



Effekten av digitale ordbøker i norsk skole

Bedre språk eller tomme ord?

Fredrik Haga og Espen Thorsen Stræte

Veileder: Kjell G. Salvanes

Masterutredning i economics

NORGES HANDELSHØYSKOLE

Dette selvstendige arbeidet er gjennomført som ledd i masterstudiet i økonomi- og administrasjon ved Norges Handelshøyskole og godkjent som sådan. Godkjenningen innebærer ikke at Høyskolen eller sensorer inntår for de metoder som er anvendt, resultater som er fremkommet eller konklusjoner som er trukket i arbeidet.

Sammendrag

I denne oppgaven har vi forsøkt å finne ut om bruk av digitale ordbøker fra iFinger fører til bedre karakterer i språkfag i norsk skole. iFinger har blitt innført på videregående skoler og ungdomsskoler ved forskjellige tidspunkter de siste 8 årene. Dette har latt oss analysere effekten av iFinger med differences-in-differences-metoden og Granger-testen. Vi har tatt for oss tre populasjoner: Ungdomsskoleelever, elever ved påbygg og yrkesfag på VGS og elever ved studiespesialiserende linjer på VGS.

For ungdomsskoleelever har vi ikke funnet noen effekter. Vi tror imidlertid dette kan endre seg når ungdomsskolen nærmer seg VGS sitt nivå av digitalisering. Vi har ikke kunnet anvende vår metode til å si noe om effekten på engelsk i ungdomsskolen.

For påbygg- og yrkesfagelever har vi heller ikke funnet noen effekt av iFinger. Disse elevene bruker digitale ordbøker lite til tross for at det er god tilgang til datamaskiner i den videregående skolen. Vi har ikke grunn til å tro at iFinger skal få en effekt for denne gruppen i fremtiden.

For elever på studiespesialiserende linjer har vi funnet betydelige forbedringer av karakterer i fremmedspråk ved bruk av iFinger i forhold til Ordnett. For hovedmål og sidemål har vi ikke kunnet anvende vår metode til å si noe om effekten av iFinger. Vi finner at iFinger forbedrer karakterene i fremmedspråk med omtrent 0,3 standardavvik, hvilket representerer en karakterforbedring på mellom 0,14 og 0,18 karakterpoeng på en skala fra 1 til 6. Tatt i betraktning kostnaden ved programvaren synes iFinger å være en god investering for skolene.

Disse funnene er klarere enn det man har funnet i tidligere forskning på effekten av utdanningsteknologi i skolen. Denne forskningen er imidlertid gjort i land som er mindre digitaliserte enn Norge. I tillegg ser vårt studie på en mer konkret behandling enn mange andre studier.

Innholdsfortegnelse

1	Introduksjon	9
1.1	Motivasjon.....	9
1.2	Problemstilling.....	10
1.3	Digitale ordbøker i norsk skole.....	10
1.4	Eksisterende forskning på digitalisering av utdanning	12
1.4.1	Effekten av generelle IKT-investeringer i utdanning	12
1.4.2	Effekten av individuelle skole-PCer	14
1.4.3	Effekten av målrettet programvare i utdanning.....	15
1.4.4	Oppsummering av eksisterende forskning på området	16
2	Metode.....	17
2.1	Kausalitet.....	17
2.2	Differences-in-differences (DD)	17
2.2.1	DD på regresjonsform.....	18
2.2.2	Local Average Treatment Effect (LATE)	19
2.2.3	Forutsetninger for DD.....	19
2.2.3.1	Felles trender	19
2.2.3.2	Omitted Variable Bias (OV) og entydig treatment.....	20
2.3	Granger-test	21
2.3.1	Sjekk av felles trender	21
2.3.2	Kausalitet.....	22
2.4	Clusterede standardfeil	22
3	Databeskrivelse	24
3.1	Populasjon	24
3.2	Datainnsamling	24
3.3	Avhengige variabler: Eksamenskarakterer.....	26
3.3.1	Ungdomsskole	28
3.3.2	Yrkesfag og påbygg VGS.....	29
3.3.3	Obligatoriske fag studiespesialiserende VGS.....	30
3.3.4	Tredjespråk VGS	32
3.4	Treatment-variabel: iFinger.....	33

4	Deskriptiv statistikk	34
4.1	Ungdomsskole	34
4.1.1	Tolkning av treatment-variabelen for US	34
4.1.2	Deskriptiv statistikk av US-utvalg til analyse	35
4.2	Videregående skole.....	37
4.2.1	Tolkning av treatment-variablene for VGS	37
4.2.2	Unntakene Finnmark og Sogn og Fjordane	40
4.2.3	Deskriptiv statistikk av VGS-utvalg til analyse	41
5	Statistisk analyse	47
5.1	Test av "felles trend"-forutsetningen	47
5.1.1	Felles trender: Ungdomsskole	48
5.1.2	Felles trender: Videregående skole	50
5.2	Modellenes estimater på effekten av iFinger	53
5.2.1	Differences-in-differences	53
5.2.1.1	DD: Ungdomsskole.....	53
5.2.1.2	DD: Yrkesfag og påbygg VGS	54
5.2.1.3	DD: Studiespesialiserende VGS	54
5.2.2	Granger-test	55
5.2.2.1	Granger-plot: Engelsk studiespesialiserende	57
5.2.2.2	Granger-plot: Tysk 2.....	58
5.2.2.3	Granger-plot: Spansk 2	59
6	Kausale tolkninger i modellene	60
6.1	Intern validitet	60
6.1.1	Omitted variable bias	60
6.1.2	LATE	62
6.1.3	Utvalg og utvalgsstørrelse	63
6.1.4	Multipelt sammenligningsproblem	65
6.2	Kausale tolkninger: Ungdomsskole.....	66
6.3	Kausale tolkninger: Yrkesfag og påbygg VGS	66
6.4	Kausale tolkninger: Obligatoriske fag studiespesialiserende VGS	67
6.5	Kausale tolkninger: Tredjespråk VGS.....	68
6.6	Forskjellene mellom fremmedspråkene	70
6.7	Oppsummering av kausale tolkninger	72
7	Praktiske implikasjoner og konklusjon	73

7.1	Implikasjonene av Ordnett som kontrollgruppe	73
7.2	Samfunnsøkonomiske implikasjoner	74
7.3	Ekstern validitet	75
7.4	Konklusjon	76
8	Bibliografi	78
9	Appendiks	83
9.1	Granger-plots: Ungdomsskole	83
9.2	Granger-plots: Videregående skole	84
9.3	Stata	88

Liste over figurer

Figur 2.A – Difference-in-difference	18
Figur 3.A – Karakterutvikling engelsk, hovedmål og sidemål US	29
Figur 3.B – Karakterutvikling engelsk yrkesfag, hovedmål yrkesfag, hovedmål påbygg og sidemål påbygg VGS	30
Figur 3.C – Karakterutvikling engelsk, hovedmål og sidemål studiespesialiserende VGS	31
Figur 3.D – Karakterutvikling fransk, spansk og tysk tredjespråk VGS	32
Figur 4.A – Antall observasjoner av eksamenskarakterer betinget på treatment-tilstand US	35
Figur 4.B – Fylkesavtaler for digitale ordbøker over tid VGS	38
Figur 4.C – Antall videregående skoler per fylke	39
Figur 4.D – Antall observasjoner av eksamenskarakterer betinget på treatment-tilstand VGS	42
Figur 5.A – Granger-plot: Engelsk studiespesialiserende VGS	57
Figur 5.B – Granger-plot: Tysk 2 VGS	58
Figur 5.C – Granger-plot: Spansk 2 VGS	59
Figur 6.A - Illustrasjon av observasjonsnivå	64
Figur 9.A – Granger-plot: Engelsk US	83
Figur 9.B – Granger-plot: Hovedmål US	84
Figur 9.C – Granger-plot: Sidemål US	84
Figur 9.D - Granger-plot: Engelsk yrkesfag VGS	85
Figur 9.E - Granger-plot: Hovedmål yrkesfag VGS	85
Figur 9.F - Granger-plot: Hovedmål påbygg VGS	86
Figur 9.G - Granger-plot: Sidemål påbygg VGS	86
Figur 9.H - Granger-plot: Hovedmål studiespesialiserende VGS	87
Figur 9.I - Granger-plot: Sidemål studiespesialiserende VGS	88

Liste over tabeller

Tabell 4.A – Deskriptiv karakterstatistikk: Ungdomsskole	36
Tabell 4.B – Deskriptiv karakterstatistikk: VGS.....	43
Tabell 5.A – Granger-test: Ungdomsskole.....	49
Tabell 5.B – Granger-test: Yrkesfag og påbygg VGS	51
Tabell 5.C – Granger-test: Studiespesialiserende VGS	52
Tabell 5.D – DD: Ungdomsskole	53
Tabell 5.E – DD: Yrkesfag og påbygg VGS	54
Tabell 5.F – DD: Studiespesialiserende VGS	55
Tabell 6.A – Oppsummering regresjonsresultater for fremmedspråk VGS.....	71

1 Introduksjon

1.1 Motivasjon

Det norske skolesystemet har gjennomgått en omfattende digitalisering de siste tiårene. Som følge av dette har norske skoleelever blitt blant Europas mest privilegerte når det kommer til å ha en digitalisert utdanning. Allerede i 2013 hadde elever på norske videregående skoler én PC per elev, mens norske ungdomsskoler hadde én PC for hver tredje elev. Ingen andre land i Europa hadde bredere dekning av PCer i skolen (Europakommisjonen, 2013).

Norge er samtidig det landet som bruker tredje mest penger på utdanning per elev i OECD (Utdanningsdirektoratet, 2014b). I 2013 utgjorde utgifter til grunnskole og videregående skole 91 milliarder kroner (Utdanningsdirektoratet, 2014b), hvilket utgjorde 6,8% av de totale utgiftene til offentlig forvaltning (Statistisk Sentralbyrå, 2014). En stor andel av midlene brukes på digitalisering; det ble estimert at norske kommuner og fylker ville bruke over 3 milliarder kroner på IKT-utstyr mellom 2006 og 2009 (Mjelva & Hoaas, 2008).

Når Norge bruker milliarder av kroner på digitalisering av skolen, går det på bekostning av andre tiltak som flere lærere, mer etterutdanning, bedre skolebygg eller varm skolelunsj. En annen implikasjon av digitaliseringen er at stadig hyppigere bruk av digitale hjelpemidler går på bekostning av andre former for undervisning; 70% av elevene ved studiespesialiserende linjer svarte i 2014 at de brukte datamaskin i norskundervisningen "alltid" eller "ofte". På ungdomstrinnet var tilvarende andel 27% (Språkrådet, 2014). Man kan spørre seg om dette er hensiktsmessig bruk av både penger og undervisningstid.

I et samfunnsøkonomisk perspektiv er det av interesse hvordan knappe ressurser anvendes. Dersom man gjør IKT-investeringer, burde dette føre til bedre læring enn andre investeringer i skolen. Dersom man bruker mer tid på PC i undervisningen, burde dette føre til bedre læring enn alternative former for undervisning. Vi vil i denne oppgaven prøve å finne ut om investeringer i det spesifikke virkemiddelet digitale ordbøker er en *god* anvendelse av ressurser. Dette er av verdi for beslutningstakere i skolen som står overfor valg mellom forskjellige investeringstiltak og undervisningsformer, gitt budsjetter og antall

skoletimer tilgjengelig.

Vi har fått i oppdrag å skrive denne oppgaven av iFinger, et norskeid EdTech¹-selskap. Et aktuelt tema i samfunnsdebatten er hvilke næringer som skal ta over sysselsetting og verdiskapning i Norge etter oljeprisfallet i 2014. En potensiell vekstnæring er nettopp EdTech. Blant annet var denne næringen viet en egen dag under Oslo Innovation Week 2015 (Oslo Innovation Week, 2015). I mai 2015 hadde norske EdTech-selskaper allerede over 60 millioner brukere globalt (IKT-Norge, 2015). I så måte kan denne oppgaven, ved siden av å gi faktagrunnlag til beslutningstakere i skolen, belyse verdien av en norsk kunnskapsnæring i vekst.

1.2 Problemstilling

Vi vil i denne oppgaven prøve å besvare følgende problemstilling:

Fører bruk av digitale ordbøker fra iFinger til bedre karakterer i språkfag i norsk skole?

Vi ønsker å presisere at det ikke har noen implikasjoner for objektiviteten til denne besvarelsen at vi skriver på oppdrag fra iFinger. Vi ønsker naturligvis å opprettholde en høy forskningsmessig integritet. Utover å hente informasjon fra deres faktureringsystemer og ansatte, har vi ikke jobbet sammen med iFinger om utformingen av oppgaven. Vi har også vært tydelige overfor dem om at funnene i oppgaven kan være negative for selskapet, og at det ikke vil påvirke vår besvarelse på noen måte.

1.3 Digitale ordbøker i norsk skole

I denne oppgaven brukes begrepet *digitale ordbøker* om programvare som lar brukere søke i kvalitetssikrede, tradisjonelle ordbøker ved hjelp av datamaskiner. Dette innebærer at brukeren får tilgang til en helhetlig språktjeneste med oversettelser, ordforklaringer, eksempler, ordklasseinformasjon og synonymer som kan brukes på eksamen. I tillegg kan en

¹ EdTech står for Educational Technology og favner fagområdet hvor pedagogikk og teknologi møtes (Universitetet i Oslo, 2015).

få oppslag når en søker på bøyde ord, i motsetning til i en papirordbok. I lys av dette tror vi digitale ordbøker kan forbedre brukernes språklige fremstillingsevne, forståelse og rettskrivningsferdigheter.

Her går for eksempel ikke Google Translate under vår definisjon av digitale ordbøker, fordi tjenesten ikke er helhetlig som nevnt over og ikke kan brukes på eksamen. Tilsvarende vil ikke det å søke etter ord på internett for å finne en enkel oversettelse eller ordforklaring kvalifiseres som tilgang til digitale ordbøker.

TNS Gallup gjorde i 2014 undersøkelsen "Digitale ordbøker i bruk" for Språkrådet. Her svarte 16 % av elevene på ungdomstrinnet at de søker i digitale ordbøker ukentlig. For VGS studiespesialiserende retning var tilsvarende tall 33%, og på VGS yrkesfag var tallet 20 %. Andelen elever som sa at de sjelden eller aldri brukte digitale ordbøker var 51% på ungdomstrinnet, 18 % på VGS studiespesialiserende og 35 % på VGS yrkesfag (Språkrådet, 2014).

De aller fleste videregående skoler i Norge har avtale med enten iFinger eller Ordnett om tilgang til digitale ordbøker. En del ungdomsskoler har de samme produktene, selv om dekningen av digitale ordbøker her er betydelig lavere. Vi vil nå redegjøre for de to største aktørene i det norske markedet for digitale ordbøker.

Ordnnett

Ordnnett er eid av Kunnskapsforlaget og er leverandør av digitale ordbøker til norsk skole. Med Ordnnett kan man søke i ordbøker fra Kunnskapsforlaget og Oxford University Press. Da Ordnnett ble lansert i 2004, var det som navnet tilsier, en nettbasert tjeneste. I 2007 lanserte Kunnskapsforlaget programvaren Ordnnett Pluss, som lar brukere søke uten internettilgang. Videregående skoler som har inngått avtale med Ordnnett bruker hovedsakelig Ordnnett Pluss (Ordnnett, 2014).

iFinger

iFinger er et norsk programvareselskap som leverer digitale ordbøker. Programvaren lastes ned for bruk uten internetttilgang. Gjennom iFinger kan man søke i ordbøker fra forlag Cappelen Damm, Samlaget, Vega Forlag og Collins.

1.4 Eksisterende forskning på digitalisering av utdanning

Bulman og Farlie (2015) oppsummerer empirisk forskning rundt teknologi og utdanning i IZA-artikkelen *Technology and Education: Computers, Software and the Internet*. Studiene diskutert her viser tvetydige resultater omkring effekten av teknologi på prestasjoner i utdanning. Hovedtendensen er imidlertid at det ikke finnes noen sammenheng. Bulman og Farlie presiserer at årsaken til de fragmenterte resultatene kan forklares med at både type treatment, lengden av treatment, intensiteten av treatment, klassetrinn som treates eller hvilke fag som måles varierer mellom studier.

Vi vil nå redegjøre for studier som er omtalt i den nevnte artikkelen av Bulman og Farlie, inndelt i i tre deler; (1) effekten av generelle IKT-investeringer i utdanning, (2) effekten av individuelle skole-PCer og (3) effekten av målrettet programvare i utdanning.

1.4.1 Effekten av generelle IKT-investeringer i utdanning²

Få studier finner at generelle IKT-investeringer har positive effekter på akademiske prestasjoner. Dette til tross for at de fleste studiene finner at IKT-investeringer øker bruken av datamaskiner i skolen, og at treatment ofte kommer som supplement til tradisjonelle læringsmetoder. Det kan da spekuleres i om IKT-basert undervisning som et substitutt for tradisjonell undervisning ville gitt negative effekter (Bulman & Fairlie, 2015). Bulman og Farlie bemerker også at få av studiene spesifiserer *hvordan* IKT-investeringene ble brukt i utdanningen.

² Med generelle IKT-investeringer menes investeringer i informasjons- og kommunikasjonsteknologier som data-hardware og internettforbindelser (Bulman & Fairlie, 2015). Dette omfavner som nevnt ikke investeringer i individuelle skole-PCer.

Fuchs og Woessmann (2004) undersøker korrelasjonen mellom datatilgang i skoler (og i private hjem) og prestasjonen på PISA-tester. Studiet tok for seg 32 land. Når de kontrollerer for familiebakgrunn og skolekarakteristikker, som trolig skaper bias i estimatene om de er utelatt fra regresjonene, finner de ingen signifikante sammenhenger (Bulman & Fairlie, 2015).

Goolsbee og Guryan (2002) undersøkte med differences-in-differences-metoden om effekten av et subsidie til internettinvesteringer på skoler i California førte til bedre akademiske utfall. Subsidien var satt opp slik at skoler med flere fattige elever fikk mer penger. Det er verdt å merke seg at dette førte til både en inntekt- og substitusjonseffekt; skoler som fikk større subsidier hadde samtidig insentiver til å forskyve mer midler til tradisjonell undervisning (Bulman & Fairlie, 2015). Til tross for dette fant de kun at subsidien førte til økt internettbruk, og ikke hadde noen signifikant effekt på akademiske utfall.

Belo, Ferreira og Rahul (2014) undersøkte implikasjonene av økt bredbåndbruk på skoler i Portugal. Dette gjorde de ved å bruke nærhet til internettleverandøren som et instrument for kvaliteten på internettilinjen. Resultatene viste at økt bredbåndbruk førte til negative effekter på akademiske resultater. Forfatterne tror at økt bredbåndbrukt generer en distraksjon for elevene.

Cristia, Czerwonko og Garofalo (2014) undersøkte innføringen av et omfattende IKT-program i Peru mellom 2001 og 2006. Effekten av programmet, som gav et utvalg skoler hardware og software som Microsoft Office, ble undersøkt med differences-in-differences-metoden. Forfatterne fant ingen effekter av IKT-programmet, verken på hvor mye elever måtte gjennomføre fag på nytt, droppet ut av skolen eller gikk videre til neste skoletrinn.

Det finnes imidlertid unntak fra disse ikke-signifikante funnene. Machin, McNally og Silva (2007) undersøkte om økte IKT-bevilgninger i England førte til bedre karakterer. Bakgrunnen for studiet var at IKT-bevilgninger til skoledistrikter både ble økt og allokert på en ny måte fra 2001. Denne endringen gjorde at mange distrikter fikk endret størrelsen på sine bevilgninger. Det gav opphav til en eksogen variasjon som lot forfatterne avdekke en kausal effekt med instrumentell variabel-metoden. De fant at økte IKT-bevilgninger førte til en

signifikant økning i engelskkarakterer på barneskolenivå. De fant også en svakt signifikant økning i naturfagkarakterer, men ingen signifikant endring i matematikkarakterer. Machin, McNally og Silva presiserer imidlertid at IKT-midlene ble rettet mot distrikter som hadde høyest marginalnytte av dem, noe som er en svakhet ved studiet.

1.4.2 Effekten av individuelle skole-PCer

I motsetning til studiene over, ser vi her på studier som spesifikt tar for seg innføringen av individuelle skole-PCer. Denne formen for treatment er sannsynligvis mye mer intens (Bulman & Fairlie, 2015). Empirien viser at slike tiltak syntes å ha noe mer positive effekter på akademiske utfall.

Maine Education Policy Research Institute (2007) undersøkte en av de første storskala innføringene av individuelle skole-PCer. I 2002 fikk samtlige 7. og 8.-klassinger, samt lærere, i delstaten hver sin skole-PC. Til analysen ble tester avlagt for hånd i år 2000 sammenliknet med tester avlagt på PC i 2005. De finner at gjennomsnittsscoren på en skrivetest øker med 0,33 standardavvik, en differanse som er signifikant med en t-test. De finner ingen tilsvarende økning i en generell "assessment-test". Videre finner de med en ANOVA-test at forbedringen på skrivetesten var størst for elever som rapporterte at de brukte skole-PC til alle skriveoppgaver på skolen. De sammenliknet også gjennomsnittsscoren til et utvalg elever som tok skrivetesten i 2005 for hånd mot et utvalg som tok den på PC. I artikkelen brukes dette som et bevis på konklusjonen "...skole-PC hjalp elevene med å bli bedre til å skrive generelt". Vi merker oss imidlertid at dette trolig er en feilaktig kausal slutning; de enkle sammenlikningene av gjennomsnitt som er brukt i studiet kan ikke tolkes kausalt.

Grimes og Warschauer (2008) undersøkte prestasjonen til skoler som deltok i et "one-to-one laptop program" i California fra 2005 til 2006. Treatment-gruppen ble sammenliknet med et utvalg skoler som ikke deltok i programmet i et Differences-in-differences rammeverk. De fant at testresultatene i matematikk og engelsk for treatment-skolene falt i treatment-året. Et interessant funn er imidlertid at prestasjonene ble bedre året etter.

The Texas Center for Educational Research (2009) undersøkte effekten av "The Texas Laptop Program", hvor hver elev fikk hver sin egen skole-PC. 21 treatment-skoler ble sammenliknet mot en kontrollgruppe i et Differences-in-differences rammeverk. Studiet finner blant annet at innføringen av individuelle skole-PCer ikke hadde noen signifikant utslag på leseresultatene til 7.- og 8.-klassinger. Det var imidlertid en signifikant effekt for utvalget 9. klassinger som gikk videre til videregående skole. Det ble ikke funnet noen signifikante effekter av skole-PC på skriveferdigheter.

1.4.3 Effekten av målrettet programvare³ i utdanning

Studier innen målrettet programvare i utdanning varierer i om tiltakene er et supplement eller en substitutt til tradisjonell undervisning. Dette har implikasjoner for hvor store effekter man kan forvente å finne. En tendens i forskningen på slike tiltak er at effekten er størst i utviklingsland. En forklaring på dette kan være at programvarene er substitutter for undervisning av en lavere kvalitet i disse landene (Bulman & Fairlie, 2015).

Rouse og Krueger (2004) undersøkte i et randomisert studie effekten av lese- og skriveprogrammet "Fat ForWord" på fire skoler i USA, med høy andel elever som ikke har engelsk som morsmål. Treatment var her at elever ble tatt ut av normal undervisning for å gjennomføre 90-100 minutter øving med programvaren; treatment var et supplement til tradisjonell undervisning. Studien fant marginalt positive til ingen effekter av programvaren på lese- og skriveprøver. Rouse og Krueger argumenterte for at læring på PC ikke er like effektivt som tradisjonell klasseromsundervisning.

The U.S. Department of Education and Mathematica Policy Research (2007, 2009) evaluerte seks programvarer rettet mot leseferdigheter og fire programvarer rettet mot mattekunnskaper for elever i grunnskolen og på videregående skole med et randomisert eksperiment. 9 av de 10 programvarene hadde ingen signifikant effekt. Den ene signifikante programvaren hadde en positiv effekt på lesing for fjerdeklassinger (Bulman & Fairlie, 2015).

³ Med målrettet programvare menes den engelske betegnelsen Computer Assisted Instruction (CAI), eller e-learning. Dette innebærer programvare på datamaskiner som ofte er rettet mot spesifikke fagområder (Bulman & Fairlie, 2015).

Barrow, Markman og Rouse (2009) undersøkte effekten av programvare som ble benyttet i datalaber for å forbedre algebra-forståelse i grunnskolen. Eksperimentet ble randomisert innad på skoler i USA, på klassenivå. Studiet fant at programvaren øker resultatene på algebratester med 0,17 standardavvik sammenliknet med en kontrollgruppe. Forfatterne merker seg at effekten var størst for større klasser. Dette kan bety at programvaren var et bedre substitutt for tradisjonell læring når lærerne ikke hadde tid til å hjelpe alle elevene godt (Bulman & Fairlie, 2015).

Carillo, Onofa og Ponce (2010) undersøkte effekten av programvare rettet mot å forbedre matematikk- og språkforståelse for barneskoleelever i Ecuador. Treatment innebar tre timer ukentlig individuell matematikk- og språktrening med programvaren. Det ble gjennomført et randomisert eksperiment hvor 8 skoler fikk treatment i 2007, og 8 kontrollskoler fikk treatment i 2009. Regresjonsanalysen viste at programvaren økte testresultater i matematikk med 0,3 standardavvik. Det ble også funnet negative, men ikke-signifikante effekter på språktester. Treatment-gruppen viste seg imidlertid å være et skjevt utvalg som de ikke klarte å kontrollere for. Derfor kontrollerte en med en test som ble gjennomført noen måneder *etter* innføringen av treatment. At denne kontrollvariabelen er generert etter innføringen av treatment gjør at en kan stille spørsmål ved den kausale tolkningen av disse funnene.

1.4.4 Oppsummering av eksisterende forskning på området

Oppsummert finner forskningen på generelle IKT-investeringer få signifikante effekter på akademiske utfall. Innføring av individuelle skole-PCer synes imidlertid å ha noen positive effekter, selv om en tendens også her er at det ikke finnes noen signifikante sammenhenger. Det samme gjelder for målrettede programvarer. Tvetydige funn kan her trolig forklares med at treatment i forskjellige studier har vært forskjellige programvarer (Bulman & Fairlie, 2015).

Bulman og Fairlie presiserer at man må være forsiktig med å trekke generelle slutninger om studier på dette området, da egenskaper ved treatment kan variere sterkt.

2 Metode

I denne seksjonen vil vi redegjøre for hva vi mener med kausaleffekter og hvordan disse, gitt informasjonen vi besitter, kan avdekkes med Differences-in-differences-metoden (DD).

Deretter vil vi redegjøre for DD og forutsetningene som må oppfylles for at funn med denne metoden skal kunne tolkes kausalt.

2.1 Kausalitet

Anta at en elev enten kan få iFinger eller ikke få iFinger. Dette gir to potensielle karakterutfall. Kausaleffekten av å få iFinger er, dersom alt annet holdes likt, differansen mellom disse to utfallene. Det åpenbare problemet vi møter i virkeligheten er at bare ett av utfallene kan observeres. Utfallet som ikke observeres er det *kontrafaktiske utfallet* (Angrist & Pischke, 2015).

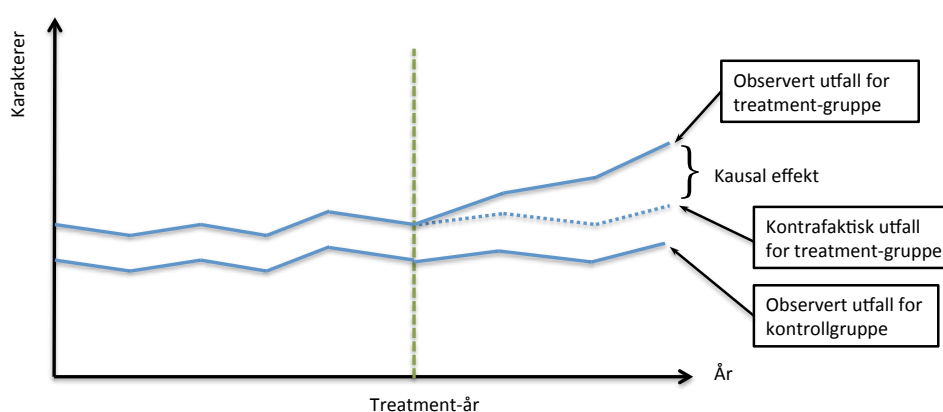
Ideelt sett løses dette problemet ved å observere et tilfeldig utvalg elever som tilfeldig deles i to, der én gruppe bruker iFinger og den andre ikke bruker noen digital ordbok. Ser en på situasjonen som et tradisjonelt klinisk forskningsstudie kan tilgang til iFinger omtales som "treatment". Når utvalgene er store nok til at alle forskjeller (utover iFinger) mellom gruppene jevner seg ut i gjennomsnitt, vil den eneste forskjellen mellom gruppene være iFinger. Hvis vi da ser at gruppen som har iFinger får bedre karakterer vil det være den gjennomsnittlige kausale effekten av iFinger (Angrist & Pischke, 2015).

Det er mange grunner til at vi ikke får gjort et ideelt eksperiment som beskrevet over. Blant annet er iFinger allerede i bruk i mange av Norges skoler, og fordelingen av iFinger har ikke vært tilfeldig; skoler, kommuner og fylker velger selv å kjøpe programvaren. Dette betyr også at tildelingen ikke skjer på elevnivå. Det er altså flere avvik fra tilfeldig tildeling av treatment. Følgelig må vi anvende andre metoder for å avdekke en eventuell kausal effekt.

2.2 Differences-in-differences (DD)

Når vi ikke har en tilfeldig tildeling av iFinger samt at treatment- og kontrollgruppen trolig er forskjellige, kan en potensiell kausal effekt avdekkes ved å bruke differences-in-differences-metoden. Denne metoden forutsetter først og fremst at skoler har en parallell utvikling i karakterer. Når noen skoler da får iFinger, vil et avvik fra denne felles trenden være et estimat på kausaleffekten. Vi bruker da kontrollgruppen til å estimere det kontrafaktiske utfallet til treatment-gruppen. Prinsippet kan sees av figur Figur 2.A under.

Figur 2.A – Difference-in-difference



2.2.1 DD på regresjonsform

DD gjennomføres best med regresjon (Angrist & Pischke, 2015). Med regresjon vil man få estimater med standardfeil, som kan generere t-statistikker og gi statistisk inferens. I tillegg kan man sammenlikne flere enheter over flere perioder (Pischke, 2005). Oppsettet for DD-regresjon med forklaringer av variabler og notasjoner er gitt i *Mastering 'Metrics* (Angrist & Pischke, 2015). Under vises dette oppsettet, tilpasset vår situasjon:

$$KARAKTER_{st} = \alpha + \delta_{DD} iFinger_{st} + \sum_{k=2}^h \beta_k SKOLE_{ks} + \sum_{j=2}^g \beta_j \text{\AA}R_{jt} + e_{st}$$

Notasjonene s og t står for henholdsvis skole og år i datasettet. Vi har h antall skole-dummies k som tar verdien 1 når k=s, og 0 ellers. β_k er koeffisienten tilhørende skole-dummy k. Skole k=1 vil være base-skolen. På samme måte har vi g antall års-dummies for

hvert år j , med koeffisienten β_j tilhørende års-dummy j . Her vil år $j=1$ være base-året (Angrist & Pischke, 2015).

Skole-dummyene kontrollerer for forskjeller mellom skoler som er konstante over tid. Et eksempel på dette kan være at skoler i ressurssterke områder systematisk har høyere karakterer enn skoler i ressursvake områder. Års-dummyene kontrollerer for effekter som er felles for alle skoler og spesifikke for hvert år. Et eksempel på dette kan være at Utdanningsdirektoratet gir en eksamen et år som er spesielt enkel eller vanskelig. $iFinger_{st}$ er DD-spesifikasjonen. Denne er tilsvarende et interaksjonsledd mellom skoler i treatment-gruppen og hvilke perioder disse er treated. Denne variabelen tar derfor verdien 1 dersom skole s har treatment i år t , og 0 ellers. Den tilhørende koeffisienten δ_{DD} gir estimatet på kausaleffekten av iFinger på karakterer (Angrist & Pischke, 2015).

2.2.2 Local Average Treatment Effect (LATE)

DD-spesifikasjonen tar normalt verdien 1 eller 0. Vi skiller da ikke mellom det å bli tildelt iFinger og det å faktisk bruke iFinger. Regresjonsestimatene kan da tolkes som en "reduced form". Dette vil si at vi estimerer effekten av "intention to treat"; effekten av å få tilgang til iFinger. Man kan da justere reduced form ved å dele på "first stage". First stage er andelen som har tilgang til iFinger gjennom sin skole og som faktisk benytter seg av programvaren. Dette vil gi "local average treatment effect" (LATE); effekten av å faktisk bruke iFinger.

2.2.3 Forutsetninger for DD

Vi vil nå redegjøre for forutsetningene om felles trender og entydig treatment. Begge er avgjørende for å kunne tolke den estimerte effekten kausalt.

2.2.3.1 Felles trender

Dersom treatment- og kontrollgruppene har felles trender før treatment, er det en indikasjon på at de ville hatt felles trender også etter treatment i fravær av treatment. Vi har

da et estimat på det kontrafaktiske utfallet til treatment-gruppen. Flere studier, som for eksempel Card og Krueger (1994), bruker kun én periode før treatment til å sjekke for felles trender. Vi vil foretrekke å sjekke over flere perioder; jo lengre før treatment vi kan bekrefte felles trender, jo sikrere kan vi være på at felles trender faktisk er tilfellet (Angrist & Pischke, 2015).

Dersom gruppene ikke har felles trender før treatment har vi ikke belegg for å si at utviklingen til kontrollgruppen er et estimat på det kontrafaktiske utfallet til treatment-gruppen. Vi kan da *ikke* tolke δ_{DD} som en kausal effekt (Angrist & Pischke, 2015). Det er verdt å merke seg at felles trender og en ikke-signifikant δ_{DD} er en indikasjon på at det ikke eksisterer noen kausal effekt.

Den enkleste formen for å verifisere felles trender er grafisk inspeksjon. Når man har flere treatment- og kontrollgrupper og flere treatment-tidspunkter, slik tilfellet er for oss, vil grafisk inspeksjon være praktisk umulig. Man kan da gjøre en sjekk for felles trender med Granger-testen (Pischke, 2005). Vi vil redegjøre for denne i del 2.3.

2.2.3.2 Omitted Variable Bias (OVB) og entydig treatment

Dersom vi har utelatte skole- og årsspesifikke variabler som både er korrelert med $iFinger_{st}$ og påvirker karakterer, vil DD-estimatet lide av omitted variable bias (OVB) (Angrist & Pischke, 2015). Årsaker til OVB i δ_{DD} kan for eksempel være innføringen av skole-PC eller IKT-basert eksamen. Hvis dette er tilfellet må disse variablene kontrolleres for. Ofte kan det være vanskelig eller umulig å skaffe data for å sjekke om vi har OVB. Da er beste tilnærming til problemet å gjøre en "kvalifisert gjetting" på sammenhengene nevnt over, for så å vurdere om den estimerte koeffisienten er for stor eller liten (Angrist & Pischke, 2015).

Merk at samvariasjonen mellom $iFinger_{st}$ og eventuelle samtidige effekter ikke kan være perfekt; hvis for eksempel innføringen av skole-PC og iFinger skjer samtidig for alle skoler, vil det være umulig å skille de to effektene som da blir estimert av δ_{DD} .

2.3 Granger-test

I utgangspunktet bruker man Granger-testen til å sjekke om årsaker skjer før konsekvenser, og ikke vice versa (Angrist & Pischke, 2009). En utvidet tolkning av Granger-testen gir imidlertid informasjon om vi har felles trender og hvordan effekten av treatment utvikler seg over tid (Pischke, 2005). Dette er blant annet gjort av Autor (2003).

Oppsettet til Granger-testen med tilhørende forklaringer er gitt av Pischke (2005). Vi har her tilpasset oppsettet i henhold til vår situasjon og notasjonene i DD-modellen fra del 2.2.1.

$$KARAKTER_{st} = \alpha + \sum_{r=-m}^q \beta_r D_{st}(t = p + r) + \sum_{k=2}^h \beta_k SKOLE_{ks} + \sum_{j=2}^g \beta_j \text{\AA}R_{jt} + e_{st}$$

Som vi ser er skole- og årsspesifikke faktorer kontrollert for på samme måte som i DD-modellen. Forskjellen fra DD-modellen er imidlertid at DD-spesifikasjonen er byttet ut med et sett dummy-variabler som representerer hvert enkelt år i en bestemt avstand fra treatment-året til de enkelte skolene. Disse avdekker om det er systematisk utslag på karakterer gitte år i forhold til året p hvor iFinger innføres. For eksempel er verdien av $D_{s,p-2}$ lik 1 for skole s 2 år før iFinger innføres, og 0 ellers. Dette er en lead-variabel, og vi har m antall lead-variabler. På samme måte er verdien av $D_{s,p+1}$ lik 1 for skole s 1 år etter iFinger innføres, og 0 ellers. Dette er en lag-variabel, og vi har q antall lag-variabler (Pischke, 2005).

2.3.1 Sjekk av felles trender

Granger-testen kan som sagt benyttes for å sjekke om forutsetningen om felles trender holder. Dersom $\beta_r = 0$ for alle $r < 0$ kan vi ikke forkaste hypotesen om at skolene som skal få treatment har samme karakterer som skoler som ikke skal få treatment, betinget på skol- og årsspesifikke effekter. Dette er da en indikasjon på at vi har felles trender (Pischke, 2005). Ved signifikante leads forkaster vi hypotesen om felles trender. Det er viktig å være oppmerksom på at denne testen ikke kan brukes som *bevis* for felles trender, men heller er et mislykket forsøk på å motbevise felles trender.

2.3.2 Kausalitet

Dersom vi har felles trender, vil både DD-spesifikasjonen og lags i Granger-testen ha en kausal tolkning. Altså, om β_r er signifikant forskjellig fra 0 for en $r \geq 0$, tilsier det at koeffisienten β_r er den kausale effekten av iFinger i år r etter treatment. Med Granger-modellen kan vi derfor fange opp hvordan effekten av treatment utvikler seg over tid. Dette står i kontrast til DD-modellen som bruker én dummy med verdi 1 for *alle* år med treatment (Pischke, 2005).

2.4 Clusterede standardfeil

Et vanlig problem i paneldatasett er seriekorrelasjon. Seriekorrelasjon betyr at en observasjon er korrelert med foregående observasjoner. For eksempel vil en skole med høye karakterer i år t , sannsynligvis ha høye karakterer i år $t+1$, $t+2$ osv. Dette er et avvik fra tilfeldigheter i observasjonene, og følgelig inneholder dataen vår mindre informasjon (Angrist & Pischke, 2015).

Datasett brukt til DD har ofte (1) lengre tidsserier, (2) seriekorrelasjon i avhengig variabel og (3) en treatment-variabel som endrer seg lite langs tidsserien til en observasjonsenhet. Disse tre faktorene forsterker hverandre på en måte som gjør at standardfeilene i DD-modellen kan bli underestimerte (Bertrand, Duflo, & Sendhil, 2004). Følgelig øker sannsynligheten for å forkaste en sann nullhypotese. I paneldatasett som våre finnes det i tillegg en risiko for å få en tilsvarende bias ved at seriekorrelasjonen til skoler kan være korrelert i tverrsnittet. Eksempler på dette kan være sjokk som er felles for alle skoler innad i en kommune eller et fylke.

Dersom man har seriekorrelasjon mellom observasjonene til en skole, kan man hankses med dette problemet ved å samle feilleddene til hver skole i egne *cluster*e, og bruke variasjonen mellom disse clusterene til å estimere standardfeil (Bertrand, Duflo, & Sendhil, 2004). Angrist og Pischke (2009) kaller disse modifiserte standardfeilene for *clusterede standardfeil*. For at de clusterede standardfeilene skal være konsistente kreves det imidlertid at at clusterene ikke er korrelerte med hverandre (Bertrand, Duflo, & Sendhil, 2004). Vi er nemlig

ute etter tilfeldige observasjonsenheter i feilledet. Dersom vi clusterer på skolenivå, må korrelasjonen mellom skole-clustre være tilfeldig. Hvis dette ikke er tilfellet må vi clustre på et høyere nivå, som for eksempel ved sjokk på fylkes- eller kommunenivå. Kostnaden ved å clustre på et høyere nivå er at vi får færre clustere, hvilket impliserer færre tilfeldige grupper av residualer til utregning av standardfeil (Angrist & Pischke, 2009).

Når antallet tilfeldige observasjoner i feilledet går mot null vil utregningen av standardfeil bryte sammen. En konsekvens av dette er at vi ikke kan stole på inferensen i regresjonene. Det viser seg at den nedre grensen for å få konsistente clusterede standardfeil er ved ca. 40 clustere (Abadie, Athey, Imbens, & Wooldridge, 2014). Angrist og Pischke setter i *Mostly Harmless Econometrics* (2009) den nedre grensen til 42 clustere. Siden vår oppgave i stor grad er bygget på deres litteratur, vil vi i denne oppgaven operere med 42 clustere som nedre grense.

3 Databeskrivelse

Vi vil benytte oss av to paneldatasett fra norsk skole, ett for ungdomsskoler (US) og ett for videregående skoler (VGS). Vi vil først definere populasjonene for studiet. Deretter vil vi forklare hvordan datasettene er generert, før vi redegjør for eksamenskarakterer og treatment.

3.1 Populasjon

I lys av problemstillingen definerer vi tre forskjellige populasjoner av norske skoleelever for dette studiet. Først har vi (1) ungdomsskoleelever, som går 8.-10. trinn i grunnskolen. Deretter har vi (2) elever ved yrkesfag- og påbygg på VGS-nivå. Disse er mer praktisk rettede, og fokuserer trolig mindre på språkfag. Den siste populasjonen er (3) elever på studiespesialiserende linjer på VGS-nivå. Disse er mer teoretisk rettede, hvor elevene trolig har mer fokus på språkfag.

Påbygg er en studiespesialiserende linje hvor en har egne eksamener. Når vi i denne oppgaven omtaler studiespesialiserende linjer omfatter det ikke påbygg, som er plassert i populasjon (2).

3.2 Datainnsamling

Karakterdata

Vi har hentet karakterdata fra Skoleporten.no sin rapportbygger, et nettsted underlagt Utdanningsdirektoratet (Utdanningsdirektoratet, 2015). Denne inneholder data for alle ordinære elever ved ordinære skoler. Vi har med andre ord ikke data for privatister eller rene skoler for privatister. Observasjonsnivået i Skoleporten, og følgelig våre datasett, er på skolenivå. En karakterobservasjon er da gjennomsnittskarakteren til elevene ved en gitt skole som avla eksamen i et gitt fag et gitt år. Til VGS-datasettet har vi hentet ut karakterdata for alle tilgjengelige skoler, mens vi til US-datasettet har hentet ut karakterdata for skoler vi har digital ordbok-informasjon om.

I praksis har vi hentet ut data fra Skoleporten som 1-2 Excel-dokumenter per fylke. Vi har så laget makro-koder som transformerer datasettene til ønsket form før de har blitt importert inn i US eller VGS-datasettet. Deretter har vi manuelt søkt etter feil som kan ha oppstått under transformeringen. Ingen feil er oppdaget.

Datasettene inneholder mange tomme verdier for karakterobservasjoner. Dette er en naturlig konsekvens av eksamensordningen; det kan forekomme at ingen elever ved en skole trekkes ut til en gitt eksamen et år. At trekket er tilfeldig gjør at disse tomme verdiene ikke fører til utvalgsskjevheter, og derfor ikke skaper problemer med inferensen. Mange tomme karakterobservasjoner kan også forklares med at elever ved yrkesfag-skoler ikke tar studiespesialiserende fag og vice versa.

Videre inneholder Skoleporten kun data for skoler som eksisterer per 2015. For eksempel finnes ikke historisk data for nylig nedlagte Bergen Handelsgymnasium. Slike "hull" i datasettet er mest sannsynlig tilfeldige i forhold til karakterer og iFinger, og følgelig er også dette uproblematisk for vår inferens.

Informasjon om treatment

Informasjon om hvorvidt skoler har hatt iFinger, Ordnett eller ingen digital ordbok har vi fått fra iFinger. For iFinger-kunder har vi hentet skole- og fylkesavtaler fra iFinger sine faktureringsystemer. Utover dette har iFinger-ansatte gjennom intervjuer fortalt når ulike skoler og fylkeskommuner har hatt ulike avtaler for digitale ordbøker. Siden fylkeskommunene-avtalene er ute på anbud (med unntak av Oslo og Telemark, se 4.2.1) har ansatte hos iFinger god oversikt over når fylker har valgt Ordnett som leverandør i anbudsrundene. Det de har fortalt er inkludert i datasettene. Informasjonen fra intervjuene er ikke av subjektiv art, men vi er bevisste på at den kan være upresis. Dette kan føre til feil i datasettet som i verste fall kan gi feilaktige resultater. En betryggende faktor er at eventuelle feilrapporteringer trolig er helt tilfeldige, slik at de ikke skaper systematiske skjevheter i utvalgene våre. Intervjuobjektene kjenner nemlig ikke til kausalitetsrammeverket som benyttes i denne oppgaven, og har trolig ikke ikke kjennskap til hvilke skoler som har gode eller dårlige resultater i gitte år. En bevisst manipulering av datasettet vil da være

usannsynlig. Vi merker oss at vi ikke har noen grunn til å tro at intervjuobjektene har ønske om å bevisst feilrapportere om når skoler hadde iFinger, Ordnett eller ingen digital ordbok.

3.3 Avhengige variabler: Eksamenskarakterer

Karakterer brukes som et mål på elevenes oppnåelse av kompetanse i fag. Eksempler på kompetansemål er at elevene skal kunne *"uttrykke seg med et presist og nyansert ordforråd og mestre språklige formkrav"* (norsk, studiespesialiserende linjer), *"skrive tekster med tema og fagterminologi som er tilpasset eget utdanningsprogram"* (norsk, yrkesfag), *"forstå innholdet i skriftlige og muntlige autentiske tekster"* (engelsk, alle utdanningsprogram) og *"skrive sammenhengende tekster i ulike sjangere"* (fremmedspråk). Dette gjør karakterer til en god proxy for språklig fremstillingsevne, forståelse og rettskrivningsferdigheter. Merk at samtlige kompetansemål er hentet fra de respektive fags utdanningsplaner, som er gitt i bibliografien.

Karakterer settes på en ordinal skala fra 1 til 6. Forskrift til opplæringslova (2009) paragraf 4-4 sier at de enkelte karakterene skal representere følgende oppnåelser av kompetansemålene:

- a) karakteren 6 uttrykkjer at eleven har framifrå kompetanse i faget
- b) karakteren 5 uttrykkjer at eleven har mykje god kompetanse i faget
- c) karakteren 4 uttrykkjer at eleven har god kompetanse i faget
- d) karakteren 3 uttrykkjer at eleven har nokså god kompetanse i faget
- e) karakteren 2 uttrykkjer at eleven har låg kompetanse i faget
- f) karakteren 1 uttrykkjer at eleven har svært låg kompetanse i faget

I videregående opplæring svarer karakteren 1 til "ikke bestått" (Forskrift til opplæringslova, 2009, paragraf 3-4).

En fordel med en slik bunden skala er at samtlige kompetansenivåer en elev kan oppnå faller i intervallet 1 til 6. Vi kan da være sikre på at datasettet ikke inneholder ekstremverdier. Vi velger å kun benytte karakterer fra skriftlige eksamener i vår analyse. Grunnen til vi kun

bruker vurderinger av skriftlige besvarelser er at ordbøker i mindre grad kan forbedre en elevs prestasjon på muntlige tester. En kandidat til muntlig eksamen i norsk testes for eksempel i større grad i litteraturhistorie enn rettskrivning.

Vi vil ikke bruke standpunktkarakterer i vår analyse. Det er tre grunner til dette. (1) Hvordan karakterskalaen anvendes kan variere mellom skoler og fylker. Dette betyr at det kan være problematisk å sammenligne standpunktkarakterer. (2) En lærer vil kun eksponeres for en skole av gangen. Om iFinger har en effekt kan det hende læreren observerer systematisk bedre besvarelser på sin skole etter treatment. Da er det en fare for at karakterskalaen blir flyttet innad på skolen snarere enn at skolen blir flyttet innad i karakterskalaen. Hvis dette skjer vil vi ikke kunne observere den reelle effekten av iFinger med karakterskalaen. (3) Vi ønsker å ha et så objektivt mål som mulig på elevenes språkferdigheter. Det er åpenbart at standpunktkarakterer kan være påvirket av subjektive og personlige forhold mellom lærere og elever.

Skriftlige eksamener utarbeides sentralt av 3-5 fagpersoner utnevnt av Utdanningsdirektoratet. På denne måten vil hver kandidat i Norge få en like utfordrende oppgave, uavhengig av hvilken skole vedkommende går på. Hver enkelt fylkeskommune og kommune har ansvar for å gjøre et tilfeldig trekk av hvilke elever på henholdsvis VGS og US som skal opp til eksamen (Utdanningsdirektoratet u.å. a,b). Det kreves at trekket fordeler seg jevnt på fag og skoler over tid (Utdanningsdirektoratet, 2014a). Med tilstrekkelig mange observasjoner får vi da et representativt utvalg av norske skoleelever som tar eksamen i hvert fag.

Skriftlige eksamensbesvarelser leveres anonymt og vurderes av to forskjellige sensorer. Hver sensor deltar på sensorskoleringer før vurderingsprosessen, slik at alle har en felles forståelse av hva som kreves av elevene. Dette foregår på fylkesnivå. Formålet med sensorskoleringen er at eksamenskarakteren ikke skal være avhengig av hvilken sensor som har vurdert de enkelte besvarelsene. Etter at hver sensor har sensurert, møtes de til "fellessensuren" hvor de skal fastsette endelig karakter på hver besvarelse. Dersom sensorene ikke blir enige om hvilken karakter en besvarelse skal få, vil den endelige avgjørelsen bli tatt av lederen av fellessensuren (Utdanningsdirektoratet, 2013).

Vi ser at det generelle formålet med sentralt gitte eksamener er at alle elever skal vurderes likt opp mot kompetansemålene. Derfor føler vi oss trygge på at eksamenskarakterer er et godt og objektivt mål på språkferdigheter. Vi merker oss i tillegg at det ikke er mulig for sensor å se om en besvarelse er levert av en kandidat som bruker iFinger, Ordnett eller ingen digital ordbok. Vedkommende kan kun se om en besvarelse er levert digitalt eller ikke. På denne måten vil ikke selve *vurderingen* av språkferdigheter ved sensur være påvirket av hvorvidt en kandidat har fått iFinger eller ikke.

Vi vil nå redegjøre for karakterene i fagene som benyttes i analysen. Når vi refererer til årstallet for enkelte karakterer, menes kalenderåret eksamen er avlagt. En karakterobservasjon for år 2010 er da fra eksamen som gjennomføres i slutten av skoleåret 2009-2010. Vi minner om at en observasjon er gjennomsnittskarakteren for kandidatene fra en gitt skole et gitt år.

3.3.1 Ungdomsskole

For ungdomsskoleelever vil vi se på følgende fag:

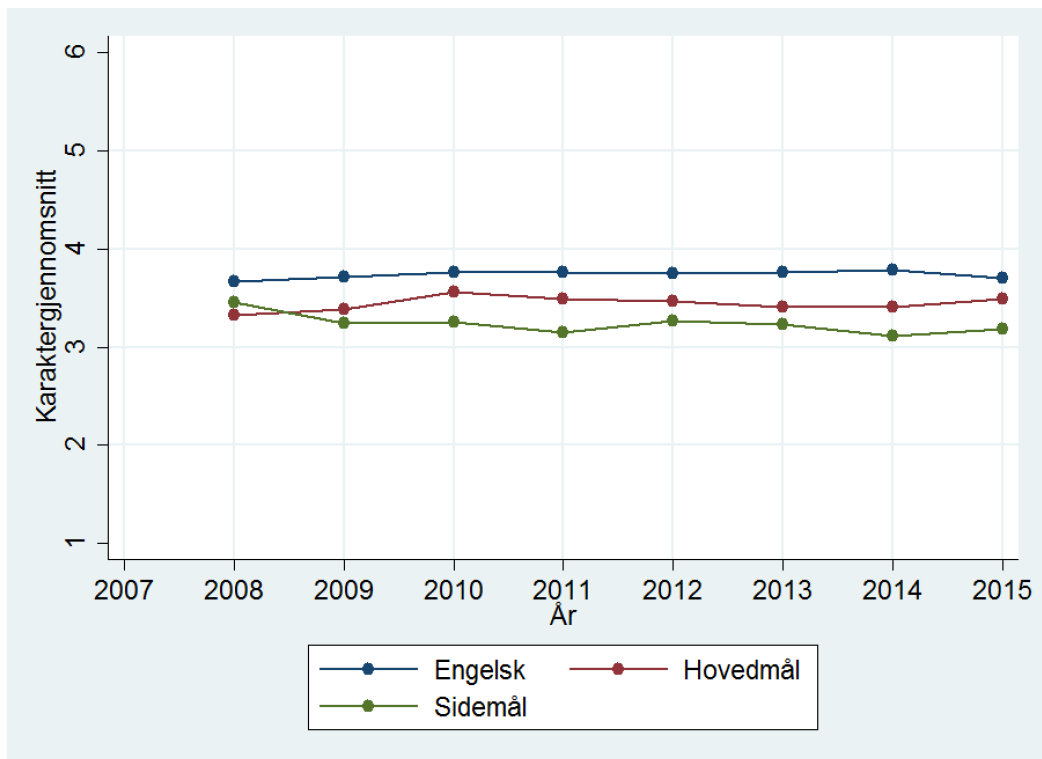
Engelsk ungdomsskoleeksamen (EUE)

Hovedmål ungdomsskoleeksamen (HUE)

Sidemål ungdomsskoleeksamen (SUE)

Disse eksamenene avlegges på 10. trinn. Da skal elevene trekkes ut til én skriftlig og én muntlig eksamen. Engelsk, hovedmål og sidemål er da blant fagene hver elev kan trekkes ut til (Utdanningsdirektoratet, 2014a).

Figur 3.A – Karakterutvikling engelsk, hovedmål og sidemål US



Karakterene for ungdomsskoleeksamener ser ut til å være stabile over tid, uten tidsavhengige trender. Engelsk ligger høyest, hovedmål noe lavere og sidemål ligger lavest.

3.3.2 Yrkesfag og påbygg VGS

For yrkesfag og påbygg ser vi på følgende fag:

Engelsk yrkesfaglig eksamen (EYE)

Norsk yrkesfaglig eksamen (HYE)

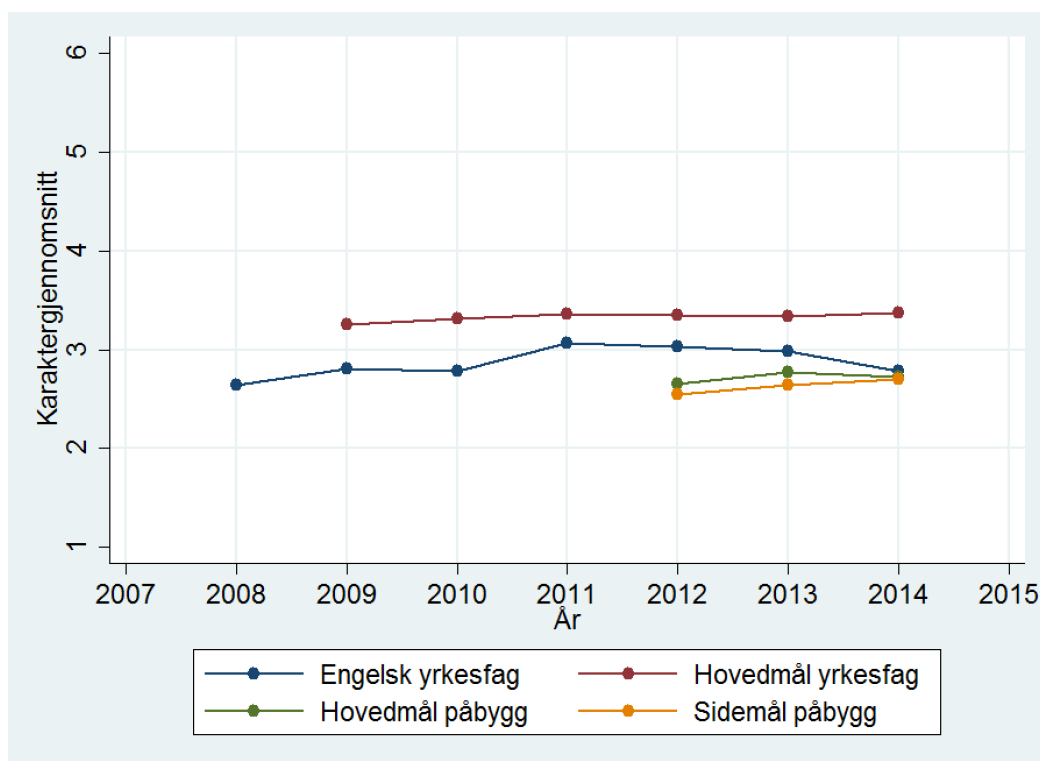
Hovedmål påbygg eksamen (HPE)

Sidemål påbygg eksamen (SPE)

20% av elevene skal i løpet av sin toårsperiode på yrkesfag tilfeldig trekkes ut til eksamen i ett fag utover yrkesfag. Engelsk yrkesfag og hovedmål yrkesfag er blant fagene de kan bli trukket ut til. Etter et toårig program på yrkesfag kan elever velge å ta ett år med påbygg for å få generell studiekompetanse. Samtlige elever skal her opp til eksamen i hovedmål. I tillegg

skal samtlige elever trekkes ut til én skriftlig og én muntlig-praktisk eksamen. Av de skriftlige eksamenene kan enkelte elever da trekkes ut til sidemål (Utdanningsdirektoratet, 2014a).

Figur 3.B – Karakterutvikling engelsk yrkesfag, hovedmål yrkesfag, hovedmål påbygg og sidemål påbygg VGS



For yrkesfagene ser karakterene ut til å utvikle seg forholdsvis flatt. Karakterene i påbygg øker noe, men her har vi så få år med observasjoner at dette kan være helt tilfeldig. Det behøver ikke nødvendigvis å finnes noen form for oppadgående trend for påbyggkarakterer.

3.3.3 Obligatoriske fag studiespesialiserende VGS

For å gjøre plottene lettere å tolke har vi her delt opp fagene for populasjonen av elever på studiespesialiserende linjer i to; obligatoriske fag presenteres her, mens tredjespråkene presenteres i del 3.3.4.

De obligatoriske fagene for studiespesialiserende linjer vi skal se på er:

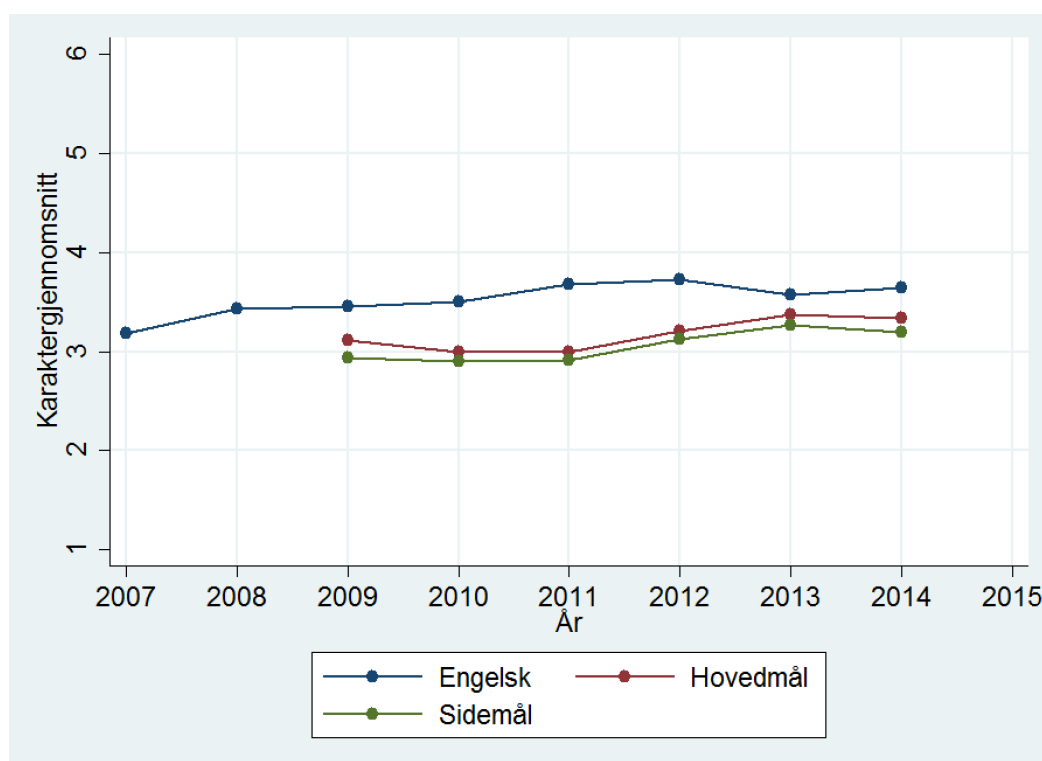
Hovedmål studiespesialiserende eksamen (HSE)

Engelsk studiespesialiserende eksamen (ESE)

Sidemål studiespesialiserende eksamen (SSE)

Samtlige elever på de studiespesialiserende linjene skal opp i hovedmål. Elever på ren studiespesialiserende linje trekkes i tillegg tilfeldig ut til to skriftlige eksamener. Elever på formgivningsfag eller idrettsfag med studiespesialisering skal trekkes ut til tre eksamener; skriftlig, muntlig og/eller skriftlig-praktisk. Elever på musikk, dans og drama med studiespesialisering skal trekkes ut til to skriftlige, muntlige og/eller skriftlig-praktiske eksamener. Blant eksamenene disse elevene kan trekkes ut til har vi engelsk og sidemål (Utdanningsdirektoratet, 2014a).

Figur 3.C – Karakterutvikling engelsk, hovedmål og sidemål studiespesialiserende VGS



Karaktersnittet for engelsk ligger generelt noe høyere enn karakterene for hovedmål og sidemål. Hovedmål og sidemål har hatt en svak økning i karakterer de siste tre årene med observasjoner. Det ser imidlertid ikke ut til å være noen tidsavhengige trender.

3.3.4 Tredjespråk VGS

For tredjespråk VGS ser vi på følgende fag:

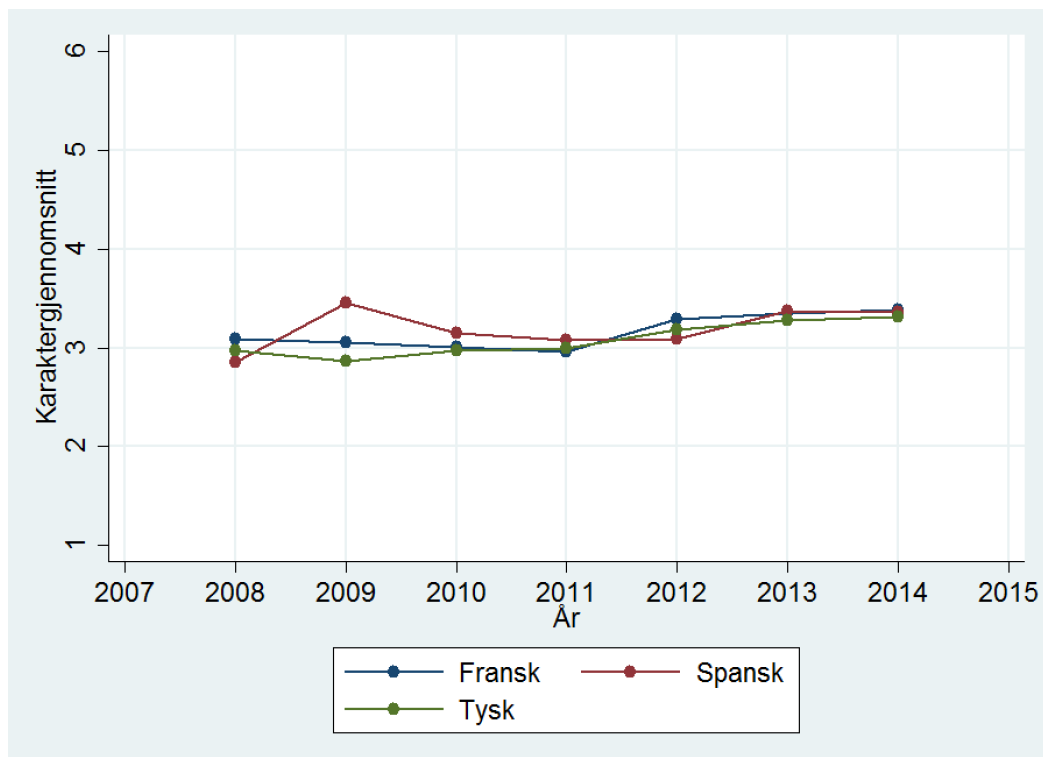
Fransk 2 eksamen (F2E)

Spansk 2 eksamen (S2E)

Tysk 2 eksamen (T2E)

Elevene på studiespesialiserende velger ett tredjespråk. Eksamen i disse fagene trekkes ut blant elever på VG2. Samtlige elever på VG2 skal opp i enten ett skriftlig, ett muntlig eller ett muntlig-praktisk fag. Elever som har fransk, spansk eller tysk som valgfag, kan bli trukket ut til skriftlig eksamen i disse (Utdanningsdirektoratet, 2014a).

Figur 3.D – Karakterutvikling fransk, spansk og tysk tredjespråk VGS



Karaktersnittene i fransk og tysk ligger forholdsvis stabilt rundt 3 frem til 2011. Spansk har en midlertidig hopp i 2012. Samtlige fag har karaktersnitt på snaut 3,5 i 2013 og 2014. Noe

av denne økningen kan skyldes endret beregningsmetode (Utdanningsdirektoratet, 2010). Slike skole-uavhengige inngripener blir kontrollert for som årsspesifikke effekter i våre modeller.

3.4 Treatment-variabel: iFinger

Når en skole begynner å bruke iFinger, vil verdien av DD-spesifikasjonen være 1 ved første eksamen i kalenderåret *etter* at denne skolen ble fakturert første gang. Deretter tar DD-spesifikasjonen verdien 1 for alle påfølgende år med iFinger. Hvis for eksempel en skole ble fakturert for iFinger i oktober 2010, vil datasettet vise at treatment skjer første gang under eksamen våren 2011. Blir en skole fakturert første gang i april 2013, vil datasettet vise at treatment skjer første gang under eksamen våren 2014.

Ved å sette treatment-variabelen på ovennevnte måte er vi sikre på at iFinger ble utrullert til elevene innen eksamensdagen og at elevene har hatt minst ett semester til sette seg inn i programvaren før eksamensdagen. Dette er viktig fordi en potensiell effekt av programvaren trolig vil være et produkt av bruk over tid. DD-spesifikasjonen er bygget på samme måte for både US- og VGS-datasettet.

Merk at vi senere vil introdusere forskjellige DD-variabler. Disse har lik tolkning av verdien 1, men de har forskjellige tolkninger av det kontrafaktiske utfallet, altså av verdien 0. Dette vil spesifiseres i del 3.4.1.1 og 3.4.2.1.

4 Deskriptiv statistikk

På bakgrunn av innsamlet data har vi som sagt to paneldatasett; ett for US og ett for VGS. Hver observasjon i datasettene tilhører en gitt skole i et gitt år. For eksempel vil karakterobservasjonen HSE_{st} gi gjennomsnittskarakteren på hovedmål studiespesialiserende eksamen for skole s i år t . Vi vil nå redegjøre for hvert av de to datasettene, og hvilke utvalg fra de vi vil benytte i analysen.

4.1 Ungdomsskole

Datasettet for US består av de samme 97 skolene over 8 år, fra 2008 til 2015. 45 av skolene får iFinger i løpet av tidsperioden og utgjør treatment-gruppen. De resterende 52 skolene er i kontrollgruppen. Videre har hver skole en tidsserie for karakterutviklingen i hvert fag. Som vi vil se i del 4.1.2 er ikke nødvendigvis disse tidsseriene komplette.

Datasettet er generert basert på informasjon fra salgsdokumenter og ansatte i iFinger; alle skoler hvor vi kan bekrefte bruk av iFinger eller ingen bruk av digitale ordbøker fra 2008 til 2015 er inkludert. Dette har resultert i at de 97 skolene i utvalget ligger i fylkene Aust-Agder, Buskerud, Finnmark, Hedmark, Oppland, Oslo, Rogaland, Troms og Østfold. Vi har da et bredt spekter skoler etter demografi, som blant annet byfylket Oslo, distriktsfylkene Hedmark og Finnmark samt storbyfylkene Rogaland og Troms. Basert på dette tror vi at utvalget er representativt for populasjonen, som er samtlige elever ved norske ungdomsskoler.

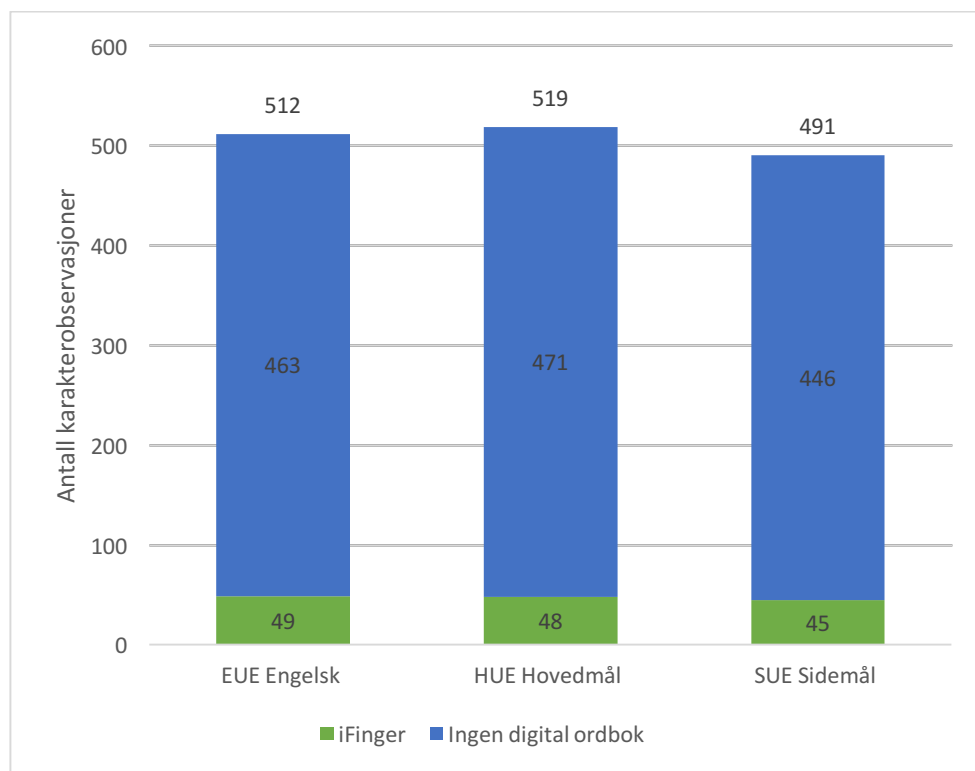
4.1.1 Tolkning av treatment-variabelen for US

DD-spesifikasjonen for US heter $iFingerUS_{st}$ og tar verdien 1 dersom en skole i et gitt år har iFinger. Dersom variabelen tar verdi 0, betyr det at vi observerer en skole som ikke har noen avtale for digitale ordbøker i det bestemte året. I US-datasettet har vi ingen tomme verdier for DD-spesifikasjonen, hvilket betyr at samtlige karakterobservasjoner vil bli benyttet i regresjonene.

4.1.2 Deskriptiv statistikk av US-utvalg til analyse

Vi vil nå undersøke karakterobservasjonene fra de 97 skolene i ungdomsskoleutvalget. I tabellen under ser vi at skolegjennomsnitt av eksamenskarakterer i hvert av de tre fagene er rapportert ca. 500 ganger. Dette tilsier at vi har tomme verdier av karakterobservasjoner for samtlige fag; hvis ikke ville vi hatt $97 \times 8 = 776$ observasjoner per fag. Årsaken til dette er at eksamen trekkes tilfeldig, noe som kan gjøre at det forekommer at ingen elever ved en skole avlegger eksamen i et gitt fag et gitt år.

Figur 4.A – Antall observasjoner av eksamenskarakterer betinget på treatment-tilstand US



I stolpediagrammene over er observasjoner merket "iFinger" fra treatment-gruppen etter treatment. Observasjoner med merket "Ingen digital ordbok" kommer fra kontrollgruppen eller treatment-gruppen før treatment. Tidsseriene for hvert av de tre fagene går fra 2008 til 2015, mens ungdomsskolene anskaffet iFinger først mellom 2012 og 2015. Dette forklarer den lave andelen observasjoner fra treatment-skoler som er treated. Datasettet tilsier at vi kan sjekke for felles trender i 4 år før innfasinger av iFinger. Vi vil derfor inkludere Lead 4 til Lead 1 i Granger-testen for disse fagene.

I Tabell 4.A, som sees under, er karakterobservasjonene delt inn etter treatment- og kontrollgrupper. Ved å gjøre en slik inndeling vil vi få et inntrykk av forskjellene mellom skoler i utvalget som anskaffer iFinger og ikke.

Tabell 4.A – Deskriptiv karakterstatistikk: Ungdomsskole

EUE, engelsk	Observasjoner	Gjennomsnitt	Standardavvik	Min	Max
Totalt	512	3,744	0,341	2,6	4,7
Treatment-gruppe	183	3,827	0,363	2,6	4,7
Kontrollgruppe	329	3,698	0,319	2,8	4,7
<i>Differanse</i>	<i>-146</i>	<i>0,129</i>	<i>0,044</i>	<i>-0,2</i>	<i>0</i>
HUE, hovedmål					
Totalt	519	3,453	0,310	2,2	4,4
Treatment-gruppe	186	3,492	0,347	2,2	4,2
Kontrollgruppe	333	3,430	0,286	2,7	4,4
<i>Differanse</i>	<i>-147</i>	<i>0,062</i>	<i>0,060</i>	<i>-0,5</i>	<i>-0,2</i>
SUE, sidemål					
Totalt	491	3,212	0,337	2,3	4,2
Treatment-gruppe	168	3,267	0,355	2,4	4
Kontrollgruppe	323	3,184	0,325	2,3	4,2
<i>Differanse</i>	<i>-155</i>	<i>0,083</i>	<i>0,030</i>	<i>0,1</i>	<i>-0,2</i>

Kontrollgruppene har nesten dobbelt så mange observasjoner av karakterer som treatment-gruppene for alle fagene. Spredningen i gjennomsnittskarakterer for hver skole er ca. 2 for både treatment- og kontrollgruppene. At ekstremverdiene ikke er 1 eller 6 forklares med at observasjonene er gjennomsnitt av skolegjennomsnitt, og ikke individer. Standardavviket til hvert fag ligger mellom 0,30 og 0,35 karakterpoeng.

En interessant observasjon er at gjennomsnittskarakterene til treatment-gruppene er høyere enn kontrollgruppene for hvert fag. Per nå kan vi ikke si at treatment-skoler tenderer å være bedre enn kontrollgruppen på grunn av innfasing av iFinger, eller om skolene som anskaffer iFinger er bedre i utgangspunktet. En slik potensiell utvalgsskjevhet vil vi prøve å skille ut i DD-analysen.

4.2 Videregående skole

Datasettet for VGS består av 418 videregående skoler over 8 år, fra 2007 til 2014. For hver skole har vi informasjon om eieform (offentlig/privat) og hvilket fylke den ligger i. Også her har hver skole en tidsserie for karakterutviklingen for hvert fag. Sammenliknet med US-datasettet har VGS-datasettet enda flere tomme verdier for karakterer. Årsaken til dette er at flere av fagene ikke er felles for alle elever samt at vi har både rene yrkesfagsskoler og rene studiespesialiserende skoler i datasettet.

4.2.1 Tolkning av treatment-variablene for VGS

Det er viktig å merke seg at VGS-datasettet har to forskjellige DD-variabler; $iFinger_{st}$ og $iFinger2_{st}$. Begge tar verdien 1 dersom en skole i et gitt år bruker iFinger, som forklart i del 3.4. $iFinger_{st}$ tar verdien 0 dersom en skole bruker Ordnett, mens $iFinger2_{st}$ tar verdien 0 dersom en skole ikke har noen digital ordbok. Bakgrunnen for dette er at enkelte skoler i datasettet har hatt enten iFinger eller Ordnett i hele tidsperioden 2007 til 2014, mens andre skoler har hatt enten iFinger eller ingen digital ordbok i samme tidsperiode.

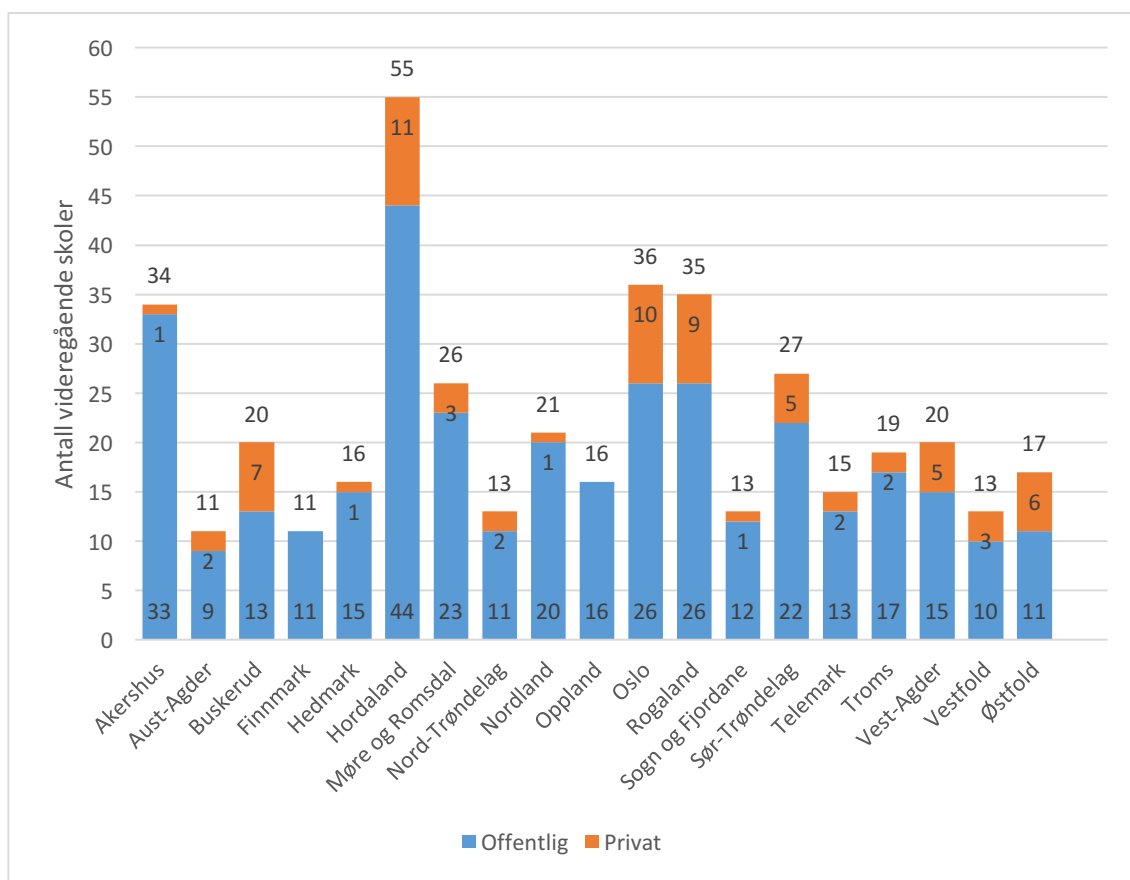
De fleste videregående skoler i Norge er offentlige, og underlagt hver sin fylkeskommune. Samtlige fylkeskommuner, med unntak av Oslo og Telemark, inngår digitale ordbok-avtaler på vegne av alle offentlige skoler i fylket gjennom anbudsrunder. Offentlige skoler i Oslo og Telemark, samt privatskoler, inngår egne avtaler direkte med iFinger eller Ordnett dersom de har digitale ordbøker. Informasjonen vi besitter om fylkesavtaler for digitale ordbøker kan sees i Figur 4.B under.

Figur 4.B – Fylkesavtaler for digitale ordbøker over tid VGS

Fylke	2007	2008	2009	2010	2011	2012	2013	2014
Akershus	Ordnett	Ordnett	Ordnett	Ordnett	iFinger	iFinger	iFinger	Ordnett
Aust-Agder		Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett
Buskerud		Ordnett	Ordnett	iFinger	iFinger	iFinger	iFinger	iFinger
Finnmark	Ingen ordbøker			iFinger	iFinger	iFinger	iFinger	iFinger
Hedmark		Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett
Hordaland		Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	iFinger	iFinger
Møre og Romsdal		Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett
Nord-Trøndelag	iFinger	iFinger	iFinger	iFinger	Ordnett	Ordnett	Ordnett	Ordnett
Nordland		Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett
Oppland		Ordnett	Ordnett	iFinger	iFinger	iFinger	iFinger	iFinger
Oslo	Abonnementer på skolenivå							
Rogaland		Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett
Sogn og Fjordane	Ingen ordbøker				iFinger	iFinger	iFinger	iFinger
Sør-Trøndelag		Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett
Telemark	Abonnementer på skolenivå							
Troms		Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett
Vest-Agder		Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett
Vestfold		Ordnett	Ordnett	iFinger	iFinger	iFinger	iFinger	iFinger
Østfold		Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett	Ordnett

Figur 4.B leses best i sammenheng med Figur 4.C. Vi får da et inntrykk av hvor mange skoler som inngår i hver fylkeskommune.

Figur 4.C – Antall videregående skoler per fylke



Basert på denne informasjonen kan vi dele inn de 418 skolene i 3 forskjellige grupper:

(1) Offentlige skoler i Finnmark og Sogn og Fjordane

Disse skolene sin treatment-status beskrives med $iFinger2_{st}$. Her er det kontrafaktiske utfallet til iFinger å ikke ha digitale ordbøker. Dette gjelder 23 skoler.

(2) Offentlige skoler i Oslo og Telemark samt privatskoler

Disse skolene er verken tilknyttet $iFinger_{st}$ eller $iFinger2_{st}$, da vi kun har informasjon om årene hvor en DD-identifikasjon ville hatt verdien 1. Uten å vite hva en nullverdi av DD-spesifikasjonen impliserer kan vi verken sjekke for felles trender eller vite hvilken kausaleffekt vi estimerer. Følgelig vil ikke disse 110 skolene anvendes i analysen.

(3) Resterende offentlige skoler

Disse skolene er tilknyttet $iFinger_{st}$, hvor alternativet til å ha iFinger er å ha Ordnett. Dette utgjør 285 skoler.

To forskjellige treatment-identifikasjoner lar oss estimere to forskjellige effekter av iFinger for VGS med DD. Med $iFinger_{st}$, kan vi estimere effekten av iFinger sammenliknet med Ordnett. Med $iFinger2_{st}$ kan vi estimere effekten av iFinger sammenliknet med å ikke ha noen digital ordbok. Vi vil nå redegjøre for utvalgene som tilhører hver av de to DD-identifikasjonene.

4.2.2 Unntakene Finnmark og Sogn og Fjordane

Offentlige videregående skoler i Finnmark og Sogn og Fjordane er tilknyttet treatment-variabelen $iFinger2_{st}$. Disse skolene går som sagt fra å ikke ha digitale ordbøker til å bruke iFinger, og må analyseres separat.

Her kunne vi potensielt estimert en kausal effekt av iFinger med DD for videregående skoler i Finnmark i 2010, med Sogn og Fjordane som estimat på det kontrafaktiske utfallet, ref. Figur 4.B. Dette utvalget består imidlertid bare av 23 offentlige skoler. Vi ville da hatt for få cluster for å ha tiltro til standardfeilene og den statistiske inferensen i en regresjonsanalyse.

Vi kunne også potensielt brukt "regression discontinuity" (RD) på dette utvalget. Vi ville da studert karaktertidsseriene for hvert fag, og sett om det var en "cutoff" i karakterutviklingen når iFinger innføres, gitt at iFinger er eneste treatment i treatment-årene. Denne cutoffen ville da vært et estimat på den kausale effekten av iFinger (Angrist & Pischke, 2015). Vi har prøvd å anvende denne metoden med forskjellige spesifikasjoner av RD-modellen. De forskjellige spesifikasjonene gir svært varierende RD-estimer. Dette gjør at vi har liten tiltro til RD-modellen (Angrist & Pischke, 2015). Vi konkluderer derfor med at RD ikke egner seg for å estimere kausaleffekter på dette datautvalget.

Følgelig vil vi ikke bruke data fra Finnmark og Sogn og Fjordane i resten av oppgaven. Siden disse to fylkene kun har 23 offentlige videregående skoler, taper ikke mye informasjon av dette.

4.2.3 Deskriptiv statistikk av VGS-utvalg til analyse

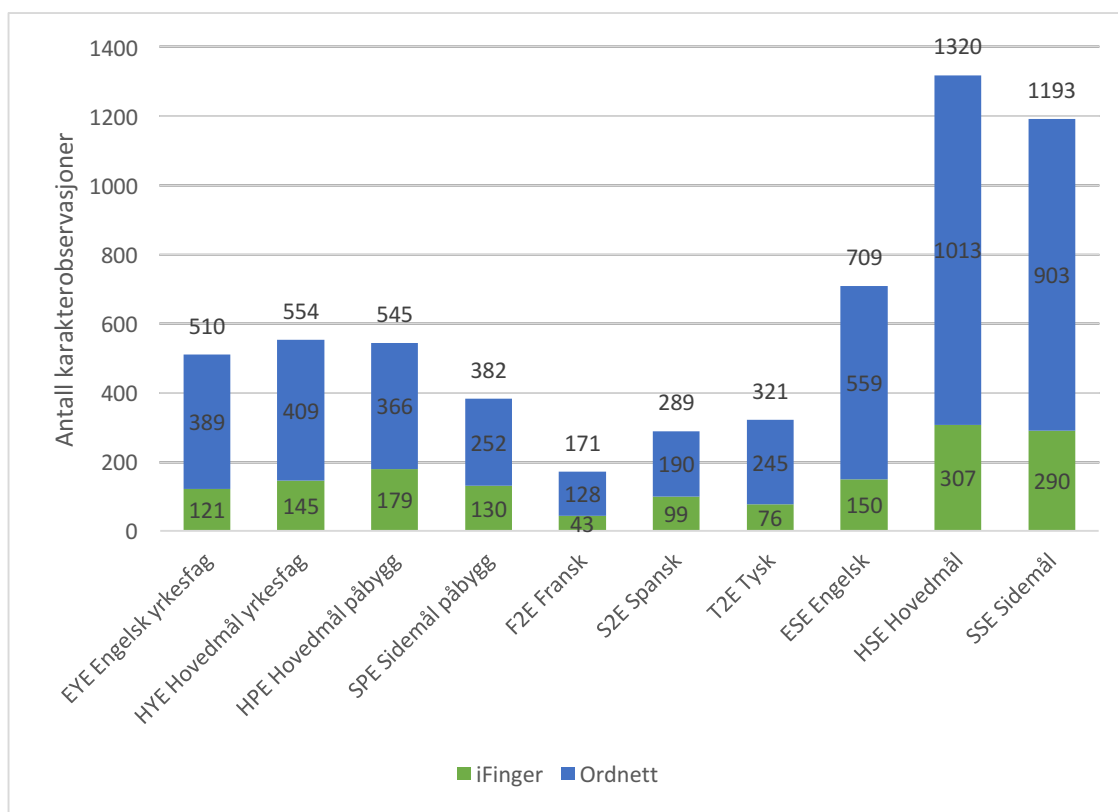
Vi står nå igjen med et utvalget skoler som er tilknyttet DD-spesifikasjonen $iFinger_{st}$, altså gruppe (3). Dette er som nevnt skoler som mellom 2007 og 2014 enten har hatt Ordnett eller iFinger. Observasjoner fra offentlige skoler i Nord-Trøndelag vil dog ikke bli benyttet, siden de i motsetning til resten av skolene ikke har noen perioder hvor vi kan sjekke for felles trender før treatment, ref. Figur 4.B. Dette reduserer bredden i utvalget med 11 skoler. Fylkene som ikke er med i utvalget vi anvender videre i oppgaven er altså Finnmark, Nord-Trøndelag, Oslo, Sogn og Fjordane og Telemark.

Vi har da et utvalg på 274 skoler videregående skoler som vil bli benyttet i analysedelen. 116 av disse skolene får iFinger i løpet av tidsperioden, og er treatment-gruppen. De 158 skolene som gjennom hele tidsperioden bruker Ordnett er kontrollgruppen.

Vi har som nevnt definert to populasjoner av VGS-elever i denne oppgaven; yrkesfag- og påbyggelever og elever ved studiespesialiserende linjer. Med utgangspunkt i et datasett som dekker begge populasjonene, har vi i utvalget til analyse *utelatt* fylker av alle typer demografi, fra urbaniserte Oslo til distriktspregede Finnmark. Privatskoler, som stort sett er private grunnet religiøse filosofier eller toppidrettslinjer, er også utelatt. Vi har ikke grunn til å tro at elever her har spesielle faglige egenskaper. Derfor tror vi at de 274 skolene som benyttes i analysedelen gir oss to utvalg som er representative for de to populasjonene.

Vi vil nå undersøke karakterobservasjonene fra disse videregående skolene i utvalget på samme måte som vi undersøkte ungdomsskole-utvalget.

Figur 4.D – Antall observasjoner av eksamenskarakterer betinget på treatment-tilstand VGS



I Figur 4.D ser vi at VGS-eksamener har stor variasjon i antall observasjoner. Dette har forklaring i fagenes natur. Eksamenene i tredjespråk er rapportert sjeldent. Dette er et resultat av at færre elever per skole tar disse fagene, noe som øker sannsynligheten for at ingen elever på en gitt skole et gitt år blir trukket ut til en slik eksamen. Samtlige elever som tar studiespesialiserende linjer må opp i hovedmål i 3. klasse, noe som fører til at karakterer i dette faget er rapportert flest ganger. Studiespesialiserende sidemål er rapportert noe sjeldnere siden det er et trekkfag på VG3. Det er imidlertid et fag alle elever på studiespesialiserende linjer må ta. Det gjelder også engelsk, men dette er et trekkfag i på VG1, hvor færre elever blir trukket ut enn på VG3. Yrkesfag- og påbygg-eksamenene er rapportert omtrent 500 ganger hver. Dersom alle skolene hadde rapportert alle fag i alle tidsperioder, ville hver stolpe i Figur 4.D bestått av $274 \times 8 = 2192$ observasjoner. Vi har derfor mange tomme karakterobservasjoner. Antallet observasjoner for de forskjellige fagene er imidlertid gjennomgående høyt. Vi merker oss dog at tredjespråkene har betraktelig færre observasjoner enn de andre fagene.

Vi ser av Figur 4.D at antall karakterobservasjoner fra treatment-gruppen etter de har fått treatment (merket iFinger) er betydelig lavere enn antall karakterobservasjoner fra kontrollgruppen og treatment-gruppen før treatment. Dette er en naturlig konsekvens av at alle skolene vi benytter har Ordnett når vårt datasett begynner i 2008. iFinger innføres gradvis, som vist i Figur 4.B. En implikasjon av dette er at vi har et betydelig antall observasjoner før treatment som kan anvendes til å sjekke for felles trender.

Vi så på plottene av karakterutviklinger i Figur 3.B til Figur 3.D at dataseriene for hvert fag er av forskjellige lengder. Alle dataserier strekker seg frem til 2014, men for forskjellige fag har vi første observasjon på forskjellige år. Første observasjon av engelsk yrkesfag, hovedmål yrkesfag og påbygg-fagene er i henholdsvis 2008, 2009 og 2012. Tredjespråkene er observert først i 2008. Engelsk studiespesialiserende er observert først i 2007, mens hovedmål studiespesialiserende og sidemål studiespesialiserende er observert først i 2009. Som vi ser av Figur 4.B har samtlige fylkeskommuner i utvalget vårt en observasjon av $iFinger_{st}$ fra og med 2008, og iFinger-treatment skjer mellom 2010 og 2013. Følgelig er det naturlig å sjekke for felles trender i 2 år før treatment for VGS-eksamenene. På denne måten vil samtlige fag få et estimat av Lead1 og Lead2 basert på et høyt antall observasjoner. Vi merker oss at påbygg-fagene kun kan sjekkes for felles trender i én periode. Vi kommer tilbake til implikasjonene av dette i analysen.

Vi vil nå undersøke karakterobservasjonene til utvalget ytterligere, her delt inn etter treatment- og kontrollgrupper.

Tabell 4.B – Deskriptiv karakterstatistikk: VGS

EYE, engelsk yrkesfag	Observasjoner	Gjennomsnitt	Standardavvik	Min	Max
Totalt	510	2,865	0,569	1,5	4,5
Treatment-gruppe	220	2,941	0,530	1,7	4,5
Kontrollgruppe	290	2,806	0,592	1,5	4,3
Differanse	-70	0,135	-0,062	0,2	0,2

HYE, hovedmål yrkesfag	Observasjoner	Gjennomsnitt	Standardavvik	Min	Max
Totalt	554	3,310	0,476	1,8	5,3
Treatment-gruppe	248	3,307	0,458	2	4,8
Kontrollgruppe	306	3,313	0,491	1,8	5,3
<i>Differanse</i>	<i>-58</i>	<i>-0,006</i>	<i>-0,033</i>	<i>0,2</i>	<i>-0,5</i>

HPE, hovedmål påbygg

Totalt	545	2,717	0,371	1,8	4,2
Treatment-gruppe	229	2,750	0,377	1,9	4,2
Kontrollgruppe	316	2,692	0,365	1,8	3,8
<i>Differanse</i>	<i>-87</i>	<i>0,058</i>	<i>0,012</i>	<i>0,1</i>	<i>0,4</i>

SPE, sidemål påbygg

Totalt	382	2,615	0,425	1,5	4,2
Treatment-gruppe	167	2,674	0,411	1,6	4
Kontrollgruppe	215	2,570	0,432	1,5	4,2
<i>Differanse</i>	<i>-48</i>	<i>0,104</i>	<i>-0,021</i>	<i>0,1</i>	<i>-0,2</i>

ESE, engelsk

studiespesialiserende

Totalt	709	3,557	0,465	2,2	5,1
Treatment-gruppe	336	3,590	0,472	2,2	4,9
Kontrollgruppe	373	3,528	0,458	2,3	5,1
<i>Differanse</i>	<i>-37</i>	<i>0,061</i>	<i>0,014</i>	<i>-0,1</i>	<i>-0,2</i>

HSE, hovedmål

studiespesialiserende

Totalt	1 320	3,153	0,391	1,8	4,6
Treatment-gruppe	554	3,195	0,396	1,9	4,6
Kontrollgruppe	766	3,123	0,385	1,8	4,5
<i>Differanse</i>	<i>-212</i>	<i>0,072</i>	<i>0,011</i>	<i>0,1</i>	<i>0,1</i>

SSE, sidemål					
studiespesialiserende	Observasjoner	Gjennomsnitt	Standardavvik	Min	Max
Totalt	1 193	3,012	0,430	1,4	4,3
Treatment-gruppe	515	3,071	0,447	1,5	4,2
Kontrollgruppe	678	2,967	0,411	1,4	4,3
<i>Differanse</i>	<i>-163</i>	<i>0,104</i>	<i>0,036</i>	<i>0,1</i>	<i>-0,1</i>
F2E, fransk					
Totalt	171	3,112	0,627	1,5	4,9
Treatment-gruppe	92	3,108	0,644	1,5	4,6
Kontrollgruppe	79	3,116	0,610	1,8	4,9
<i>Differanse</i>	<i>13</i>	<i>-0,009</i>	<i>0,034</i>	<i>-0,3</i>	<i>-0,3</i>
S2E, spansk					
Totalt	289	3,167	0,627	1,7	5,5
Treatment-gruppe	142	3,258	0,626	1,7	5
Kontrollgruppe	147	3,080	0,618	1,7	5,5
<i>Differanse</i>	<i>-5</i>	<i>0,178</i>	<i>0,009</i>	<i>0</i>	<i>-0,5</i>
T2E, tysk					
Totalt	321	3,079	0,574	1,5	4,8
Treatment-gruppe	156	2,991	0,635	1,5	4,7
Kontrollgruppe	165	3,161	0,499	1,6	4,8
<i>Differanse</i>	<i>-9</i>	<i>-0,170</i>	<i>0,136</i>	<i>-0,1</i>	<i>-0,1</i>

Som vi ser av Tabell 4.B er antall karakterobservasjoner i treatment-gruppen gjennomgående mindre enn i kontrollgruppen, unntatt for fransk. Dette er i tråd med at den totale treatment-gruppen er på 158 skoler, mens kontrollgruppen er på 116 skoler. Standardavvikene ligger mellom 0,371 og 0,627. Karakterobservasjonene for tredjespråkene (fransk, spansk og tysk) har generelt høyere standardavvik enn de andre fagene, med opp mot 0,627. Dette er trolig en konsekvens av at det er et lavere antall elever per skole som

står bak hver karakterobservasjon. Til sammenlikning har hovedmål et standardavvik på 0,391, som er betraktelig lavere.

Som for ungdomsskoleutvalget, observerer vi også her at treatment-gruppen tenderer til å ha høyere karakterer enn kontrollgruppen. I analysen vil vi da prøve å skille karakterforskjeller som er en potensiell kausaleffekt av iFinger og karakterforskjeller som kommer av utvalgsskjevheter.

5 Statistisk analyse

Her vil vi presentere statistiske utskrifter fra DD og Granger-testene. I alle de kommende modellene benytter vi standardiserte karakterdata. Hvert fag har da gjennomsnittskarakter 0 og et standardavvik lik 1. Koeffisientestimatene i regresjonsutskriftene tolkes som antall standardavvik.

I regresjonene vil vi clustre standardfeilene på skolenivå. Grunnen til dette er at det trolig forekommer seriekorrelasjon langs tidsseriene til hver skole; hvis en skole har høye karakterer i et gitt år, har den trolig høye karakterer det neste året også. Videre kan hver skole få eksogene sjokk som gir utslag i karakterer over flere perioder, som for eksempel en ny skolebygning eller nye faglærere. Grunnen til at vi ikke clusterer på et høyere nivå er at de enkelte skolenes seriekorrelasjon trolig ikke er korrelert med hverandre. Det er lite sannsynlig at det forekommer sjokk eller systematiske utslag som gjenspeiles i karakterer over flere perioder, og er felles for alle skoler i en kommune eller et fylke.

Videre vil vi rapportere regresjonsutskrifter i egne tabeller for hver populasjon vi analyserer. Vi har da egne tabeller for ungdomsskoleeksamener, påbygg –og yrkesfagseksamener og eksamener på studiespesialiserende linjer. Utover dette vil vi presentere ytterligere figurer for spesifikke eksamener når disse viser interessante funn. Når tilsvarende figurer for andre eksamener ikke viser interessante funn er de plassert i appendiks.

Karakterskalaen har et forholdsvis snevert utfallsrom fra 1 til 6. En implikasjon av dette er at det skal mye til for å øke en elevs karakter fra 3 til 4. Hadde skalaen for eksempel gått fra 1 til 100 ville flere nyanser kommet frem av datasettet. Siden det kan være vanskelig å avdekke slike små karaktervariasjoner vil vi i utgangspunktet anvende et 10%-signifikansnivå.

5.1 Test av "felles trend"-forutsetningen

Vi vil først gjøre Granger-testen for å sjekke om vi kan motbevise hypotesen om felles trender for hvert fags karakterserier. Dette er som kjent en kritisk forutsetning for at vi skal kunne tolke effekten av iFinger kausalt. Her er vi interessert i Leads-variablene. Som forklart

i 4.1.2 og 4.2.3 vil vi estimere 4 og 2 leads for henholdsvis US-fag og VGS-fag. Dersom samtlige leads er ikke-signifikante, kan vi ikke avvise en hypotese om at treatment- og kontrollgruppene har felles trender før treatment, gitt skole- og årsspesifikke effekter. For hvert fag vil vi drøfte t-tester om hver enkelt lead er ikke-signifikant, og F-tester for om samtlige leads er ikke-signifikante.

En variabel for treatment-år (Lag0) og variabler for de tre påfølgende periodene (Lag1-Lag3) er også inkludert i modellene. Lag-estimatene, som ved felles trender kan tolkes kausalt, vil drøftes i del 5.2.2.

5.1.1 Felles trender: Ungdomsskole

Tabell 5.A – Granger-test: Ungdomsskole

	(1) EUE Engelsk	(2) HUE Hovedmål	(3) SUE Sidemål
Lead4	-0.285 (0.242)	0.352 (0.263)	0.461 (0.355)
Lead3	-0.676** (0.301)	0.0685 (0.318)	0.171 (0.333)
Lead2	-0.213 (0.240)	0.438 (0.343)	0.236 (0.286)
Lead1	-0.529** (0.246)	0.190 (0.314)	0.135 (0.298)
Lag0	-0.401 (0.265)	0.436 (0.304)	0.259 (0.313)
Lag1	-0.272 (0.407)	0.373 (0.449)	0.719 (0.484)
Lag2	-0.917 (0.621)	0.277 (0.365)	0.0872 (0.440)
Lag3	-0.352 (0.575)	0.530 (0.777)	-0.129 (1.273)
r ²	0.429	0.459	0.394
N	512	519	491
Clustere	86	89	87
F-test leads, p-verdi	0.155	0.463	0.768

Standardfeil i parentes

MERK: Alle modeller har dummies for skole og år

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Vi ser at engelsk har to signifikante leads, hvilket tilsier at vi kan forkaste en hypotese om at gruppene er like før treatment gitt skole og år-spesifikke effekter. F-testen har p-verdi 0,155 som kunne tilsagt felles trend, men med to klart signifikante leads anser vi felles trend-forutsetningen for brutt. Det ser ut til at treatment-gruppen har dårligere karakterer enn

kontrollgruppen i år 1 og 3 før treatment, og at dette forårsakes av effekter vi ikke kontrollerer for. Følgelig kan ikke DD-estimerer eller lags for dette faget tolkes kausalt.

For hovedmål og sidemål har vi ingen signifikante leads, og begge har høye p-verdier for F-test av leads. Følgelig kan vi ikke forkaste en hypotese om at treatment-gruppen og kontrollgruppen for disse fagene er like før treatment, gitt skole- og årsspesifikke effekter. Disse resultatene tilsier at det er felles trender mellom gruppene før treatment.

5.1.2 Felles trender: Videregående skole

Tabell 5.B – Granger-test: Yrkesfag og påbygg VGS

	(1) EYE Engelsk yrkesfag	(2) HYE Hovedmål yrkesfag	(3) HPE Hovedmål påbygg	(4) SPE Sidemål påbygg
Lead2	-0.0568 (0.272)	0.0765 (0.276)	0 (.)	0 (.)
Lead1	-0.268 (0.255)	-0.233 (0.291)	-0.344 (0.429)	0.557 (0.582)
Lag0	-0.0822 (0.286)	-0.120 (0.252)	-0.202 (0.452)	0.670 (0.539)
Lag1	-0.113 (0.266)	-0.302 (0.196)	-0.0572 (0.244)	0.0855 (0.337)
Lag2	0.0718 (0.373)	0.173 (0.361)	0.0775 (0.199)	-0.0708 (0.284)
Lag3	-0.160 (0.388)	0.352 (0.357)	0.128 (0.336)	-0.0327 (0.395)
r2	0.568	0.449	0.632	0.689
N	510	554	545	382
Clustere	180	180	191	163
F-test leads, p-verdi	0.567	0.596	kun én lead	kun én lead

Standardfeil i parentes

MERK: Alle modeller har dummies for skole og år

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Vi ser at det ikke finnes noen signifikante leads for yrkesfagene og vi har høye p-verdier for F-test på leads. Med samme logikk som for ungdomsskole tilsier dette at vi har felles trender. Vi merker oss at vi ikke har noen observasjoner for Lead2 til påbyggfagene, som vi også så i deskriptiv statistikk. Vi har da ingen god test for om disse fagene har felles trender og kausale tolkninger av disse fagene vil da være svært usikre.

Tabell 5.C – Granger-test: Studiespesialiserende VGS

	(1) ESE Engelsk	(2) HSE Hovedmål	(3) SSE Sidemål	(4) F2E Fransk	(5) S2E Spansk	(6) T2E Tysk
Lead2	0.157 (0.171)	-0.205** (0.0965)	-0.0350 (0.122)	-0.174 (0.433)	0.574 (0.546)	-0.231 (0.314)
Lead1	0.0246 (0.180)	-0.120 (0.0919)	-0.241** (0.112)	-0.0574 (0.475)	-0.132 (0.504)	-0.0606 (0.329)
Lag0	0.0790 (0.159)	-0.0947 (0.0909)	-0.210* (0.112)	0.205 (0.447)	0.276 (0.362)	0.347 (0.236)
Lag1	0.369* (0.221)	-0.0598 (0.109)	-0.0166 (0.114)	-0.0913 (0.413)	0.205 (0.282)	0.264 (0.350)
Lag2	0.338 (0.209)	-0.0723 (0.109)	-0.132 (0.135)	-0.0241 (0.527)	0.356 (0.297)	0.915** (0.410)
Lag3	0.345 (0.221)	-0.128 (0.176)	-0.202 (0.159)	-0.237 (0.531)	0.626** (0.283)	0.820** (0.384)
r ²	0.518	0.698	0.651	0.689	0.698	0.649
N	791	1320	1193	171	289	321
Clustere	201	251	241	69	112	114
F-test leads, p- verdi	0.629	0.079*	0.084*	0.920	0.285	0.759

Standardfeil i parentes

MERK: Alle modeller har dummies for skole og år

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

For hovedmål og sidemål på studiespesialiserende linjer ser vi at det finnes to signifikante lead-variabler. Begge har også lave p-verdier for F-test på leads. Vi har altså ikke felles trender for disse fagene. Utover dette har ingen av de andre fagene signifikante leads, noe som indikerer felles trender.

Følgende fag ser da ut til å ha felles trender før treatment:

Ungdomsskole: Hovedmål (HUE) og sidemål (SUE)

VGS, yrkesfag og påbygg: Engelsk yrkesfag (EYE), hovedmål yrkesfag (HYE), hovedmål påbygg (HPE) og sidemål påbygg (SPE)

VGS, studiespesialiserende: Engelsk (ESE), fransk (F2E), spansk (S2E) og tysk (T2E)

5.2 Modellenes estimater på effekten av iFinger

I denne delen vil vi se på estimater for DD-spesifikasjonene og lags fra Granger-testene. Lagvariablene skiller seg som sagt fra DD-spesifikasjonene ved at de viser hvordan eventuelle kausale effekter utvikler seg over tid.

Denne kommende presentasjonen vil være generisk. Testenes validitet og hvorvidt resultatene gir mening i lys av logikk og fakta vil drøftes i seksjon 6.

5.2.1 Differences-in-differences

Vi vil nå presentere DD-resultatene for samtlige fag. Det er svært viktig å merke seg at estimatene for fagene som ikke har felles trender *ikke* kan tolkes kausalt.

5.2.1.1 DD: Ungdomsskole

Tabell 5.D – DD: Ungdomsskole

	(1) EUE Engelsk	(2) HUE Hovedmål	(3) SUE Sidemål
iFingerUS	-0.0431 (0.215)	0.174 (0.170)	0.104 (0.187)
r ²	0.416	0.453	0.386
N	512	519	491
Clustere	86	89	87

Standardfeil i parentes

MERK: Alle modeller har dummies for skole og år

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

For hovedmål og sidemål, som har felles trender, kan vi ikke forkaste en hypotese om at det ikke finnes en kausal sammenheng mellom iFinger og eksamenskarakterer.

Engelsk har ikke felles trender før treatment. Derfor kan *ikke* DD-estimatet tolkes som at det ikke er noen kausal sammenheng.

5.2.1.2 DD: Yrkesfag og påbygg VGS

Tabell 5.E – DD: Yrkesfag og påbygg VGS

	(1) EYE Engelsk yrkesfag	(2) HYE Hovedmål yrkesfag	(3) HPE Hovedmål påbygg	(4) SPE Sidemål påbygg
iFinger	0.0331 (0.201)	-0.117 (0.198)	0.161 (0.167)	0.0701 (0.203)
r ²	0.565	0.438	0.632	0.682
N	510	554	545	382
Clustere	180	180	191	163

Standardfeil i parentes

MERK: Alle modeller har dummies for skole og år

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

For yrkesfagene, hvor vi hadde felles trender, finner vi ingen kausale effekter av iFinger på eksamenskarakterer. For påbyggfagene, hvor vi kun har svært svake indikasjoner på felles trender, finner vi heller ingen effekter av iFinger.

5.2.1.3 DD: Studiespesialiserende VGS

Tabell 5.F – DD: Studiespesialiserende VGS

	(1)	(2)	(3)	(4)	(5)	(6)
	ESE	HSE	SSE	F2E	S2E	T2E
	Engelsk	Hovedmål	Sidemål	Fransk	Spansk	Tysk
iFinger	0.232*	0.0223	0.0578	-0.0330	0.105	0.591**
	(0.134)	(0.0818)	(0.0811)	(0.283)	(0.281)	(0.235)
r ²	0.522	0.697	0.647	0.684	0.685	0.632
N	709	1320	1193	171	289	321
Clustere	196	251	241	69	112	114

Standardfeil i parentes

MERK: Alle modeller har dummies for skole og år

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Her ser vi at engelsk og tysk, hvor vi fant felles trender, har statistisk signifikante DD-estimer med t-verdiene 1,73 og 2,51. Estimaten tilsier da at innføringen av iFinger øker karakterene i ESE og T2E med henholdsvis 0,232 og 0,591 standardavvik.

Fransk og tysk har felles trender. De ikke-signifikante DD-spesifikasjonene for disse fagene indikerer da at det ikke er noen kausal effekt av iFinger.

For hovedmål og sidemål fant vi ingen felles trender, slik at DD-estimatene til disse fagene ikke kan si oss noe om kausale sammenhenger.

Følgende fag har da signifikant estimat av DD-spesifikasjonen.

Ungdomsskole: Ingen

VGS, yrkesfag og påbygg: Ingen

VGS, studiespesialiserende: Engelsk (ESE) og tysk (T2E)

Både engelsk og tysk har felles trender før treatment.

5.2.2 Granger-test

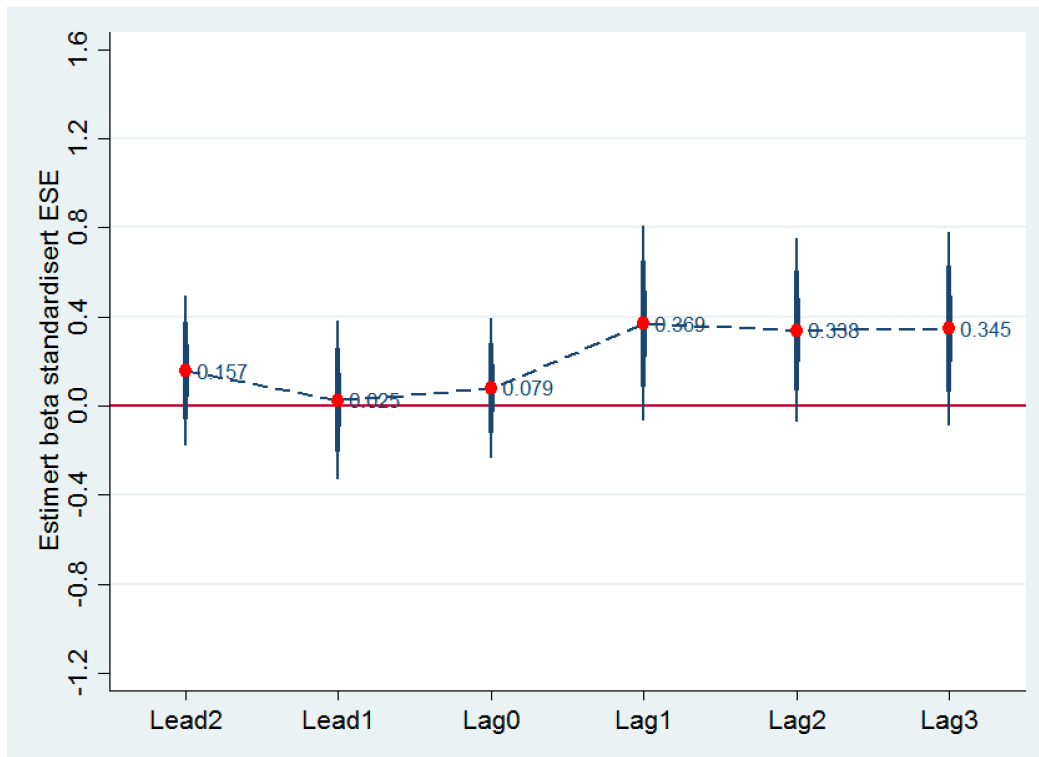
Vi vil her drøfte Lag-variablene i Granger-testene fra Tabell 5.A til Tabell 5.C.

Studiespesialiserende engelsk og tysk har signifikante lag-estimer, som er i samsvar med hva vi fant i DD-modellen. Dette styrker tilliten til de kausale funnene. Spansk har et signifikant Lag3-estimat med t-verdi 2,14 til tross for et ikke-signifikant DD-estimat med t-verdi 0,37. Dette øker usikkerheten til hvilken effekt iFinger har på eksamenskarakterer i spansk. Ingen av de resterende fagene har noen signifikante lag-estimer.

Siden Granger-testen lar oss undersøke den kausale effekten av iFinger over tid, vil vi se nærmere på lead- og lag-estimatene fra Tabell 5.C for engelsk, tysk og spansk. Disse er visualisert i Figur 5.A - Figur 5.C. under. Her ligger de respektive lead- og lag-variablene langs x-aksen med første lead lengst til venstre og siste lag lengst til høyre. Estimateret på koeffisienten til hver variabel leses av y-aksen. Vi minner her om at karakterdata er standardisert og at y-aksen leses i antall standardavvik differanse mellom treatment- og kontrollgruppe. De vertikale linjestykkene representerer standardfeilen til estimatene, hvor tykkelsen fra tynnest til tykkest viser standardfeil som svarer til henholdsvis 95%, 80% og 50% konfidensintervaller.

5.2.2.1 Granger-plot: Engelsk studiespesialiserende

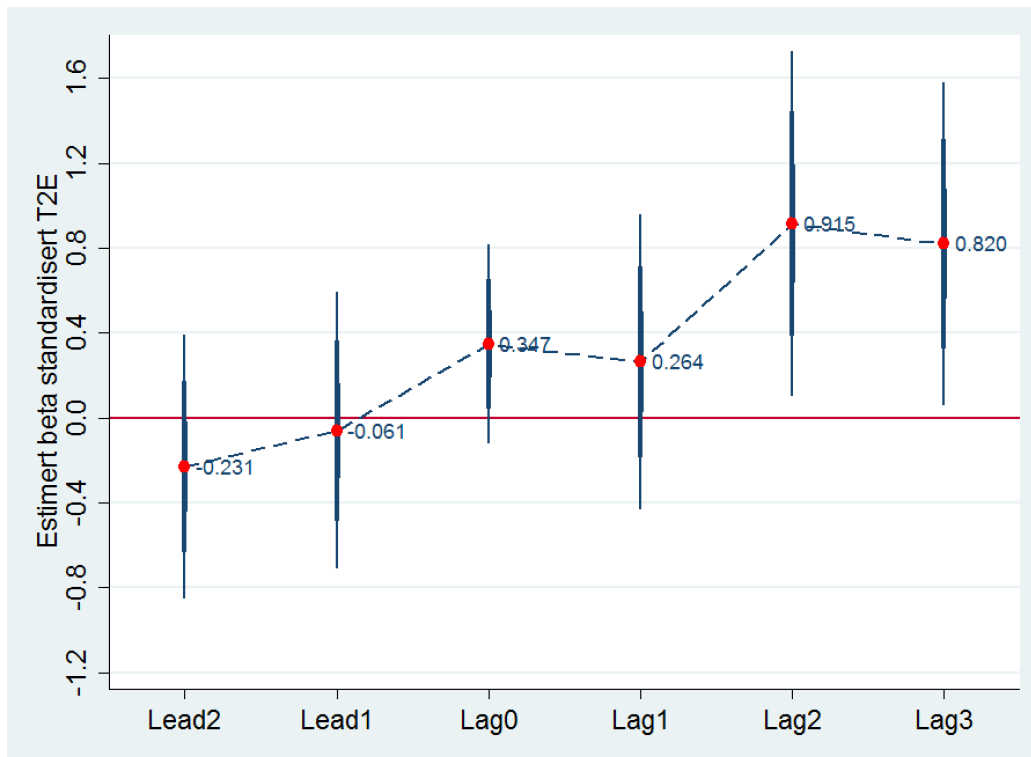
Figur 5.A – Granger-plot: Engelsk studiespesialiserende VGS



Som nevnt tidligere kan vi ikke forkaste en hypotese om felles trender. iFinger ser ikke ut til å gi noen effekt i treatment-året. Året etter treatment får imidlertid treatment-gruppen en økning i karakterer på 0,37 standardavvik sammenliknet med kontrollgruppen. Denne forskjellen er signifikant med t-verdi er 1,67. Effekten synes å vare, dog med gradvis mindre signifikante t-verdier på 1,62 og 1,56. Disse resultatene tilsier at det er en kausal effekt av iFinger på karakterer i engelsk studiespesialiserende eksamen, men at effekten ikke kommer i første treatment-år. Vi minner om at den estimerte DD-spesifikasjonen var 0,23 med t-verdi 1,73.

5.2.2.2 Granger-plot: Tysk 2

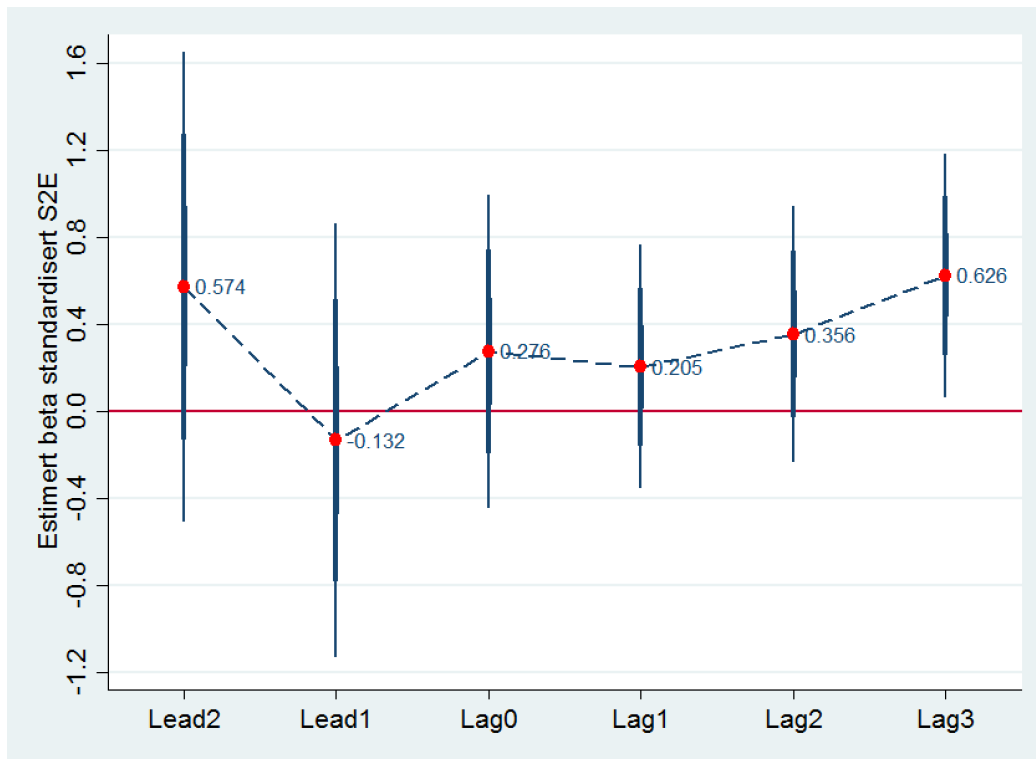
Figur 5.B – Granger-plot: Tysk 2 VGS



Heller ikke for tysk kan hypotesen om felles trender forkastes. Det ser ut til å være en økning på 0,347 standardavvik for treatment-gruppen allerede i treatment-året med t-verdi 1,47. Deretter faller den estimerte forskjellen noe og gir en ikke-signifikant Lag1 med t-verdi 0,75. Deretter ser vi en sterk og signifikant effekt av iFinger i år to og tre etter treatment med t-verdier på henholdsvis 2,23 og 2,14. Her ser effekten av iFinger ut til å være helt opp mot et helt standardavvik. Vi minner her om at den estimerte DD-spesifikasjonen var 0,59, med t-verdien 2,51.

5.2.2.3 Granger-plot: Spansk 2

Figur 5.C – Granger-plot: Spansk 2 VGS



Spansk, som også har felles trender, har gitt tvetydige estimerer på den kausale effekten av iFinger. Estimert DD-spesifikasjon er 0,105 med t-verdi 0,37. Estimert Lag3 i Granger er 0,626 med t-verdi 2,21. Lag0, Lag1 og Lag2 er ikke-signifikante med t-verdier på henholdsvis 0,76, 0,73 og 1,20. Lag-estimatene er gjennomgående positive og økende, noe som *kan* indikere en positiv effekt av iFinger. Her må en naturligvis være veldig oppmerksom på at dette ikke er signifikant; det er en forholdsvis høy sannsynlighet for at de estimerte forskjellene skyldes tilfeldigheter.

6 Kausale tolkninger i modellene

Vi vil nå gjøre en vurdering av den interne validiteten til modellene, og drøfte hvorvidt vi har funnet kausale effekter. De kausale drøftingene vil gjøres i lys av forutsetninger, kombinert med fakta og forståelse for de underliggende prosessene tilknyttet tildeling og bruk av iFinger.

6.1 Intern validitet

Vi vil nå drøfte hvorvidt våre design klarer å avdekke kausale effekter for populasjonene av interesse i vårt studie, altså om de er internt valide (Angrist & Pischke, 2009). Diskusjonene vil gjelde for alle regresjonene vi har kjørt.

6.1.1 Omitted variable bias

Estimatene i regresjonsutskriftene kan lide av OVB. Dette forekommer hvis vi har variabler som ikke er inkludert i modellene som påvirker karakterer og er korrelerte med innføringen av iFinger (se del 2.2.3.2). Dersom dette er tilfellet vil de rapporterte koeffisientene være over- eller underestimerte. Vi minner om at skole- og årssummies forklarer henholdsvis skole- og årsspesifikke effekter, hvilket reduserer antall potensielt utelatte variabler betydelig. Våre to største OVB-bekymringer er da innføringen av individuelle skole-PCer og IKT-basert eksamen.

Individuelle skole-PCer

Som vi så i del 1.4.4, synes *ikke* innføringen av individuelle skole-PCer å ha en effekt på akademiske resultater (Bulman & Fairlie, 2015). Når en går til innkjøp av individuelle skole-PCer er det naturlig at en vil ha passende programvare til. Dette tilsier at det kan være en korrelasjon mellom innkjøp av datautstyr og programvare. For VGS begynte de fleste fylkesvise skole-PC-ordningene mellom 2007 og 2009 (Knudsen, 2008). iFinger ble hovedsakelig innfaset fra 2010 og 2011. Det kan bety at det finnes en korrelasjon mellom innføringen av iFinger og individuelle skole-PCer.

Selv om det kan være en korrelasjon mellom innføringen av iFinger og individuelle skole-PCer, er vi ikke veldig bekymret for OVB i VGS-estimatene på grunn av at individuelle skole-PCer trolig ikke påvirker karakterer.

Vi har ikke informasjon om innføringen av individuelle skole-PCer for ungdomsskolene i utvalget vårt, men er ikke bekymret for OVB av samme grunn som for VGS.

IKT-basert eksamen

IKT-basert eksamen betyr at elever kan benytte PC under eksamen i stedet for å bruke penn og papir.

Det er grunn til å tro at IKT-basert eksamen har en positiv effekt på *eksamensprestasjon*. Det er for eksempel lettere å strukturere en besvarelse digitalt. Om det fører til bedre karakterer er likevel usikkert. Utdanningsdirektoratet har strenge retningslinjer for at eksamen skal være rettferdig (Utdanningsdirektoratet, 2013), og en gradvis innføring av IKT-basert eksamen gir kandidater ulike forutsetninger for å skrive gode eksamensbesvarelser. Sensorene burde da øke kravene til blant annet struktur når de retter en digitalt levert besvarelse. I så fall vil ikke IKT-basert eksamen ha effekt på *karakterer*.

Vi har vage indikasjoner på at innføringen av IKT-basert eksamen på VGS og iFinger samvarierer. iFinger ble hovedsakelig innført fra 2010 og 2011 på VGS og i 2015 for US. IKT-basert eksamen har vært valgfritt for videregående skoler siden 2008. 80% av besvarelsene i hovedmål studiespesialiserende ble imidlertid levert digitalt i 2011, og fra 2012 var IKT-basert eksamen planlagt å være normen for alle skriftlige språkeksamener (Skolelederforbundet, 2011). Vi har ikke informasjon om innfasingen av IKT-basert eksamen for ungdomsskoler.

Vi er ikke særlig bekymret for OVB som følge av IKT-basert eksamen for verken VGS eller US. Bakgrunnen for dette er Utdanningsdirektoratets sterke ønske om rettferdighet ved sensur, slik at IKT-basert eksamen ikke vil påvirke karakterer.

Vi er da relativt sikre på at våre estimer av effekten av iFinger ikke lider av OVB.

6.1.2 LATE

Det er viktig å merke seg at våre datasett kun inneholder informasjon om når skoler har *tilgang* til iFinger. Vi har ikke informasjon om faktisk bruk av programvaren. Følgelig er resultatene våre på "reduced form", som estimerer effekten av "intention to treat" (se 2.2.2). Det kan være mange ledd mellom sentrale innkjøpsorgan i fylkene, hver enkelt lærers omfavelse av digitale ordbøker og elevenes "take-up". Dette tilsier at fylkenes innkjøp av iFinger ikke nødvendigvis betyr at alle elever i fylkeskommunen bruker programvaren. Det samme argumentet gjelder for US, dog med potensielt færre ledd mellom innkjøpsorgan og lærer.

For å estimere LATE, den kausale effekten av å *bruke* iFinger, må vi dele "reduced-form"-estimatene på andelen elever som faktisk bruker iFinger ved hver skole. Denne andelen tilsvarer "first stage", og er alltid 100% eller lavere. Dersom iFinger fører til bedre karakterer vil elever i treatment-gruppen som ikke bruker iFinger trekke den estimerte effekten i "reduced form" ned. Vi har grunn til å tro at "first stage" i vårt tilfelle er godt under 100% (Språkrådet, 2014). Hvis for eksempel halvparten av elevene ved hver treated skole brukte iFinger, ville koeffisientene til "LATE" vært dobbelt så høye som "reduced-form"-estimatene vi har rapportert. Vi har ikke presise data for "first stage", og kan derfor ikke estimere effekten av å faktisk bruke iFinger. Vi bemerker i stedet at våre reduced form-estimerer er en nedre grense for "LATE".

Vi merker oss i tillegg at faktisk bruk i realiteten ikke er et binært fenomen. Enkelte elever kan søke et par ganger mens andre søker hundrevis av ganger. Et forholdstall mellom antall søk og antall elever per skole kunne blitt inkludert i "first stage" for å få korrekt estimat av "LATE". Den prinsipielle diskusjonen om forholdet mellom bruk og tilgang gjort i avsnittene ovenfor gjelder imidlertid uavhengig av denne nyanseringen.

Til tross for "LATE"-diskusjonen over vil vi presisere at estimerer på "reduced form" er av høy praktisk verdi. Brukervennligheten til iFinger-programvaren vil nemlig påvirke omfanget av bruk. En brukervennlig programvare vil bli brukt mer enn en programvare som ikke er

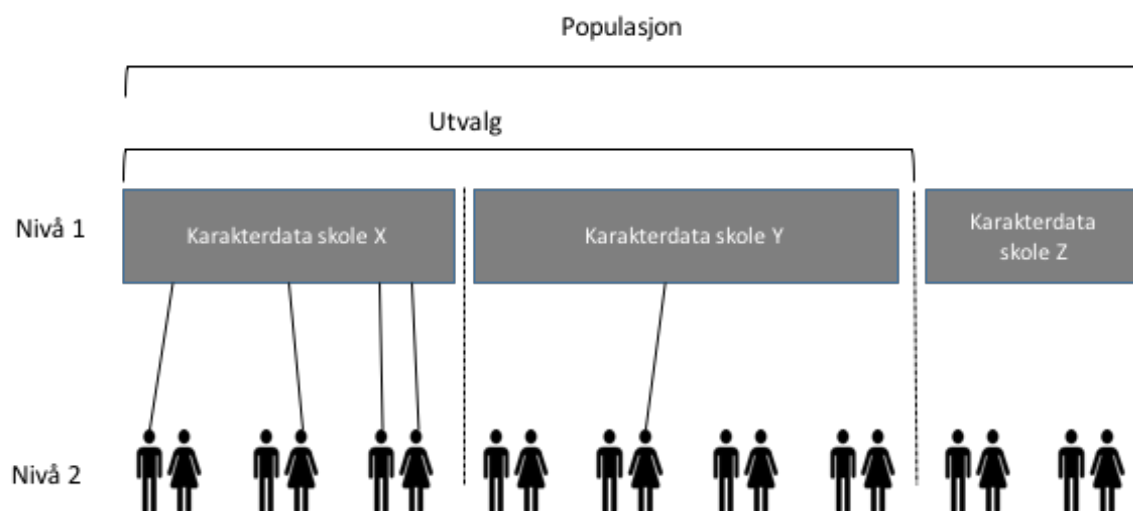
brukervennlig. Et "LATE"-estimat tar ikke hensyn til dette. Derfor vil "intention to treat"-estimatene være viktige for beslutningstakere; når de kjøper programvaren har de nettopp en *intention to treat*.

6.1.3 Utvalg og utvalgsstørrelse

Som forklart i del 2.4 krever vi 42 clustere per regresjon for å ha tillit til standardfeilene og inferensen. Med flere clustere kan vi stole mer på inferensen fra regresjonene, gitt at utvalget er representativt. Denne reduserte usikkerheten kommer til uttrykk ved at standardfeilene blir mindre.

Vi har som nevnt definert tre populasjoner av elever: Ungdomsskoleelever, yrkes- og påbyggelever og elever som går studiespesialiserende. Vi merker oss først at for samtlige populasjoner er eksamenstrekket tilfeldig. Vi sitter da igjen med to forhold som må adresseres. (1) Om skolene vi anvender i modellene er representative for populasjonen. I figur Figur 6.A (under dette avsnittet) betyr dette om skole X og Y representerer populasjonen. (2) Videre må gruppene med tilfeldig valgte eksamenskandidater være store nok til å være representative for skolene de tilhører. I Figur 6.A betyr dette om elevene som er trukket ut til eksamen for skole Y er representativ for de andre elevene på skole Y. Dersom vi kan bekrefte disse to punktene, vil vi ha tillit til inferensen i regresjoner når de er gjort med minst 42 clustere.

Figur 6.A - Illustrasjon av observasjonsnivå



Figuren illustrer hvordan observasjonene våre er bygget opp. Vi har eksamenskarakterobservasjoner på nivå 1. Hver observasjon på nivå 1 er gjennomsnittet av resultatene for eksamenskandidater på nivå 2.

Strekene fra individene illustrerer hvem som er trukket ut til eksamen.

Ungdomsskole

Som forklart i deskriptiv statistikk del 4.1 tror vi at utvalget ungdomsskoler er representativt for populasjonen. At utvalget er gjort tilnærmet tilfeldig og uavhengig av denne oppgaven, styrker troen på dette. Videre er US-fagene vi ser på obligatoriske, slik at mange elever trekkes ut til eksamen per skole. For ungdomsskolefagene engelsk, hovedmål og sidemål har vi henholdsvis 86, 89 og 82 clustere. Følgelig stoler vi på inferensen i regresjonsutskriftene.

VGS

Vi forklarte i del 4.2.3 at vi tror utvalget videregående skoler er representativt for alle norske videregående skoler, som dekker begge populasjonene. Derfor anser vi forhold (1) som oppfylt for VGS.

VGS, yrkesfag- og påbygg

Fagene vi ser på er obligatoriske, og trekket gjøres blant alle elever på disse linjene. Vi har derfor grunn til å tro at utvalget elever bak hver eksamensobservasjon for yrkesfag- og påbygg er representativt for skolen. Når vi i tillegg har mellom 150 og 200 clustere i hver regresjon, har vi grunn til å stole på den statistiske inferensen.

VGS, studiespesialiserende linjer

Siden hovedmål-, sidemål- og engelskeksamener trekkes ut på samme måte som yrkesfag- og påbyggeksamener, tror vi at utvalgene våre representerer skolene godt. Med over 200 clustere for hvert fag har vi grunn til å stole på den statistiske inferensen også for disse fagene.

Forhold (2) er ikke nødvendigvis oppfylt for tredjespråkene fransk, spansk og tysk. Disse fagene er valgfrie, og følgelig tar færre elever per skole disse fagene. Vi er da bekymret for at det ligger få eksamensavleggelser bak hver karakterobservasjon. Denne bekymringen stammer fra de store standardavvikene i karakterobservasjonene som vi så i 4.2.3. Konsekvensen av dette kan være at karakterdataen vi besitter ikke er representativ for skolen de tilhører. Utdanningsdirektoratet presiserer selv at dette kan være tilfellet (Utdanningsdirektoratet, 2010). Antall clustere for de tre fagene er henholdsvis 69, 112 og 114, hvilket burde være tilstrekkelig til å kunne stole på våre t-statistikker, *dersom* gjennomsnittskarakterene er representative for skolene de stammer fra.

For alle fag, utenom tredjespråk VGS, føler vi oss trygge på at vi besitter representative utvalg med nok clustere til å stole på inferensen. For tredjespråk kan det være at karakterdataen vi besitter ikke nødvendigvis er representativ for de tilhørende skolene. Vi vil diskutere dette nærmere i del 6.6.

6.1.4 Multipelt sammenligningsproblem

Fra datasettet bruker vi 13 forskjellige utfallsvariabler som representerer 13 forskjellige eksamenskarakterer. Når man gjør mange hypotesetester på utfallsvariabler kan vi få et problem med p-verdiene våre. Dette kommer av at det å gjøre mange tester øker faren for at en finner signifikante forskjeller mellom treatment og kontrollgruppe som egentlig bare er tilfeldigheter. Sjansen for å gjøre type 1-feil øker. I slike tilfeller kan en bruke forskjellige metoder til å korrigere p-verdiene. Jo høyere korrelasjon det er mellom utfallsvariablene jo mindre fare er det for å gjøre type 1-feil (Mckenzie, 2012).

I vårt tilfelle ser vi på tre forskjellige populasjoner: ungdomsskole-elever, yrkesfags-elever og elever på studiespesialiserende VGS. Disse har henholdsvis tre, fire og seks utfallsvariabler hver. Fordi vi ser på forskjellige populasjoner kan dette anses som forskjellige studier. Bekymringen rundt multiple hypoteser gjelder altså kun for maks seks utfallsvariabler av gangen. Videre er det grunn til å tro at utfallsvariablene i hver populasjon er forholdsvis sterkt korrelerte. Elever på en bestemt skole er ofte dyktige i flere enn bare et fag, hvilket betyr at det vil være korrelasjon mellom fagene.

På grunn av at vi sjekker maks seks hypoteser og fordi det er god grunn til å tro at det er høy korrelasjon mellom de forskjellige karakterene tror vi at vi ikke vil ha betydelig problemer med multippelt sammenligningsproblem.

6.2 Kausale tolkninger: Ungdomsskole

For hovedmål- og sidemålseksamen på US-nivå finner vi ingen signifikante effekter av iFinger. Siden vi har felles trender før treatment for disse fagene, kan vi påstå at det ikke er noen kausal effekt av iFinger på eksamenskarakterer i disse fagene. Engelsk har ikke felles trender før treatment. Følgelig kan vi ikke bekrefte eller avkrefte noen kausal effekt av iFinger for dette faget.

Når vi ikke finner kausale effekter av iFinger på US-nivå, er det lite overraskende sett i lys av diskusjonen i LATE-kapittelet 6.1.2. IKT-bruken på disse klassetrinnene er svært lav (Senter for IKT i utdanningen, 2013), og bruken av iFinger tilsvarende (Språkrådet, 2014). At DD og Granger-testen samsvarer styrker troen på funnene våre.

6.3 Kausale tolkninger: Yrkesfag og påbygg VGS

For yrkesfagene tilsier testene at det ikke finnes noen kausal sammenheng mellom innføring av iFinger og eksamenskarakterer i språkfag. Også dette er som forventet; undersøkelsen fra Språkrådet om digitale ordbøker viser at elever på yrkesfaglige studieretninger bruker digitale ordbøker markant mindre enn elever på studiespesialiserende linjer (Språkrådet, 2014).

For påbygg-fagene har vi indikasjoner på at iFinger ikke har noen kausale effekt på eksamenskarakterer i hovedmål og sidemål. Det kan tenkes at bruken også her er for lav til å avdekke en eventuell effekt. Vi presiserer at dette er svært usikre kausale påstander siden vi ikke kan bekrefte felles trender i mer enn én periode før treatment.

6.4 Kausale tolkninger: Obligatoriske fag studiespesialiserende VGS

Vi minner om at forutsetningen om felles trender ser ut til å holde for eksamenskarakterer i engelsk på studiespesialiserende linjer, mens den ikke gjør det for hovedmål og sidemål. Likevel vil vi nå foreta subsidiære diskusjoner av resultatene for hovedmål og sidemål, hvor vi diskuterer fraværet av signifikante funn gitt at forutsetningene hadde holdt.

Engelsk (ESE)

Forutsetningen om felles trender holder for engelsk skriftlig eksamen. Regresjonsanalysene kan da tolkes som at iFinger har en positiv kausal effekt på eksamenskarakterer i dette faget. I DD er den estimerte effekten 0,23 standardavvik, mens Granger-testen estimerer en effekt på ca. 0,35 standardavvik i gjennomsnitt. Dette tilsier en forbedring på henholdsvis 0,1 og 0,16 karakterpoeng, som er av klar praktisk betydning.

De engelske ordbøkene som tilbys fra Collins og Vega forlag til VGS er svært omfattende. Elevene får tilgang til mange og fylldige oppslag i English Dictionary, English Thesaurus og engelsk/norsk – norsk/engelsk. Det syntes rimelig at disse produktene kan løfte evnen til å tolke og skrive engelsk tekst for elever i VGS studiespesialiserende. Vi har mange observasjoner for engelskeksamen, hvilket gjør oss trygge på funnene. Vi merker oss at effekten ikke blir signifikant før året etter treatment. Dette er ikke overraskende, da det sannsynligvis vil ta noe tid for lærere og elever å vende seg til rutiner for fast bruk.

At både DD og Granger-testen estimerer signifikante effekter gjør oss trygge på de estimerte funnene.

Hovedmål (HSE)

Vi presiserer at hovedmål *ikke* har felles trender før treatment. Derfor kan ikke de ikke-signifikante estimatene av DD-spesifikasjonen eller lagsene i Granger si noe om kausal effekt.

I hovedmål, som er morsmålet til de fleste elevene, ligger trolig elevene på en så høy språklig forståelse at marginalnyttene av å slå opp i en ordbok er lavere enn for fremmedspråk. Som hjelp til hovedmål vil iFinger kunne tilby synonymer, ordforklaring og eksempler på kontekst for ord. Vurderingen i hovedmål går mye lenger i å kreve ferdigheter som refleksjon, setningsoppbygging, tekstopbygging og tolking av tekster. Derfor er nytten av iFinger trolig større når mer fundamental språkforståelse testes. Det er derfor ikke veldig overraskende *dersom* det ikke finnes en kausal effekt av iFinger på karakterer i hovedmål.

Vi presiserer igjen at vi ikke kan trekke kausale slutninger siden vi ikke har felles trender.

Sidemål (SSE)

Som for hovedmål har vi ikke felles trender for sidemål. Derfor kan vi heller ikke her tolke de ikke-signifikante estimatene som at iFinger ikke har kausal effekt på eksamenskarakterer.

Vi hadde forventet å finne en effekt av iFinger for sidemål. iFinger tilbyr en bokmål-nynorsk ordbok fra Cappelen Damm som er svært populær på skolemarkedet. Om iFinger mister en skole som kunde er det vanlig at flere elever selv kjøper denne spesifikke ordboken privat direkte fra iFinger. Dette tyder på at det er et produkt som brukes og som elevene opplever at kan hjelpe dem i sidemål. Vi merker oss her at vi ikke anvender data der et fylke går fra iFinger til Ordnett og følgelig vil ikke enkeltkjøp av iFinger skape problemer med treatment i kontrollgruppen vår.

Om iFinger har en kausal effekt på karakterer i sidemål eller ikke, kan vi ikke si noe om på grunn av brutt forutsetning om felles trender.

6.5 Kausale tolkninger: Tredjespråk VGS

Forutsetningen om felles trender ser ut til å være oppfylt for samtlige tredjespråk. Derfor kan DD-estimatene og Granger-lags tolkes kausalt. Vi merker oss at alle tredjespråkene er

språk hvor elevene ligger på et mer grunnleggende nivå enn i hovedmål og engelsk. Det skulle tilsi at en trolig har verdi av å bruke digitale ordbøker her.

Tysk

Tysk er faget hvor vi ser en klarest effekt av iFinger. DD estimerer at iFinger fører til betydelige 0,591 standardavvik økte karakterer, med t-verdi 2,51. I Granger-testen ser vi at denne effekten øker over tid. I treatment-året og året etter er iFinger estimert til å øke karakterene med ca. 0,3 standardavvik. Disse estimatene er dog ikke-signifikante. I år 2 og 3 etter innføringen av iFinger er den estimerte effekten ca. 0,9 standardavvik forbedring i karakterer, med t-verdier 2,23 og 2,14.

At DD-identifikasjonen er estimert til å ligge omtrent mellom de estimerte Granger-lagsene er forståelig, da DD estimerer effekten av iFinger over samtlige treatment-år. At det er samsvar mellom DD og Granger-testen styrker vår tro på at vi har avdekket en kausal sammenheng.

Tysk er et språk med relativ komplisert grammatikk, blant annet med fire kasus. Derfor syntes det rimelig at marginalnyttens av digitale ordbøker er høy for tysk. Likevel er den estimerte effekten på 0,9 standardavvik mistenkelig høy. Nesten et standardavvik bevegelse på karakterskalaen er ekstremt betydningsfullt, og det syntes rett og slett ikke plausibelt at digitale ordbøker skal føre til så kraftige endringer i karakterer.

Spansk

Spansk har gitt tvetydige statistiske resultater. DD-koeffisienten er insignifikant. Kun Lag3 på Granger-testen er signifikant, med en estimert økning på 0,626 standardavvik. Vi merker oss at Lag0, Lag1 og Lag2 har estimerer som ligger høyere enn kontrollgruppen, men at disse ikke er statistisk signifikante.

Om det faktisk finnes en kausal effekt her kan det være at det har tatt lengre tid før elever og lærere i spansk har tatt i bruk programvaren aktivt enn i tysk og engelsk. I så fall vil denne effekten komme så sent at DD-estimatet trekkes ned av de uendrede karakterene til treatment-gruppen de første årene etter treatment.

Vi har indikasjoner på at iFinger har en kausal effekt på karakterer i S2E, men at denne først inntreffer tre år etter treatment. Siden kun ett kausalt estimat er signifikant er denne indikasjonen imidlertid usikker.

Fransk

Det siste faget vi har sett på er fransk, hvor forutsetningen om felles trend også ser ut til å holde. Her finner vi verken signifikant estimat på DD-identifikasjonen eller signifikante lag-estimer. Vi har tilsynelatende ikke grunnlag for å si at det finnes en kausal effekt. Siden vi har avdekket en kausal effekt for tysk og engelsk samt en mulig effekt for spansk, er det noe overraskende at vi ikke finner noen effekt for fransk.

6.6 Forskjellene mellom fremmedspråkene

Vi vil nå diskutere funnene for fremmedspråkene nærmere. Grunnen til dette er at vi avdekker signifikante effekter av iFinger for engelsk og tysk, usikre effekter for spansk og ingen effekter for fransk. Vi ville forventet å finne de samme effektene for disse fagene, spesielt for tredjespråkene. Legg her merke til at alle fire språkene er *fremmedspråk*, mens tysk, fransk og spansk er *tredjespråk*.

Tabell 6.A – Oppsummering regresjonsresultater for fremmedspråk VGS

	ESE	T2E	S2E	F2E
	Engelsk	Tysk	Spansk	Fransk
Felles trend	x	x	x	x
Signifikante Granger-lags	x	x	x	
Signifikant DD-variabel	x	x		
Antall observasjoner	709	321	289	171
Antall clustere	201	114	112	69
Gj.snitt. kausalt estimat	0,35	0,67	0,40	-
Granger (målt i st.avvik)				
Kausalt estimat DD	0,23	0,59	-	-
(målt i st.avvik)				

En mulig forklaring på hvorfor tredjespråkene gir så forskjellige resultater kan være at de forskjellige elevene og lærerne er ulike mellom fagene. Mens engelsk, sidemål og hovedmål er felles for alle, velger den enkelte elev selv hvilket tredjespråk en ønsker å ta. Dette kan være av statistisk betydning fordi det kan forårsake systematiske forskjeller mellom utvalgene i hvert av disse fagene. Vi tror imidlertid ikke at forskjeller mellom elevene i de forskjellige språkene kan være så store at de forårsaker forskjellene vi ser i estimatene.

Som vi merket oss i del 3.3.4 er spredningen per karakter for tredjespråk stor, siden det trolig er få elever bak hver observasjon av skole-år gjennomsnitt. En mulig forklaring på forskjellene mellom språkene er da at de som har blitt trukket ut til tysk-eksamen tilfeldigvis har brukt iFinger mer enn de som har blitt trukket ut i fransk. Denne forklaringen virker plausibel og vi tror at denne skjevheten i bruk skyldes tilfeldigheter i hvem som har blitt trukket ut til hver enkelt eksamen, snarere enn at det er systematiske forskjeller på elever som tar tysk og fransk.

Vi tror altså at dataen vi besitter ikke gir oss et representative utvalg for elevene som har tredjespråk ved hver skole. Det kan da hende at den høye *estimerte* effekten på 0,6 i tysk er

noe høyere enn den *faktiske* effekten i populasjonen grunnet tilfeldigheter i utvalget. Av samme grunn tror vi at det kan være en tilfeldighet at vi ikke finner noen som helst effekt på fransk. Alt dette tilsier at *estimatene* for tredjespråkene er over- og underdrivelser av den *faktiske* kausale effekten. Vi har følgelig grunn til å tro at den faktiske kausale effekten for alle fremmedspråkene ligger omtrent på samme nivå som effekten vi finner for engelsk på 0,3 standardavvik karakterforbedring.

6.7 Oppsummering av kausale tolkninger

Vi kan ikke trekke noen kausale slutninger om engelsk på ungdomsskolen eller hovedmål og sidemål på VGS på grunn av at forutsetningen om felles trender ikke er oppfylt.

Vi har ikke funnet noen signifikante kausaleffekter av iFinger for hovedmål og sidemål på ungdomsskolenivå. Datamaskin-bruk og følgelig iFinger-bruk har vært lav i ungdomsskolen. Vi anser det som sannsynlig at bruken kan øke i fremtiden og at en da potensielt kan få en effekt på fremmedspråk av iFinger.

For yrkesfag og påbygg VGS finner vi heller ingen kausal effekt. Det kan trolig forklares av lav bruk av iFinger, på tross av god tilgang til IKT-utstyr. Det er ikke spesiell grunn til å tro at dette vil endre seg i fremtiden. Vi merker oss at konklusjonene for påbygg-fagene er svært usikre, da vi kun har kunnet sjekke for felles trender i én periode.

Vi har grunn til å tro at det finnes en positiv kausal effekt på rundt 0,3 standardavvik for fremmedspråk på studiespesialiserende retning. Vi har funnet en klar positiv effekt for engelsk, en veldig stor positiv effekt for tysk, litt tvetydige resultater i spansk og ingen effekt for fransk. Vi tror vi *ikke* har representative utvalg bak hver eksamensobservasjon for tredjespråkene, og at dette fører til tilfeldigheter i estimatene. Derfor har vi tro på at den reelle effekten trolig ligger på rundt 0,3 standardavvik. Det sammenfaller med estimatet for engelsk, som vi har stor tro på, og ligger mellom estimatene for tysk og fransk som virker ekstreme i hver sin retning. Vi merker oss også at effekten ikke ser ut til å komme i selve behandlingsåret. Dette tilsier at bruk over tid er viktig for å få utbytte av programvaren.

7 Praktiske implikasjoner og konklusjon

Vi vil i denne siste delen løfte blikket fra analysene og drøfte de praktiske implikasjonene av funnene våre. Deretter avslutter vi med å besvare problemstillingen i en helhetlig og oppsummerende konklusjon.

7.1 Implikasjonene av Ordnett som kontrollgruppe

Det er viktig å være oppmerksom på at kontrollgruppen for VGS bruker Ordnett; de signifikante resultatene her tilsier at iFinger fører til bedre karakterer enn Ordnett. Kjernefunksjonaliteten til de to produktene er som kjent den samme; å kunne søke i digitaliserte ordbøker. Spørsmålet blir hva som forårsaker de økte karakterene når skoler begynner å bruke iFinger. Vi ser for oss tre mulige årsaker: (1) iFinger kan ha et bedre brukergrensesnitt som gjør at programvaren brukes hyppigere. iFinger inngår blant annet som et lite søkefelt som fester seg i tittellinjen på tekstbehandlingsprogrammer. Brukeren kan med det slå opp med en hurtigtast direkte fra alle vanlige programmer i Windows. (2) Alle videregående skoler vi anvender i våre regresjoner hadde Ordnett før iFinger. Kanskje det å kjøpe nye digitale ordbøker fører til at lærere og elever i større grad blir oppmerksomme på at man har tilgang til en digital ordbok, og at det fører til økt bruk av slik programvare. (3) iFinger kan ha bedre faglig innhold. Ordnett er eid av Kunnskapsforlaget og leverer deres ordbøker, samt fra Oxford University Press. iFinger tilbyr ordbøker fra blant annet Cappelen Damm og Collins.

Siden innholdsproduktene som tilbys fra de forskjellige forlagene i stor grad er like, anser vi forklaring (1) og (2) som mest sannsynlige. Dette tilsier at det er den faktiske bruken av digitale ordbøker, uavhengig av om innholdet er fra Kunnskapsforlaget eller Cappelen Damm, som fører til bedre karakterer.

Årsaken til at iFinger fører til bedre karakterer på VGS er i så fall at andelen iFinger-elever som faktisk bruker iFinger er større enn andelen Ordnett-elever som faktisk bruker Ordnett. Sett i et LATE-rammeverk fører iFinger da til en høyere "take-up" som følge av bedre brukergrensesnitt og/eller at det å bytte programvare gjør skolene oppmerksomme på at

produktet burde brukes. Vi merker oss også at den estimerte effekten av iFinger trolig hadde vært enda større om alternativet var *ingen digitale ordbøker*.

7.2 Samfunnsøkonomiske implikasjoner

I introduksjonen påpekte vi hvordan beslutningstakere i skolen står overfor bruk av de knappe ressursene tid og penger, og at disse burde anvendes på best mulig måte.

Gitt at skoleelever allerede bruker PC i språkfag, fortrenger ikke innføringen av iFinger undervisningstid som kunne vært brukt annerledes. Det er sannsynlig at digitale ordbøker faktisk *frigjør* tid som ellers ville vært brukt på å slå opp ord i en fysisk ordbok manuelt. I tilfellet for VGS, hvor man tidligere har hatt Ordnett, kan det være at programvaren til iFinger er mer effektiv, slik at tid også her er frigjort til undervisning og skolearbeid. Selv om ikke mange sekunder frigjøres per søk, vil dette akkumuleres til å ha betydning for en elev gjennom et skoleår. I så måte effektiviserer iFinger skolehverdagen.

Hva angår budsjetter er iFinger en marginal investering relativt til de totale utdanningsutgiftene i det offentlige skolesystemet. Den årlige prisen for iFinger er mellom 20 og 50 kr per elev. Til sammenlikning kostet hver elev på VGS det offentlige om lag 141 000 kr i 2014 (Utdanningsdirektoratet, 2014b). Åpenbart er en stor del av dette kostnader som uansett må påløpe, som drift av skolebygg og lønninger. Likevel er en årlig kostnad på 20-50 kr per elev svært lav for å øke gjennomsnittskarakterene i fremmedspråk med hele 0,3 standardavvik.

Vi tror som sagt at en kausal effekt av iFinger forutsetter at adaptasjonen av digitale læremidler er utbredt. Dette er trolig ikke tilfellet for ungdomsskoler. En digitalisering av ungdomsskoler er imidlertid påbegynt. Dersom denne digitaliseringen gjennomføres (som på VGS), vil trolig ungdomsskolene være bedre rustet til å få positive effekter av iFinger. Det er også verdt å merke seg at ungdomsskoleelever stort sett skal begynne på VGS. Dersom de venner seg til å bruke digitale hjelpemidler som iFinger på ungdomsskolen, vil de trolig få en høyere nytte av skole-PC-ene når de begynner på VGS.

Ved siden av å frigjøre tid i utdanningen, merker vi oss også at effekten av iFinger aldri er negativ, vi finner enten ingen effekt eller en positiv effekt. Derfor kan beslutningstakere i skolen være trygge på at det å anskaffe programvaren ikke har andre nedsider enn kostnadene ved å kjøpe programvaren. Vi føler vi trygt kan påstå at det å anskaffe iFinger er en *god* bruk av skolens ressurser.

7.3 Ekstern validitet

Vi ser ut til å ha funnet en effekt av iFinger på eksamenskarakterer i fremmedspråk blant norske VGS-elever. Gitt innholdet i kompetansemålene og hvordan karakterer skal gjenspeile kompetansen i de forskjellige språkene, har vi grunn til å tro at iFinger har en effekt på språklig fremstillingsevne, forståelse og skriveferdigheter. Dette gir grunnlag for å drøfte den eksterne validiteten til funnene våre, som er hvorvidt iFinger kan ha en tilsvarende effekt i andre domener (Angrist & Pischke, 2009).

Først og fremst krever iFinger at man har tilgang til PC. Vi kan derfor utelukke at iFinger har en effekt i domener hvor man ikke bruker PC. Det at vi finner en effekt av iFinger for VGS studiespesialiserende og ikke på ungdomsskoler, kan tilsa at en effekt av iFinger forutsetter en relativt høy grad av digitalisering.

Hva gjelder skoleelever i andre land vil iFinger trolig ikke ha noen effekt der elevene ikke bruker PC, eller bruker PC sjeldent. Vi har imidlertid grunn til å tro at iFinger kan ha effekt på fremmedspråk i land hvor skolesystemene er like digitaliserte som norske videregående skoler. Per 2013 er imidlertid Norge det eneste landet i Europa hvor hver elev på VGS har sin egen PC. Danmark hadde 2 elever per PC på nivå tilsvarende norsk VGS. Resterende velutviklede land hadde en dekning på 3-6 elever per PC (Europakommisjonen, 2013). Følgelig er digitaliseringen av europeiske skoler på VGS-nivå trolig for lav per 2013 til at iFinger skal ha en effekt.

Det er imidlertid sannsynlig at PC-dekningen kan øke ellers i verden fremover. Da kan våre funn tilsa at iFinger vil ha en kausal effekt på ferdigheter i fremmedspråk i andre land. Vi

merker oss igjen at våre signifikante funn viser den kausale effekten av iFinger mot Ordnett. Det kan tenkes at land som går fra ingen digital ordbok til iFinger vil ha en enda større effekt.

Hva gjelder andre domener enn skolelever, vil den eksterne validiteten være betinget på samme logikk. iFinger vil trolig bedre språklig fremstillingsevne, forståelse og skriveferdigheter i fremmedspråk for personer som skriver og/eller leser mye på PC. Dette kan for eksempel være konsulenter, journalister, jurister eller lignende.

Vi har i denne oppgaven kun sett på digitale ordbøker til PC. iFinger lanserte i 2015 en applikasjon som lar brukere søke i de samme ordbøkene ved hjelp av smarttelefoner og nettbrett. Dette lar både skoleelever og "folk flest" bruke programvaren når de for eksempel er i utlandet, leser avisen eller skriver meldinger. Våre funn *kan* tilsa at bruk av denne applikasjon fører til bedre språk og språkforståelse i hverdagen.

7.4 Konklusjon

Vi har i denne oppgaven prøvd å svare på problemstillingen:

Fører bruk av digitale ordbøker fra iFinger til bedre karakterer i språkfag i norsk skole?

For å gjøre dette definerte vi tre populasjoner av norske skoleelever: Ungdomsskoleelever, elever ved yrkesfag og påbygg på VGS-nivå og elever ved studiespesialiserende linjer på VGS-nivå.

Ungdomsskoleelever

For ungdomsskoleelever har vi ikke funnet noen kausal effekt av iFinger på karakterer. Vi anser det som sannsynlig at dette kommer av at utbredelsen av IKT i ungdomsskolen har vært lav. Dette endres raskt i disse tider, og det kan være grunn til å tro at iFinger kan få en positiv effekt på karakterer i ungdomsskolen i fremtiden.

Elever ved yrkesfag og påbygg på VGS-nivå

For elever ved yrkesfag har vi ikke funnet noen kausal effekt av iFinger på karakterer, og trolig ikke for påbygg heller. Disse elevene bruker digitale ordbøker lite til tross for at det er god tilgang til datamaskiner i den videregående skolen. Vi tror at elever ved disse linjene vil fortsette å fokusere på yrkesfagene sine i fremtiden, og følgelig har vi ingen god grunn til å tro at disse resultatene vil endre seg i fremtiden.

Elever ved studiespesialiserende linjer på VGS-nivå

For elever ved studiespesialiserende linjer har vi funn som tilsier at iFinger fører til en karakterforbedring på 0,3 standardavvik i fremmedspråk, sammenliknet med Ordnett. Det er likevel mange elever som har tilgang til iFinger, men ikke bruker programvaren fast. Om flere velger å bruke det vil en trolig kunne se enda større effekt på karakterene. For hovedmål og sidemål studiespesialiserende har vi ikke kunnet anvende våre metoder til å si noe om kausale effekter.

Våre funn i forhold til tidligere forskning

Vi merker oss her at vi har jobbet med en veldig konkret problemstilling med en tydelig "intention to treat". Videre er den norske videregående skolen som kjent den mest digitaliserte i Europa. Denne digitale modenheten kan kanskje være en forutsetning for å kunne se effekt av utdanningsprogramvare. Alt dette er nok mye av grunnen til at vi har funnet klarere positive effekter enn andre tidligere studier som tar for seg utdanningsteknologi.

Praktiske implikasjoner og samfunnsøkonomisk betydning

Vi finner at iFinger forbedrer fremmedspråkkarakterer på studiespesialiserende VGS med omkring 0,3 standardavvik i forhold til Ordnett. På karakterskalaen som går fra 1 til 6 representerer dette en forbedring på mellom 0,14 og 0,18 karakterpoeng. Videre vet vi at iFinger koster mellom 20 og 50 kr per elev årlig, mens det brukes omkring 140 000 kr per elev i norsk VGS årlig. Våre funn tilsier derfor at iFinger gir en betydelig økning i karakterer for en lav kostnad.

8 Bibliografi

- Abadie, A., Athey, S., Imbens, G., & Wooldridge, J. (2014, juli). Finite Population Standard Errors. *NBER Working Paper*(20325).
- Angrist, J. D., & Pischke, J.-S. (2015). *Mastering 'Metrics*. 08540, NJ, USA: Princeton University Press.
- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics*. 08540, NJ, USA: Princeton University Press.
- Autor, D. (2003). Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. *Journal of Labor Economics*(21).
- Belo, R., Ferreira, P., & Rahul, T. (2010, juli 8). *Broadband in School: Evidence from Portugal*. Hentet desember 1, 2015 fra Social Science Research Network: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1636584
- Bertrand, M., Duflo, E., & Sendhil, M. (2004, Februar). How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics*, ss. 249-275. Hentet fra MIT Economics: <http://economics.mit.edu/files/750>
- Bulman, G., & Fairlie, R. (2015). *Technology and Education: Computers, Software, and the Internet*. Bonn: Forschungsinstitut zur Zukunft der Arbeit Institute for the Study of Labor (IZA).
- Card, D., & Krueger, A. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review*(84), ss. 772-793.
- Carillo, P., Onofa, M., & Ponce, J. (2010, Desember). Information Technology and Student Achievement: Evidence from a Randomized Experiment in Ecuador. *IDB Working Paper Series*(IDB-WP-223).
- Cristia, J., Czerwonko, A., & Garofalo, P. (2014, mai). Does Technology in Schools Affect Repetition, Dropout and Enrollment? Evidence From Peru. *Journal of Applied Economics*(17), ss. 89-112.
- Europakommisjonen. (2013, februar). *Survey of Schools: ICT in Education*. Hentet november 23, 2015 fra Europakommisjonen: <https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/KK-31-13-401-EN-N.pdf>

- Forskrift til opplæringslova. (2009). Hentet 5. oktober 2015, fra <https://lovdata.no/forskrift/2006-06-23-724>.
- Fuchs, T., & Woessmann, L. (2004, november). Computers and Student Learning: Bivariate and Multivariate Evidence on the Availability and Use of Computers at Home and at School. *CESifo Working Paper*(1321).
- Goolsbee, A., & Guryan, J. (2002, august). The Impact of Internet Subsidies in Public Schools. *NBER Working paper*(9090).
- Grimes, D., & Warschauer, M. (2008). Learning With Laptops: A Multi-Method Case Study. *Journal of Educational Computing Research*(38), ss. 305-332.
- IKT-Norge. (2015, mai 12). *Fremtidens skole - er politikerne klare?* Hentet november 15, 2015 fra IKT-Norge: <https://www.ikt-norge.no/2015/05/fremtidens-skole-er-politikerne-klare>
- Knudsen, O. (2008, september 1). *Her er PC-tilbudet i ditt fylke*. Hentet desember 3, 2015 fra Tek.no: http://www.tek.no/artikler/her_er_pc-tilbudet_i_ditt_fylke/83752
- Machin, S., McNally, S., & Silva, O. (2007, juli). New Technology in Schools: Is There a Payoff? *The Economic Journal*(117), ss. 1145–1167.
- Maine Education Policy Research Institute. (2007). *Maine's Middle School Laptop Program: Creating Better Writers*. Maine Education Policy Research Institute, University of Southern Maine.
- Mathematica Policy Research, Inc. (2007, mars). *Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort*. Jessup: The U.S. Department of Education and Mathematica Policy Research.
- Mathematica Policy Research, Inc. (2009). *Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts*. Jessup: The U.S. Department of Education and Mathematica Policy Research.
- Mckenzie, D. (2012, oktober 21). *Tools of the Trade: A quick adjustment for multiple hypothesis testing*. Hentet november 25, 2015 fra The World Bank: <http://blogs.worldbank.org/impactevaluations/tools-of-the-trade-a-quick-adjustment-for-multiple-hypothesis-testing>
- Mjelva, H., & Hoaas, K. (2008, juni 15). *Håver inn på skolenes PC-boom*. Hentet september 10, 2015 fra Bergens Tidende: Hentet fra <http://www.bt.no/nyheter/lokalt/Haver-inn-pa-skolenes-PC-%20boom-1890394.html>

- Ordnett. (2014). *Ordnett 10 år*. Hentet september 20, 2015 fra Ordnett:
<http://www.ordnett.no/info/ordnett10aar>
- Oslo Inovation Week. (2015). *EdTech - The Norwegian Classroom*. Hentet november 10, 2015 fra Oslo Inovation Week: http://oiw.no/program_2015/ed-tech/
- Pischke, J.-S. (2005, oktober). *Empirical Methods in Applied Economics Lecture Notes*. Hentet oktober 17, 2015 fra London School of Economics, Department of Economics:
<http://econ.lse.ac.uk/staff/spischke/ec524/evaluation5.pdf>
- Rouse, C., & Krueger, A. (2004). Putting computerized instruction to the test: a randomized evaluation of a “scientifically based” reading program. *Economics of Education Review*(23), ss. 323–338.
- Senter for IKT i utdanningen. (2013, 12 14). *Monitor Skole 2013*. Oslo: Senter for IKT i utdanningen. Hentet 11 30, 2015 fra Senter for IKT i utdanningen:
<http://iktsenteret.no/ressurser/monitor-skole-2013#.VOIFq8Zz53s>
- Skolelederforbundet. (2011, mai). *Skolelederforbundet*. Hentet november 9, 2015 fra Eksamen: Farvel til penn og papir: <http://www.nslf.no/pc-1275-150-Farvel-til-penn-og-papir.aspx>
- Språkrådet. (2014, november). *Digitale ordbøker i bruk*. Oslo: Språkrådet.
- Statistisk Sentralbyrå. (2014, mai 27). *Offentlig forvaltnings inntekter og utgifter, 2013*. Hentet desember 2, 2015 fra Statistisk Sentralbyrå: <https://www.ssb.no/offentlig-sektor/statistikker/offinnut/aar/2014-05-27>
- The Texas Center for Educational Research. (2009). *Evaluation of the Texas Technology Immersion Pilot: Final Outcomes for a Four-Year Study (2004-05 to 2007-08)*. Austin: The Texas Center for Educational Research.
- Universitetet i Oslo. (2015, april 30). *EdTech-UiO*. Hentet november 27, 2015 fra Universitetet i Oslo: <http://www.uio.no/tjenester/it/utdanning/edtech/edtech.html>
- Utdanningsdirektoratet. (2010). *Karakterstatistikk for videregående opplæring skoleåret 2009- 2010*. Oslo: Utdanningsdirektoratet. Hentet desember 4, 2015 fra Utdanningsdirektoratet:
http://www.udir.no/globalassets/upload/statistikk/karakterer/karakterstatistikk_vgo_2009_2010.pdf
- Utdanningsdirektoratet. (2013, april 9). *Retningslinjer for sentral sensur*. Hentet desember 4, 2015 fra Utdanningsdirektoratet:

<http://www.udir.no/Vurdering/Eksamen/Retningslinjer-for-sentral-sensur-ved-sentralt-gitt-skriftleg-eksamen-i-grunnskolen-grunnskoleopplaring-for-vaksne-og-vidaregaande-opplaring/>

Utdanningsdirektoratet. (2014a, januar 14). *Trekkordning ved eksamen i Kunnskapsløftet Udir-1-2009*. Hentet desember 4, 2015 fra Utdanningsdirektoratet:

<http://www.udir.no/Regelverk/Finn-regelverk-for-opplaring/Finn-regelverk-etter-tema/eksamen/Udir-1-2009-Trekkordning-ved-eksamen-i-Kunnskapsloftet/>

Utdanningsdirektoratet. (2014b). *Utdanningsspeilet 2014*. Oslo: Utdanningsdirektoratet.

Hentet fra Utdanningsdirektoratet:

<http://www.udir.no/globalassets/upload/rapporter/utdanningsspeilet-2014-utskriftsversjon.pdf>

Utdanningsdirektoratet. (2015). *Lag dine egne rapporter*. Hentet september 2015 fra

Skoleporten: <https://skoleporten.udir.no/rapportbygger/>

Utdanningsdirektoratet. (u.å.a). *Eksamen grunnskole*. Hentet desember 4, 2015 fra

Utdanningsdirektoratet: <http://www.udir.no/Vurdering/Eksamen-grunnskole/>

Utdanningsdirektoratet. (u.å.b). *Eksamen videregående*. Hentet desember 4, 2015 fra

Utdanningsdirektoratet: <http://www.udir.no/Vurdering/Eksamen-videregaende/>

Utdanningsdirektoratet. (u.å.c). *Læreplan i engelsk - kompetansemål*. Hentet desember 4,

2015 fra Utdanningsdirektoratet: [http://www.udir.no/kl06/ENG1-](http://www.udir.no/kl06/ENG1-03/Kompetansemaal?arst=1858830315&kmsn=-915046718)

[03/Kompetansemaal?arst=1858830315&kmsn=-915046718](http://www.udir.no/kl06/ENG1-03/Kompetansemaal?arst=1858830315&kmsn=-915046718)

Utdanningsdirektoratet. (u.å.d). *Læreplan i fremmedspråk*. Hentet desember 4, 2015 fra

Utdanningsdirektoratet: <http://www.udir.no/kl06/FSP1-01/Hele/Vurdering>

Utdanningsdirektoratet. (u.å.e). *Læreplan i fremmedspråk - kompetansemål*. Hentet

desember 4, 2015 fra Utdanningsdirektoratet: [http://www.udir.no/kl06/FSP1-](http://www.udir.no/kl06/FSP1-01/Kompetansemaal?arst=1858830315&kmsn=-1464567501)

[01/Kompetansemaal?arst=1858830315&kmsn=-1464567501](http://www.udir.no/kl06/FSP1-01/Kompetansemaal?arst=1858830315&kmsn=-1464567501)

Utdanningsdirektoratet. (u.å.f). *Læreplan i norsk - kompetansemål studiespesialiserende vg1*

og yrkesfag vg2. Hentet desember 4, 2015 fra Utdanningsdirektoratet:

<http://www.udir.no/kl06/NOR1-05/Kompetansemaal?arst=1858830315&kmsn=-2089340600>

Utdanningsdirektoratet. (u.å.g). *Læreplan i norsk - kompetansemål studiespesialiserende*

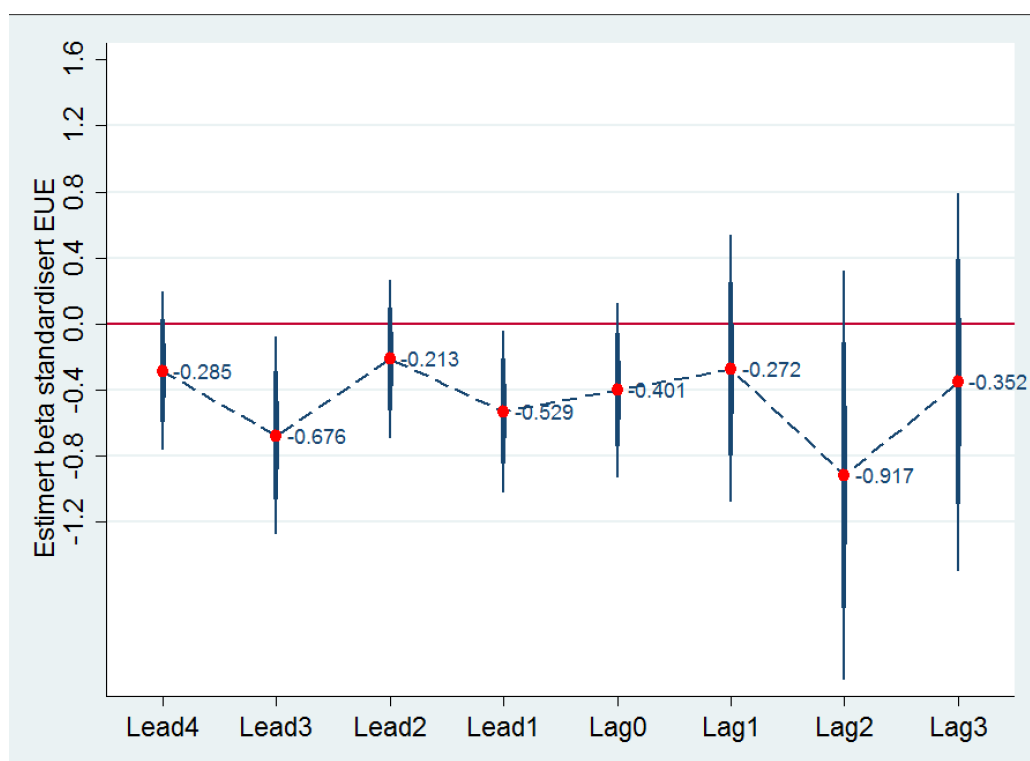
vg3. Hentet desember 4, 2014 fra Utdanningsdirektoratet:

<http://www.udir.no/kl06/NOR1-05/Kompetansemaal?arst=1858830314&kmsn=-1569321230>

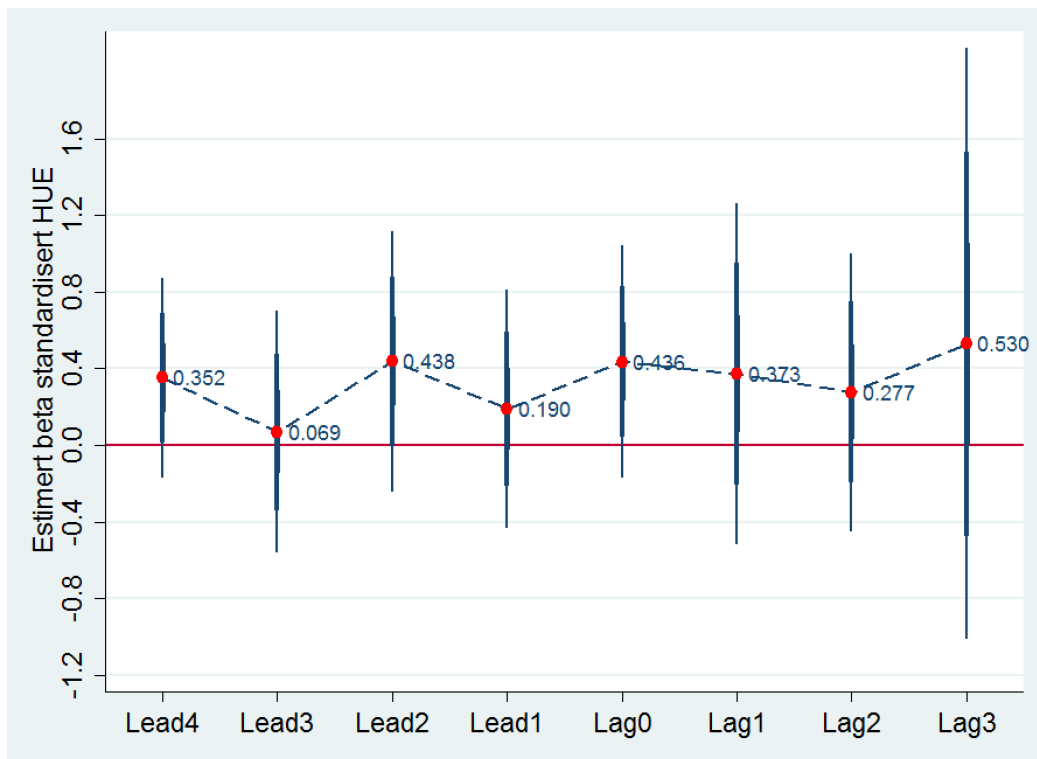
9 Appendiks

9.1 Granger-plots: Ungdomsskole

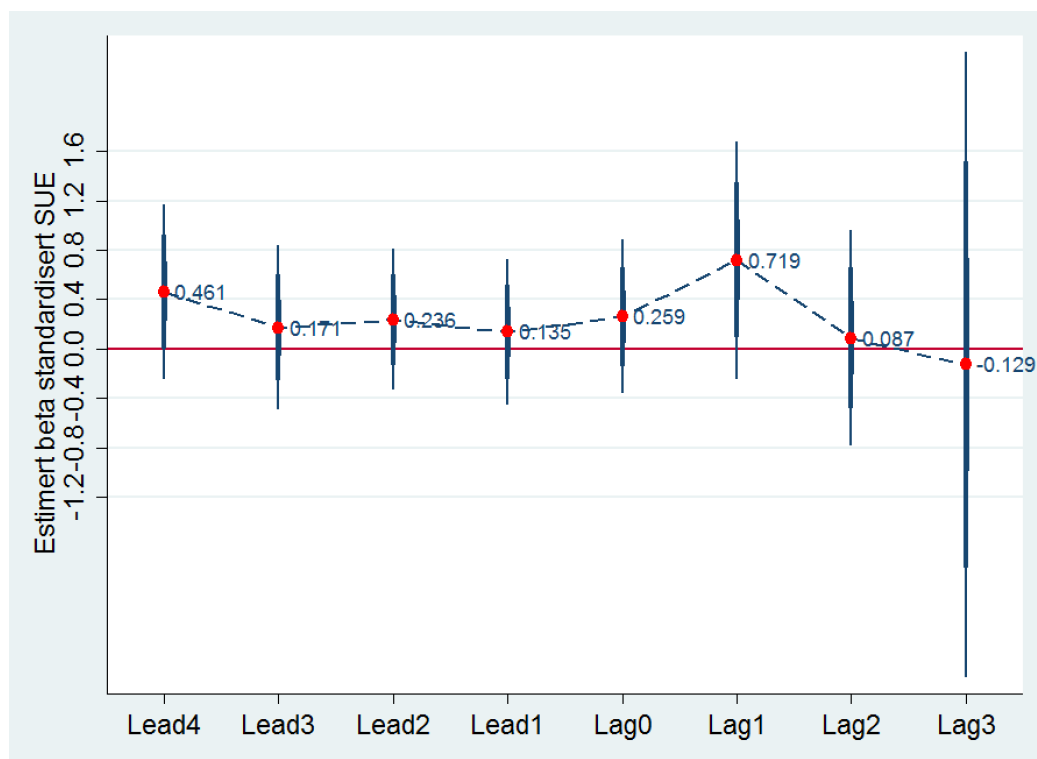
Figur 9.A – Granger-plot: Engelsk US



Figur 9.B – Granger-plot: Hovedmål US

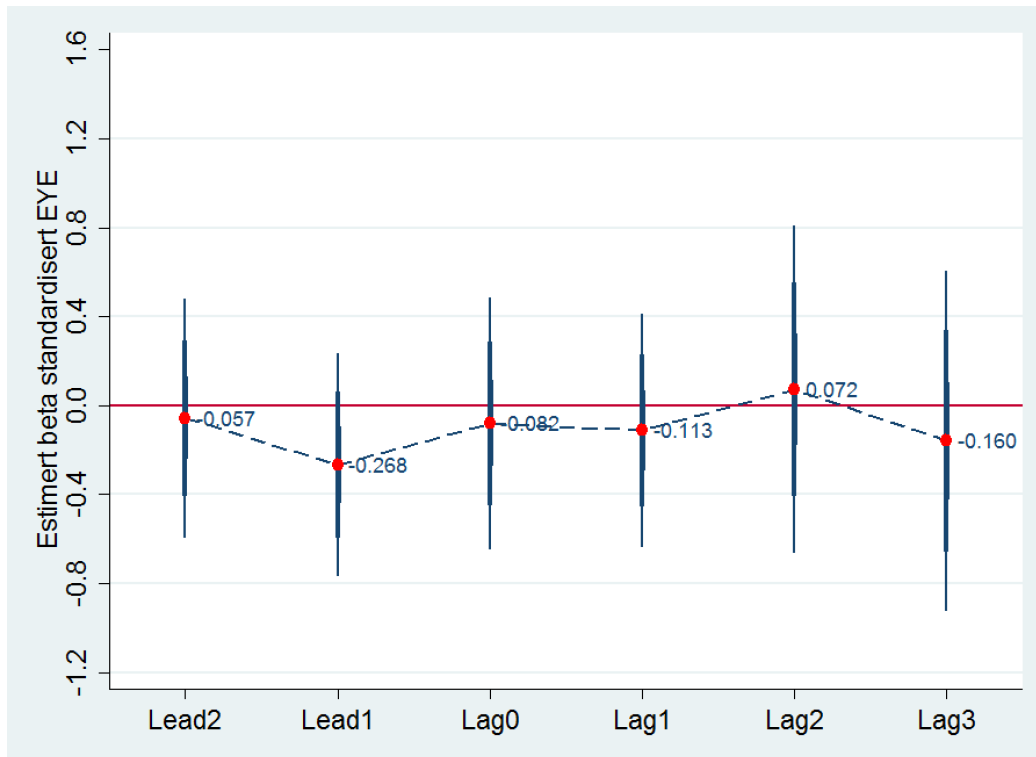


Figur 9.C – Granger-plot: Sidemål US

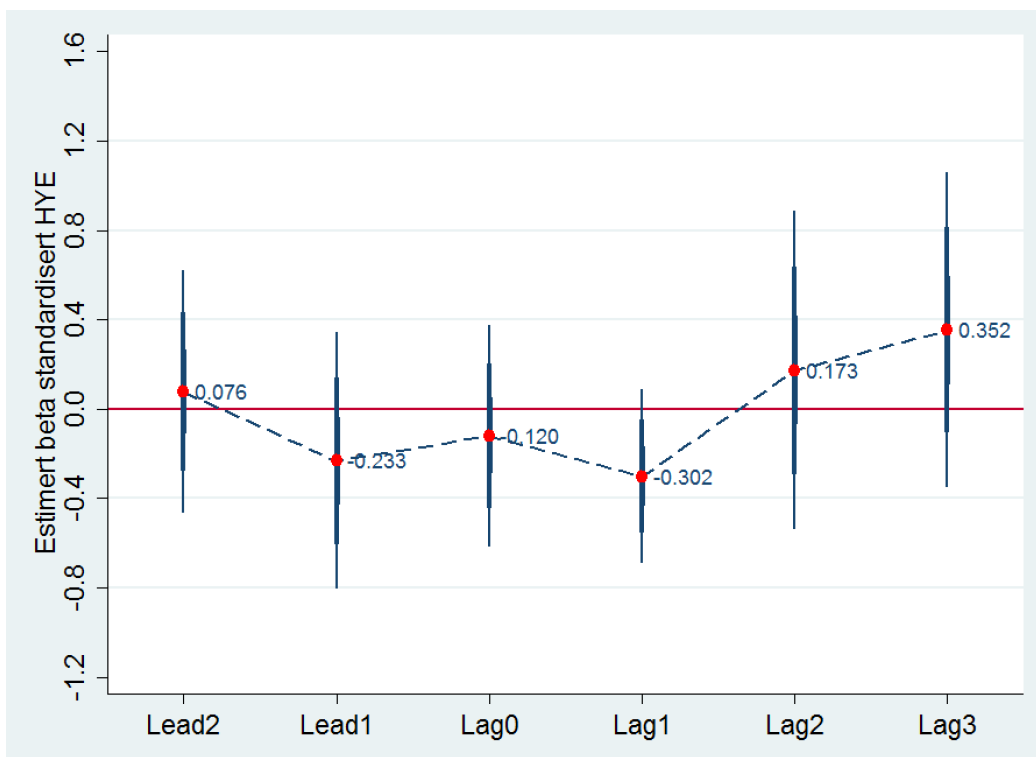


9.2 Granger-plots: Videregående skole

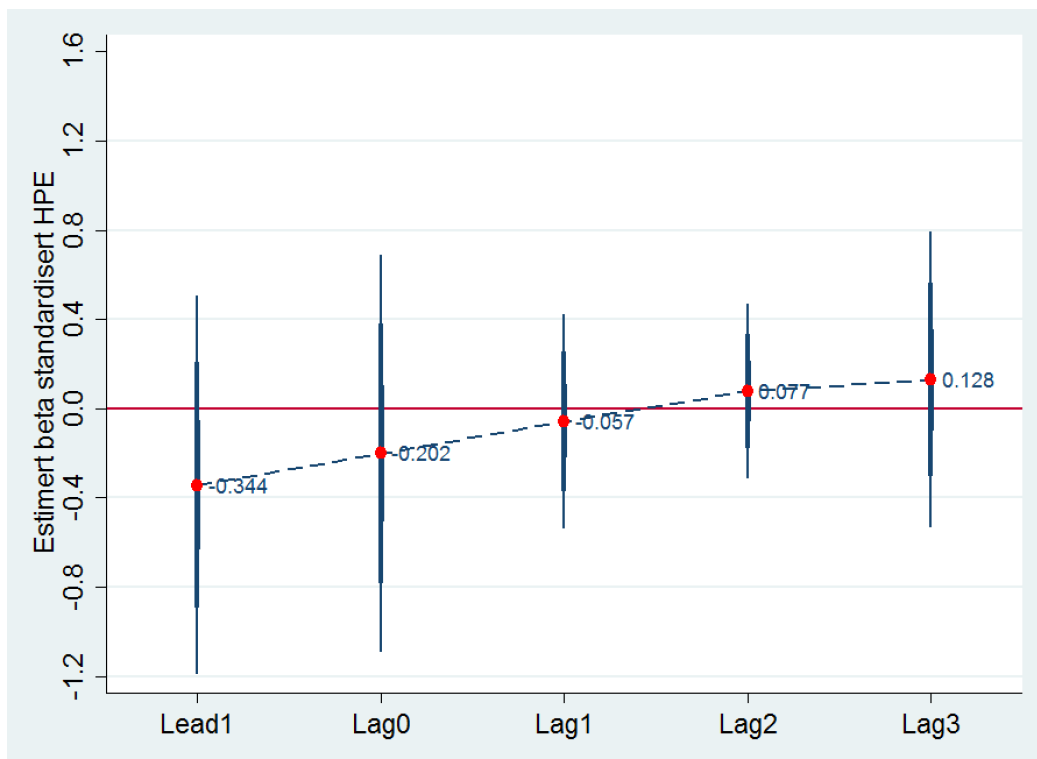
Figur 9.D - Granger-plot: Engelsk yrkesfag VGS



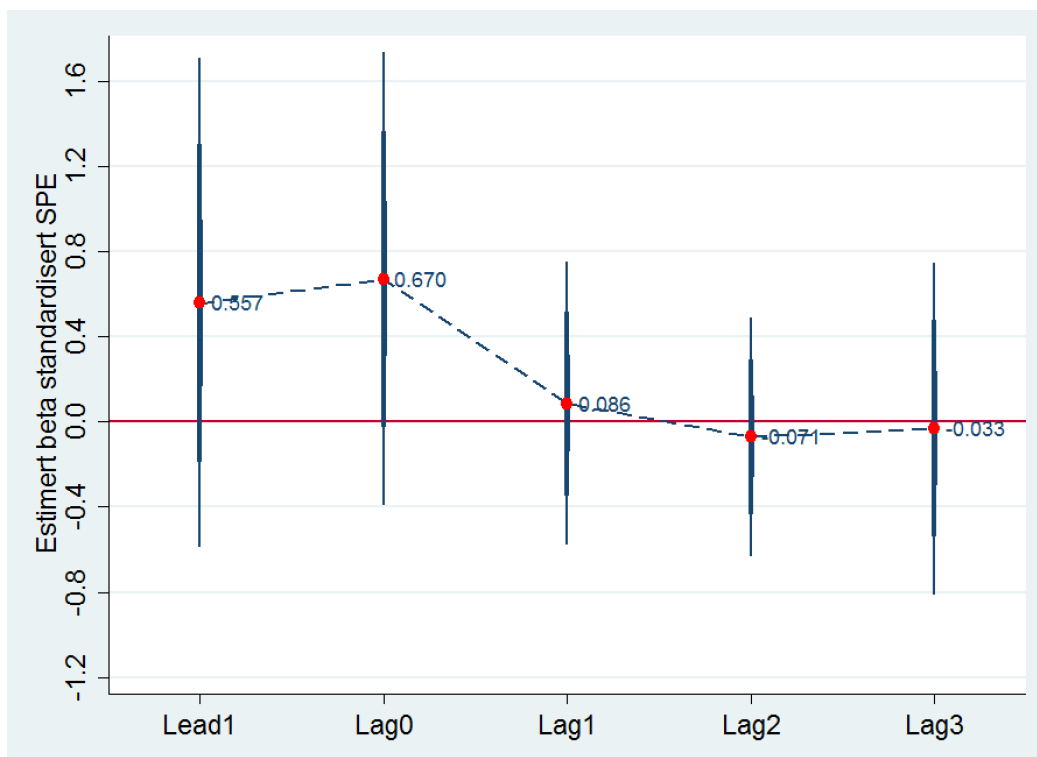
Figur 9.E - Granger-plot: Hovedmål yrkesfag VGS



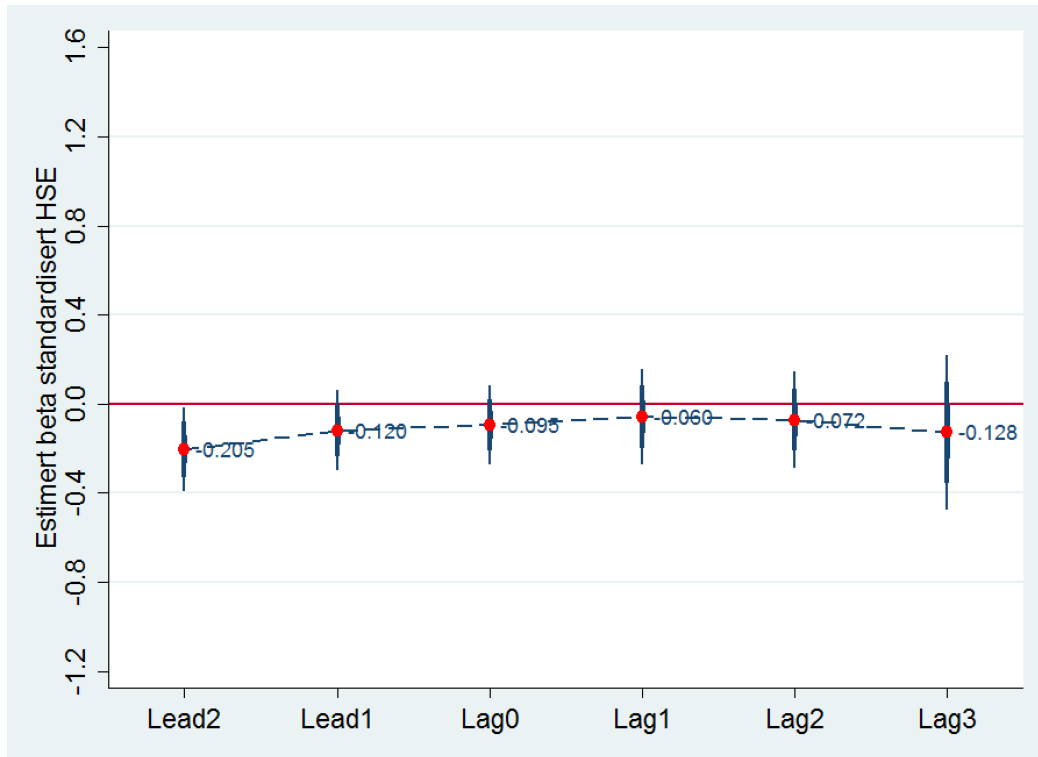
Figur 9.F - Granger-plot: Hovedmål påbygg VGS



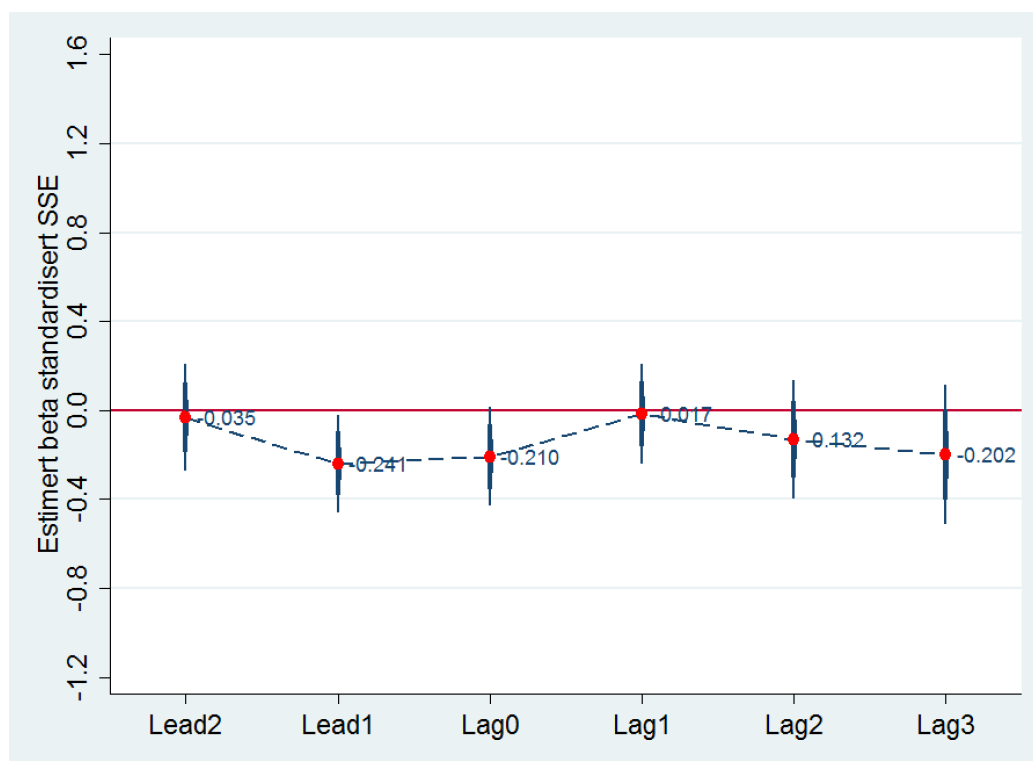
Figur 9.G - Granger-plot: Sidemål påbygg VGS



Figur 9.H - Granger-plot: Hovedmål studiespesialiserende VGS



Figur 9.I - Granger-plot: Sidemål studiespesialiserende VGS



9.3 Stata

Stata-kode og tilhørende datasett ligger vedlagt besvarelsen i egne filer.