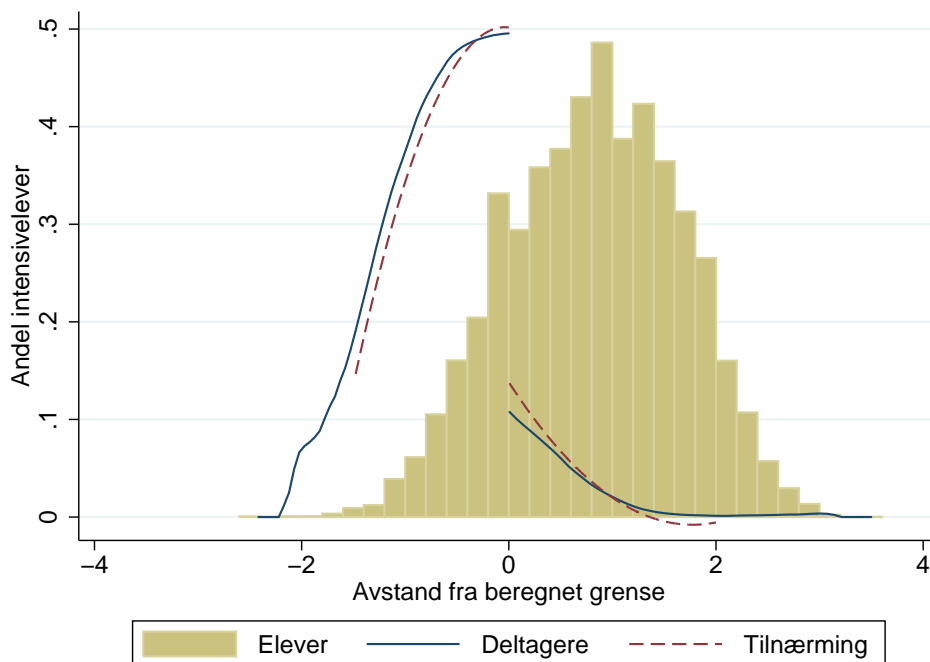


Figur A.2. Fordelingen av skoler på størrelse etter første deltakelse i intensivopplæringen, per fylke



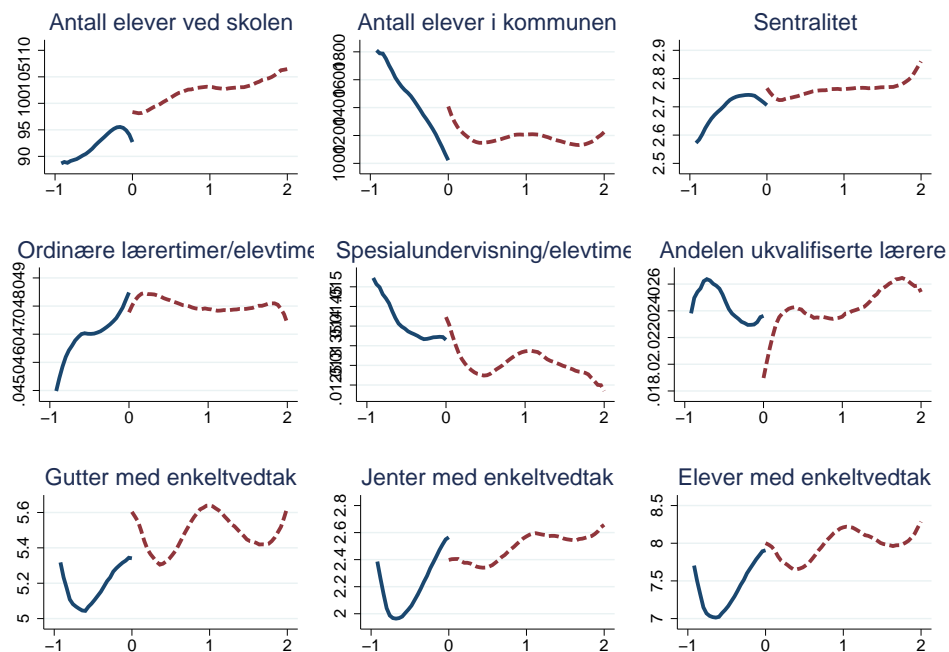
Elevkjennetegn som samvarierer med deltakelse i intensivopplæringen

Figur A.3. Tilnærming av elever rundt beregnet grense (skolenivå)



Forskjeller i skole- og kommunekjennetegn rundt tilordningsgrenser

Figur A.4. Forskjeller i skolekjennetegn etter avstand fra beregnet skolespesifikk grense



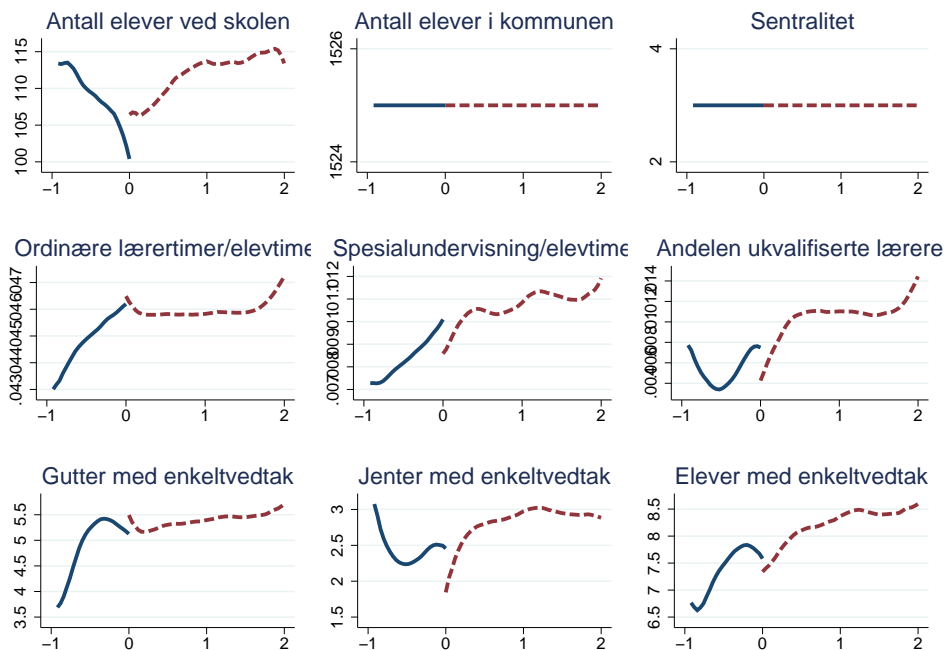
Figurene viser hvordan forskjellige kjennetegn varierer med avstand til skolespesifikk karaktergrense for deltakelse i intensivopplæringen. x-aksen viser avstanden (karakterpoeng) til denne grensen, se avsnitt 3.1.2 for mer detaljer om datamaterialet og beregning av avstand. Se fotnote 16 for mer detaljer om estimeringen av gjennomsnittlige kjennetegn for forskjellig avstand til karaktergrensen.

Tabell A.1. Forskjeller i skolekjennetegn etter avstand fra beregnet skolespesifikk grense

	(1)	(2)	(3)	(4)
	0,25	0,50	1,00	1,50
Antall elever ved skolen	-7,230 (6,246)	-5,367 (4,269)	-1,652 (3,149)	-1,855 (2,742)
Antall elever i kommunen	-347,401 (131,659)	-203,250** (85,478)	-78,143 (59,355)	-47,447 (50,538)
Sentralitet	-0,103 (0,101)	-0,062 (0,076)	-0,006 (0,057)	0,016 (0,050)
Ordinære lærertimer/elevertime	0,001 (0,001)	0,001 (0,001)	0,000 (0,001)	0,000 (0,001)
Spesialundervisning/elevertime	0,001 (0,002)	0,001 (0,001)	0,001* (0,001)	0,001** (0,001)
Andelen ukvalifiserte lærere	0,008 (0,005)	0,007* (0,004)	0,001 (0,003)	0,001 (0,003)
Gutter med enkeltvedtak	0,339 (0,544)	-0,018 (0,390)	0,220 (0,286)	0,213 (0,248)
Jenter med enkeltvedtak	0,712 (0,375)	0,399 (0,268)	0,303 (0,195)	0,285* (0,168)
Elever med enkeltvedtak	1,051 (0,833)	0,380 (0,588)	0,524 (0,426)	0,498 (0,369)
Antall observasjoner	4849	4 849	4 849	4 849
Test om noen forskjeller er signifikante (χ^2) ..	18,267	14,603	9,720	10,912
p-verdi for test	0,032	0,102	0,374	0,282

Se avsnitt 3.1.2 for mer detaljer om datamaterialet og karaktergrenser, og fotnote 16 for mer detaljer om estimeringsmetode. Beregnede standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

Figur A.5. Forskjeller i skolekjennetegn etter avstand fra beregnet kommunespesifikk grense



Figurene viser hvordan forskjellige kjennetegn varierer med avstand til kommunespesifikk karaktergrense for deltakelse i intensivopplæringen. x-aksen viser avstanden (karakterpoeng) til denne grensen, se avsnitt 3.1.2 for mer detaljer om datamaterialet og beregning av avstand. Se fotnote 16 for mer detaljer om estimeringen av gjennomsnittlige kjennetegn for forskjellig avstand til karaktergrensen.

Tabell A.2. Forskjeller i skolekjennetegn etter avstand fra beregnet kommunespesifikk grense

	(1)	(2)	(3)	(4)
	0,25	0,50	1,00	1,50
Antall elever ved skolen	-9,014 (9,738)	-6,816 (7,181)	-3,089 (5,427)	-2,969 (4,762)
Antall elever i kommunen	0,000 (.)	0,000 (0,000)	0,000 (0,000)	0,000 (.)
Sentralitet	0,000 (.)	0,000 (.)	0,000 (0,000)	0,000 (.)
Ordinære lærertimer/elevertime	-0,001 (0,001)	-0,000 (0,001)	0,000 (0,001)	0,000 (0,001)
Spesialundervisning/elevertime	0,001 (0,002)	0,002 (0,001)	0,000 (0,001)	0,000 (0,001)
Andelen ukvalifiserte lærere	-0,000 (0,004)	0,003 (0,003)	0,001 (0,002)	-0,001 (0,002)
Gutter med enkeltvedtak	-0,739 (0,807)	-0,331 (0,568)	0,065 (0,409)	0,197 (0,354)
Jenter med enkeltvedtak	0,585 (0,640)	0,519 (0,459)	0,163 (0,339)	0,042 (0,298)
Elever med enkeltvedtak	-0,154 (0,887)	0,188 (0,645)	0,228 (0,479)	0,239 (0,422)
Antall observasjoner	1450	1450	1450	1450
Test om noen forskjeller er signifikante (χ^2) ..	3,817	5,135	1,112	1,492
p-verdi for test	0,801	0,644	0,993	0,983

Se avsnitt 3.1.2 for mer detaljer om datamaterialet og karaktergrenser, og fotnote 16 for mer detaljer om estimeringsmetode. Beregnede standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

Forskjeller i utvikling mellom skoler

Tabell A.3. Resultater over tid, relativt til avgangskullet 2010 og fase to og tre (samme resultater som Figur 4.5): Karakterer fra grunnskolen

	(1)	(2)	(3)	(4)	(5)	(6)
	Andel intensiv- elever	Snitt skr. eksamen	Snitt standpunkt	Stp., skr. fag	Stp., mun. fag	Stp., fag uten eksamen
2002-kullet	0,000* (0,000)	0,046* (0,027)	0,013 (0,020)	0,002 (0,024)	0,010 (0,027)	0,034* (0,019)
2003-kullet	0,000 (0,000)	-0,006 (0,033)	0,009 (0,017)	0,005 (0,019)	0,010 (0,024)	0,017 (0,018)
2004-kullet	0,000 (0,000)	0,011 (0,040)	0,017 (0,019)	0,003 (0,022)	0,021 (0,026)	0,028 (0,018)
2005-kullet	0,000 (0,000)	0,005 (0,039)	0,015 (0,018)	-0,010 (0,021)	0,028 (0,022)	0,024 (0,018)
2006-kullet	0,000 (0,000)	0,021 (0,022)	0,000 (0,015)	-0,006 (0,016)	0,012 (0,020)	-0,002 (0,016)
2007-kullet	0,000 (0,000)	0,004 (0,028)	0,013 (0,015)	0,009 (0,015)	0,023 (0,019)	0,008 (0,016)
2008-kullet	0,000 (0,000)	-0,003 (0,022)	-0,002 (0,015)	-0,003 (0,019)	-0,006 (0,019)	0,005 (0,016)
2009-kullet	0,000 (0,000)	0,010 (0,030)	0,004 (0,016)	-0,000 (0,020)	0,007 (0,019)	0,004 (0,017)
2011-kullet	0,095* (0,004)	0,038* (0,021)	0,001 (0,015)	-0,003 (0,018)	-0,004 (0,020)	0,014 (0,018)
N	445709	424 097	443 281	438 345	439 439	442 998
R ²	0,094	0,130	0,202	0,200	0,183	0,151
F-test: Ingen forskjeller for gsk-kullene 2002-2009	0,773	0,910	0,324	0,357	0,607	1,015
p-verdi	0,626	0,508	0,957	0,942	0,773	0,424

Beregnede standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

Tabell A.4. Resultater over tid, relativt til avgangskullet 2010 og fase to og tre (samme resultater som Figur 4.5): Resultater fra videregående

	(1)	(2)	(3)	(4)	(5)
	Fullført Vg1 år 1	Reg. i Vg2 år 1	Reg. i VGO år 1	Vg-poeng	Mangler vgpoeng
2002-kullet	0,012 (0,010)	0,002 (0,007)	0,005 (0,004)		
2003-kullet	0,012 (0,009)	0,002 (0,007)	0,005 (0,004)		
2004-kullet	0,007 (0,007)	0,005 (0,006)	0,002 (0,004)		
2005-kullet	0,004 (0,007)	-0,006 (0,008)	-0,001 (0,004)		
2006-kullet	-0,013 [*] (0,006)	-0,001 (0,005)	-0,002 (0,004)	-0,042 (1,181)	0,008 ^{**} (0,003)
2007-kullet	-0,008 (0,010)	-0,006 (0,006)	0,001 (0,005)	0,069 (0,243)	0,004 (0,003)
2008-kullet	0,003 (0,009)	0,001 (0,008)	-0,002 (0,004)	0,243 (0,219)	-0,003 (0,007)
2009-kullet	0,003 (0,008)	-0,006 (0,007)	-0,005 (0,004)	-0,016 (0,201)	-0,002 (0,003)
2011-kullet	0,003 (0,007)	0,006 (0,006)	0,003 (0,004)	0,142 (0,214)	-0,004 (0,003)
<i>N</i>	445329	445329	445329	268748	278170
<i>R</i> ²	0,041	0,018	0,003	0,086	0,009
<i>F</i> -test: Ingen forskjeller for gsk-kullene 2002-2009	1,433	0,491	0,893	0,523	3,647
<i>p</i> -verdi	0,180	0,863	0,523	0,719	0,006

Beregnete standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

Tabell A.5. Resultater over tid, relativt til avgangskullet 2010 og fase to og tre (samme resultater som Figur 4.5): Resultater fra Elevundersøkelsen

	(1)	(2)	(3)	(4)
	Motivasjon (skolesnitt)	Motivasjon i NG-målgruppen	Elevenes vurd. av lærer	Vurd. av lærer i NG-målgruppen
2007-kullet	-0,026 (0,022)	0,035 (0,060)	0,005 (0,026)	-0,031 (0,066)
2008-kullet	0,008 (0,023)	-0,009 (0,062)	-0,012 (0,030)	-0,091 (0,067)
2009-kullet	-0,000 (0,023)	0,082 (0,060)	-0,017 (0,026)	-0,002 (0,057)
2011-kullet	0,029 (0,020)	0,079 (0,061)	0,027 (0,025)	0,111 [*] (0,061)
<i>N</i>	4 223	3 771	4 223	3 758
<i>R</i> ²	0,282	0,073	0,329	0,108
<i>F</i> -test: Ingen forskjeller for gsk- kullene 2007-2009	1,057	0,782	0,390	0,894
<i>p</i> -verdi	0,367	0,504	0,761	0,444

Beregnete standardfeil i parentes, * $p < 0,10$, ** $p < 0,05$

Effekter på intensivelever og ringvirkninger

Tabell A.6. Forskjeller mellom snittresultat fase én og fase to og tre, relativt til 2010, etter termin én (samme resultater som Figur 4.10)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Andel intensiv-elever	Fullført Vg1 år 1	Reg. i Vg2 år 1	Reg. i VGO år 1	Snitt skr. eksamen	Stp., skr. Snitt stp.	Stp., skr. fag	Stp., mun. fag	Stp., fag uten eksamen	Vg-poeng	Mangler vgpoeng
Mangler termin 1	0,073** (0,029)	-0,005 (0,085)	-0,025 (0,078)	-0,017 (0,070)	-0,504 (0,352)	-0,019 (0,172)	-0,058 (0,247)	0,055 (0,262)	-0,131 (0,174)	-0,117 (5,005)	0,048 (0,086)
Desil 1	0,421 (0,019)	0,016 (0,027)	-0,004 (0,027)	-0,012 (0,017)	0,035 (0,044)	-0,002 (0,022)	0,003 (0,023)	0,001 (0,028)	-0,008 (0,042)	-1,142* (0,666)	-0,035** (0,012)
Desil 2	0,312 (0,027)	-0,017 (0,023)	-0,009 (0,023)	-0,005 (0,012)	0,019 (0,048)	-0,000 (0,019)	-0,006 (0,021)	0,005 (0,027)	-0,025 (0,036)	-1,117* (0,595)	-0,011 (0,009)
Desil 3	0,166 (0,019)	-0,030 (0,024)	-0,017 (0,022)	0,004 (0,009)	0,066 (0,044)	0,019 (0,015)	0,022 (0,023)	0,014 (0,023)	0,015 (0,031)	0,418 (0,498)	0,016** (0,008)
Desil 4	0,065 (0,010)	0,029 (0,017)	0,021 (0,016)	0,009 (0,008)	0,018 (0,043)	0,014 (0,014)	0,003 (0,020)	0,024 (0,024)	0,012 (0,020)	1,281** (0,425)	-0,004 (0,005)
Desil 5-10	0,007 (0,001)	0,003 (0,005)	0,010 (0,006)	0,007 (0,006)	0,024 (0,025)	-0,013 (0,012)	-0,024 (0,018)	-0,021 (0,017)	0,009 (0,015)	0,158 (0,217)	-0,002 (0,002)
N	202756	202682	202682	202682	192720	201311	198720	199277	201189	196396	202756
R ²	0,295	0,229	0,118	0,025	0,376	0,747	0,662	0,715	0,591	0,373	0,100
F-test: Ingen forskjeller for gsk-kullene 2007-2009	69,772	2,361	3,132	1,591	1,874	1,661	1,888	1,869	2,484	3,143	3,624
p-verdi	0,000	0,000	0,000	0,040	0,008	0,028	0,008	0,009	0,000	0,000	0,000
F-test: Ingen forskjeller 2007-2010 for desil 1-10	0,856	1,069	2,337	1,516	1,217	1,001	0,405	0,824	2,018	2,768	2,338
p-verdi	0,614	0,384	0,003	0,096	0,256	0,453	0,977	0,650	0,014	0,000	0,003

Beregnete standardfeil i parentes, * p < 0,10, ** p < 0,05

Tabell A.7. Forskjeller mellom snittresultat fase én og fase to og tre, relativt til 2010, etter NP8

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Andel intensiv-elever	Fullført Vg1 år 1	Reg. i Vg2 år 1	Reg. i VGO år 1	Snitt skr. eksamen	Stp., skr. Snitt stp.	Stp., skr. fag	Stp., mun. fag	Stp., fag uten eksamen	Vg-poeng	Mangler vgpoeng
Manglerr NP8	0,072** (0,017)	-0,043 (0,056)	0,024 (0,049)	0,016 (0,027)	-0,061 (0,189)	-0,076 (0,105)	0,057 (0,150)	-0,053 (0,144)	-0,085 (0,093)	-1,687 (1,749)	0,037 (0,046)
Desil 1	0,279 (0,015)	-0,011 (0,022)	-0,011 (0,026)	0,018 (0,017)	0,039 (0,040)	-0,013 (0,026)	-0,007 (0,025)	-0,019 (0,031)	-0,038 (0,039)	-0,920 (0,806)	-0,002 (0,014)
Desil 2	0,245 (0,016)	-0,002 (0,025)	0,011 (0,019)	0,007 (0,011)	-0,004 (0,044)	-0,009 (0,033)	-0,010 (0,037)	-0,035 (0,044)	0,009 (0,031)	0,099 (0,485)	-0,019 (0,008)
Desil 3	0,173 (0,012)	0,016 (0,021)	0,005 (0,017)	0,004 (0,010)	0,013 (0,043)	0,008 (0,025)	-0,001 (0,023)	0,008 (0,031)	0,018 (0,034)	0,093 (0,483)	-0,008 (0,007)
Desil 4	0,117 (0,013)	0,045 (0,015)	0,033 (0,014)	0,004 (0,009)	-0,008 (0,035)	-0,001 (0,025)	-0,006 (0,028)	-0,031 (0,034)	0,028 (0,029)	1,129 (0,362)	-0,003 (0,008)
Desil 5-10	0,032 (0,004)	-0,005 (0,007)	-0,000 (0,007)	0,000 (0,006)	0,035 (0,022)	-0,009 (0,014)	-0,027 (0,018)	-0,013 (0,019)	0,013 (0,016)	0,064 (0,207)	-0,002 (0,002)
N	93547	93505	93505	93505	89134	92874	91472	91737	92802	90788	93547
R ²	0,171	0,136	0,058	0,006	0,351	0,474	0,511	0,447	0,293	0,240	0,088

Beregnete standardfeil i parentes, * p < 0,10, ** p < 0,05

Figurregister

3.1. Elever ved fase én-skoler: Antall elever totalt og andel intensivelever etter snitt første termin	19
3.2. Andelen deltakere i fase én-skoler etter plass i fordelingen av snittkarakterer til første termin	20
3.3. Fordeling fagkombinasjoner	21
3.4. Tilordningsgrenser på skole- og kommunenivå. Persentilene som best forklarer tilordning av elever til intensivopplæringen på henholdsvis skole- og kommunenivå ..	22
3.5. Grad av streng tilordning, skole- og kommunenivå	23
3.6. Sannsynlighet for å delta i intensivopplæringen, fordelt etter kategorier av "strenghet", skole- og kommunenivå	23
3.7. Sannsynlighet for å delta intensivopplæringen i skoler i Stavanger etter snittkarakter til første termin, grense satt på kommunenivå	24
3.8. Geografisk fordeling av kommuner med første deltakelse i fase én, to og tre	27
3.9. Fordelingen av kommuner på størrelse etter første deltakelse i intensivopplæringen ..	28
3.10. Fordelingen av skoler på størrelse etter første deltakelse i intensivopplæringen	28
4.1. Elevsammensetning rundt grensen for deltakelse. Forskjeller i elevkjennetegn etter avstand fra beregnet skolespesifikk grense.....	34
4.2. RD-effekter av intensivopplæringen: Forskjeller i utfall etter avstand fra beregnet skolespesifikk grense	35
4.3. Forskjeller i elevkjennetegn etter avstand fra beregnet kommunespesifikk grense. Skoler i Stavanger	37
4.4. Forskjeller i resultater etter avstand fra beregnet kommunespesifikk grense. Skoler i Stavanger	38
4.5. Forskjeller mellom fase én- og fase to og tre-elever over tid	40
4.6. Forskjeller mellom fase én- og fase 2- og 3-elever over tid – utfall fra Elevundersøkelsen	41
4.7. Resultater over tid, relativt til avgangskullet 2010 og fase to og tre.....	43
4.8. Forskjeller mellom snittresultat fase én-skoler og snittresultat matchede skoler, relativt til 2010	46
4.9. Forskjeller mellom snittresultat fase én-skoler og snittresultat matchede skoler, relativt til 2010, elevvektet	47
4.10. Forskjeller mellom snittresultat fase én og fase to og tre, relativt til 2010, etter termin én	49
A.1. Fordelingen av kommuner på størrelse etter første deltakelse i intensivopplæringen, per fylke.....	59
A.2. Fordelingen av skoler på størrelse etter første deltakelse i intensivopplæringen, per fylke	59
A.3. Tilnærming av elever rundt beregnet grense (skolenivå)	60
A.4. Forskjeller i skolekjennetegn etter avstand fra beregnet skolespesifikk grense	60
A.5. Forskjeller i skolekjennetegn etter avstand fra beregnet kommunespesifikk grense ...	61

Tabellregister

3.1. Hvem var intensivelevne i fase én? Sammenligning av kjennetegn ved intensivelever og andre elever i fase én-skoler	18
3.2. Sammenheng mellom utvalgte elevkjennetegn og deltakelse i intensivopplæringen, elever ved fase én-skoler og gitt første termin-karakterer	26
3.3. Forskjeller mellom fase én og fase to og tre skoler før og etter matching	29
4.1. Alternative utfallsmål. Sammenhenger mellom fullføring etter 5 år og tidligere resultatmål. Grunnskolekullene 2002-2004	31
4.2. Resultater i grunnskole- og Vg1, ubetingede og betingede forskjeller mellom intensivelever og andre elever.....	32
4.3. Elevsammensetning rundt grensen for deltakelse. Forskjeller i elevkjennetegn etter avstand fra beregnet skolespesifikk grense.....	35
4.4. RD-effekter på ulike utfallsmål. Forskjeller i utfall etter avstand fra beregnet skolespesifikk grense	36
4.5. Forskjeller i elevkjennetegn etter avstand fra beregnet kommunespesifikk grense. Skoler i Stavanger	37
4.6. Forskjeller i resultater etter avstand fra beregnet kommunespesifikk grense. Skoler i Stavanger	39
4.7. Effektestimater fra sammenligning av utvikling over tid, fase én i 2011 relativt til tidligere avgangskull og fase to og tre	43
4.8. Effekter på flere utfall, estimert ved sammenligning av resultater for 2011-kullet i fase én-skoler og matchede skoler.....	44
4.9. Effekter på fullføring av Vg1 til normert tid, estimert ved å sammenligne utviklingen i fase én-skoler og matchede skoler i fase to/tre	45
4.10. Effekter på flere utfall, estimert ved å sammenligne resultater i fase én og matchede fase to/tre-skoler før og etter innføringen av intensivopplæringen	46
A.1. Forskjeller i skolekjennetegn etter avstand fra beregnet skolespesifikk grense	61
A.2. Forskjeller i skolekjennetegn etter avstand fra beregnet kommunespesifikk grense ...	62
A.3. Resultater over tid, relativt til avgangskullet 2010 og fase to og tre (samme resultater som Figur 4.5): Karakterer fra grunnskolen	62
A.4. Resultater over tid, relativt til avgangskullet 2010 og fase to og tre (samme resultater som Figur 4.5): Resultater fra videregående	63
A.5. Resultater over tid, relativt til avgangskullet 2010 og fase to og tre (samme resultater som Figur 4.5): Resultater fra Elevundersøkelsen.....	63
A.6. Forskjeller mellom snittresultat fase én og fase to og tre, relativt til 2010, etter termin én (samme resultater som Figur 4.10).....	64
A.7. Forskjeller mellom snittresultat fase én og fase to og tre, relativt til 2010, etter NP8 ..	64

