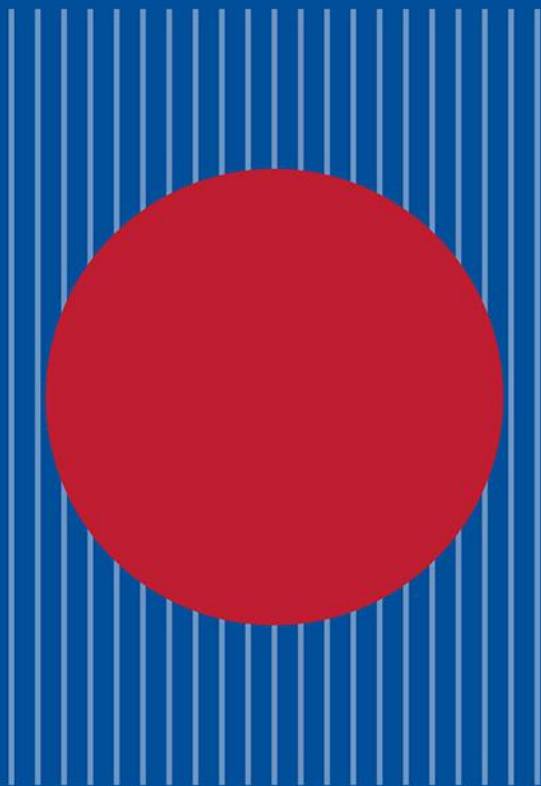


Analyse af REFUD-relaterede effekter af klassestørrelse



Analyse af REFUD-relaterede effekter af klassestørrelse

© VIVE og forfatterne, 2022

e-ISBN: 978-87-7582-088-7

Forsidefoto: Ole Bo Jensen/VIVE

Projekt: 302355

Finansiering: Børne- og Undervisningsministeriet

VIVE

Det Nationale Forsknings- og Analysecenter for Velfærd

Herluf Trolles Gade 11

1052 København K

www.vive.dk

VIVEs publikationer kan frit citeres med tydelig kildeangivelse.



Forord

Denne rapport undersøger effekten af klassestørrelsen i folkeskolens 0.-9. klassetrin på de udfaldsmål, der indgår i Børne- og Undervisningsministeriets Regnemodel for Uddannelsesinvesteringer til beregning af budgetøkonomiske konsekvenser af investeringer i dagtilbud og grundskolen (herefter benævnt REFUD). Det er hensigten, at rapportens resultater skal kunne indgå i ministeriets videre udvikling af modellen.

Analysen baseres på en anerkendt mikroøkonometrisk metode til analyse af effekten af klassestørrelse på elevers udfaldsmål; 2SLS-model baseret på et fuzzy RD-design. Metoden er anvendt i både danske og internationale studier og bygger på et relativt stort forskningsområde og metodeudvikling gennem flere årtier.

Analysen er gennemført i perioden fra 10. maj til 20. juni 2022. Det har været en forudsætning for analysen, at den udføres på eksisterende data fra REFUD, ligesom det var en præmis for analysen, at VIVE skulle undersøge klassestørrelseeffekter på samtlige effektmål og klassetrin i SØM-modellen. En række metodevalg og definitionen følger derfor REFUD. Rapporten har gennemgået kvalitetssikring af to fagfæller (en intern og en ekstern). Analysen er finansieret af Børne- og Undervisningsministeriet.

Carsten Strømbæk Pedersen

Forsknings- og analysechef for VIVE Børn og Uddannelse



Indholdsfortegnelse

DEL 1 Afrapportering	6
----------------------	---

Hovedresultater	7
-----------------	---

1	Indledning	10
---	------------	----

2	Analyse og resultater	12
2.1	Effekter af klassestørrelse på trivsel	18
2.2	Effekter af klassestørrelse på SØM-konsekvenser	20
2.3	Persistens	31
2.4	Robusthed	31
2.5	Effekternes størrelsesorden i forhold til andre studier	35
2.6	Implementering i REFUD	38

3	Konklusion	40
---	------------	----

DEL 2 Dokumentation	44
---------------------	----

4	Metode	45
4.1	Identifikationsstrategi og øvrige metodevalg	45
4.2	Analyse af validitet	48
4.3	Analyse af persistens af indsatseffekter	54
4.4	Analyse af robusthed	56
4.5	Multiple hypotesetestning	65

5	Data	66
5.1	Analysedata	66
5.2	Antal årgange, der kan måles udfaldsmål for	68

Litteratur	71
Bilag 1 Figurer	75
Bilag 2 Tabeller	82
Bilag 3 REFUD udfaldsmål	93



DEL 1

Afrapportering

Hovedresultater

Investeringer på grundskoleområdet i form af fx lavere klassekvotienter optager både politikere, fagfolk og lokale beslutningstagere såvel som den generelle offentlighed. Omkostningerne ved sådanne initiativer er ofte markante, mens der er mindre viden om de budgetøkonomiske fordele af investeringerne. I samarbejde med VIVE har Børne- og Undervisningsministeriet derfor udviklet REFUD – en kommunal budgetøkonomisk regnemodel, der tilbyder brugeren et skøn på de kortsigtede økonomiske konsekvenser af investeringer på grundskole- og dagtilbudsområdet.

Formålet med denne rapport er at bidrage med estimater på den kausale effekt af at reducere klassestørrelsen i grundskolens 0.-9. klasse på elevernes trivsel og forbrug af REFUDs budgetøkonomiske udfaldsmål samt at vurdere kvaliteten af disse estimater som grundlag for beregninger i REFUD.

Det er vigtigt at bemærke, at der kun er undersøgt effekter på de udfaldsmål, der indgår i REFUD, og som – ud over trivsel – alene er budgetøkonomiske mål. Andre udfaldsmål, som man kunne forvente påvirkes direkte af klassestørrelsen (fx læring), er ikke undersøgt i denne rapport.

Effekt af klassestørrelse på trivsel og REFUD-konsekvenser

Analyseresultaterne for trivsel viser signifikante kausale effekter på 1 ud af 5 af trivselsindikatorerne fra den nationale trivselsmåling (indekset for Ro og orden). En elev mindre i klassen forbedrer således elevernes opfattelse af ro i klassen og lærernes klasserumsledelse. Effekten dokumenteres på alle klassetrin fra 4. til 9. klasse, hvor dette trivselsmål er tilgængeligt.

Resultaterne for REFUD-konsekvenserne er blandede og er usikkert estimeret. Blandt de interessante og troværdige tendenser er sammenhængen med elevernes skolevalg: Når klassestørrelsen reduceres, ses et øget forbrug af folkeskolen, hvorimod forbrug af friskoler/private grundskoler omkring overgangen fra indskolingen til mellemtrinnet og omkring 6.-7. klasse mindskes. Der er også interessante tendenser på elevens forbrug af sundhedsydelse, men disse er i højere grad sporadiske, fx forbindes en klassestørrelsesreduktion med færre skadestuebesøg i 5. klasse især.

Der er ikke fundet overbevisende effekter på elevernes forbrug af specialtilbud ved en ændring i klassestørrelsen. Dette udfaldsmål var på forhånd udvalgt til yderligere robusthedscheck og er således undersøgt nærmere, uden at det ændrer på konklusionen.

Rapportens metode og forbehold

Effekterne er estimeret ved at overføre en allerede kendt metode fra klassestørrelseslitteraturen til data for den danske folkeskole i perioden efter Folkeskolereformen i 2014 og de udfaldsmål, der er specificeret af Børne- og Undervisningsministeriet (BUVM).

Validiteten af rapportens resultater bør vurderes på linje med den relaterede forskning på området. Overordnet vurderes det, at metoden er overførbart og valid, da metodens bagvedliggende antagelser, i det omfang de kan efterprøves, i overvejende grad bekræftes på baggrund af den samme type af empiriske tests, som foretaget i studier fra andre lande. Endvidere er størrelsesordenen på analysens resultater sammenlignelige med resultater fra internationale studier.

På denne baggrund vurderes det, at rapportens resultater udgør den bedst tilgængelige evidens på betydningen af klassestørrelse for børns udvikling og trivsel i folkeskolen, der er tilgængelige for danske data efter Folkeskolereformen.

Tre forbehold skal dog bemærkes ved læsning af rapportens mange analyse-resultater (og eventuel implementering i REFUD):

- For det første er præmissen for analysen, at den valgte estimationsmodel skal kunne håndtere evaluering af klassestørrelses effekter på samtlige klassetrin og typer af udfaldsmål, og modelvalget er dermed baseret på en afvejning mellem bias og statistisk præcision inden for opgavens rammer, og hvad datatilgængeligheden tillader. Det betyder blandt andet, at metoden generelt vurderes mindre gyldig desto ældre klassetrin, der betragtes. Derudover er der særlige forhold vedrørende klassedannelsen i 0. og 9. klasse, der gør, at metoden vurderes mindre gyldig for de klassetrin.
- For det andet analyseres 59 udfald (for op til 10 klassetrin og 4 år), hvilket udsætter analysen og effektestimerne for stor risiko for multiple-hypotese-testnings bias. Dermed er det sandsynligt, at nogle effektestimer vil være 'falsk positive', og såfremt disse anvendes direkte i REFUD, kan det betyde, at resulterende beregninger overestimerer afkastet ved investering

i klassestørrelsesreduktioner. Rapportens analyseafsnit indeholder en diskussion af forskellige tilgange til signifikanskorrektioner for at tage højde for denne problemstilling. På grund af omfanget af analysens udfaldsmål og det tilgængelige datagrundlag vurderes de tilstrækkelige korrektioner ikke at være hensigtsmæssige, og de implementeres ikke i denne rapport.

- For det tredje er datagrundlaget for nogle SØM-konsekvenser sparsomt. Det betyder blandt andet, at analysen kan mangle statistisk styrke til at kunne påvise, om effekter på enkeltvise SØM-konsekvenser er signifikante. Dette bliver en endnu større udfordring, hvis man samtidig vil tage højde for ovenstående forbehold om signifikanskorrektion (hvilket rapporten derfor undlader for REFUD-konsekvenserne generelt).

På baggrund af ovenstående forbehold opfordres læseren til at betragte rapportens resultater som en helhed snarere end enkeltvist og ikke lægge stor vægt på enkeltvise signifikante estimater.

Rapportens anbefalinger i forhold til implementering i REFUD

- Der er positive effekter af klassestørrelsesændringer på trivselsindekset Ro og orden, og det vurderes, at dette trivselsmål kan benyttes som såkaldt link outcome i REFUD-modellen. Vurderingen er baseret på, at effekterne overbevisende overlever robusthedstest og signifikanskorrektion for multiple hypotesetestning. Endvidere er størrelsesorden af effekten på Ro og orden inden for den forventede og sammenlignelig med størrelsesorden på trivselseffekter, der er dokumenteret i andre lande.
- For flere REFUD-konsekvenser er der enkeltvise signifikante resultater af klassestørrelsesændringer. Af flere grunde er disse resultater dog ikke ligeså overbevisende, som resultaterne på trivsel (fx er de kun signifikante for enkelte ud af 10 klassetrin, eller de er upræcist estimerede). Da analysen desuden betragter et stort antal udfaldsmål og klassetrin samtidig, kan disse enkeltvis signifikante resultater ikke, ligesom trivselsresultaterne, overleve en tilsvarende omfattende signifikanskorrektion på dette datagrundlag og kan dermed ikke benævnes som 'effekter'. Vurderingen er derfor, at resultaterne på de budgetøkonomiske konsekvenser i REFUD ikke på nuværende tidspunkt bør implementeres direkte i modellen. I stedet kan de anvendes til at underbygge og kvalitetssikre modellens eksisterende estimater for klassestørrelseseffekter.

1 Indledning

VIVE har efter opdrag fra Børne- og Undervisningsministeriet (BUVM) udviklet en socialøkonomiske investeringsmodel til beregning af budgetøkonomiske konsekvenser af indsatser inden for BUVM's område (REFUD). Modellen udvikles og videreudvikles løbende.

Blandt de indsatser, der indgår i modellens første version, er "klassestørrelse". Det vil sige, at modellen beregner de forventede budgetøkonomiske konsekvenser ved at ændre klassestørrelsen på en lang række af offentlige ydelser og services. Der foreligger i dag en relativ stor international forskningslitteratur inden for uddannelsesøkonomi om effekter af klassestørrelse, men det er også et felt, der fortsat har stor interesse blandt både praktikere og forskere. Der sker derfor også fortsat udvikling både metodemæssige og fortolkningsmæssigt. Se fx seneste studier fra Angrist m.fl. (2019), Leuven & Løkken (2020) og Connolly & Haecck (2022). Der foreligger desuden flere danske analyser af klassestørrelsens effekt på elevers læring (fx Bingley m.fl. 2005, Browning & Heinesen 2007, Nandrup 2016, Beuchert & Nandrup, under udgivelse), men ikke analyser af den direkte sammenhæng mellem klassestørrelse og de monetære effektmål, som indgår i REFUD, ligesom effekterne på forskellige aspekter af trivsel fortsat er relativt ubelyste.

Formålet med denne rapport er at overføre disse metoder til evaluering af effekterne af klassestørrelse på samtlige af de udfald, der indgår i REFUD, for på sigt at kunne implementere dem i modellen. Der anvendes derfor et kendt analysedesign, der kan operationaliseres ved hjælp af de tilgængelige administrative registre fra Danmarks Statistiks grunddatabank. Metodevalget tager udgangspunkt i den nyere litteratur på området (fx Fredriksson m.fl. 2013, 2016; Leuven & Løkken 2020; Connolly & Haecck 2022), dog tilpasset tilgængelige data og den elevpopulation, som indgår i REFUD. Effektestimaterne på klassestørrelse i denne sammenhæng er såkaldte *marginale effekter* og fortolkes som den gennemsnitlige effekt af at øge klassestørrelsen med en enkelt elev.

Analysen indeholder dermed REFUD-modellens 54 budgetøkonomiske konsekvenser samt fem mål for elevtrivsel (herefter anvendes 'udfaldsmål' generelt for både SØM-konsekvenser og trivselsmål).

Derudover er formålet med opgaven at undersøge og vurdere persistensen af effekten af klassestørrelsen over en tidshorizont på op til 4 år. Dette skal ligeledes anvendes til at kvalificere første version af REFUD.

Analyserne af de SØM-relaterede effekter af klassestørrelse består således af følgende undersøgelsesspørgsmål:

- Hvad er effekten af klassestørrelse på hvert klassetrin fra 4.-9. klasse på de 5 trivselsmål i REFUD?
- Hvad er effekten af klassestørrelse på hvert klassetrin fra 0.-9. klasse på de 54 monetære konsekvenser i REFUD i førstkommande budgetår?
- Hvad er persistensen af effekten af klassestørrelse 2, 3 og 4 år efter en ændring i klassestørrelsen på de 54 monetære konsekvenser i REFUD?

Rapporten estimerer og analyserer således i alt 59 udfaldsmål på op til 10 klassetrin¹, hvoraf de 54 budgetmæssige konsekvenser desuden estimeres over en 4-årig tidshorisont. Dette udsætter analysens enkelte resultater for en stor risiko for type I-fejl, hvor usande effekter detekteres som signifikante. Idet datagrundlaget for studiet allerede er sparsomt i forhold til de forventede effektstørrelser, signifikansjusteres resultaterne på budgetmæssige SØM-udfald ikke. En tilstrækkelig omfattende korrektion vil reducere styrken i et omfang, der (bortset fra effekter på trivselsudfaldene) ikke tillader detektion af signifikante effekter af en realistisk størrelsesorden. Med dette in mente, behandler rapportens resultat afsnit alene elevtrivselsudfaldene samt tendenser i de primære default SØM-konsekvenser, der forventes at være direkte relateret til evt. effekter af klassestørrelse på elevernes fagligheder og trivsel. Læsere opfordres desuden til at betragte rapportens resultater som en helhed snarere end enkeltvist.

Rapporten består af to dele. Rapportens Del 1 præsenterer metoden og udvalgte resultater og effekter af klassestørrelsen på tværs af 0.-9. klassetrin. Rapportens Del 2 – Dokumentation indeholder metodebilag, der uddyber metoden og dokumenterer metodens validitet og effekternes robusthed.

¹ Indikatorerne for elevtrivsel eksisterer alene for 4.-9. klassetrin, mens de budgetøkonomiske konsekvenser kun analyseres for relevante årgange. For eksempel undersøges udfald relateret til ungdoms- og videregående uddannelse kun umiddelbart efter 9. klassetrin, mens udfald i grundskolen kun undersøges til og med 9. klasse.

2 Analyse og resultater

Dette kapitel præsenterer de beregnede effekter af klassestørrelse i folkeskolen. Analysen er foretaget separat for hvert klassetrin i 0.-9. klasse, men her beskrives og redegøres for de generelle fund og tendenser på tværs af klassetrin.

Kapitlet beskriver indledningsvist klassestørrelsen i danske folkeskoler, de forventede effekter af klassestørrelsen, data og analysemetode og fortolkningsmæssige forhold, som er vigtige for læseren at have med sig i den efterfølgende præsentation og analyse af effekter.

Afsnit 2.1 præsenterer de beregnede effekter af klassestørrelse på trivsel. Afsnit 2.2 præsenterer de beregnede sammenhænge mellem klassestørrelse og SØM-konsekvenser målt op til 4 år efter, og afsnit 2.3 diskuterer fortolkningen af persistensen over tid. Derefter diskuterer afsnit 2.4 og 2.5 fortolkningen af effekterne, men med to forskellige tilgange. Afsnit 2.4 kvalificerer effekterne ved at undersøge og diskutere deres robusthed over for forskellige metodevalg, og afsnit 2.5 kvalificerer effekternes størrelsesorden ved at inddrage en sammenligning med effekter fundet i andre danske og internationale studier. Slutteligt diskuterer afsnit 2.6 estimaternes implementering i REFUD.

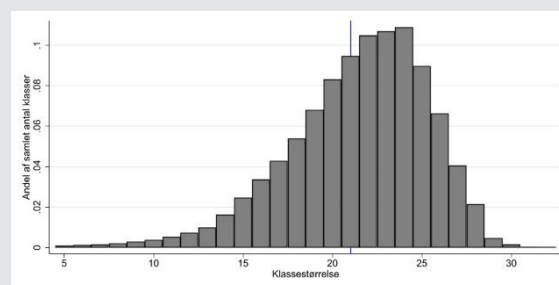
Klassestørrelse

Figur 2.1 viser fordelingen af klassestørrelse for alle fuldt årgangspdelte 0.-9. klasser i folkeskolen i analyseperioden 2014/2015-2018/2019. Figuren viser, at flest klasser har en størrelse på 20-25 elever, og den gennemsnitlige klassestørrelse er 21.04 elever.

Halvdelen af klasserne har højst 22 elever, 90 % af klasserne har højst 26 elever, og 99 % af klasserne har højst 28 elever.

Analysens metodestrategi tager afsæt i klasseloftsreglen, som foreskriver et loft på 28 elever i en klasse. Som Figur 2.1 viser, er det meget få klasser, der overskrider dette loft. Metodestrategien beskrives kort i Boks 2.1, hvor også sampleudvælgelsen præsenteres. Idet analysens estimater forventes at være mindre biased desto nærmere årgangsstørrelsen er på klasseloftstærsklerne,

Figur 2.1 Klassestørrelse, 0.-9. klasse



Anm.: Blå linje markerer den gennemsnitlige klassestørrelse.

Data indeholder alle fuldt årgangspdelte 0.-9. klasser i perioden 2014/15-2018/19. Af diskretionshensyn er data trunke ret ved klassestørrelser under 5 og over 32 elever. Data er beskrevet i kapitel 5.

samt at identifikationsstrategien primært er valid omkring de tre laveste tærskler (28, 56 og 84), er den gennemsnitlige klassestørrelse for klasserne i estimationssamplen 20,97 set i forhold til periodens gennemsnitlige klassestørrelse på 21,04 elever. Der er dog underliggende forskelle i fordelingen af klassestørrelser som følge af udvælgelseskriterierne for estimationssamplen, hvor især andelen af klasser med 21-25 elever er relativt lavere end i den fulde population.

Forventninger til effekter på rapportens udfaldsmål

Overordnet forventes en reduktion af klassestørrelsen *alt andet lige* at have betydning for elevernes læring; færre elever i klassen giver bedre tid til den enkelte og muliggør en højere grad af undervisningsdifferentiering. For eksempel giver flere interaktioner mellem lærer og elev øgede muligheder for at vurdere den enkeltes behov, tage individuelle hensyn og hermed understøtte den enkelte elevs læring (Filges m.fl. 2018). Dertil kommer, at lærer/elev-relationer også understøttes bedre, hvis den enkelte lærer har mere tid med den enkelte elev.

Det forventes samtidig, at der vil være mindre uro og konflikter, fordi klasserumsledelse er mindre kompliceret i en lille gruppe end i en større: Der er en højere lærer-elev-ratio, færre elever, som kan forstyrre hinanden, og færre kammerater at forholde sig til. Det er dog ikke entydigt, at færre elever i klassen nødvendigvis altid er til fordel for gruppedynamikken i en klasse. Større klasser medfører alt andet lige flere potentielle muligheder for sociale relationer, ligesom der i større klasser kan opleves en større faglig trivsel. For eksempel er der i større klasser bedre muligheder for gruppearbejde med andre med samme faglige interesser og fagligt niveau. Dermed er det ikke på forhånd givet, i hvilken retning effekterne på det generelle elevtrivselsindeks og indeksene for social og faglig elevtrivsel forventes at have, mens der forventes en positiv effekt af mindre klasser på trivselsindeksene Ro og orden og 'Støtte og inspiration' (se Boks 2.2 om trivselsudfaldsmål).

Klassens størrelse har særlig betydning i klassens første leveår (Filges m.fl. 2018), hvor der er flere dynamikker på spil, dels i forhold til at læreren skal lære de nye elever at kende og deres faglige og sociale styrker og svagheder, dels i forhold til at eleverne skal forholde sig til hinanden. På denne baggrund forventes lidt større effekter af klassestørrelsen i 0.-1. kl. og i 7. kl., hvor der typisk dannes nye klasser.

I litteraturen studeres effekterne af klassestørrelse oftest med henblik på elevers umiddelbare faglighed (typisk målt ved test scores eller karakterer) (Filges m.fl. 2018; se også BUVMs vidensdatabase), mens der i de senere år har været fokus på at omsætte disse til senere udfald på arbejdsmarkedet (Fredriksson m.fl. 2013; Leuven & Løkken 2020) samt forskellige dimensioner

af ikke-kognitive evner (fx Connolly & Haeck 2022). Overordnet set finder litteraturen negative, men beskedne, effekter af en marginal stigning i klassestørrelse på umiddelbar faglighed. Fredriksson m.fl. (2013) dokumenterer på svensk data, at disse umiddelbare effekter medfører betydelige negative løn-effekter i voksenlivet af en klassestørrelsesstigning, mens Leuven & Løkken (2020) baseret på norsk data estimerer præcise nul-effekter.

Budgetøkonomiske effekter måles som ændringer i borgernes forbrug af offentlige services og ydelser (se Boks 2.3). Det kan både være offentlige services og ydelser målrettet eleverne men også af deres forældre og lærere. Der er begrænset viden om klassestørrelseseffekter på disse typer af udfaldsmål i den eksisterende litteratur, som vi kan basere vores forventninger på.

I denne rapport undersøges de elevudfald, der er indeholdt i REFUD. Disse udfald aggregeres for eleven i hvert kalenderår. Det bemærkes, at en stor del af disse udfald tæller relativt sjældne begivenheder, hvorfor variationen i disse er betragteligt mindre end fx elevkarakterer eller testscores, og der stilles derfor større datamæssige krav i analysen for at kunne dokumentere statistisk signifikante effekter. Sådanne krav kan ikke nødvendigvis indfries i det tilgængelige datamateriale, og det understreges derfor, at fraværet af konsistente og signifikante effekter af klassestørrelse ikke nødvendigvis er ensbetydende med, at disse ikke eksisterer, men kan skyldes lav præcision i estimatet som følge af manglende variation i outcome.

Boks 2.1 Metode

Metodisk udfordring: Klassestørrelse er endogen

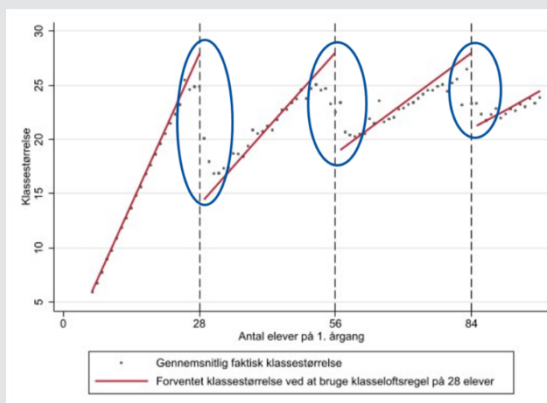
Elevens udfaldsmål og den klassestørrelse, som elever går i, kan være påvirket af en lang række fælles forhold, som betyder, at en simpel sammenligning mellem klassestørrelse og udfaldsmål vil være biased. For at tage højde for denne endogenitet i klassestørrelsesvariablen anvendes et instrument for klassestørrelse til at skabe eksogen variation i variabelen, hvorved den kausale effekt af *faktisk* klassestørrelse kan estimeres. Således udløser klassestørrelsesloftet på 28 elever under visse antagelser en eksogent drevet ændring i antallet af og deraf størrelsen på klasser, når årgangsstørrelsen krydser multipla af 28.

Faktisk klassestørrelse (sorte punkter i figuren)

Den faktiske klassestørrelse opgøres på baggrund af UDSP-data, som opgøres i april et givent skoleår. Det er derfor udtryk for den faktiske klassestørrelse, som antallet af elever udgør i slutningen af skoleåret.

Forventet klassestørrelse baseret på klassestørrelsesloftet

Den røde linje illustrerer den *forventede klassestørrelse* givet antallet af elever på årgangen ved skoleårets start på baggrund af Elevregisteret, som opgøres hvert år i september. Den forventede klassestørrelse beregnes under antagelse af et klassestørrelsesloft på 28 elever, og at årgangen opdeles i klasser af nøjagtig samme størrelse. Denne er stigende indtil 28 elever, hvorefter der er et skarpt fald, når årgangens elevantal når 29, og den forventede klassestørrelse derfor sænkes til 14,5 elever i hver af de nu to klasser på årgangen. Ligeledes er der et skarpt fald omkring tærskelværdierne 56 og 84, hvorefter årgangen deles op i hhv. 3 eller 4 klasser.



Klassestørrelsesloftet skaber et Fuzzy RD-design

Figuren illustrerer således to forhold, som vi anvender metodisk: 1) Der er en stærk sammenhæng mellem den forventede og faktiske klassestørrelse, og 2) der sker et eksogent fald i klassestørrelse omkring hver tærskelværdi pga. klasseloftsreglen. Idet klasseloftet er nationalt bestemt, kan denne betragtes som eksogent givet, og variationen skabt heraf betragtes dermed også som eksogen (se metodebilag, Del 2 Dokumentation).

Sammenligner elever på skoler, der ligger tæt på tærskelværdierne (blå cirkler)

Vi udvælger de skoler, der ligger tæt på tærskelværdierne (enten lige over eller lige under). Den præcise afstand bestemmes på baggrund af en afvejning af de forhold, der gælder for analysepopulationerne baseret på de forskellige afstande – nærmere bestemt præcision i forhold til gyldighed af det resulterende effektestimat (se metodebilag, Del 2 Dokumentation). Antagelsen er, at især skoler lige omkring tærskelværdierne er sammenlignelige, bortset fra at klassestørrelsesloftet udløser oprettelsen af en ekstra klasse og dermed mindre klassestørrelser for årgangstørrelser lige over hhv. lige under tærsklerne.

Denne udvælgelse af skoleårgange danner en analysepopulation med en indsats-indikatorer for hver tærskelværdi (hhv. 'Over 28', 'Over 56' og 'Over 84'). Indsats-indikatorerne anvendes som instrumenter for faktisk klassestørrelse i en 2SLS-estimation af effekten af faktisk klassestørrelse på udfaldsmål. Uddybes i metodebilag, afsnit 4.1.

Data

Som datagrundlag for analysen indgår alle normalklasseelever i fuldt årgangsopdelte klasser i de danske folkeskolors 0.-9. klasse i skoleårene 2014/2015-2018/2019.

Elevudfald måles i kalenderår, hvorfor de estimerede marginale klassestørrelses effekter skal fortolkes som effekten af at øge klassestørrelsen med én elev på et budgetøkonomisk udfald i førstkomende hele kalenderår. Undtaget herfra er effekterne på trivsel, der måles i sidste del af skoleåret, og derfor fortolkes som den marginale effekt af klassestørrelse på elevtrivslen i samme skoleår. Data, analysepopulation, og hvilke udfaldsmål der kan måles over tid, er uddybet i kapitel 5.

Metode til analyse af den kausale effekt af klassestørrelse

Idet sammenhængen mellem klassestørrelse og elevudfald, foruden den direkte årsagssammenhæng, formodes at være påvirket af en lang række fælles, men uobserverbare, faktorer, baseres analysen på en veletableret mikroøkonometrisk metode til estimation af den kausale effekt af klassestørrelse på elevudfald.

Metoden benytter klassestørrelsesloftet på 28 elever ved skoleårets begyndelse i et såkaldt Fuzzy RD-design til at skabe variation i den målte klassestørrelse, der ikke påvirkes af øvrige faktorer (se illustration i Boks 2.1). Ideen er, at klassestørrelsesloftet medfører, at nogle skoler må oprette flere, men mindre klasser på grund af reglen, mens skoler med næsten – men ikke helt –

samme årgangsstørrelse kan nøjes med færre men større klasser. Under antagelse af, at reglen binder, og at årgangsstørrelse (i visse mindre intervaller omkring klasseloftstærsklerne) ikke manipuleres, kan denne variation udnyttes til at estimere årsagssammenhængen mellem klassestørrelse og elevudfald for skoler, der er stort set ens i størrelse, men som af den ene eller anden grund er placeret lige over henholdsvis lige under klasseloftstærsklerne. Selve identifikationsstrategien samt implementeringen og gyldigheden heraf gennemgås detaljeret i kapitel 4.

Estimater af sådanne analyser på observationsdata angiver såkaldte *total policy effects*, dvs. de indeholder også læreres, forældres og øvrige aktørers reaktion på en ændring i klassestørrelsen. Dette kan have betydning for den sociale lighed i implementeringen af skolebaserede indsatser som fx ændringer i klassestørrelse. Flere studier finder, at især forældre justerer deres adfærd i det omfang, de har mulighed for, som reaktion på de forhold, der gælder for deres børns skolegang (fx Fredriksson m.fl. 2016; Pop-Eleches & Urquiola 2013). Hvis forældre med højere socioøkonomisk status har større mulighed for at kompensere deres børn for variation i mængden af skoleressourcer, vil deres børn være relativt mindre påvirkede af en klassestørrelsesstigning alt andet lige. Ligeledes kan skolen kompensere for klassestørrelsesstigninger ved fx at tildele klassen en mere erfaren lærer eller tilføje tolærerordninger i klassen (Andersen m.fl. 2020).

Analysen inkluderer desuden information om elevernes demografiske baggrund samt forældres uddannelsesniveaue og arbejdsmarkeds- og indkomstmæssige forhold. Ligeledes inkluderes basale skolekarakteristika til understøttelse af identifikationen i robusthedsanalyser.

Den marginale effekt af klassestørrelse

Den analyserede klassestørrelseseffekt er en *marginal effekt* i den forstand, at det for den udvalgte population er den gennemsnitlige effekt af at øge klassestørrelsen med én elev.

Analysen estimerer således ikke den 'optimale' klassestørrelse, men giver i stedet et indblik i, om der er forbedringer i vente ved at reducere (øge) klassestørrelsen under det nuværende regime (jf. klasseloftet på 28 elever, og at den gennemsnitlige klassestørrelse er 21 elever).

2.1 Effekter af klassestørrelse på trivsel

Dette afsnit præsenterer de estimerede marginale effekter af klassestørrelse på STILs fire trivselsdimensioner i folkeskolens nationale trivselsmåling for elever i 4.-9. klasserne: Faglig trivsel, Social trivsel, Ro og orden samt Støtte og inspiration samt generel elevtrivsel, der indeholder alle anvendte spørgsmål i de fire øvrige trivselsmål (STIL 2020).

Boks 2.2 Udfaldsmål: Trivsel

Der undersøges effekter på elevernes nationale trivselsmålinger gennemført i 4.-9. klasse. Vi anvender de 5 trivselsindikatorer fra Styrelsen for It og Læring (STIL):

- Faglig trivsel
- Social trivsel
- Støtte og inspiration
- Ro og orden
- Generel skoletrivsel

Faglig trivsel indeholder spørgsmål omhandlende elevernes oplevelse af egne faglige evner, koncentrationsevne og problemløsningsevne. Social trivsel omhandler elevernes opfattelse af deres egen tilhørsforhold til skolen, klassen og fællesskabet, samt tryghed og mobning. Ro og orden omhandler elevernes oplevelse af ro og støj i klassen samt klasseledelse. Støtte og inspiration omhandler elevernes oplevelse af motivation og medbestemmelse i timerne, samt af lærernes hjælp og støtte.

Ex ante er det ikke åbenlyst, at en større klasse vil påvirke de fire trivselsdimensioner i samme retning, hvilket forstyrrer fortolkningen af effektestimaterne på den samlede generelle elevtrivselsscore. Trivselsdimensionerne Støtte og inspiration samt Ro og orden forventes at være negativt påvirkede af en stigning i klassestørrelse, idet sandsynligheden for afbrydelser af undervisningen stiger (Lazear 2001), og læreren alt andet lige får mindre tid per elev. Derimod er det muligt, at en større klasse forbedrer både social og faglig trivsel, idet der forefindes øgede muligheder for sociale interaktioner, samt for at klassen indeholder flere elever på samme undervisningsniveau – særligt for elever på meget høje og meget lave niveauer (Filges m.fl. 2018).

Den nationale trivselsmåling udføres normalt i januar til marts hvert år, og måler således elevtrivsel i sidste halvdel af skoleåret. Dermed fortolker vi de estimerede klassestørrelses effekter på elevtrivsel som effekten af at øge klassestørrelsen med én elev på elevtrivslen i samme skoleår *alt andet lige*.

Tabel 2.1 rapporterer de estimerede trivselseffekter af en marginal ændring i klassestørrelse for hvert klassetrin separat.

Med undtagelse af trivselsdimensionen Ro og orden, er effektestimaterne meget små (under 0,5 % af en standardafvigelse, SD) og relativt præcise nul-effekter. Der er et enkelt positivt effektestimater af en marginal stigning i klassestørrelse på social trivsel i 6. klasse på 0,006 SD, men den er i sig selv ikke overbevisende, idet effektestimaterne på Social trivsel for de omkringliggende klassetrin er meget små, insignifikante og skifter fortegn, og er derfor meget muligt et udtryk for en type-I-fejl, hvor egentlige nul-effekter detekteres som signifikante.

Tabel 2.1 Standardiserede effekter af klassestørrelse på elevtrivsel, år 1.

Udfaldsmål	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
Generel elevtrivsel						
Effektestimater	-0,005	-0,005	0,000	-0,001	-0,005	-0,007
Standardfejl	0,004	0,004	0,004	0,005	0,005	0,006
p-værdi	0,228	0,238	0,984	0,879	0,350	0,232
Antal observationer	62.468	61.313	62.207	54.776	53.116	46.391
F-test for 1.stage	148,1	117,0	145,2	100,3	73,8	60,6
Faglig trivsel						
Effektestimater	0,000	-0,001	-0,002	0,003	0,000	-0,006
Standardfejl	0,003	0,003	0,003	0,003	0,004	0,004
p-værdi	0,908	0,645	0,555	0,435	0,958	0,166
Antal observationer	62.306	61.234	62.159	54.749	53.128	46.439
F-test for 1.stage	147,5	117,3	145,4	100,3	73,7	60,6
Social trivsel						
Effektestimater	0,000	-0,001	0,006	0,002	-0,003	-0,004
Standardfejl	0,003	0,003	0,003	0,004	0,004	0,005
p-værdi	0,941	0,834	0,048	0,686	0,505	0,410
Antal observationer	62.463	61.302	62.210	54.801	53.184	46.495
F-test for 1.stage	148,1	117,2	145,3	100,2	74,0	60,8
Ro og orden						
Effektestimater	-0,015	-0,018	-0,011	-0,015	-0,013	-0,013
Standardfejl	0,004	0,004	0,004	0,005	0,006	0,006
p-værdi	0,000	0,000	0,011	0,006	0,022	0,039

Udfaldsmål	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
Antal observationer	62.479	61.308	62.194	54.743	53.085	46.344
F-test for 1.stage	148,1	117,2	145,4	100,4	73,7	60,5
Støtte og inspiration						
Effektestimat	-0,007	-0,003	-0,001	0,001	-0,003	-0,001
Standardfejl	0,004	0,004	0,004	0,005	0,005	0,006
p-værdi	0,085	0,462	0,815	0,848	0,551	0,847
Antal observationer	62.304	61.224	62.113	54.715	53.025	46.302
F-test for first stage	147,8	117,2	145,3	100,1	73,7	60,2

Anm: F-test for first stage angiver F-testet for fælles signifikans af modellens instrumenter (indikatorer for årgangsstørrelse over klasselofts-tærskler) i first stage-regressionen, dvs. om instrumenterne har relevans i forhold til at forklare den observerede klassestørrelse.

2SLS-estimerne er baseret på populationen af folkeskoleelever i fuldt årgangssopdelte normalklasser i skoleårgange med årgangsstørrelser +/- fem elever omkring de tre nederste klasselofts-tærskler i skoleårene 2014/15-2018/19. Trivselsmålene er standardiserede. 2SLS-regressionerne kontrollerer desuden for årgangsstørrelsessegment, år for måling af klassestørrelse, alder og kvadreret alder, køn, ikkevestlig baggrund, antallet af ældre søskende, moderens og faderens uddannelse, alder ved fødsel og kvadreret alder ved fødsel, samt indikatorer for, om moderen og faderen er diagnosticeret med angst og depression hhv. opmærksomhedsforstyrrelser. Standardfejl er klyngekorrigeret på skoleniveau. Effektestimater, der er signifikante på et 5 %-niveau, er markeret med **fed**.

Derimod er effektestimaterne for Ro og orden konsekvent negative og i cirka samme størrelsesorden over klassetrinnene. En tilføjelse af en ekstra elev i klassen sænker således elevernes opfattelse af ro i klassen og god klasseledelse med 0.01-0.02 SD i 4.-9. klasse. Ydermere fastholdes signifikansen af effektestimaterne generelt ved justeringen af P-værdierne til skarpe q-værdier ved at kontrollere for *false detection rate* (FDR) og således tage højde for multipel inferens i de mange trivselsestimater (fx Anderson 2008) (se også metodebilag, afsnit 4.5). Undtaget herfra er effektestimatet for 9. klasse, der dog stadig har en skarp q-værdi under 0,10.

På baggrund heraf vurderes det, at trivselsdimensionen Ro og orden kan anvendes som linkoutcome for beregninger af klassestørrelsesindsatser i REFUD.

2.2 Effekter af klassestørrelse på SØM-konsekvenser

I alt måler REFUD på 54 budgetmæssige konsekvenser (se Boks 2.3), hvoraf 26 af dem betegnes som førsteordensafledte (default) udfald.² Dette udsætter analysen for en relativt stor risiko for type I-fejl, hvor usande effekter detekteres som signifikante, idet den statistiske signifikans betragtes individuelt for

² REFUD's default konsekvenser tæller de udfald, der forventes umiddelbart at blive påvirkede af skole-baserede indsatser gennem elevers faglighed, trivsel og lærernes arbejdsmiljø, dvs. udfald omhandlende elevens videre skolegang, beskæftigelse og risikofyldt adfærd i form af kontakt med sundhedsvæsen samt politi og retsvæsen. Herudover indgår forældrenes lønindkomst til måling af eventuel beskæftigelsesrespons til indsatsens direkte elev effekter.

hvert estimat. Men selv med fokus alene på de 26 primære default konsekvenser (estimeret på op til 10 klassetrin i en 4-årig tidshorisont, hvor muligt), er det umuligt at kontrollere familywise error rate (og selv false detection rate), uden at styrken reduceres i sådan en grad, at det vil være usandsynligt at dokumentere signifikante effekter af en realistisk størrelsesorden for en marginal ændring i klassestørrelse, se afsnit 4.5 for yderligere detaljer. Derfor er der i udarbejdelsen af denne rapport ikke brugt ressourcer herpå. Med dette in mente, behandler rapportens resultat afsnit alene tendenser i estimater på de primære default konsekvenser. Læseren opfordres desuden til at betragte resultaterne som en helhed snarere end enkeltvist.

Bilagstabel 2.1 giver et overblik over antallet af signifikante effekter, der er estimeret for hvert klassetrin i hhv. 1., 2., 3. og 4. budgetår efter ændring i klassestørrelse, herunder om effekten er positiv eller negativ. Selvom den umiddelbare effekt af klassestørrelse naturligvis er interessant i sig selv, er persistensen af sådanne effekter i efterfølgende år ligeledes af interesse i vurderingen af fordelagtigheden af at implementere en ændring i klassestørrelse. En stor forventet fade out af effekterne vil således gøre det mindre attraktivt på længere sigt at reducere klassestørrelsen. Empirisk er persistensen i effekten af en klassestørrelsesændring i populationsdata dog svær at adskille fra effekten af senere klassestørrelser, der ofte vil være relativt sammenlignelige, så længe eleven fortsætter i samme klasse (udbydes i afsnit 2.3).

I relation til forståelsen af estimaterne som *total policy effects*, fremhæves desuden den konceptuelle forskel mellem punkttestimaterne, der gælder samme kohorter i samme år, men på tværs af tidshorisonten. Således gælder punkttestimaterne for 3. klassetrin i år 2 og 4. klasse i år 1 begge for 4. klassetrin, men med den afgørende forskel, at alle individer, der gik i folkeskolens normalklasse på 3. klassetrin, inkluderes i førstnævnte uanset deres skolegang i det efterfølgende år, mens sidstnævnte betinger på, at individerne skal gå i folkeskolens normalklasse på 4. klassetrin.

Af oversigten i Bilagstabel 2.1 ses det, at der er enkeltvis signifikante estimater på en række af default udfaldsmål. For nogle udfaldsmål vurderes disse som interessante og relevante tendenser, blandt andet fordi de er konsistente på tværs af klassetrin og år i forhold til forventningen til disse udfaldsmål, mens mere eller mindre sporadiske signifikante estimater for andre udfaldsmål på tværs af klassetrin og år vurderes som usikre og mindre relevante.³

I det følgende har vi udvalgt seks udfaldsmål, der illustrerer forskellige mønstre i punkttestimaterne, for en generel præsentation af resultaterne.

³ Bilagstabel 2.2 rapporterer tilsvarende overblikstabel for ikke-default udfaldsmål.

Boks 2.3 Udfaldsmål: SØM-konsekvenser

Budgetøkonomiske effekter opstår af ændringer i borgernes forbrug af offentlige services og ydelser. Det kan både være offentlige services og ydelser målrettet eleverne men også af deres forældre og lærere.

- For eleverne omfatter det forbrug af grundskole og specialundervisning, 10. klasse og ungdomsuddannelser, sociale foranstaltninger til børn og unge under 18 år, efterværn (til 18-22-årige), sundhedsydelser, kontakt med retsvæsen og kriminalforsorg, indkomstoverførsler, beskæftigelsesindsatser, beskæftigelse og sociale serviceydelser. Flere af disse er først relevante fra eleven fylder 15-18 år.
- For elevernes forældre omfatter det mødre og fædres indkomstoverførsler, beskæftigelsesindsatser og beskæftigelse.
- For elevernes lærere omfatter det deres personalefravær.

Tendenser på udfaldsmål vedrørende grundskole

De følgende tabeller viser estimatet på den marginale effekt af klassestørrelse i førstkomende budgetår (år 1) og de efterfølgende budgetår (hhv. år 2, 3 og 4) på fire udvalgte udfaldsmål vedrørende skolevalg og specialtilbud.

Tabel 2.2 viser estimater på forbrug af hhv. kommunal grundskole (folkeskole) og frie/private grundskoler. Dette er interessant, fordi de giver et indblik i forældre og elevers reaktion på klassestørrelse i forhold til skolevalg (Frederiksson m.fl. 2016, Johansen, Mølbæk & Nandrup 2020).

Tabel 2.2 Effekter på grundskole, udvalgte

Udfaldsmål	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
Antal uger indskrevet i kommunal grundskole										
År 1										
Effektestimat	0,004	-0,005	0,001	-0,001	-0,004	0,002	-0,006	0,004	0,008	0,001
Standardfejl	0,003	0,003	0,003	0,003	0,002	0,003	0,005	0,004	0,004	0,002
p-værdi	0,146	0,079	0,842	0,744	0,052	0,595	0,224	0,316	0,057	0,628
År 2										
Effektestimat	0,000	-0,004	-0,001	-0,004	-0,005	-0,001	-0,005	0,003	0,007	
Standardfejl	0,003	0,002	0,003	0,003	0,002	0,003	0,005	0,004	0,004	
p-værdi	0,964	0,131	0,691	0,150	0,065	0,878	0,306	0,398	0,083	
År 3										
Effektestimat	-0,003	-0,003	-0,003	-0,007	-0,006	0,002	-0,001	0,007		
Standardfejl	0,003	0,003	0,003	0,003	0,003	0,005	0,004	0,005		
p-værdi	0,409	0,352	0,248	0,026	0,054	0,663	0,899	0,152		
År 4										
Effektestimat	0,001	-0,004	-0,008	-0,006	-0,006	0,004	0,000			
Standardfejl	0,004	0,004	0,004	0,005	0,004	0,004	0,004			
p-værdi	0,869	0,342	0,042	0,215	0,218	0,422	0,942			
Antal uger indskrevet i fri- og privat grundskole										
År 1										
Effektestimat	-0,001	0,003	0,002	0,006	0,004	-0,002	0,007	-0,004	-0,003	
Standardfejl	0,003	0,002	0,002	0,003	0,002	0,003	0,005	0,004	0,003	
p-værdi	0,649	0,153	0,422	0,036	0,063	0,491	0,126	0,280	0,271	
År 2										
Effektestimat	0,001	0,005	0,002	0,006	0,004	0,000	0,006	-0,004	-0,003	
Standardfejl	0,003	0,003	0,003	0,003	0,002	0,004	0,005	0,004	0,003	
p-værdi	0,726	0,035	0,439	0,039	0,127	0,897	0,207	0,232	0,314	
År 3										
Effektestimat	0,003	0,004	0,004	0,008	0,007	-0,003	0,000	-0,004		
Standardfejl	0,003	0,003	0,003	0,003	0,003	0,005	0,004	0,004		
p-værdi	0,320	0,183	0,186	0,021	0,043	0,606	0,971	0,237		
År 4										
Effektestimat	0,000	0,002	0,006	0,004	0,006	-0,004	-0,002			
Standardfejl	0,004	0,004	0,004	0,005	0,005	0,005	0,004			
p-værdi	0,913	0,669	0,105	0,375	0,223	0,443	0,603			

Anm.: Estimater baseret på under 100 elever er udeladt fra tabellen.

2SLS-estimaterne er baseret på populationen af folkeskoleelever i fuldt årgangsopdelte normalklasser i skoleårgange med årgangsstørrelser +/- fem elever omkring de tre nederste klasseloftstærskler i skoleårene 2014/15-2018/19. 2SLS-regressionerne kontrollerer desuden for årgangsstørrelsessegment, år for måling af klassestørrelse, alder og kvadreret alder, køn, ikkevestlig baggrund, antallet af ældre søskende, moderens og faderens uddannelse, alder ved fødsel og kvadreret alder ved fødsel, samt indikatorer for, om moderen og faderen er diagnosticeret med angst og depression hhv. opmærksomhedsforstyrrelser. Standardfejl er klyngekorrigeret på skoleniveau. Effektestimater, der er signifikante på et 5 %-niveau, er markeret med fed.

Estimaterne indikerer en tendens til, at en større klasse medfører et øget forbrug af fri- og privat grundskole i 3.-4. kl., hvilket tidsmæssigt stemmer overens med, at forældre revurderer skolevalg omkring overgangen fra indskoling til mellemtrinnet på baggrund af deres skoleerfaringer, herunder klassestørrelsen, i løbet af indskoling. Enkeltvis er estimaterne for 3. klasse signifikante på 5 %-signifikansniveau (4. klasse på 10 %-signifikansniveau) og

persistente over den 4-årig tidshorizont (dog, insignifikant i år 4). Samtidig medfører en marginal stigning i klassestørrelse et tilsvarende reduceret forbrug af kommunal grundskole for de samme årgange (dog estimeret med generel mindre præcision).

Denne tendens bekræftes af estimaterne for 2. klasse, om end lidt svagere (der er signifikant mindre forbrug af folkeskolen i år 4, hvilket stemmer overens med resultaterne for 3.-4. klasse).⁴

⁴ Skoleskift relateret til klassestørrelsesloftet kan udfordre validiteten af estimationsstrategien (se afsnit 4.1 og 4.2). I særdeleshed, hvis skoleadministratorer afviser bestemte elevtyper for at undgå oprettelsen af nye klasser, eller forældre vælger skole på baggrund af årgangsstørrelsen, der udløser en bestemt klassestørrelse. Dette understreger vigtigheden af at betragte årgangsstørrelser i smalle intervaller omkring klasseloftstærsklerne, idet præcis manipulation af årgangsstørrelsen her er svære og relativt mere omkostningsfyldt.

Tabel 2.3 viser estimater på antal uger indskrevet i hhv. almenklasse med specialundervisning (min. 9 timer/uge) og på specialskole. Disse udfaldsmål er interessante, da tildeling af specialundervisning, ud over en forventet reaktion på klasse- og læringsmiljøet, særligt i almenklasserne, kan fungere som direkte kompensation for ændret klassestørrelse.

Tabel 2.3 Effekter på specialundervisning og specialskole, udvalgte

Udfaldsmål	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
Specialundervisning i almenklasse, antal uger indskrevet										
År1										
Effektestimat	0,004	0,000	0,005	0,001	0,005	0,001	0,003	-0,005	0,000	0,004
Standardfejl	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,004
p-værdi	0,181	0,905	0,107	0,789	0,129	0,610	0,232	0,052	0,898	0,359
År2										
Effektestimat	0,006	0,000	0,007	0,000	0,003	0,002	0,002	-0,002	-0,001	
Standardfejl	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,002	0,003	
p-værdi	0,052	0,873	0,038	0,999	0,304	0,485	0,576	0,505	0,712	
År3										
Effektestimat	0,003	0,002	0,004	0,002	-0,002	0,001	0,001	-0,005		
Standardfejl	0,003	0,004	0,004	0,004	0,004	0,004	0,003	0,003		
p-værdi	0,371	0,542	0,276	0,582	0,567	0,764	0,792	0,136		
År4										
Effektestimat	0,006	0,003	0,001	0,001	0,000	0,005	0,000			
Standardfejl	0,004	0,005	0,004	0,005	0,005	0,004	0,003			
p-værdi	0,132	0,521	0,706	0,824	0,991	0,183	0,852			
Antal uger indskrevet på specialskole										
År1										
Effektestimat	-0,002	0,000	0,001	0,000	0,001	-0,001	0,005	0,006	0,000	0,000
Standardfejl	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,003	0,002
p-værdi	0,390	0,863	0,568	0,829	0,524	0,774	0,019	0,011	0,929	0,829
År2										
Effektestimat	-0,003	0,000	-0,001	-0,002	0,003	-0,001	0,002	0,004	0,000	
Standardfejl	0,003	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002	
p-værdi	0,226	0,904	0,723	0,332	0,107	0,772	0,243	0,060	0,972	
År3										
Effektestimat	-0,001	0,000	-0,002	-0,001	0,002	-0,001	0,001	0,000		
Standardfejl	0,003	0,002	0,002	0,002	0,002	0,002	0,002	0,002		
p-værdi	0,774	0,870	0,391	0,663	0,441	0,648	0,723	0,984		
År4										
Effektestimat	0,000	0,004	0,000	0,001	0,001	-0,003	0,000			
Standardfejl	0,003	0,003	0,003	0,003	0,003	0,003	0,002			
P-værdi	0,883	0,201	0,897	0,720	0,574	0,384	0,839			

Anm.: Estimater baseret på under 100 elever er udeladt fra tabellen.

2SLS-estimaterne er baseret på populationen af folkeskoleelever i fuldt årgang opdelt i normalklasser i skoleårgange med årgangsstørrelser +/- fem elever omkring de tre nederste klasseloftstærskler i skoleårene 2014/15-2018/19. 2SLS-regressionerne kontrollerer desuden for årgangsstørrelsessegment, år for måling af klassestørrelse, alder og kvadreret alder, køn, ikkevestlig baggrund, antallet af ældre søskende, moderens og faderens uddannelse, alder ved fødsel og kvadreret alder ved fødsel, samt indikatorer for, om moderen og faderen er diagnosticeret med angst og depression hhv. opmærksomhedsforstyrrelser. Standardfejl er klyngekorrigeret på skoleniveau. Effektestimater, der er signifikante på et 5 %-niveau, er markeret med fed.

Estimaterne af en marginal ændring i klassestørrelse på tildeling af specialundervisning i almenklasse er generelt meget små og insignifikante i det førstekommende budgetår (år 1) og efterfølgende år (år 2-4).

Der er marginalt signifikante positive estimater i år 1 og 2 efter en klassestørrelsesændring i 2. klasse, der angiver, at en stigning i klassestørrelse forårsager mere forbrug af specialundervisning i år 1 og 2. Idet de omkringliggende estimater er tættere på nul og usikkert estimerede, vil vi ikke tillægge estimaterne for 2. klasse for stor vægt. Selvom det virker sandsynligt, at visitation til specialundervisning først foretages i slutningen af indskolingen og derfor kan forklare, at der ikke er effekter før i 2. klasse, er der heller ingen effekter for 3. klasse og senere, hvilket man ville forvente. I afsnit 2.4 uddybes disse estimaters robusthed for at styrke denne konklusion.⁵

Estimaterne antyder desuden, at en marginal stigning i klassestørrelsen flytter elever over i specialskolen ved overgangen til udskolingen, om end denne tendens er svag og kun lokaliseret omkring 6.-7. klassetrin.

Effekter på kontakt med sundhedsvæsen

Tabel 2.4 viser den marginale effekt af klassestørrelse i førstekommende budgetår (år 1) og de efterfølgende budgetår (hhv. år 2, 3 og 4) på antal psykiatriske ambulante- og skadestuebesøg og somatiske skadestuebesøg.

Disse udfaldsmål er relevante i forhold til elevernes trivsel og adfærd i klassen. Ambulant og skadestuekontakt til psykiatrien kan både være et udtryk for øget opmærksomhed på og opsporing af elevernes potentielle udfordringer, men kan også være afledt af konkrete hændelser i eller uden for skolen. Somatiske skadestuebesøg kan ligeledes være et udtryk for hændelser i eller uden for skolen, hvor en hypotese kan være, at større klasser medfører mere uro og flere konflikter, som potentielt kan ende i ulykker eller voldelige hændelser.

Estimaterne på besøg i psykiatrien er generelt små og upræcist estimeret (standardfejl på størrelse med eller større end estimatet). Tabellen viser samtlige estimater på 'antal psykiatriske skadestuebesøg', skønt estimaterne for 0.-5. klasse er baseret på så lille et datagrundlag, at vi ellers anbefaler helt at udelade disse. De er rapporteret for at illustrere, at man kan overveje at samle flere af de meget specifikke SØM-konsekvenser (fx samle de psykiatriske ambulantebesøg og skadestuebesøg) for at opnå et bedre datagrundlag og potentielt mere præcise estimater. Det er dog uden for rammerne af denne rapport at definere andre udfaldsmål end de eksisterende i REFUD.

⁵ Herunder estimatet på 7. klasse, der angiver, at en stigning i klassestørrelse forårsager mindre forbrug af specialundervisning (p-værdi=0,052). Robusthedsanalysen konkluderer, at det ikke er gyldigt.

Der er sporadiske, overvejende positive effekter af en marginal stigning i klas-
sestørrelse på antallet af somatiske skadestuebesøg. Her kan også diskuteres,
om udfaldsmålet er for specifikt konstrueret, og man med fordel kan samle
flere udfaldsmål for at øge sample og den statistiske styrke.

Dette eksempel illustrerer afvejningen mellem anvendelighed og statistisk sik-
kerhed og bias i forhold til inklusionen af punktestimater til beregning i REFUD
i de tilfælde, hvor der er signifikante men usystematiske effekter over klasse-
trin og år. BUVM kan fx overveje, om man på baggrund af signifikante effekter
i 0. og 5. klasse vil inkludere disse alene, delvist eller helt udelade estimer
for dette udfaldsmål. For eksempel fastholdes signifikansen af punktestima-
terne på somatiske skadestuebesøg for år 1 og 3 ved justeringen af P-værdi-
erne til skarpe q-værdier i forsøg på at kontrollere FDR over den 4-årige tids-
horisont for 5. klassetrin alene (punktestimatet for år 2 bliver derimod margi-
nalt signifikant). 0. klasseestimerne overlever ikke en tilsvarende justering.
Ingen af punktestimaterne forbliver signifikante, hvis justeringen udvides til at
omfatte både klassetrin og tidshorisont. Alternativt, kan estimerne for de
mellemliggende klassetrin medtages ud fra en helhedsbetragtning upåagtet
deres præcision.

Tabel 2.4 Effekter på kontakt med sundhedsvæsen, udvalgte

Udfaldsmål	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
Antal psykiatrisk ambulante besøg										
År1										
Effektestimat	-0,001	0,001	-0,003	0,002	-0,003	-0,001	0,002	0,003	0,001	0,001
Standardfejl	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,003	0,003
p-værdi	0,634	0,704	0,258	0,513	0,102	0,642	0,194	0,265	0,779	0,723
År2										
Effektestimat	0,006	0,001	0,000	-0,003	-0,002	-0,002	0,004	-0,002	0,002	-0,001
Standardfejl	0,004	0,002	0,003	0,003	0,002	0,002	0,002	0,003	0,003	0,003
p-værdi	0,114	0,700	0,896	0,426	0,287	0,522	0,086	0,450	0,635	0,800
År3										
Effektestimat	-0,001	0,002	-0,005	0,000	-0,003	0,003	0,001	-0,003	0,008	0,002
Standardfejl	0,003	0,004	0,003	0,003	0,003	0,003	0,003	0,004	0,006	0,004
p-værdi	0,805	0,543	0,101	0,914	0,223	0,198	0,789	0,472	0,138	0,596
År 4										
Effektestimat	0,013	0,002	-0,007	0,004	0,002	0,003	0,001	-0,001	-0,002	-0,004
Standardfejl	0,007	0,004	0,005	0,004	0,003	0,005	0,004	0,008	0,005	0,007
p-værdi	0,077	0,553	0,142	0,387	0,435	0,522	0,893	0,879	0,742	0,523
Antal psykiatriske skadestuebesøg										
År1										
Effektestimat	-0,003	0,001	0,002	0,000	0,000	0,003	0,000	-0,001	0,000	0,000
Standardfejl	0,002	0,002	0,002	0,000	0,000	0,003	0,000	0,001	0,000	0,000
p-værdi	0,118	0,547	0,477	0,229	0,381	0,308	0,831	0,367	0,877	0,261

Udfaldsmål	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
År2										
Effektestimat	0,002	-0,001	0,004	0,000	0,001	0,001	0,000	0,003	0,001	-0,001
Standardfejl	0,003	0,002	0,003	0,000	0,000	0,002	0,000	0,002	0,001	0,002
p-værdi	0,523	0,784	0,147	0,851	0,011	0,624	0,815	0,118	0,390	0,601
År3										
Effektestimat	-0,002	0,004	0,000	0,000	0,000	0,002	0,004	-0,004	0,003	-0,002
Standardfejl	0,002	0,002	0,002	0,001	0,003	0,003	0,002	0,004	0,005	0,003
p-værdi	0,490	0,118	0,857	0,770	0,96	0,499	0,068	0,332	0,484	0,500
År4										
Effektestimat	0,011	0,002	0,002	0,004	0,004	0,009	0,005	-0,011	0,004	-0,001
Standardfejl	0,007	0,003	0,004	0,004	0,004	0,004	0,004	0,007	0,007	0,003
p-værdi	0,107	0,584	0,596	0,331	0,429	0,025	0,254	0,128	0,535	0,681
Antal somatiske skadestuebesøg										
År1										
Effektestimat	0,003	0,002	0,001	0,002	0,001	0,005	0,002	-0,001	-0,001	-0,002
Standardfejl	0,002	0,002	0,002	0,002	0,002	0,002	0,001	0,002	0,002	0,002
p-værdi	0,047	0,233	0,771	0,261	0,508	0,011	0,242	0,378	0,556	0,271
År2										
Effektestimat	0,006	0,003	-0,001	0,007	0,002	0,002	0,001	-0,003	-0,002	0,004
Standardfejl	0,004	0,002	0,002	0,003	0,003	0,002	0,002	0,002	0,002	0,005
p-værdi	0,100	0,234	0,719	0,035	0,565	0,145	0,410	0,099	0,374	0,411
År3										
Effektestimat	0,001	0,004	0,003	0,003	0,002	0,005	-0,001	0,005	0,001	-0,005
Standardfejl	0,004	0,002	0,003	0,003	0,002	0,002	0,002	0,003	0,003	0,003
p-værdi	0,834	0,058	0,411	0,254	0,279	0,012	0,736	0,084	0,789	0,077
År4										
Effektestimat	-0,002	0,013	-0,011	0,007	0,000	0,001	-0,003	-0,003	0,000	-0,012
Standardfejl	0,007	0,006	0,005	0,004	0,004	0,002	0,003	0,005	0,006	0,008
p-værdi	0,828	0,024	0,025	0,122	0,924	0,720	0,351	0,578	0,931	0,147

Anm.: De røde tal indikerer, at disse estimater er baseret på under 100 elever og derfor vurderes for usikre og bør udelades.

2SLS-estimaterne er baseret på populationen af folkeskoleelever i fuldt årgangspdelte normalklasser i skoleårgange med årgangsstørrelser +/- fem elever omkring de tre nederste klasseloftstærskler i skoleårene 2014/15-2018/19. 2SLS-regressionerne kontrollerer desuden for årgangsstørrelsessegment, år for måling af klassestørrelse, alder og kvadreret alder, køn, ikkevestlig baggrund, antallet af ældre søskende, moderens og faderens uddannelse, alder ved fødsel og kvadreret alder ved fødsel, samt indikatorer for om moderen og faderen er diagnosticeret med angst og depression hhv. opmærksomhedsforstyrrelser. Standardfejl er klyngekorrigeret på skoleniveau. Effektestimater, der er signifikante på et 5 %-niveau, er markeret med fed.

Effekter på kriminalitet

Tabel 2.5 viser den marginale effekt af klassestørrelse i førstkomende budgetår (år 1) og de efterfølgende budgetår (hhv. år 2, 3 og 4) på ét udvalgt udfaldsmål omkring kriminalitet: 'Antal fængselsdage'. Det kan måles for 15-årige og ældre, og derfor først relevant for de ældste klassetrin eller senere tidshorisonter (se Tabel 2.5). Eftersom der naturligt nok er få elever, der dømmes og fængsles, er der her en væsentlig dataudfordring, som Tabel 2.5 er taget med for at eksemplificere.

Tabel 2.5 Effekter på kriminalitet, udvalgte

Udfaldsmål	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
Antal fængselsdage										
År1										
Effektestimat								0,000	0,000	0,000
Standardfejl								0,000	0,002	0,002
p-værdi								0,387	0,865	0,910
År2										
Effektestimat							0,000	0,000	0,006	0,001
Standardfejl							0,000	0,003	0,003	0,002
p-værdi							0,849	0,895	0,042	0,776
År3										
Effektestimat						-0,001	-0,001	0,001	0,005	0,004
Standardfejl						0,001	0,001	0,003	0,003	0,003
p-værdi						0,312	0,298	0,765	0,078	0,241
År4										
Effektestimat					0,001	-0,001	0,000	-0,001	0,008	0,002
Standardfejl					0,001	0,002	0,002	0,003	0,004	0,004
p-værdi					0,316	0,637	0,987	0,863	0,054	0,679

Anm.: De røde tal indikerer, at disse estimater er baseret på under 100 elever og derfor vurderes for usikre og bør udelades.

2SLS-estimaterne er baseret på populationen af folkeskoleelever i fuldt årgangsoopdelte normalklasser i skoleårgange med årgangsstørrelser +/- fem elever omkring de tre nederste klasseloftstærskler i skoleårene 2014/15-2018/19. 2SLS-regressionerne kontrollerer desuden for årgangsstørrelsessegment, år for måling af klassestørrelse, alder og kvadreret alder, køn, ikkevestlig baggrund, antallet af ældre søskende, moderens og faderens uddannelse, alder ved fødsel og kvadreret alder ved fødsel, samt indikatorer for om moderen og faderen er diagnosticeret med angst og depression hhv. opmærksomhedsforstyrrelser. Standardfejl er klyngekorrigeret på skoleniveau. Effektestimater, der er signifikante på et 5 %-niveau, er markeret med fed.

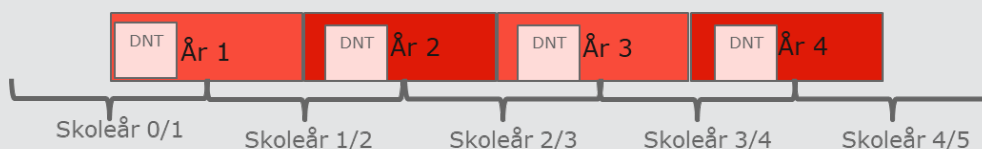
Estimaterne er positive, hvilket indikerer, at en øget klassestørrelse kan øge kriminalitet, her målt som antal fængselsdage. Estimaterne er tilmed signifikante på et 10 %-niveau for 8. klasse i alle årene 2-4 (hvor de er gamle nok til at blive dømt for kriminalitet). Men, jf. de generelle kriterier opsat, udelades alle estimater (undtagen 9. klasse målt i år 4), da der er færre end 100 elever med en positiv værdi på udfaldsmålet (markeret med rødt i tabellen), og dermed drives estimatet af ganske få individer. Med andre ord, anbefales det, at estimater på kriminalitet (i deres nuværende form) ikke direkte benyttes i REFUD.

Kriminalitet er dog en kategori af udfaldsmål, hvor der ikke kan forventes meget større datagrundlag. I fremtidige analyser bør man derfor overveje at lave en særskilt analyse om kriminalitet og undersøge muligheder for small-sample estimationsmetoder og korrektion, hvilket er uden for rammerne for denne analyse, hvor formålet er at etablere en generel model til analyse af klassestørrelses effekter på tværs af alle klassetrin og udfaldsmål.

Boks 2.4 Fortolkning og timing

Forskydning i skoleåret i forhold til budgetåret

Skoleåret starter i august og løber frem til juli det efterfølgende kalenderår. Kommunernes budgetår følger derimod typisk kalenderåret, og beregningerne i REFUD følger derfor kalenderåret. Dvs. at forbrug af skoleydelse mv. er opgjort per år (pba. registerdata).



Note: DNT: De nationale trivselsmålinger.

Eksempel på tidshorisont for årgangen af elever i 2. klasse i skoleår 2014/15

- År 1: Deres klassestørrelse måles i skoleår 2014/15 og deres udfaldsmål måles i 2015
- År 2: Deres klassestørrelse målt i 2. klasse i skoleår 2014/15, udfaldsmål måles i 2016
- År 3: Deres klassestørrelse målt i 2. klasse i skoleår 2014/15, udfaldsmål måles i 2017
- År 4: Deres klassestørrelse målt i 2. klasse i skoleår 2014/15, udfaldsmål måles i 2018

Betydning for fortolkning af effektestimater

Denne forskydning i måling af modellens komponenter betyder, at der er et lag på et halvt skoleår fra indsatsen igangsættes til måling af SØM-konsekvenser.⁶ Det vurderes rimeligt under antagelse af, 1) at der er et lag i response (det tager lidt tid førend et udfaldsmål påvirkes af den ændrede klassestørrelse, og 2) at der er persistens i klassestørrelser og ens klassestørrelseseffekter. Antagelse 1) betyder også, at forbrug af fx specialundervisning ikke reagerer umiddelbart på klassestørrelse (altså, at der går lidt tid, før skolen ændrer på specialundervisning).

Det vurderes, at antagelserne er rimelige, og at klassestørrelseseffekterne fortsat har berettigelse, idet der formodes at være et vist lag i response fra indførelse af en ændring i klassestørrelse, til det har en reel betydning for størstedelen af de SØM-relaterede udfaldsmål: En klassestørrelsesindsats må forventes først at skulle påvirke klassemiljøet, trivsel og/eller læring, førend der forventes effekter på de budgetøkonomiske udfaldsmål.

⁶ Idet REFUD har en tidshorisont på fire kalenderår (/budgetår), og vi antager, at den målte klassestørrelse gælder skoleåret ud, betyder den halve års forskydning mellem klassestørrelsesindsats og budgetøkonomiske konsekvenssmål i fire efterfølgende kalenderår, at klassestørrelsesindsatsen reelt vil implementeres (og omkostningerne dermed afholdes) et halvt år før modelhorisonten, medmindre effekttestimatet på første år skaleres til at omfatte en halvårlig klassestørrelsesindsats.

2.3 Persistens

Det er værd at notere sig, at efterfølgende effektstørrelser ikke alene kan tilskrives klassestørrelsen i det enkelte år, idet der for mange klasser er relativt lidt variation i klassestørrelse over en kortere årrække. Denne træghed betyder, at efterfølgende klassestørrelser i nogen grad vil opsamles i den 1-årige klassestørrelsesvariabel, og de deraf estimerede klassestørrelses effekter, men også at det ikke er muligt at inkludere flere på hinanden følgende klassestørrelsesvariable i analysen, idet samvariationen mellem disse er for stor til at resultere i meningsfyldte punkttestimater.

Persistensfortolkningen sløres yderligere af, at de budgetøkonomiske konsekvenser måles per kalenderår, hvoraf skoleåret kun dækker halvdelen (se **Fejl! Henvisningskilde ikke fundet.** Effekttestimaterne kan således fortolkes som effekten af en marginal ændring i klassestørrelse på de budgetøkonomiske konsekvenser i det førstkommende (og efterfølgende) kalenderår, om end det er sandsynligt, at disse til en vis grad også påvirkes af klassestørrelsen i det efterfølgende skoleår, der sammenfalder med kalenderåret for konsekvensmåling. Dette uddybes i Del 2 Dokumentation.

2.4 Robusthed

Vi foretager forskellige robusthedstest for at udforske gyldigheden af rapportens identifikationsstrategi og deraf også analysens resultater. Set i lyset af de mange udfald, der behandles i analysen, valgte vi på forhånd effekttestimaterne på Specialundervisning i almenklasse samt elevtrivselsfaktoren Ro og orden i klasselokalet som grundlag for robusthedsanalyserne.

Specialundervisning i almenklasse blev valgt, idet denne forventedes at være relativt påvirket af evt. lærings- og trivselsniveau på baggrund af klassestørrelse og desuden ofte forekommende i de danske folkeskoleklasser, hvorved effekttestimatet kunne forventes at opnå en vis præcision. Imidlertid modtager kun omkring 0,5 % af normalklasseeleverne på hver årgang i den danske folkeskole i årene 2015-2019 specialundervisning, hvilket til dels skyldes, at specialundervisning kun registreres for undervisningsbehov på ni eller flere timer ugentligt, hvilket hæmmer præcisionen af effekttestimaterne.

Risikoen for afbrydelser i undervisningen stiger i stadig større grad jo flere elever, der er i klasselokalet, jf. blandt andet Lazears forstyrrelsesmodel (2001). Med dette in mente udvalgte vi derfor elevtrivselsfaktoren Ro og orden

fra den nationale trivselsmåling i folkeskolen, der omhandler elevernes opfattelse af ro og støj i klasse samt klasseledelse (se Krassel m.fl., 2022b; STIL, 2020). Trivselsfaktoren måles på 4.-9. klassetrin.

Figur 2.2 og Figur 2.3 viser estimerede sammenhænge mellem klassestørrelse og henholdsvis specialundervisning i almenklasse og trivselsfaktoren Ro og orden for hvert klassetrin under forskellige modelspecifikationer og valgt sample størrelse omkring klasseloftstærskler (kaldet bandwidth, BW, se de blå cirkler i figuren i Boks 2.1).

Vores foretrukne modelspecifikation er 2SLS-model med et samplevalg på +/- 5 elever omkring de tre nederste klasseloftstærskler (BW5) og udvalgte elevkontroller.⁷ Vi tester derfor alternativer til dette, hvor et eller flere valg justeres:

- OLS med fuld vektor af elevkontrolvariable på BW14-samplen⁸
- OLS med fuld vektor af elevkontrolvariable på BW5-samplen
- 2SLS uden elevkontrolvariable på BW5-samplen
- 2SLS med udvalgte kontrolvariable på BW5-samplen (den foretrukne specifikation)
- 2SLS med den fuld vektor af elevkontrolvariable fra REFUD-estimationerne⁹ samt skolekontrolvariable: skolestørrelse, kvadreret skolestørrelse, kommunetype for skolens administrationskommune og årgangsstørrelse for tilsvarende årgang i skoleåret før.

Af Figur 2.3 for trivselsfaktoren Ro og orden fremgår det for især mellemtrinene tydeligt, at OLS-estimatet er biased mod nul i forhold til 2SLS-estimatene, hvilket antyder positiv selektion ind i større klasser (elever i både større og mindre klasser rapporterer generelt samme niveau af uro og støj i klasseværelset). Ydermere ser forskellen mellem 2SLS- og OLS-estimatene ikke ud til at være drevet af sampleudvælgelsen til BW5-samplen, ligesom gradvis inklusion af kontroller på betryggende vis stort set ikke ændrer på de estimerede 2SLS-klassestørrelses effekter.¹⁰ I fald den gradvise inklusion af kontrolvariable drastisk ændrer på effektestimaterne, antyder dette, at der er selektion ind i årgangsstørrelsen (fx ved manipulation af denne, se afsnit 4.3), og at

⁷ Imidlertid forventes instrumentet kun at skabe virkelig eksogen variation i et mindre interval omkring klasseloftstærsklerne, hvor aktører ikke præcist kan forudsige og manipulere årgangsstørrelsen til at udløse de ønskede klassestørrelser. Baseret på bandwidth-analysen i Kapitel 4 vælges vinduet på +/- fem elever omkring de tre nederste klasseloftstærskler som foretrukken estimationssample

⁸ Den såkaldte fulde sample med alle elever på skoler omkring de tre nederste klasseloftstærskler, hvor vi har frasortet de helt små skoler med 14 eller færre elever på årgangen. BW14-samplen består således af skoler med årgangsstørrelser på 15-98 elever.

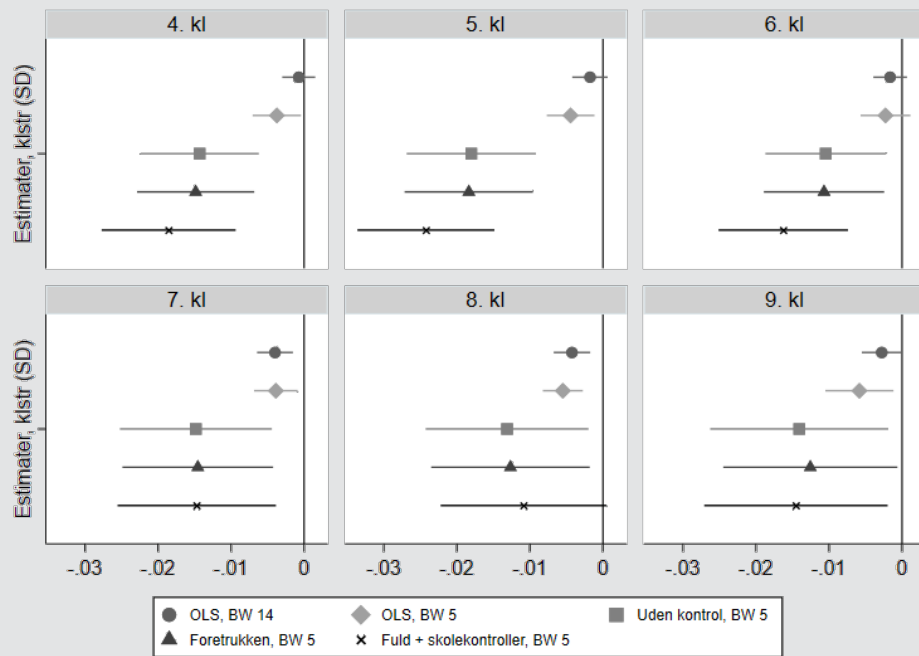
⁹ Dog uden lags af baseline variable, se Krassel m.fl. (2022a,b).

¹⁰ Imidlertid er det overraskende, at de inkluderede kovariate forklarer så relativt lidt af variationen i udfaldsmål, at præcisionen af de estimerede effekter ikke øges. Inklusionen af skolekontroller øger endda standardfejlene en smule, men bidrager generelt til forklaringsgraden af modellen.

analysestrategien dermed ikke er gyldig. Hvis der ikke er balance i inklusionen af de observerbare kontrolvariable, har vi ingen grund til at tænke, at der er på de uobserverbare, og denne selektion vil vi ikke kunne tage højde for.

Figur 2.2 Effekter af klassestørrelse på trivselsfaktor Ro og orden, år 1, OLS og 2SLS-specifikationer

Robusthedsanalyse i forhold til valg af modelspecifikation og kontrolvariable.

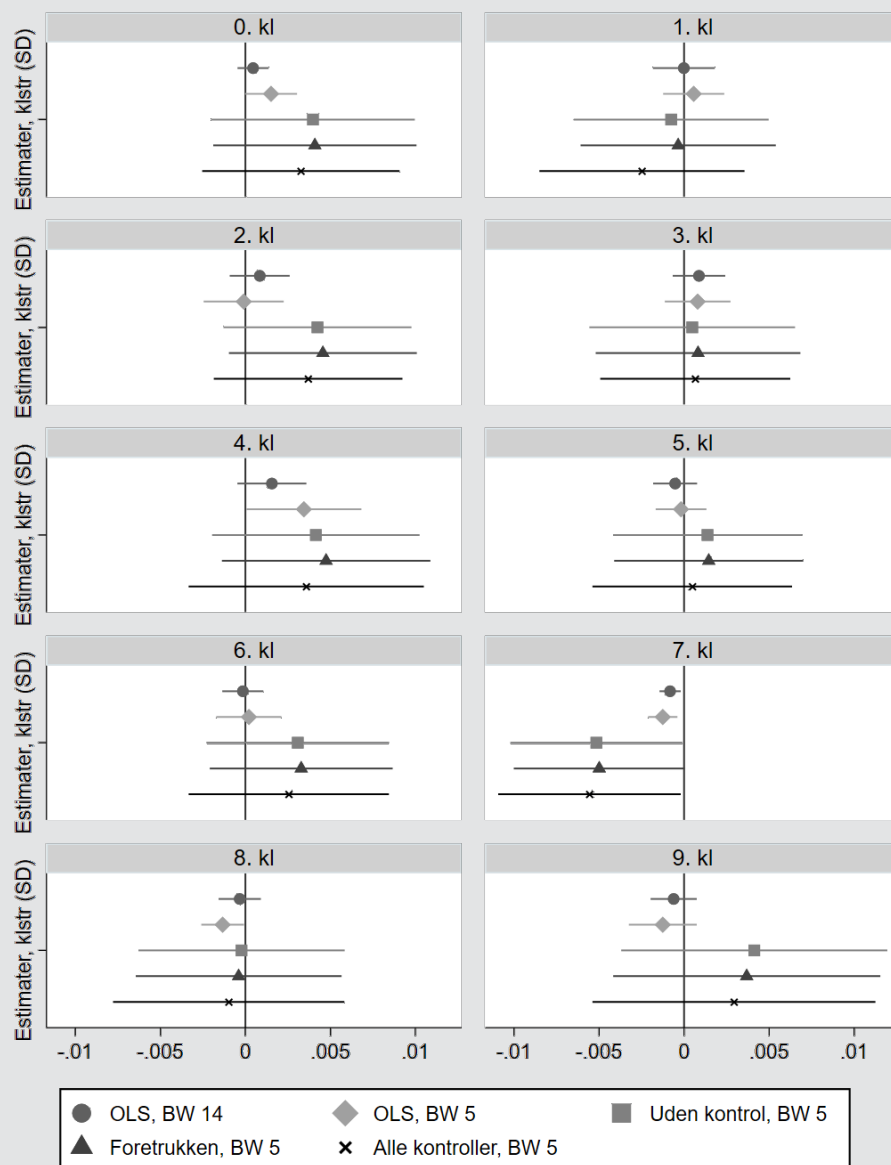


Anm.: Konfidensintervallerne angiver signifikans på 5 %-niveau. Standardfejl er klyngekorrigeret på skoleniveau.

Estimaterne er baseret på populationen af folkeskoleelever i fuldt årgangsopdelte normalklasser i skoleårge med årgangsstørrelser på +/- BW-elever omkring de tre nederste klasseloftstærskler i skoleårene 2014/15-2018/19. Trivselsmålet er standardiseret. Foruden de i teksten oplyste regressorer kontrollerer alle regressioner for årgangsstørrelsessegment og år for måling af klassestørrelse. Den foretrukne 2SLS-regression kontrollerer desuden for alder og kvadreret alder, køn, ikkevestlig baggrund, antallet af ældre søskende, moderens og faderens uddannelse, alder ved fødsel og kvadreret alder ved fødsel samt indikatorer for, om moderen og faderen er diagnosticeret med angst og depression hhv. opmærksomhedsforstyrrelser.

Figur 2.3 Effekter af klassestørrelse på forbrug af specialundervisning i almenklasse, år 1, OLS og 2SLS-specifikationer

Robusthedsanalyse i forhold til valg af modelspecifikation og kontrolvariable.



Anm.: Konfidensintervallerne angiver signifikans på 5 %-niveau. Standardfejler er klyngekorrigeret på skoleniveau.

Estimerne er baseret på populationen af folkeskoleelever i fuldt årgangspdelte normalklasser i skoleårgange med årgangsstørrelser på +/- BW-elever omkring de tre nederste klasseløftstærskler i skoleårene 2014/15-2018/19. Forbrug af specialundervisning er standardiseret. Foruden de i teksten oplyste regressorer kontrollerer alle regressioner for årgangsstørrelsessegment og år for måling af klassestørrelse. Den foretrukne specifikation kontrollerer desuden for alder og kvadreret alder, køn, ikkevestlig baggrund, antallet af ældre søskende, moderens og faderens uddannelse, alder ved fødsel og kvadreret alder ved fødsel samt indikatorer for, om moderen og faderen er diagnosticeret med angst og depression hhv. opmærksomhedsforstyrrelser.

Figur 2.3 afslører et mere uklart mønster for udfaldet Specialundervisning i almenklasse. Særligt er der i langt højere grad overlap mellem konfidensintervallerne for OLS- og 2SLS-estimer, der i øvrigt er relativt præcise estimerede nul-estimer, og hvor 2SLS-estimerne igen kun påvirkes af den gradvise inklusion af kontrolvariable i begrænset omfang. Undtaget herfra er de marginalt signifikante negative klassestørrelses effekter i 7. klasse, der angiver, at en stigning i klassestørrelse forårsager mindre forbrug af specialundervisning. Idet de omkringliggende estimer ikke udviser et lignende mønster, mener vi ikke, at dette estimat er overbevisende (se også diskussionen af samtidige hypotesetest i afsnit 4.5).¹¹

For øvrige robusthedsanalyser og validering af instrumentet henvises læseren til rapportens Del 2 Dokumentation.

2.5 Effekternes størrelsesorden i forhold til andre studier

Dette afsnit opsummerer den eksisterende litteratur om effekter af klassestørrelse for at kvalificere rapportens resultater.

Indsigter fra tidligere studier – faglige resultater

Vi beskriver her klassestørrelses effekter på elevers faglige resultater, da størrelsesordenen af disse effekter kan være en pejling for den forventede størrelsesorden af effekter på andre typer af udfaldsmål direkte påvirket af klassestørrelsen (jf. trivsel) eller afledte effekter (jf. budgetøkonomiske konsekvenser).

Effekter af klassestørrelse på elevers faglige resultater i grundskolen er bl.a. undersøgt i Norge (Iversen & Bonesrønning 2011; Leuven m.fl., 2008; Leuven & Løkken 2020), Holland (Dobbelsteen m.fl., 2002), Sverige (Fredriksson m.fl. 2013, 2016), USA (Hoxby, 2000; Connolly & Haecck, 2022), Israel (Angrist & Lavy, 1999; Angrist m.fl., 2019) og Danmark (Heinesen, 2010; Nandrup, 2016). Alle med samme metode, som anvendt i denne rapport. Opgøres disse estimer på samme skala, varierer effekterne fra små og insignifikante op til omkring 0,04 SD pr. elevreduktion.¹²

¹¹ Om end denne også kan ses i sammenhæng med de positive effekter af større klassestørrelser på forbrug af specialskole og fri- og privatskoler, og således i højere grad være et udtryk for den udvdring fra folkeskoleklassen, som er observeret på tidligere klassetrin.

¹² Se litteraturgennemgang fra Beuchert (under udgivelse), der opgør effekter fra 16 kvasi-eksperimentelle studier. Inkluderer også søskende- og tvillingestudier (Bingley m.fl., 2005; Gerritsen m.fl., 2017).

Det danske studie fra Nandrup (2016) estimerer signifikante positive effekter af klassestørrelsen på elevernes faglige resultater på 2.-6. klassetrin i størrelsesorden omkring 0,01-0,02 SD pr. elevreduktion. Effekterne måles efter ét år med en given klassestørrelse. Studiet undersøger også, om der kan være en akkumuleret effekt af mindre klasser på mellemtrinnet, men denne er ikke betydeligt større end effekten målt over ét skoleår (samme konklusion finder Leuven & Løkken, 2020).¹³ Studiet bidrager således også til test af persistensen i klassestørrelseseffekter (jf. afsnit 2.3).

En nyere metaanalyse fra VIVE Campbell, baseret på 10 studier¹⁴, konkluderer ingen eller meget små effekter af klassestørrelse (Filges m.fl., 2018). Antager vi for illustrationens skyld, at effekten af klassestørrelsen er lineær og omregner effekten målt i metaanalysen til effekten pr. elev, som i de kvasi-eksperimentelle studier, fås en effekt på omkring 0,02 SD i læsning.¹⁵ Resultaterne fra hhv. kvasi-eksperimentelle studier og lodtrækningsstudier giver således omtrent samme billede af størrelsesordenen på effekten af at reducere klassestørrelsen.

På forhånd forventedes effekter på trivsel (omregnet til effektstørrelser) at være i samme størrelsesorden som effektstørrelser på faglige resultater eller en smule større, hvis det antages, at elevtrivsel påvirkes mere umiddelbart end faglige resultater. Set i lyset af de øvrige marginale klassestørrelseseffekter i litteraturen er effektestimaterne på elevtrivsel således relativt sammenlignelige.

Indsigter fra tidligere studier – effekter på fx uddannelse og lønindkomst

Studier af STAR (Student Teacher Achievement Ratio) projektet viser langsigtede effekter på elevernes ikke-kognitive evner og uddannelses-/arbejdsmarkeds-udfaldsmål, fx større sandsynlighed for at påbegynde en universitetsuddannelse (Chetty m.fl., 2011; Dynarski m.fl., 2013). Effekten er dog ikke kun er drevet af klassens størrelse, men også af dens elevsammensætning og lærerens erfaring (Chetty m.fl., 2011).

Nordiske forskere argumenterer også for, at der kan være langsigtede økonomiske gevinster af klassereduktioner. Fredriksson m.fl. (2013) finder, at reduceret klassestørrelse i 1.-6. klasse på svenske skoler har en positiv effekt på elevernes senere løn og indkomst. Modsat finder et norsk studie af Leuven & Løkken (2020) ingen positive indkomsteffekter af en reduceret klassestørrelse. Denne rapport finder heller ingen effekter på elevernes lønindkomst 4 år efter en ændring i klassestørrelse (Bilagstabel 2.1), hvilket dog også er en kort tidshorisont og primært vil udgøres af studie/sabbatårs-job sammenlignet

¹³ Den gennemsnitlige effekt for elever på 6. klassetrin, der har gået i en lidt mindre klasse i 3 år end andre elever i 6. klasse, er lille og kun signifikant på 10 %-niveau.

¹⁴ Fire studier baseret på STAR og seks andre, heraf ét lodtrækningsforsøg. Disse studier måler effekten af at gå i hhv. 'små klasser' eller 'store klasser', hvor den gns. forskel er 5,1 elev (Filges m.fl., 2018).

¹⁵ Egen beregning baseret på Filges m.fl. (2018): Gns. effekt 0,11 SD/5,1 gns. elevreduktion = 0,02 SD effekt pr. elev.

med de to øvrige skandinaviske studier, hvor lønindkomst i højere grad måles efter endt uddannelse.

Et antal danske studier har vist, at en reduceret klassestørrelse i 8.-9. klasse har positiv effekt på, om eleverne færdiggør en ungdomsuddannelse (Browning & Heinesen, 2007) og øger længden af videre uddannelse efter folkeskolen (Bingley m.fl., 2005).¹⁶ I tråd med dette viser denne rapport en tendens til, at en øget klassestørrelse i 9. klasse mindsker og forskyder elevernes efterfølgende forbrug af ordinær ungdomsuddannelse fra de to følgende år til år fire (Bilagstabel 2.1).

Indsigter fra tidligere studier – trivsel og adfærd

Beuchert & Nandrup (under udgivelse)'s foreløbige resultater peger på små, negative effekter af klassestørrelsen i indskolingen på trivsel og personlige kompetencer, afhængigt af hvilke dimensioner der undersøges. Frederiksson m.fl. (2016) har adgang til elevsurveys på 4.- 6. klassetrin og finder, at elever i mindre grad oplever, at det er let at forstå, når læreren forklarer foran hele klassen, når klassestørrelsen øges (effektstørrelse på 0,0055 SD) – hvilket er i overensstemmelse med denne rapport's resultater på Ro og orden, der dog i højere grad har fokus på støj og uro. Frederiksson m.fl. (2016) viser samtidig, baseret på data fra et lærersurvey, at lærere i større klasser tildeler eleverne mere ansvar for egen læring, hvilket de forklarer med mindre tid til den enkelte elev.

Der findes desuden studier om effekten af en reduceret klassestørrelse på elevernes ikke-kognitive færdigheder, der, selvom de ikke er direkte overlappende, er relateret til elevtrivsel og -adfærd. Et nyt amerikansk studie finder fx positive effekter af klassereduktioner på udsatte elevers sociale kompetencer, emotionelle modenhed, kommunikationsevner og et generelt indeks for udvikling (Connolly & Haeck, 2022). Effekterne er i størrelsesordenen 0,007-0,013 SD, til sammenligning finder de effekter på læsning og matematik på 0,016-0,017 SD. Denne rapport finder effekter på trivselsindekset Ro og orden på 0,011-0,018 SD, hvilket er i samme størrelsesorden, om end to meget forskellige trivselsmål.

Man kan også forestille sig andre afledte effekter af mindre klasser, fx afledte effekter gennem forældres kompensering og skoleskift. Rapportens resultater, der indikerer, at øget klassestørrelse får forældrene til at flytte deres børn til en fri-/privatskole, er i overensstemmelse med resultater for skoleskift i det svenske studie af Fredriksson m.fl. (2016).

¹⁶ Bingley m.fl. (2005) opstiller også en cost-benefit-analyse, hvor de dog ikke kan vise, at gevinsterne ved mindre klasser (i 8. klasse) kan opveje omkostningerne.

Endvidere kan der være afledte effekter på eller gennem lærerne, fx hvis sænkning af klassestørrelsen påvirker udbuddet af erfarne lærere (Jepsen & Rivkin, 2009) eller mere erfarne lærere, der søger mod mindre klasser (Muel-ler, 2013). Denne rapportes resultater indikerer, at større klasser mindsker sy-gefraværet blandt lærerne (Bilagstabel 2.1). I den danske kontekst kan det for-klares, hvis skolen tildeler større klasser en mere erfaren lærer eller andre ekstra ressourcer, fx tolærerordninger, der mindsker arbejdspresset i den større klasse. I så fald bør de øvrige resultater fortolkes i lyset heraf.

2.6 Implementering i REFUD

Analyserne i denne rapport er os bekendt det mest omfattende forsøg på at danne et overblik over de direkte potentielle effekter af klassestørrelse på for-skellige elevudfald. I den eksisterende REFUD-model baseres forventningen til klassestørrelseseffekter på SØM-konsekvenser på indirekte evidens (hvor ef-fekten af at sænke klassestørrelsen evalueres på faglige testscores, der korre-leres med empiriske sammenhænge mellem testscores og SØM-konsekvenser). Analyserne i denne rapport er således en øvelse i at forbedre disse forventnin-ger ved at tilbyde supplerende troværdige estimater på effektstørrelserne.

Imidlertid efterlader omfanget af analysen dog resultaterne med især to svag-heder i forhold til at implementere disse i REFUD: 1) de mange udfald og klas-setrin i analysen betyder, at sandsynligheden for at detektere nul-effekter som signifikante, er anselig¹⁷, og 2) analysedesignet skal kunne anvendes på alle klassetrin, hvilket potentielt svækker gyldigheden af identifikationen i nogle tilfælde.

Udfordringen forbundet med 1) er af særlig betydning: hvis man fejlagtigt for-kaster en sand nulhypotese (nul-effekt), og estimatet inkluderes i REFUD-mo-dellens beregninger, vil forventningen til de budgetøkonomiske besparelser som følge af en klassestørrelsesreduktion være forkert i forhold til, hvad der er statistisk belæg for. Imidlertid er ingen af de mulige løsninger til problema-tikken med multiple hypoteser særligt attraktive på grund af analysens om-fang. Således vil signifikanskorrektioner, der vurderes som tilstrækkelige på tværs af samtlige klassetrin og tidshorizont, uundgåeligt medføre betragtelige styrketab og dermed i praksis umuliggøre påvisningen af (sande) klassestør-relseseffekter (uddybes i metodebilag, afsnit 4.5). Idet vi vurderer, at udvalgte resultater på trods af det sparsomme datagrundlag peger på interessante ten-denser af effekterne af klassestørrelse på SØM-konsekvenserne, der er til-strækkeligt overbevisende i størrelse og konsistens (om end ikke præcision),

¹⁷ Ved brug af et 5 %-signifikansniveau er sandsynligheden for at forkaste mindst én sand nulhypotese (dvs. at effekten af klassestørrelse er nul) $1 - 0.95^m$, hvor m er antallet af samtidige hypotesetest, der foretages, når estimaterne betragtes enkeltvis.

omtales disse i rapporten netop som tendenser og ikke klassestørrelseseffekter. Af samme årsag anbefaler vi ikke, at punktestimaterne for de budgetøkonomiske konsekvenser umiddelbart implementeres i REFUD. Samtidig vurderer vi heller ikke, at eventuelle signifikanskorrigerede punktestimater vil medføre retvisende modelberegninger, idet det sparsomme datagrundlag og de dertilhørende i forvejen relativt upræcist estimerede resultater i så fald vil undervurdere de potentielle effekter af en klassestørrelsesindsats.

I forhold til 2) om, at analysedesignet skal være anvendeligt på alle klassetrin, er dette en forudsætning for anvendeligheden for estimerne i REFUD-regi. Således forudsætter identifikationsstrategien, at samfundsaktørerne ikke manipulerer årgangsstørrelse omkring tærskelværdierne i analysesamplen, som fx hvis forældre flytter deres børn ud af skoler med visse størrelser (se også en detaljeret gennemgang i afsnit 4.1 og 4.2). Gyldigheden af denne antagelse kan være mindre overbevisende, jo højere klassetrin der betragtes, hvilket bør afstedkomme søgning af alternative og bedre instrumenter (se afsnit 4.1). Der er dog ikke fundet andre og mere overbevisende instrumenter, som samtidig kan anvendes for alle klassetrin.

Baseret herpå er det VIVEs vurdering, at det samtidige årgangsstørrelsesinstrument er et fornuftigt, om end ikke ideelt, grundlag for analysen. Således leveres analyseresultaterne fra denne rapport efter følgende kriterier:

Boks 2.5 Kriterier for om et estimat medtages i leverancen til REFUD

Der pålægges følgende objektive kriterier for, om et estimat medtages i leverancen til BUVM's videre arbejde med klassestørrelseseffekter i REFUD:

- Udfaldsmål er relevant for årgangen på baggrund af elevens klassetrin og alder, jf. modellens allerede fastlagte kriterie (se afsnit 5.2, der redegør for antallet af årgange, hvert udfaldsmål kan måles for).
- Hvis der er under 100 elever med et givent udfaldsmål, udelades estimatet (se afsnit 2.2, der eksemplificerer dette).

Yderligere kriterier for, hvilke estimer der medtages i REFUD, overlades til drøftelse i udviklingsarbejdet omkring modellen. Med leverancen medfølger 1) Excel-ark indeholdende samtlige estimer, P-værdier og F-test, og 2) Excel-ark med beskrivende statistik om hvert udfaldsmål til at kvalificere vurderingen af, hvilke estimer der kan medtages i REFUD.

3 Konklusion

Effekter på trivsel

Resultaterne for trivsel viser signifikante effekter af klassestørrelse på én ud af fem indikatorer for elevtrivsel: Ro og orden. Resultaterne for Ro og orden viser, at en tilføjelse af en ekstra elev i klassen sænker elevernes opfattelse af ro i klassen og god klasseledelse, og at denne effekt ses på alle klassetrin fra 4.-9. klasse. Effekternes størrelse stemmer overens med effekter fundet på andre udfaldsmål målt i klassen, fx faglige resultater og personlige kompetencer (Nandrup, 2016; Connolly & Hack 2022) og oplevelse af at kunne høre læreren (Frederiksson m.fl., 2016).

Det bemærkes endvidere, at effekterne på Ro og orden er bemærkelsesværdigt overbevisende og robuste over for alternative modelspecifikationer og korrektion for multiple hypotese testning. På denne baggrund vurderes det, at trivselsindikatoren Ro og orden med fordel kan tilføjes REFUD.

Effekter på REFUD-konsekvenser

Resultaterne for REFUD-konsekvenserne viser sporadiske signifikante estimer af en ændring i klassestørrelse på de fleste default SØM-konsekvenser. Af flere grunde er disse estimer dog ikke overbevisende (fx kun signifikant for et enkelt klassetrin eller upræcist estimeret), og vi er derfor varsomme med at benævne disse som 'effekter'. Endvidere udviser analysen en relativt stor risiko for både type-1-fejl og type-2-fejl, som det ikke har været muligt at korrigere for inden for analysens rammer.

Ud fra en helhedsbetragtning er der således begrænsede signifikante effekter på REFUD-konsekvenserne, men af de mere interessante tendenser kan nævnes, at en marginal øgning af klassestørrelsen medfører mindre forbrug af folkeskolen, mere forbrug af fri- og private grundskoler og specialskoler og flere kontakter med skadestue. Der er ikke fundet signifikante effekter på forbrug af specialundervisning i almenklassen, hvilket ellers var forventet.

Det understreges, at fraværet af konsistente og signifikante effekter af klassestørrelse i denne rapport *ikke* nødvendigvis er ensbetydende med, at disse ikke eksisterer, men i stedet kan skyldes, at datagrundlaget ikke tillader tilstrækkelig præcision i estimerne til at være overbevisende.

Anbefalinger i forhold til overførbare af metoden til samtlige klassetrin og udfaldsmål

Det var nødvendigt at opstille en relativt generel metode og model for at kunne foretage analysen for samtlige klassetrin og udfaldsmål i REFUD. Der anvendes derfor et kendt analysedesign, tilpasset de tilgængelige data i REFUD. Som denne rapport's analyse af validitet og robusthed illustrerer, er det dog vanskeligt at foretage metodevalg, der er lige gyldige på tværs af alle klassetrin, samples og udfaldsmål. Der er her en afvejning mellem på den ene side bias og statistisk præcision og på den anden side generaliserbarhed og anvendelighed i REFUD.

Generelt vurderes, at det er muligt at overføre metoden til danske data, og at analysen er på niveau med sammenlignelige danske og internationale klassestørrelsesstudier. Denne rapport er således den bedst tilgængelige evidens om klassestørrelseseffekter, der eksisterer for danske data efter Folkeskole-reformen, jf. opdraget om klassestørrelseseffekter på tværs af samtlige klassetrin og udfaldsmål i REFUD. En styrkelse af evidensniveauet vil kræve en afgrænsning på analyserede udfaldsmål og klassetrin, hvormed yderligere analyser måske vil kunne forfine instrumenter og modelspecifikationen for de respektive udfaldsmål.

Rapporten påpeger, at analysen har mindre gyldighed jo højere klassetrin, der analyseres. Derudover er der særlige forhold vedrørende klassesammenhængen i 0. og 9. klasse, der gør, at vi vurderer, at instrumentet er mindre gyldigt for disse klassetrin.

For fremtidige analyser vil det være relevant at undersøge, om der kan findes bedre instrumenter for særligt 0. og 9. klassetrin, og om SØM-konsekvenser kan aggregeres, således at sample size øges.

Anbefalinger i forhold til anvendelighed i REFUD

I udgangspunktet efterlader omfanget af analysen flere svagheder og forbehold i forhold til fortolkningen af resultaterne. Disse forbehold udfordrer den umiddelbare implementering af estimaterne i REFUD, som grundlag for beregning af forventede budgetøkonomiske gevinster af klassestørrelsesindsatser, uden en bredere skelnen til kriterier for teori, datagrundlag, kausalitet, bias, præcision og statistisk signifikans.

Vi har udeladt estimater, der var baseret på for tyndt et datagrundlag (fx hvis der er under 100 personer, der foretager en specifik aktivitet). Derudover har vi ikke opstillet øvrige kriterier for, hvorvidt de enkelte estimater kan indgå direkte i REFUD. De udvalgte resultattabeller i afsnit 2.2. illustrerer forskellige udfordringer ved fortolkning af estimaterne på SØM-konsekvenserne med

henblik på en drøftelse af, hvorvidt estimerne kan implementeres direkte i REFUD.

Rapportens største forbehold er testning af mange samtidige hypoteser. Jo flere udfaldsmål, der undersøges, og jo flere klassetrin og år hvert udfaldsmål undersøges for, jo større er bekymringen. I kombination med i visse tilfælde relativt sparsomt datagrundlag, udsætter dette rapportens analyser for betydelige risici for både type-1- og type-2-fejl. De to fejltyper har forskellige implikationer for REFUD.

På den ene side er der risiko for type-1-fejl; et estimat er statistisk signifikant, skønt der i virkeligheden ikke er en effekt, men estimatet er blot udtryk for en tilfældig sammenhæng i data (falsk positiv). I vores fortolkning af resultaterne har vi derfor været varsomme, hvis et udfaldsmål fx kun viste en signifikant effekt på ét klassetrin og/eller i et enkelt år. Hvis alle signifikante effekter accepteres som gyldige og anvendes direkte i REFUD, vil det kunne betyde, at man investerer i klassestørrelsesreduktioner, der ikke giver det ønskede afkast (for både individet og kommunen).

På den anden side er der risiko for type-2-fejl; at effekter, der findes i virkeligheden, ikke er signifikante og derfor fortolkes som nul-effekter. Denne fejltypen kan både opstå, fordi en stor del af REFUD-konsekvenserne tæller relativt sjældne begivenheder, og der derfor stilles større datamæssige krav i analysen for at kunne dokumentere statistisk signifikante effekter – særligt i 2SLS-regi, hvor standardfejlene generelt er forstørrede. Den kan også opstå pga. anden støj i data, der ikke i tilstrækkelig grad kan tages højde for med den valgte modelspecifikation. I vores fortolkning af resultaterne har vi derfor været opmærksomme på den statistiske sikkerhed på estimerne (standardfejl) for at vurdere, om estimerne var af en størrelsesorden, der tyder på en meningsfuld effekt, skønt datagrundlaget måske ikke er stærkt nok til at detektere statistisk signifikans på det konventionelle 10 %- eller 5 %-niveau. Givet de tilgængelige data skal effekter have en størrelsesorden over 0,01 SD, før vi kan detektere signifikans på et 5 %-niveau.

Implikationen af risikoen for type-2-fejl for REFUD er, at de meget små effekter, der findes på nogle udfaldsmål ikke inkluderes i SØM, og alle potentielle budgetøkonomiske effekter af klassestørrelsesændringer dermed ikke medregnes. Dette har ledt til vores henstilling til at betragte rapportens resultater som en helhed og evt. anvende estimer i REFUD, skønt de ikke er signifikante, hvis fx de tilsvarende estimer på tilstødende klassetrin er signifikante.

I forhold til beregninger i REFUD kan begge fejltyper medføre både en over- og underestimering af de samlede budgetøkonomiske gevinster ved en ændring i klassestørrelsen. Det er derfor VIVEs anbefaling at betragte rapportens resultater som den foreløbige bedste evidens om klassestørrelses effekter på

REFUD-modellens udfaldsmål, indtil fremtidige og mere forfinede analyser foreligger, der kan tage højde for flere af rapportens beskrevne forbehold. I stedet kan rapportens resultater anvendes til at underbygge og kvalitetssikre modellens eksisterende estimater for klassestørrelses effekter, samt danne grundlag for videreudvikling af estimationsmetoden i forhold til implementering af resultaterne i modellen.



DEL 2

Dokumentation

4 Metode

4.1 Identifikationsstrategi og øvrige metodevalg

Sammenhængen mellem antallet af elever i en klasse og elevernes udfaldsmål forventes at være endogen, da der er en række forhold og faktorer omkring både skolen, eleven og dennes baggrund, som kan påvirke både antallet af elever i klassen og deres udfaldsmål. Sammenlignes fx skolers gennemsnitlige klassestørrelser og karaktergennemsnit, observeres det typisk, at skoler med store klasser også opnår højere gennemsnitlige karakterer, men det er et forhold, der ikke nødvendigvis kan tilskrives, at skolen har store klasser. Det kan fx være drevet af, at gode skoler får et godt ry og derfor tiltrækker flere elever, hvilket medfører større klasser. Samtidig kan populære skoler også tiltrække flere og dygtigere lærere og har derfor tilstrækkelige ressourcer til at oprette flere og større klasser. Det kan også være udtryk for, at der er flere store skoler i byområder, hvor der samtidig også er både et socioøkonomisk stærkere elevgrundlag og rekrutteringsgrundlag.

Man kan også forestille sig andre afledte effekter af klassestørrelse, fx afledte effekter gennem forældres øget lektielæsning, og om forældre i højere grad vælger at flytte deres børn, hvis de går i store klasser end i små klasser (Fredriksson m.fl., 2016; Johansen, Mølbæk & Nandrup 2020). Endvidere kan der være mulige afledte effekter gennem lærerne. Mueller (2013) viser fx i en opfølgning på STAR, at den positive effekt af små klasser er drevet af, at små klasser undervises af mere erfarne lærere, mens Jepsen & Rivkin (2009) finder, at den positive effekt i deres afprøvning af mindre klasser blev modereret af et lavere arbejdsudbud af lærere, som betød, at der måtte ansættes flere uerfarne eller uuddannede lærere.

For at identificere effekten af selve klassens størrelse er det derfor nødvendigt at anvende en eksperimentel eller kvasi-eksperimentel metode, der kan tage højde for (noget af) den endogenitet, der opstår i størrelsen på en klasse. Således anvendes i analysen en anerkendt kvasi-eksperimentel metode til at estimere kausale effekter af klassestørrelse på de i opgavebeskrivelsen ud-specificerede SØM-relaterede udfaldsmål. Specifikt udnyttes klassestørrelsesloftet på 28 elever i danske folkeskoleklasser til at skabe eksogen variation i klassestørrelse, hvorved det kausale forhold mellem klassestørrelse og udfaldsmål kan estimeres ved hjælp af instrumental variable (IV)-metoden.

Angrist & Lavy (1999) og Hoxby (2000) var de første til at udnytte politisk bestemte lofter over antallet af elever i en klasse til at estimere den kausale effekt af klassestørrelse ved at instrumentere elevernes faktiske klassestørrelse med den forventede, baseret på antallet af elever på årgangen ved ligelig fordeling i de regelsatte antal af klasser.

Fredriksson m.fl. (2013, 2016) forfinede denne strategi ved at benytte tærsklerne i årgangsstørrelse, der opstår som følge af klassestørrelsesloftet, i et såkaldt *fuzzy regression discontinuity design*. Således benyttes kun information om, hvorvidt årgangsstørrelsen er over eller under hver tærskel, der udløser oprettelsen af en yderligere klasse, som grundlag for instrumentet og ikke den mere 'glatte' variation, der er i den forventede klassestørrelsesvariabel uden for tærsklerne.

Se Boks 2.1 for grafisk illustration af de to tilgange.

I praksis er forskellene for de estimerede effekter fra de to instrument-specifikationer dog relativt små (se fx Frederiksson m.fl., 2013, Nandrup, 2016). En forudsætning for, at instrumentet er gyldigt, er dog, at der i tilstrækkelig grad kontrolleres for RD-variablen (her: årgangsstørrelse), idet årgangsstørrelse – eller skolestørrelse – i sig selv forventes at være korreleret med elevudfaldsmål, ligesom effekttestimatet kun er fuldstændig unbiased lige omkring tærsklerne. Dette uddybes og testes i de følgende afsnit om analyse af validitet og robusthed.

Vi estimerer således følgende model ved hjælp af *two-stage least squares model* (2SLS) for hvert klassetrin i 0.-9. klasse:

Second stage:
$$y_{ics\tau} = \mathbf{X}_{ics}\boldsymbol{\alpha}_2 + \alpha_1 CS_{cs\tau} + \varphi_\tau + \varepsilon_{ics\tau}$$

Hvor klassestørrelsen ($CS_{cs\tau}$) instrumenteres af en vektor af dummy-variable, der antager værdien 1, når årgangsstørrelsen (e_s) overstiger tærsklen for oprettelse af en yderligere klasse: $above_{s\tau} = \mathbf{1}[\tau < e_s \leq \tau + 14]$, for $\tau = 28, 56, \dots$:

First stage:
$$CS_{cs\tau} = \mathbf{X}_{ics}\boldsymbol{\beta}_2 + above_{s\tau}\boldsymbol{\beta}_1 + \mu_\tau + v_{ics\tau}$$

Her angiver $y_{ics\tau}$ udfaldet for elev i i klasse c på skole s . τ angiver det segment, skolens årgang ligger på, dvs. omkring hvilken tærskel årgangsstørrelsen er. \mathbf{X} er en vektor af kontrolvariable for eleven og skolen, mens φ og μ er segment fixed-effekter. ε og v er individspecifikke fejlede. α_1 angiver den kausale effekt af en marginal ændring i klassestørrelse på udfaldet y . Idet instrumentet kun binder for årgangsstørrelser lige omkring klasseloftstærsklerne, forventes

estimationssamplen at bestå af årgange på skoler, der har en størrelse i umiddelbar nærhed heraf.^{18,19}

Instrumentets gyldighed, forbehold og alternative instrumenter

Identifikationsstrategien forudsætter, at samfundsaktørerne ikke manipulerer årgangsstørrelse omkring tærskelværdierne i analysesamplen, som fx hvis forældre flytter deres børn ud af skoler med visse størrelser.

Angrist m.fl. (2019) angiver, at klassetrinnets årgangsstørrelse kan være udsat for manipulation, og anbefaler, at man i stedet benytter antallet af børn i skolestartsalderen i hvert skoledistrikt som basis for instrumentet, idet dette er mere omkostningsfuldt for forældre og skoleledere at manipulere. Vi noterer os dette forbehold, men benytter trods alt årgangsstørrelsen på det relevante klassetrin af to årsager: 1) valget af, hvornår årgangsstørrelsen til grund for instrument for de højere klassetrin måles i en dansk kontekst, er i praksis et valg mellem gyldigheden (eksogeniteten) og styrken (korrelationen med klassestørrelsen) af instrumentet. Det står fx skolen frit for at sammenlægge klasser, når årgangsstørrelsen falder under en klasseloftstærskel. Og 2) datamæssige begrænsninger som følge af Folkeskolereformen fra 2014 gør det til en nødvendighed at anvende årgangsstørrelsen på klassetrinnet til baggrund for instrumentet (hvis instrumentet skulle defineres pba tidligere klassetrin, måtte vi anvende data fra før Folkeskolereformen). Angrist m.fl. (2019) finder i øvrigt, at deres resultater fra (1999) er robuste over for valget af det forbedrede instrument.

I forlængelse heraf kan man ligeledes argumentere for, at årgangsstørrelsen på indskolingens 0. klassetrin er det mest eksogene grundlag for instrumentet (frem for den samtidige målte årgangsstørrelse). Det er dog ikke valgt som instrument af følgende årsager:

1. Databegrænsninger som følge af Folkeskolereformen fra skoleåret 2014/2015) og modellens fireårige horisont betyder, at vi ikke kan benytte årgangsstørrelsen for 0. klasse som instrument for alle øvrige klassetrin.

¹⁸ Kontroller for den funktionelle form af årgangsstørrelse ($f_{\tau}(e_s)$ i 2. stage og $h_{\tau}(e_s)$ i 1. stage) er derfor udeladt af den foreslåede modelspecifikation, idet de kun burde være nødvendige på den store sample.

¹⁹ Vi klyngekorrigerer standardfejl på skoleniveau. I litteraturen er det en generel anbefaling, at standardfejl bør klyngekorrigeres på den løbende variabel i RD-designs (dvs. årgangsstørrelsen i denne analyse) (se fx Lee & Lemieux, 2010; Cattaneo m.fl., 2018a,b). Med et bandwidth på omkring 3-5 elever er det dog vores vurdering, at det vil betyde for få klynger til at opnå pålidelig inferens – en kritik, der også er fremsat i fx Kolesár & Rothe (2018). Vi klyngekorrigerer i stedet på skoleniveau, hvilket også typisk er anvendt enten som foretrukket eller robusthedsspecifikation i andre klassestørrelsesstudier (fx Leuven & Løkken, 2020; Connolly & Haeck, 2022).

2. Som følge af udsivning fra folkeskolen til andre undervisningstilbud er det desuden åbenlyst, at årgangsstørrelsen på 0. klassetrin har relativt mindre bid og dermed udgør et dårligere instrument, des højere klassetrin, der betragtes i analysen.
3. I det omfang, at skolerne har specifikke årgange, hvor klasserne sammenlægges, som fx ved overgangen fra mellemtrinnet til udskolingen, kunne årgangsstørrelsen fra det første år af disse benyttes som grundlag for instrumentet i indeværende skoleafdeling. Imidlertid er det empirisk ikke alle skoler, der venter med at sammenlægge klasser til særlige klassetrin. Desuden har skolerne økonomisk incitament til at sammenlægge klasser, når det er muligt.

Baseret herpå er det VIVEs vurdering, at det samtidige årgangsstørrelsesinstrument er et fornuftigt, om end ikke ideelt, grundlag for analysen.

Afsnit 4.4 beskriver desuden de forskellige robusthedsanalyser, der foretages til at underbygge gyldigheden af det valgte instrument samt af de deraf estimerede klassestørrelses effekter.

4.2 Analyse af validitet

For at analysedesignet resulterer i effektestimater af kausal karakter, skal identifikationsstrategien overholde en række krav. I særdeleshed skal instrumentet 1) være relevant i first stage-regressionen, dvs. det skal kunne forklare en stor del af variationen i den endogene variabel (klassestørrelse) og 2) være irrelevant i forklaringen af vores udfaldsmål alene (dvs. instrumentet skal være gyldigt og dermed kun have en påvirkning på udfaldet, vi er interesserede i at undersøge, gennem dets påvirkning af den endogene variabel). Af disse krav er det kun muligt at teste 1), mens 2) må antages, om end det i nogen grad er muligt at teste implikationer af overtrædelser af denne.

Test af instrumentets gyldighed er derfor med til at understøtte fortolkningen af effektestimaterne, herunder hvilke elever/klassestørrelser som instrumentet/effekterne er gyldige for (jf. generaliserbarhed). I dette afsnit beskrives testene og udvalgte tabeller/grafer rapporteres for rapportens to udvalgte hovedudfaldsmål (jf. afsnit 2.4).

Det valgte instrument er baseret på klassestørrelsesloftet på 28 elever pr. klasse, og således instrumenteres den faktiske klassestørrelse på klassetrin c , $c = 0, 1, \dots, 9$, med indikatorvariable, der antager værdien 1 for årgangsstørrelser

i et interval på 1-14 elever over klasseloftstærsklerne (15 elever over en klasseloftstærskel tildes i stedet værdien 0 som tilhørende årgangsstørrelser under næstkommende tærskel.

Test angående instrumentets styrke i 1) behandles i afsnit 4.2.4 om end non-compliance, som behandlet i afsnit 4.2.1, også vil påvirke denne), mens den resterende del af dette afsnit adresserer hvorvidt instrumentet ikke overholder 2) og dermed er ugyldigt:

- Instrumentets gyldighed i forhold til, i hvor høj grad skolerne er bundet af og lever op til klasseloftsreglen (jf. compliance)
- Instrumentets gyldighed i forhold til, i hvor høj grad skolerne kan manipulere årgangsstørrelsen og derved omgå klasseloftsreglen (jf. manipulation)
- Instrumentets gyldighed i forhold til, at der dannes to sammenlignelige elevgrupper på hver side af tærskelværdierne (jf. diskontinuitet i baggrundsvARIABLE)
- Instrumentets gyldighed i forhold til eksogenitet og elevbaggrund.

4.2.1 Om skolerne er bundet af klasseloftsreglen (compliance)

Bilagstabel 2.3 illustrerer en meget høj compliance i forhold til klassestørrelsesloftet i de danske folkeskoler. Fraregnet 0. klassestrin går mere end 97 % af eleverne på de forskellige årgange i klasser, der er registreret i uddannelsesregistre som indeholdende 28 elever eller færre.

I rapportens analyser ekskluderer vi ikke klasser, der overtræder klassestørrelsesloftet. For det første, fordi vi ikke ønsker at betinge vores sample på den endogene variabel, vi har interesse i, og for det andet, fordi det er uklart, om klassestørrelsen reelt er større, end loftet tillader (fx pga. dispensation), eller om der er tale om fejlregistreringer. Det er dog værd at notere, at disse klasser udgør en meget lille del af vores datapopulation.

Vores resultater viser derudover en stærk first stage-regression, altså en stærk empirisk sammenhæng mellem indikatorerne for årgangsstørrelser over klasseloftstærskler (instrumentet) og den observerede faktiske klassestørrelse. Det understøtter, at skolerne er bundet af klasseloftsreglen og lever op til den. Vi har testet dette ved et F-test af first stage regressionen (rapporteret i Bilagstabel 2.4). Stock og Yogo (2005) angiver F-test på mindst 10 og gerne større end 20 som tommelfingerregel for at afvise tilstedeværelsen af svage instrumenter. Bilagstabel 2.4 viser, at F-test på størstedelen af estimerne i denne rapport er langt over 30.

4.2.2 Manipulation af årgangens størrelse

Metoden bygger på en antagelse om, at skolerne står over for en eksogent givet (udefrakommende) regel for, hvor store klasser de må lave. Det betyder, at nogle skole bliver nødt til at oprette flere klasser, end de måske ellers ville, hvis der ikke var et klasseloft. I det omfang skolerne har mulighed for at påvirke (manipulere) klasseloftsreglen, svækkes antagelsen om eksogenitet og dermed gyldighed af instrumentet. Hvis nogle skoler aktivt manipulerer deres årgangsstørrelser for at ligge lige under en klasseloftstærskel, er det sandsynligt, at de også i andre henseender tilpasser deres (uobserverbare) ressourcer, hvorfor de ikke længere er sammenlignelige med skoler umiddelbart over klassestørrelsesloftet.

I dette afsnit undersøger og argumenterer vi for, at skoler og forældre kun i begrænset omfang har mulighed for at manipulere reglen og dermed klassernes størrelser.

For det første er klasseloftsreglen nationalt bestemt. Det er dog muligt at søge om dispensation fra klasseloftsreglen, og det er kun muligt at søge om dispensation op til 30 elever i en klasse. Dispensation har kunnet søges som del af frikommuneordningen, men efter forespørgsel i BUVM og STUK er det ikke muligt at fremskaffe en liste over skoler, der har ansøgt og fået dispensation. Det er derfor ikke muligt at udelade disse skoler fra analysen. Skolen kan også søge den enkelte kommune om dispensation på op til 30 elever i klassen i det tilfælde, at der kommer nye elever ind i løbet af skoleåret. I efterfølgende skoleår skal skolen omlægge klasserne, så de igen lever op til klasseloftsreglen.

For det andet betyder skoledistrikterne, at en skole skal modtage en elev, hvis denne hører til skoledistrikterne. Det betyder endvidere, at skolen ikke kan afvise en elev med henvisning til klassestørrelser, fx at en klasse vil blive for stor, og derved sikre, at klasseloftet ikke binder skolen.

For det tredje har vi foretaget en række empiriske test for at undersøge, om der i data er tegn på, at en form for manipulation finder sted. Disse analyser uddybes nedenfor.

Slutteligt er det muligt at forestille sig forældre, der forsøger at udløse en ekstra klasse og dermed mindre klassestørrelse ved at indmelde sig på en skole lige under en tærskel. Idet hverken den enkelte forælder eller skoleleder har fuld information omkring årgangens øvrige elever, vil dette dog være en risikabel strategi i mindre intervaller omkring tærsklerne, hvilket er med til at understrege vigtigheden af at begrænse analysesamplen til mindre intervaller omkring klasseloftstærsklerne.

En empirisk 'test' af ovenstående viser imidlertid, at skolens årgangsstørrelse for samme klassetrin i året før er en dårlig prædikator i forhold til den aktuelle klassestørrelse i smalle intervaller omkring klasseloftstærsklerne. F-testet for relevans af dette instrument i first stage er kun 2,5 for 0. klasse og langt under Stock og Yogo's (2005) tommelfingerregel på 10-20. Det betyder med andre ord, at forældrene før deres barns skolestart ikke kan drage konklusioner om deres barns klassestørrelse på baggrund af årgangsstørrelsen i 0. klasse året før.

Beskrivende analyse af årgangsstørrelser omkring klasseloftstærskler

Nedenfor præsenteres en empirisk analyse af fordelingen af klassestørrelse omkring hver klasseloftstærskel. Hvis skolerne har mulighed for at påvirke (manipulere) en årgangsstørrelse således, at de lige præcis undgår at kunne oprette en ekstra klasse, vil man forvente, at det vil afspejles i fordelingen.

For det første har vi visuelt undersøgt klassestørrelsesfordelingen på hvert klassetrin (Bilagsfigur 1.3 og Bilagsfigur 1.4). Der skal gerne være omtrent lige mange klasser på hver side af tærskelværdierne. For flere klassetrin er der dog en vis tendens til at have flere klasser lige under en tærskelværdi, hvilket kan indikere, at årgangens størrelse er påvirket for at være netop under en given tærskelværdi.

McCrary-test og grafer

Det kan også testes mere formelt ved at sammenligne tætheden af fordelingen på hver side af klasseloftstærskler. Denne test kan udføres ved en McCrary-test (McCrary, 2008). Resultatet er præsenteret grafisk i Bilag 1.

Bilagsfigur 1.5 præsenterer resultat af McCrary-testen grafisk på det samlede datasæt for hvert klassetrin 0.-9. Testen viser nogle diskontinuiteter, som ikke er forventede (kan også ses af en grafisk inspektion af figurerne som 'hop' mellem de to linjer hhv. under og over tærskelværdien). Vi har derfor gentaget testen for hver af de tre tærskelværdier separat (Bilagsfigur 1.6 og Bilagsfigur 1.7), hvilket viser, at en diskontinuitet i det samlede sample i de fleste tilfælde er drevet af en diskontinuitet omkring én enkelt tærskelværdi, mens der ses en fin smooth linje (uden hop) gennem de øvrige tærskelværdier. Da der ikke er en systematik i, hvilken tærskelværdi der udviser denne diskontinuitet på tværs af klassetrin, og der skal modelleres en model gældende for alle klassetrin, vurderer vi derfor at fortsætte med det pooled sample inkluderende alle tre tærskelværdier 28, 56 og 84.²⁰

²⁰ Tidligere analyser har vist, at samples omkring højere tærskelværdier (112 og højere multiple af 28) ikke bør inddrages.

Det skal i denne sammenhæng også bemærkes, at der er fremført en kritik om, at McCrary-density-testen ikke er retvisende, når den underliggende variabel (running variable) i Fuzzy RD-design er diskret (som det er tilfælde med årgangsstørrelsen, der måles i hele værdier 1, 2,...,N). I det diskrete tilfælde kan testen komme til at overestimere forskellen i tætheden på hver side af klasse-loftstærskler (Cattaneo, Jansson & Ma, 2018). For at reducere dette problem følges anbefaling fra Frandsen (2017)²¹, men som det ses af figurerne, er de fortsat kendetegnet med diskrete massepunkter, hvilket giver sig udslag i væsentlige 'hop', der altså kan være overestimerede. Af denne årsag tillægger vi ikke McCrary-testene størst betydning i vores samlede vurdering af metodens validitet (men ser mere på de øvrige argumenter og tests præsenteret).

I robusthedsanalysen foretager vi en test, der kan adressere ovenstående bekymring for, om rapportens estimerede effekter kan være drevet af det 'hop', som McCrary viser for nogle årgange og tærskelværdier. Robusthedstesten kaldes en 'donut-test'; se afsnit 4.4.

4.2.3 Test af diskontinuitet i baggrundsvariable (balancetest)

En af modellens bagvedliggende antagelser er, at andre forhold (fx vedrørende eleverne eller skolen) ikke må udvise samme diskontinuitet (hop) i årgangens størrelse. Det vil sige, der må ikke være en forskel i fx andelen af elever fra socioøkonomiske svage familier, der er sammenfaldende med, at klassens størrelse ændrer sig, når årgangens størrelse overstiger hhv. 28, 56 og 84 elever. Hvis det er tilfældet, er det vanskeligt at adskille effekten af ændringen i klassestørrelsen fra ændringen i elevernes sammensætning.

For at undersøge dette empirisk foretages en balancetest efter anvisning fra Cattaneo m.fl. (2018a,b). Balancetesten sammenligner, om elevgruppen, der ligger lige under en klasseloftstærskel, er sammenlignelig med elevgruppen, der ligger lige over en klasseloftstærskel på en lang række af baggrundsforhold. Det testes ved på samme måde som ovenfor at undersøge, om der er et 'hop' i diverse baggrundsvariable omkring årgangsstørrelsen.

Vi har anvendt to forskellige modeller til balancetestene: en OLS-model og en Fuzzy RD-model. Her præsenteres resultater fra Fuzzy RD-modellen, da den tilgang er anbefalet af Cattaneo m.fl. (2018a,b). RD-modellen køres ved brug af Statapakken *rdrobust* (Calonico m.fl., 2017). Det er en datadreven metode, der yderligere estimerer robuste og bias-korrigerede standardfejl.²² En fordel

²¹ I Statakommandoen specificeres en binsize på 1 og tærskelværdien forskydes -0.5.

²² Vi anvender Statapakken *RDrobust* og anvender følgende kommando: *rdrobust 'baggrundsvariabel' 'antal elever indskrevet på årgangen' if klassetrin = {0;9}, c('tærskel-værdi') bwselect(msetwo) fuzzy(klassestørrelse) vce(cluster 'institutionsnummer) covs('dummies for tærskelsegment' 'dummies for år')*

er endvidere, at den estimerer den optimale bandwidth, som kan bidrage til udvælgelse af bandwidth for vores foretrukne modelspecifikation.

Resultatet af balancetesten ved brug af RD-modellen for 1. klassetrin er rapporteret i Bilagstabel 2.5. Tabellen rapporterer det estimerede 'hop' og p-værdi fra en test af, om hoppet = 0. Insignifikante p-værdier indikerer, at der ikke er et signifikant hop/forskel i den respektive baggrundsvariabel mellem elever lige over og lige under klasseloftstærskler – eleverne er således sammenlignelige på den specifikke observerede baggrundsvariabel.

Tabellen viser samme balancetest omkring hver tærskelværdi separat. Her ses kun få ubalancer: to omkring tærskelværdi 28 (vedrørende elevernes alder), en omkring tærskelværdi 56 og ingen omkring tærskelværdi 84, når der anvendes et signifikansniveau på 5 %. Balancetestene omkring hver tærskelværdi viser således, at det er rimeligt at antage, at elevgruppen hhv. lige under og lige over tærskelværdien er sammenlignelige på de observerbare baggrundsforhold.

På baggrund af balancetesten vurderes det også, at det er rimeligt at samle elever omkring alle tre tærskelværdier i ét samlet estimationssample ('pooled' sample), da der ikke ses væsentlige forskelle eller ubalancer på tværs af de tre tærskelværdier, der kunne drive gennemsnitlige effekter. Et pooled datasæt har den fordel, at det øger sample size og den statistiske styrke af estimationerne.²³ Balancetesten for det pooled data ses i de sidste tre kolonner i Bilagstabel 2.5 og viser ingen signifikante p-værdier og understøtter dermed balancen og antagelsen bag vores valgte identifikationsstrategi.

Samme mønster ses for de øvrige klassetrin. Af pladshensyn er balancetesten opdelt på tærskelværdier derfor ikke rapporteret for de øvrige klassetrin.

Valget af bandwidth omkring tærskelværdierne er en afvejning mellem bias og statistisk præcision. Vores foretrukne modelspecifikation anvender en bandwidth omkring tærskelværdien på +/-5 elever på årgangsstørrelsen. Balancetesten kan bidrage til at verificere dette valg, da den samtidig estimerer den optimale bandwidth. Bilagstabel 2.5 viser, at den optimale bandwidth for 1. klassetrin estimeres til at ligge mellem 3,3 og 5,7 afhængigt af baggrundsvariabel. Dette bestyrker vores valg af bandwidth på 5.

Bilagstabel 2.6 rapporterer p-værdier fra tilsvarende balancetest foretaget for hvert klassetrin separat. Der er ingen ubalancer på 5 %-signifikansniveau, med

²³ I den foretrukne modelspecifikation kontrolleres for, hvilken tærskelværdi som eleverne ligger tættest på. Der kontrolleres for evt. niveau-effekter på tværs af tærskelværdier.

undtagelse af 4. klassetrin, hvor der er ubalance i to variable vedrørende mødres uddannelsesniveau. Der er højst to ubalancer på 10 %-signifikansniveau pr. klassetrin.

Balancetestene understøtter således metodens bagvedliggende antagelse om, at eleverne på hver side af en tærskel (når vi ser på elever, der ligger meget tæt på tærsklen) er sammenlignelige.

4.2.4 Test af instrumentets eksogenitet

En empirisk test af instrumentets eksogenitet kan udføres ved at undersøge, om der er en sammenhæng mellem elevens værdi på instrumentet (above-indikatorerne) og diverse baggrundsvARIABLE. Dette testes i en regressionsmodel, hvor above-indikatoren regresseres på det fulde sæt af baggrundskontroller. Resultatet for hvert klassetrin er rapporteret i Bilagstabel 2.7 og viser, at kun meget få af baggrundsvARIABLE er korreleret med elevens status på above-indikatoren. Nederst i tabellen er rapporteret et F-test for joint signifikans, hvilket også bekræfter, at baggrundsvARIABLENE samlet set ikke forklarer elevens status på above-indikatoren, hvilket dermed understøtter instrumentets eksogenitet.

4.3 Analyse af persistens af indsatseffekter

REFUD har en fireårig tidshorisont, og i overensstemmelse hermed behandler denne rapport de estimerede klassestørrelses effekter på de konsekvenser i førstkomende kalenderår, år 1, og til og med år 4.

I modsætning til mange andre skolebaserede indsatser er der imidlertid en betragtelig træghed i klassestørrelse, der i populationsdata besværliggør adskillelsen af en 'primær' klassestørrelses effekt fra effekten af senere klassestørrelser, når udfaldsmål betragtes over perioder over flere år. Denne træghed betyder også, at det ikke er muligt at inkludere flere på hinanden følgende klassestørrelsesvariable i samme analyse, idet samvariationen mellem disse er for stor til at give meningsfyldte punkttestimater.

En pålidelig men også omkostningstung løsning er naturligvis at udføre et troværdigt RCT-studie, der sammenligner en kontrolgruppe med indsatsgrupper, der modtager en klassestørrelsesindsats i et hhv. flere sammenhængende år, men selv det er ikke uden vanskeligheder. I det kendte klassestørrelses-RCT STAR placeredes eleverne i små (13-17 elever) vs. almindelige (22-25 elever) størrelses klasser med og uden ekstra lærerhjælp i 0.-3. klasse, men designet

blev hurtigt delvist kompromitteret af forældre, der var utilfredse med den initiale placering af deres børn og derfor flyttede deres børn både internt mellem indsatsgrupper og ud af indsatskoler (Krueger, 1999).

En løsning, der anvendes flere steder i litteraturen baseret på populationsdata, er at behandle skoleafdelingerne (indskoling, mellemtrin og udskoling) som separate institutioner og benytte den gennemsnitlige klassestørrelse til beregning af effekten af at øge den gennemsnitlige klassestørrelse på (særligt) mellemtrinnet med en elev på udfaldsmål målt i slutningen af skoletrinnet (fx Frederiksson m.fl., 2013; Leuven & Løkken, 2020). Leuven & Løkken (2020) argumenterer desuden for, at der pga. opsamlingskoler for elever i udskolingen er tilstrækkelig variation i deres klassestørrelsesvariabel og instrument til at kunne beregne effekten af den gennemsnitlige klassestørrelse på mellemtrinnet og i udskoling separat. I en REFUD-kontekst ville dette analysedesign dog betyde en gennemsnitligt endnu større forskel i timingen for måling af klassestørrelse og de budgetøkonomiske konsekvenser, der ønskes belyst.

Bilagstabel 2.8 præsenterer resultatet af en eksplorativ analyse, der erstatter udfaldsmål i vores primære specifikation med klassestørrelsen i det efterfølgende skoleår. Estimatet på klassestørrelse skal således fortolkes som den øgede klassestørrelse i år 2 af at øge klassestørrelsen med en elev i år 1 alt andet lige. Punktestimaterne afslører en høj grad af træghed i klassestørrelse for især klassetrinnene i indskolingen med en værdi tæt på 1. Effektestimaterne på udfaldsmål i år 2 vil for disse klassetrin i højere grad afspejle effekten af 'gentagen' klassestørrelsesindsats, dvs. effekten af at øge klassestørrelsen med en elev og fastholde den stigning i det efterfølgende skoleår, end den rene effekt af at øge klassestørrelsen i første skoleår. Tabellen afslører desuden, at der mellem 6. og 7. samt mellem 8. og 9. klassetrin sker sammenlægning af klasser relativt oftere og/eller højere grad af udsivning af elever, idet korrelationen mellem nuværende og efterfølgende klassestørrelse er langt lavere her.

Vi har desuden foretaget supplerende analyser, hvor vi har erstattet instrumentet baseret på samtidig årgangsstørrelser med en instrumentvektor baseret på årgangsstørrelsen, som tilfaldt eleven i skoleåret før. First stage-resultaterne fra denne analyse bekræfter mønsteret fra Bilagstabel 2.8. Det laggede instrument har generelt en smule mindre bid end det samtidige instrument, om end forskellene er små. Undtaget herfra er de estimerede first-stages for klassestørrelse i 7. og 9. klasse, hvor instrumenterne baseret på hhv. 6. og 8. klasses årgangsstørrelsen medfører en markant ringere F-test, der falder fra 103 til omkring 17 hhv. fra 70 til 24. Dette resultat antyder, at der for

7. og 9. klasse er tilstrækkelig med variation i data til, at det er muligt at bestemme klassestørrelseseffekterne for disse klassetrin, uden at denne absorberer effekter fra tidligere.²⁴

4.4 Analyse af robusthed

Robusthedsanalysen anvendes til at guide den foretrukne modelspecifikation og deraf analysens hovedresultater om effekterne af klassestørrelse på de specificerede udfald.

I tillæg til de undersøgende analyser i Afsnit 4.3, der havde til formål at afdekke gyldigheden af instrumentet, har robusthedstestene ligeledes til formål at overbevise læseren om gyldigheden og styrken af vores valg af instrument og identifikationsstrategi.

For at udforske gyldigheden af rapportens identifikationsstrategi og deraf også analysens resultater foretages forskellige robusthedstest. Set i lyset af de mange udfald, der behandles i analysen, udvalgte på forhånd effektestimaterne på *Specialundervisning i almenklasse* samt elevtrivselsfaktoren *Ro og orden i klasselokalet* som grundlag for robusthedsanalyserne.

Specialundervisning er valgt af to grunde. For det første ud fra et hensyn til 'relevans', hvor specialundervisning er et af de udfaldsmål, hvor vi forventer, at der kan være en effekt af klassestørrelsen. For det andet ud fra et empirisk hensyn, hvor det forventes, at der er en vis mængde variation i dette udfaldsmål, som vil kunne give udslag i robusthedstestene. Imidlertid modtager kun omkring 0,5 % af normalklasseeleverne på hver årgang i den danske folkeskole i årene 2015-2019 specialundervisning, hvilket til dels skyldes, at specialundervisning kun registreres for undervisningsbehov på ni eller flere timer ugentligt, hvilket hæmmer præcisionen af effektestimaterne.

Trivselsindikatoren *Ro og orden* valgtes ligeledes ud fra et relevanskriterium, idet netop elevernes opfattelse af ro, støj og klasserumsledelse i litteraturen (og i praksis) ofte relateres direkte til antallet af elever i klassen (fx Lazear, 2001). Denne dannes på baggrund af fire spørgsmål fra den nationale trivselsmåling i folkeskolens 4.-9. klasse (Krasel m.fl., 2022b; STIL 2020), hvor eleverne hver har fem svarmuligheder. Således forventedes også dette udfald at indeholde en del variation til præcis bestemmelse af klassestørrelseseffekter.

På baggrund af robusthedsanalyserne for de valgte udfaldsmål udvælges den model, som vi finder mest overbevisende, og denne anvendes til at estimere

²⁴ Af pladshensyn er resultaterne af denne supplerende analyse udeladt.

effekterne på det fulde sæt af udfaldsmål. Denne tilgang er valgt af to hensyn. For det første vil vi gerne undgå 'specification searching', som er en kritik, der med rette kan stilles, hvis modelspecifikationen først vælges, efter flere forskellige modeller er kørt på forskellige udfaldsmål i søgen efter de ønskede eller flest mulige effekter. For det andet er det ikke muligt inden for opgavens rammer at foretage en robusthedsanalyse for alle udfaldsmål og 10 klassetrin. Vi har derfor prioriteret en grundig robusthedsanalyse af to udfald.

Der er lavet følgende robusthedsanalyser:

- Valg af samplevindue omkring RD-variationen: ± 3 , 5, 7 samt 14 elevtal, idet det endelige valg er en afvejning mellem instrumentets gyldighed (tættere på RD-tærskel) og præcision i estimaterne som følge af stikprøvestørrelse (længere fra RD-tærskel). Desuden sammenlignes gyldigheden af instrumentet for hvert af populationens RD-tærskler.
- Valg af instrument-specifikation: 'Kontinuer' instrument-specifikation baseret på forventet klassestørrelse med mere fleksibel kontrol af årgangsstørrelse, fx 2. gradspolynomium.
- Donut-specifikation, hvor skoler med elevtal lige omkring RD-variationen ekskluderes for at vurdere estimaternes gyldighed i forhold til manipulation af årgangsstørrelse.
- Gradvis inklusion af baggrundsvariable til bedømmelse af effektestimaternes kausale gyldighed samt øget præcision af estimatet: a) kontroller for årgangsstørrelse og årstal, b) tilføjer elevforhold og c) tilføjer skoleforhold. Se også afsnit 2.4.

4.4.1 Valg af samplevindue (bandwidth choice)

Valget af samplevindue er et af de mest afgørende for analysens resultater og baserer sig på det velkendte trade-off mellem bias og varians.

I et RD-design som dette (udløst af klassestørrelsesloftet) forventes estimatet for indsatstildeling at være virkelig unbiased i tærskelpunktet. I særdeleshed forventes instrumentet baseret på klassestørrelsesloftet at være mere gyldigt desto mindre intervaller omkring klasseloftstærsklerne, der betragtes, idet instrumentet 'binder' her og forsøg på manipulation fra øvrige interessenter er mere omkostningsfuldt (se også afsnit 4.2.2).

På trods af disse overbevisende argumenter for kun at betragte samplen med årgangsstørrelse lige under hhv. lige over klasseloftstærsklerne er det sjældent attraktivt at begrænse analysesamplen i så stort et omfang, da den mindre sample betyder mindre præcision i modellens estimater og dermed mindre

sandsynlighed for at detektere sande effektstørrelser alt andet lige. Dette slår også igennem i analysens first stage, hvor F-testen for relevans af årgangsstørrelsesinstrumenterne er større desto mindre interval omkring klasseloftstærsklerne, der betragtes.

Slutteligt ekskluderes skoler med meget store og meget små elevtal fra analysesamplen. Klassestørrelsesloftet mister sit bid i skoler med meget store årgangsstørrelser over 98 elever, hvilket til dels også skyldes det mere sparsomme datamateriale ved sådanne skolestørrelser. De helt små skoler med årgangsstørrelser på 14 elever og derunder ekskluderes dels for at opnå symmetriske intervaller omkring klasseloftstærsklerne, dels i forventning om, at helt små skoler er underlagt ressourcebegrænsninger, der er ganske forskellige fra den øvrige del af samplen.²⁵

Figur 4.1 og Figur 4.2 præsenterer de estimerede marginale klassestørrelses effekter baseret på vores primærspecifikation (se afsnit 4.1) på henholdsvis trivselsfaktoren R_o og orden og Forbrug af specialundervisning i almenklasse baseret på samples (BW) +/-3, 4, 5, 6 og 14 elever omkring de tre nederste klasseloftstærskler. Sidstnævnte, BW 14, udgør den fulde analysesample, hvorfor denne desuden kontrollerer for et 2. gradspolynomium i årgangsstørrelse for at rense effektestimaterne for den direkte påvirkning af denne (se afsnit 4.4.2).

Af Figur 4.1 anes bias-variens-trade-off'et for BW 3-6, hvor standardfejlene gradvist mindskes med bandwidth-størrelsen. Ligeledes er der tendens til, at punktestimaterne er absolut større, desto smallere samplevindue omkring klasseloftstærsklerne. F-testet for instrumentvektorens relevans i first stage rangerer på et rimeligt niveau fra 21 på 9. klassetrin til 52 på 4. klassetrin for BW 3. For samplevinduet BW 5 rangerer F-testet for first stage mellem 74 på 8. klassetrin og 148 på 4. klassetrin. Af Figur 4.2 anes samme tendenser med F-test af omtrentlig samme størrelse.²⁶

Konceptuelt foretrækker vi det smalleste samplevindue omkring klasseloftstærsklerne, som kan bære instrumentet, hvilket i ovenstående analyse er +/-3 elever (BW3). At vi alligevel vælger BW5 som vores primære samplevindue for analyse skyldes to faktorer. Dels udviser histogrammerne for årgangsstørrelse tegn på manipulation af årgangsstørrelse lige omkring klasseloftstærsklerne (se Bilag 1), hvorfor vi ønsker at foretage såkaldte donut-test, hvor årgangsstørrelser lige omkring tærsklen udtages, dels er formålet med at levere estimater på klassestørrelses effekterne, at de skal kunne anvendes i REFUD. Modellen medtager indtil videre effektestimater, når deres signifikans overstiger

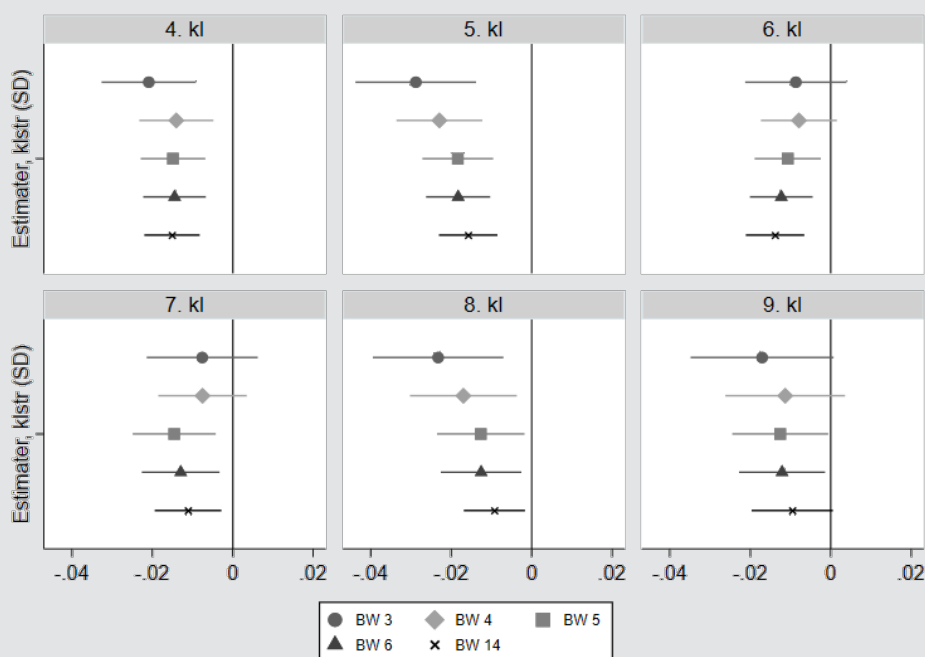
²⁵ Se også samleudvælgelsen i Nandrup (2016).

²⁶ Den eneste forskel i first stage-regressionerne skyldes eksklusionen af elever, der ikke har besvaret tiltrækkeligt med spørgsmål i den nationale trivselsmåling til at opnå en værdi i R_o og orden-faktoren.

et vist niveau, men på denne måde kan helheden i effektestimaterne over klassetrin og år gå tabt, hvis estimaterne generelt er upræcist estimerede. Således takserer vi i denne sammenhæng præcision en smule højere og er villige til at acceptere en smule mere bias i klassestørrelsesestimaterne. Af Figur 4.1 og Figur 4.2 fremgår det desuden, at der er relativt små forskelle mellem estimaterne baseret på BW3 og BW5, og at konfidensintervallerne i alle tilfælde overlapper.

Figur 4.1 Valg af samplevindue, trivselsfaktor: Ro og orden

Figuren viser 2SLS-estimater for den marginale effekt af klassestørrelse på elevtrivselsfaktoren Ro og orden (standardiseret) i forskellige samplevinduer omkring klasseloftstærsklerne 28, 56 og 84.

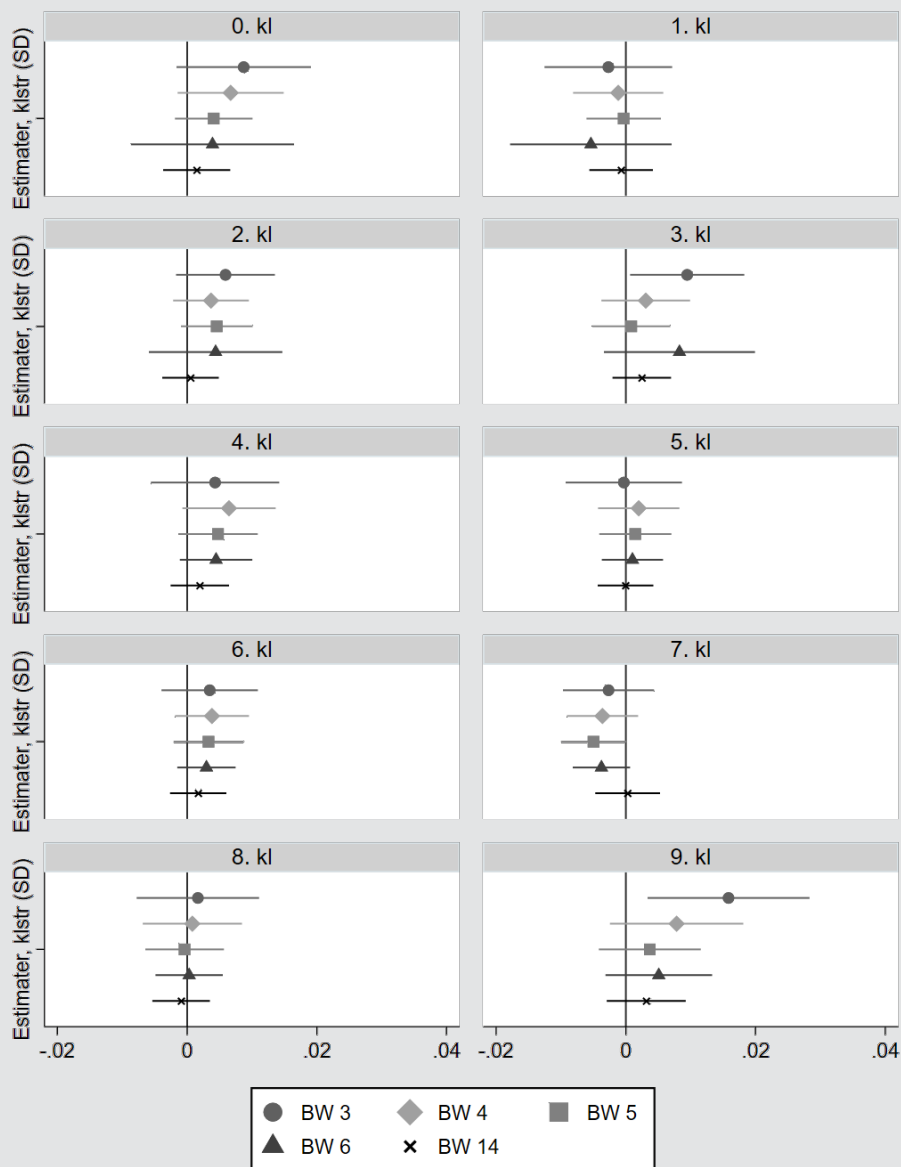


Anm.: Konfidensintervallerne angiver signifikans på 5 %-niveau. Standardfejl er klyngekorrigeret på skoleniveau.

Estimaterne er baseret på populationen af folkeskoleelever i fuldt årgangsopdelte normalklasser i skoleårsgange med årgangsstørrelser på +/- BW-elever omkring de tre nederste klasseloftstærskler i skoleårene 2014/15-2018/19. Trivselsmålet er standardiseret. Alle regressioner kontrollerer for årgangsstørrelsessegment, år for måling af klassestørrelse, alder og kvadreret alder, køn, ikkevestlig baggrund, antallet af ældre søskende, moderens og faderens uddannelse, alder ved fødsel og kvadreret alder ved fødsel samt indikatorer for, om moderen og faderen er diagnosticeret med angst og depression hhv. opmærksomhedsforstyrrelser. BW14-regressionerne inkluderer desuden 2. gradspolynomium i årgangsstørrelse.

Figur 4.2 Valg af samplevindue, Specialundervisning i almenklassen

Figuren viser 2SLS-estimater for den marginale effekt af klassestørrelse på forbrug af specialundervisning i almenklasser (standardiseret) i forskellige samplevinduer omkring klasseloftstærsklerne 28, 56 og 84.



Anm.: Konfidensintervallerne angiver signifikans på 5 %-niveau. Standardfejl er klyngekorrigeret på skoleniveau.

Estimaterne er baseret på populationen af folkeskoleelever i fuldt årgangsodelte normalklasser i skoleårge med årgangsstørrelser på +/- BW-elever omkring de tre nederste klasseloftstærskler i skoleårene 2014/15-2018/19. Udfaldsmålet er standardiseret. Alle regressioner kontrollerer for årgangsstørrelsessegment, år for måling af klassestørrelse, alder og kvadreret alder, køn, ikkevestlig baggrund, antallet af ældre søskende, moderens og faderens uddannelse, alder ved fødsel og kvadreret alder ved fødsel samt indikatorer for, om moderen og faderen er diagnosticeret med angst og depression hhv. opmærksomhedsforstyrrelser. BW14-regressionerne inkluderer desuden 2. gradspolynomium i årgangsstørrelse.

Valg af samplevindue over modellens tidshorisont

Idet vores analysepopulation består af folkeskoleelever i skoleårene 2014/15-2018/19, og at de budgetøkonomiske konsekvenser i de fleste tilfælde måles til og med kalenderår 2020, indgår relativt færre klassestørrelseskohorter i analysen, desto længere frem udfaldsmålet måles (se optælling i Tabel 5.1).

Således kan man ønske at udvide samplevinduet, når udfaldsmål længere ude i tidshorisonten betragtes, for at kompensere størrelsen af analysesamplen for tabet af klassestørrelseskohorter og dermed øge præcisionen af estimaterne, således at effektstørrelser i samme størrelsesorden vil kunne detekteres som signifikant i år 1 og helt op til år 4. Idet vi imidlertid har valgt et relativt bredt samplevindue på +/-5 elever i vores primære analysesample for år 1, ønsker vi ikke at gå yderligere på kompromis med bias i de rapporterede estimater ved at udvide samplevinduet yderligere. Vi noterer os desuden, at F-testene for BW5 fortsat er rimelige helt op til år 4 (se Bilagstabel 2.4).

4.4.2 Valg af instrumentspecifikation

På trods af, eller netop på grund af, at klassestørrelseslitteraturen ofte har benyttet klassestørrelsesloft til at skabe eksogen variation i klassestørrelse i forfølgelsen af kausale klassestørrelses effekter, findes mange forskellige afarter af identifikationsstrategien.

Specifikt benytter vi i denne rapport en vektor af indikatorer, der antager værdien 1 i intervallerne umiddelbart over klasseloftstærsklerne, og dermed fremhæver de årgangsstørrelser, der baseret på klassestørrelsesloftet bør udløse mindre klasser, $above_{st} = \mathbf{1}[\tau < e_s \leq \tau + 14]$. Dette design fremhæver Regression Discontinuity-elementet i identifikationsstrategien.

Angrist & Lavy (1999, 2019) benytter i stedet den såkaldte forventede klassestørrelse som instrument, dvs. den gennemsnitlige klassestørrelse, hvis eleverne blev delt i klasser af nøjagtig samme størrelse baseret på årgangsstørrelsen og klassestørrelsesloftet.²⁷ I modsætning hertil udnytter rapportens valgte instrumentindikatorer kun variationen i forventet klassestørrelse, der er skabt på baggrund af at krydse en klasseloftstærskel, og ikke den glatte variation i intervallerne imellem disse, om end det ikke vurderes at have den store fortolkningsmæssige forskel i smallere samplevinduer omkring tærsklerne.

²⁷ Den forventede klassestørrelse i skole s , på klassetrin k , og i skoleår t er: $f_{it} = \frac{r_{skt}}{[\ln(\frac{r_{skt}-1}{28})+1]}$, hvor r_{skt} er total antal elever på årgangen for den givne skole s , klassetrin k og skoleår t .

Dernæst er valget af kontroller for årgangsstørrelse af stor betydning for gyldigheden af identifikationsstrategien. Større skoler står alt andet lige over for andre budgetmæssige restriktioner og elevsammensætninger end mindre skoler, hvilket forårsager en empirisk sammenhæng mellem årgangsstørrelse og elevudfald. Denne sammenhæng forventes dog at være relativt 'glattere' i årgangsstørrelse – der er ikke nogen årsag til at skulle forvente, at denne skulle slå særligt ud i multipla af 28 elever. Ikke desto mindre er det dog vigtigt at kontrollere tilstrækkeligt for årgangsstørrelse for at rense effektestimaterne for støjen fra årgangsstørrelse. Angrist & Lavy (1999, 2019) nøjes med 2. gradspolynomier i årgangsstørrelse, mens Fredriksson m.fl. (2013) foretrækker et 1. gradspolynomium, der tillades forskellige hældninger både over og under klasseloftstærsklen for hvert årgangsstørrelsessegment.

Idet vores primære analysesample imidlertid kun indeholder et smalt samplevindue omkring tærsklerne, har det i tidligere studier vist sig, at det ødelægger variationen i instrumentet, hvis alt for fleksible funktioner af årgangsstørrelse inkluderes i regressionerne (se fx Nandrup, 2016). Vores primære specifikation indeholder derfor kun indikatorer for, omkring hvilken tærskel årgangsstørrelsen ligger (Figur 4.3 og Figur 4.4).

Til sammenligning præsenterer figurerne desuden de estimerede klassestørrelses effekter ved inklusion af et 2. gradspolynomium i årgangsstørrelse samt ved brug af forventet klassestørrelse som instrument på den fulde sample til at replicere Angrist & Lavy (1999)-specifikationen og instrument-indikatoren kombineret med de meget fleksible kontroller for årgangsstørrelse fra Fredriksson m.fl. (2013). Der er generelt meget små forskelle mellem estimerne i de forskellige specifikationer i Figurerne 4.3 og 4.4. Mest i øjnefaldende er den meget fleksible årgangsstørrelseskontrol i Fredriksson m.fl. (2013), der især i 7.-8. klasse ser ud til at drive effektestimaterne på trivselsfaktoren mod nul. Det er i øvrigt også værd at notere sig, hvorledes standardfejlene med denne specifikation øges markant for specialundervisningsudfaldet.

Donut-specifikation

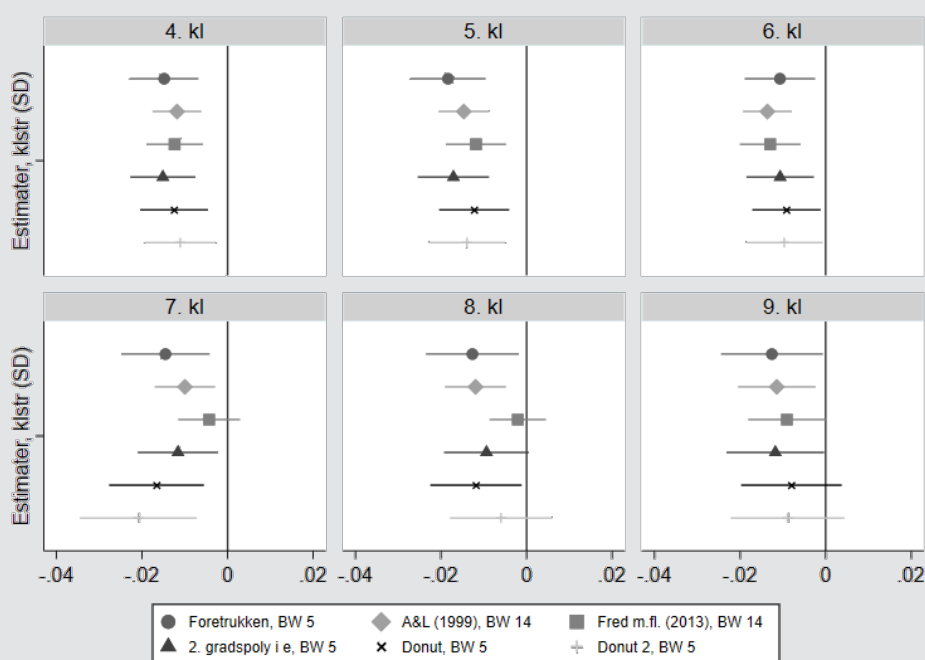
Den visuelle inspektion af histogrammerne for årgangsstørrelse (Bilag 1) afslører mulig manipulation af årgangsstørrelserne for især nederste tærskel (28) i 0.-6. klasse og tærskel 56 i udskolingen, hvilket ugyldiggør instrumentet. Histogrammerne viser således noget færre skoler med årgangsstørrelser lige over klasseloftstærsklerne i forhold til lige under. Figur 4.3 og Figur 4.4 indeholder derfor også effektestimater baseret på en såkaldt donut-specifikation, hvor årgangsstørrelse lige omkring tærsklerne udelades.

Figurerne inkluderer både en donut-specifikation, hvor årgangsstørrelser på +/-1 elev omkring tærsklerne udelades, og donut2-specifikation, der udelader årgangsstørrelser på +/-2 elever. Donut-specifikationer skal generelt fortolkes

med forsigtighed, idet vi kun forventer, at estimatet er virkelig unbiased lige omkring klasseloftstærsklerne. Derfor, hvis estimatet i en donut-specifikation ændres væsentligt i forhold til den primære analyse, kan det både være, fordi man har fjernet den ægte unbiased effekt fra donut-estimatet, eller fordi årgangsstørrelsen er blevet manipuleret lige omkring tærsklerne, hvilket vil forvrænge det primære effektestimater. I Figur 4.3 og Figur 4.4 er der dog generelt ganske stor overensstemmelse mellem de primære effektestimater og donut-estimatene.

Figur 4.3 Instrumentspecifikation, trivselsfaktor: Ro og orden

Figuren viser 2SLS-estimer for den marginale effekt af klassestørrelse på trivselsfaktoren Ro og orden (standardiseret) ved forskellige instrumentspecifikationer, samplevinduer, og kontroller for årgangsstørrelser.

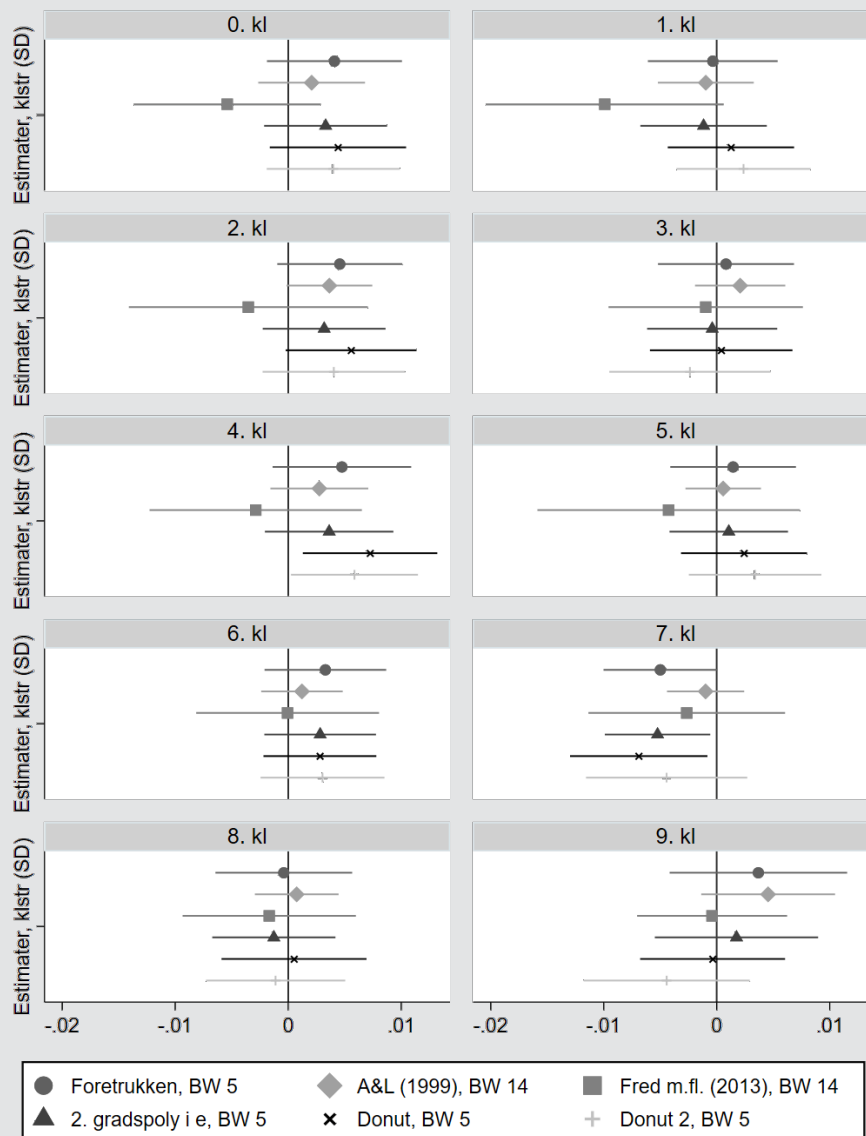


Anm.: Konfidensintervallerne angiver signifikans på 5 %-niveau. Standardfejl er klyngekorrigeret på skoleniveau.

Estimaterne er baseret på populationen af folkeskoleelever i fuldt årgangsoopdelte normalklasser i skoleårerne 2014/15-2018/19. Donut-specifikationerne udlader hhv. +/-1 og +/-2 elever omkring tærsklerne. Trivselsmålet er standardiseret. Se rapportteksten for instrumentspecifikation og inkluderede kontroller for årgangsstørrelse. Alle regressioner kontrollerer desuden for år for måling af klassestørrelse, alder og kvadreret alder, køn, ikkevestlig baggrund, antallet af ældre søskende, moderens og faderens uddannelse, alder ved fødsel og kvadreret alder ved fødsel samt indikatorer for, om moderen og faderen er diagnosticeret med angst og depression hhv. opmærksomhedsforstyrrelser.

Figur 4.4 Instrumentspecifikation, Specialundervisning i almenklasse

Figuren viser 2SLS-estimer for den marginale effekt af klassestørrelse på forbrug af specialundervisning i almenklasse (standardiseret) ved forskellige instrumentspecifikationer, samplevinduer, og kontroller for årgangsstørrelser



Anm.: Konfidensintervallerne angiver signifikans på 5 %-niveau. Standardfejl er klyngekorrigeret på skoleniveau.

Estimerne er baseret på populationen af folkeskoleelever i fuldt årgangsodelte normalklasser i skoleårge med årgangsstørrelser på +/- BW-elever omkring de tre nederste klasseloftstærskler i skoleårene 2014/15-2018/19. Donut-specifikationerne udlader hhv. +/-1 og +/-2 elever omkring tærsklerne. Se rapportteksten for instrumentspecifikation og inkluderede kontroller for årgangsstørrelse. Alle regressioner kontrollerer desuden for år for måling af klassestørrelse, alder og kvadreret alder, køn, ikkevestlig baggrund, antallet af ældre søskende, moderens og faderens uddannelse, alder ved fødsel og kvadreret alder ved fødsel samt indikatorer for, om moderen og faderen er diagnosticeret med angst og depression hhv. opmærksomhedsforstyrrelser.

4.5 Multiple hypotesetestning

Ofte løses denne problematik ved at signifikanskorrigere estimaternes statistiske betydning i forhold til antallet af hypoteser, der testes samtidigt (dvs. antallet af udfald, for hvilke klassestørrelseseffekten analyseres).

Hvis man som i denne analyse antager, at udfaldene er korrelerede, forsøger man ofte at kontrollere den såkaldte *Family-wise Error Rate (FWER)*, dvs. sandsynligheden for at begå én type-I-fejl i en familie af korrelerede udfald (forkaste en nulhypotese om ingen effekt, selvom den er sand). Ulempen ved denne tilgang er, at des flere test, der udføres samtidigt, des strengere krav stilles til estimaternes præcision.

En alternativ løsning er at kontrollere den såkaldte *False Discovery Rate (FDR)*, dvs. andelen af de afviste nulhypoteser, der er forfejlede. Dermed accepterer man tilstedeværelsen af nogle type-I-fejl for til gengæld at have større styrke til at detektere (sande) effekter. Slutteligt kan man betragte effekten af klassestørrelse for et givent udfald samlet (dvs. på tværs af klassetrin og/eller år), og således teste nulhypoteserne om, at en klassestørrelsesændring samlet er uden effekt på et givent udfald. Ulemperne ved sidstnævnte er for det første, at den testede nulhypotese ændres, hvorved det er svært at se, hvordan de enkelte estimater kan anvendes i REFUD-regi, hvor brugeren selv bestemmer de klassetrin, der anvendes i beregningerne. Og for det andet, et samlet test af klassestørrelseseffekten på tværs af alle (eller de klassetrin, der vurderes sammenlignelige) klassetrin vil stadig efterlade mindst 59 samtidige hypotesetest i analysen, hvilket i bedste fald medfører 95 % sandsynlighed for at forkaste mindst én sand nulhypotese (se fodnote 17).

Idet datagrundlaget i denne rapport allerede i udgangspunktet er tyndt i forhold til at detektere effekter af klassestørrelsesreduktioner af den forventede størrelsesorden (0,01-0,02 SD, se afsnit 2.5 om eksisterende litteratur), foretrækkes FDR-justeringen konceptuelt for at undgå yderligere styrketab (jf. rapportens FDR-justering af trivselseffekterne). I praksis er de estimerede effekter i den primære modelspecifikation på SØM-konsekvenserne dog så tilstrækkeligt upræcist estimerede, at ingen af disse vil overleve, heller ikke en signifikans-korrektion af FDR-typen, hvorfor en formel korrektion af disse ikke foretages. Idet vi vurderer, at udvalgte resultater, på trods af det sparsomme datagrundlag, peger på interessante tendenser af effekterne af klassestørrelse på SØM-konsekvenserne, der er tilstrækkeligt overbevisende i størrelse og konsistens (om end ikke præcision), omtales disse i rapporten netop som tendenser og ikke klassestørrelseseffekter.

5 Data

Kapitel 5 beskriver data, udvælgelsen af analysesamples, og hvilke samples og udfaldsmål vi forventer at kunne lave analysen for, givet det tilgængelige data fra REFUD.

5.1 Analysedata

Analysen anvender de danske administrative registre til at danne et analyse-datasæt bestående af den fulde population af fuldt årgangsopdelte normal-klasseselever i folkeskolen i perioden 2014/15-2018/19.²⁸

Der anvendes kun data fra perioden efter folkeskolereformen for at øge analysens aktualitet og overførbare til nuværende skolekontekst. Endvidere anvendes kun data om elever i normalklasser for at sikre instrumentets gyldighed.

Skoleårene 2019/20 og 2020/21 udelades, idet skolenedlukningerne i forbindelse med covid-19-pandemien sandsynligvis har påvirket flere af udfaldsmålene, hvilket vil forvrænge de estimerede klassestørrelseseffekter, såfremt påvirkningerne er korrelerede med klassestørrelse. Derudover måles de budgetøkonomiske konsekvenser kun til og med kalenderår 2020.

Analysepopulationen indeholder:

- Alle folkeskoleelever i fuldt årgangsopdelte normalklasser i 2014/15 til 2018/19
- Heraf udtages elever på skoler med årgangsstørrelse i intervallet 15-98, således at de helt små og meget store skoler ekskluderes. Dette svarer til et symmetrisk interval på +/-14 elever omkring de tre nederste klasse-loftstærskler, hvor klasseloftet binder empirisk (se figur i Boks 2.1).

Baggrundsdata om skoler og elever:

- Skolers institutionstype og indskrevne elever. Klassestørrelser for normalklasseselever i fuldt årgangsopdelte klasser dannes på baggrund af Specialundervisningsregisteret (UDSP), mens årgangsstørrelse dannes på

²⁸ De frie og private grundskoler sætter ofte deres egen (lavere) klassekvotient, hvilket både kan få betydning for analysens interne validitet (vedrørende instrumentets styrke og relevans) og eksterne validitet (generaliserbarhed), og derfor begrænses analysen til folkeskolen. Det vurderes endvidere at være mest relevant at foretage undersøgelsen alene for folkeskolen, da SØM er tiltænkt som et kommunalt beslutningsredskab.

baggrund af Det Komprimerede Elevregister (KOTRE), hvorfra antallet af specialklasseelever på årgangen (UDSP) fratrækkes.

- Elevernes alder, køn, fødselsinformationer, helbred, indvandrerbaggrund og forældres helbred og socioøkonomiske baggrund (uddannelse, indkomst og arbejdsmarkedstilknytning) dannes på baggrund af registre i Danmarks Statistiks Grunddatabank, se Krassel m.fl. (2022a) for et detaljeret overblik.

Udfaldsmål:

Analysens udfaldsmål indeholder samtlige budgetøkonomiske konsekvenser fra REFUD samt modellens fem trivselsmål. De konkrete datakilder og definitioner kan ses i Krassel m.fl. (2022a,b) og er ligeledes oplistet i Bilag 3.

De budgetøkonomiske udfaldsmål optælles pr. kalenderår til og med 2020 og omfatter:

- Forbrug af grundskole og specialundervisning, 10. klasse og ungdomsuddannelser, sociale foranstaltninger til børn og unge under 18 år, efterværn (til 18-22-årige), sundhedsydelse, kontakt med retsvæsen og kriminalforsorg, indkomstoverførsler, beskæftigelsesindsatser, beskæftigelsesindkomst og sociale serviceydelser. Flere af disse er først relevante fra eleven fylder 15-18 år.
- For elevernes forældre omfatter det mødres og fædres indkomstoverførsler, beskæftigelsesindsatser og beskæftigelsesindkomst.
- For elevernes lærere omfatter det deres personalefravær.

Derudover undersøges også elevtrivsel baseret på folkeskolens nationale trivselsmåling i 4.-9. klasse. Vi anvender de fire trivselsindikatorer dannet af STIL (2020): Faglig trivsel, social trivsel, Støtte og inspiration og Ro og orden, samt generel elevtrivsel indeholdende alle spørgsmålene fra de fire øvrige trivselsmål.

Beregning af klassestørrelser

Faktisk klassestørrelse angiver antal elever i en klasse. Det er opgjort på baggrund af UDSP-data og variabelen, der angiver klassebetegnelse.

Årgangens størrelse er opgjort på baggrund af Elevregisteret.²⁹

Instrumentet, der er baseret på årgangens størrelse, er således opgjort i starten af skoleåret hvorimod den faktiske klassestørrelse er opgjort i slutningen af skoleåret. Således tager vi højde for noget af den selektion, der kan ske i

²⁹ Hvis en elev har mere end én observation pr. skoleår i Elevregisteret anvendes den førstkomende observation som udtryk for, hvor det var forventet, at eleven var indskrevet ved skoleårets start.

løbet af et skoleår, fx skoleskift efter realisering af klassens størrelse. Dette er en forbedring af flere eksisterende studier, hvor både instrument og faktisk klassestørrelse er defineret samtidig, typisk enten i slutningen eller starten af skoleåret.

Klassestørrelsen i 0. klasse måles med mindre præcision end på de øvrige klassetrin, idet folkeskoleloven foreskriver, at skolerne kan vente med at danne og registrere den endelige klasseopdeling indtil slutningen af skoleåret i 0. klasse. Dermed registreres en del af 0. klasserne i DST's registre som indeholdende hele skolens 0. årgang, se Danmarks Statistik (2020). Det betyder både, at instrumentet for 0. klassetrin er svagere end for de øvrige klassetrin men også, at klassestørrelsesvariablen i sig selv er mindre retvisende, og således bør effektestimaterne i 0. klasse fortolkes med forbehold.

5.2 Antal årgange, der kan måles udfaldsmål for

Det tilgængelige data gør det muligt at undersøge op til fem årgange pr. klassetrin, afhængigt af udfaldsmål og tidshorisont (se Tabel 5.1).

Antallet af årgange, der kan indgå i en estimation, er vigtigt for den statistiske styrke bag estimaterne. REFUD har en tidshorisont på op til 4 år. Der vil derfor være udfaldsmål, der kan estimeres for 1) alle fire år, 2) kun for de første år eller 3) kun for de sidste år afhængigt af udfaldsmålets karakter, elevernes klassetrin/alder og datatilgængelighed.

Tabel 5.1 illustrerer fx, at udfaldsmål vedrørende grundskole og specialundervisning i år 1 er tilgængelig for 5 årgange, mens kun 3 årgange kan følges frem til år 4. Tabellen viser også, at flere udfaldsmål kun er relevante for de ældste elever i 8.-9. klasse, selvom der anvendes en tidshorisont på 4 år.

Tabellens sidste kolonne summerer det totale antal hypoteser testet, hvilket også er vigtigt at have in mente i forhold til vurdering af signifikans.

Tabel 5.1 Antal årgange, hvor udfaldsmål kan måles, for hvert klassetrin og tidshorisont

Hoveddimension	År	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.	Antal test
Grundskole og specialundervisning, 0.-9. klasse (elev) (5 udfaldsmål)	1	5	5	5	5	5	5	5	5	5	5	170
	2	5	5	5	5	5	5	5	5	5		
	3	4	4	4	4	4	4	4	4			
	4	3	3	3	3	3	3	3				
Grundskole og specialundervisning, 10. klasse (elev) (2 udfaldsmål)	1										5	8
	2										5	
	3										4	
	4										3	
Ungdomsuddannelse (elev) (3 udfaldsmål)	1										5	30
	2									5	5	
	3								4	4	4	
	4							3	3	3	3	
Sociale foranstaltninger til børn og unge under 18 år (5 udfaldsmål)	1	5	5	5	5	5	5	5	5	5	5	195
	2	4	4	4	4	4	4	4	4	4	4	
	3	3	3	3	3	3	3	3	3	3	3	
	4	2	2	2	2	2	2	2	2	2		
Efterværn (til 18-22-årige) (4 udfaldsmål)	1											24
	2										4	
	3									3	3	
	4								2	2	2	
Sundhedsydelse, fra sygesikringsregisteret (elev) (2 udfaldsmål)	1	5	5	5	5	5	5	5	5	5	5	80
	2	5	5	5	5	5	5	5	5	5	5	
	3	4	4	4	4	4	4	4	4	4	4	
	4	3	3	3	3	3	3	3	3	3	3	
Sundhedsydelse, fra Landpatientregisteret (elev) (6 udfaldsmål)	1	4	4	4	4	4	4	4	4	4	4	240
	2	3	3	3	3	3	3	3	3	3	3	
	3	2	2	2	2	2	2	2	2	2	2	
	4	1	1	1	1	1	1	1	1	1	1	
Politi, retsvæsen og kriminalforsorg (elev) (4 udfaldsmål)	1									5	5	56
	2								5	5	5	
	3							4	4	4	4	
	4						3	3	3	3	3	
Indkomstoverførsler (elev) (4 udfaldsmål)	1											24
	2										5	
	3									4	4	
	4								3	3	3	
Beskæftigelsesindsats (elev) (1 udfaldsmål)	1											6
	2										5	
	3									4	4	
	4								3	3	3	
Beskæftigelse (elev) (1 udfaldsmål)	1						5	5	5	5	5	26
	2					5	5	5	5	5	5	
	3				4	4	4	4	4	4	4	
	4			3	3	3	3	3	3	3	3	
Herberg og forsorgshjem (§ 110) (elev) (1 udfaldsmål)	1											3
	2											
	3										4	

Hoveddimension	År	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.	Antal test
Stofmisbrugsbehandling (elev) (2 udfaldsmål)	4									3	3	12
	1											
	2										5	
	3									4	4	
	4								3	3	3	
Alkoholmisbrugsbehandling (elev) (1 udfaldsmål)	1											0
	2											
	3											
	4											
Indkomstoverførsler, beskæftigelsesindsats og beskæftigelse (mødre) (6 udfaldsmål)	1	5	5	5	5	5	5	5	5	5	5	240
	2	5	5	5	5	5	5	5	5	5	5	
	3	4	4	4	4	4	4	4	4	4	4	
	4	3	3	3	3	3	3	3	3	3	3	
Indkomstoverførsler, beskæftigelsesindsats og beskæftigelse (fædre) (6 udfaldsmål)	1	5	5	5	5	5	5	5	5	5	5	240
	2	5	5	5	5	5	5	5	5	5	5	
	3	4	4	4	4	4	4	4	4	4	4	
	4	3	3	3	3	3	3	3	3	3	3	
Personalefravær (1 udfaldsmål)	1	5	5	5	5	5	5	5	5	5	5	37
	2	4	4	4	4	4	4	4	4	4	4	
	3	3	3	3	3	3	3	3	3	3		
	4	2	2	2	2	2	2	2	2			

Anm.: Denne tabel opgør antal årgange, der kan estimere et givent udfaldsmål, jf. på forhånd fastlagte kriterier fra REFUD baseret på elevens klassetrin og alder. Sidste kolonne angiver det totale antal hypotesetests, der foretages givet det tilgængelige data (antal klassetrin * antal udfaldsmål * antal år). Der er ikke taget højde for, i hvilket omfang eleverne har en værdi for det specifikke udfaldsmål – alene om det er muligt, at de kan have det, givet deres alder.

Litteratur

- Andersen, S. C., Beuchert, L., Nielsen, H. S. & Thomsen, M. K. (2020). The Effect of Teacher's Aides in the Classroom: Evidence from a Randomized Trial. *Journal of the European Economic Association*, 18(1), 469-505.
- Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Pre-school, and Early Training Projects. *Journal of the American Statistical Association*, 103(484), 1481-1495.
- Angrist, J. D. & Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics*, 114(2), 533-575.
- Angrist, J. D., Lavy, V., Leder-Luis, J. & Shany, A. (2019). Maimonides' Rule Redux. *American Economic Review: Insights*, 1(3), 309-324.
- Beuchert, L. (under udgivelse). *Litteraturgennemgang af effekter af klassestørrelse*. Forventes udgivet i 2022. VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.
- Beuchert, L. & Nandrup, A. (under udgivelse). *Class size, Well-being and Early School Life*. Forventes udgivet i 2022. VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.
- Bingley, P., Jensen, V. M. & Walker, I. (2005). *The Effects of School Class Size on Length of Post-Compulsory Education: Some Cost-Benefit Analysis*. IZA Discussion Paper No. 1605.
- Browning, M. & Heinesen, E. (2007). Class Size, Teacher Hours and Educational Attainment. *The Scandinavian Journal of Economics*, 109(2), 415-438.
- Calonico, S., Cattaneo, M. D., Farrell, M. H. & Titiunik, R. (2017). rdrobust: Software for Regression Discontinuity Designs. *The Stata Journal*, 17(2), 372-404.
- Cattaneo, M., Idrobo, N. & Titiunik, R. (2018a). A Practical Introduction to Regression Discontinuity Designs: Volume I. *Cambridge Elements: Quantitative and Computational Methods for the Social Science*. Cambridge University Press.

- Cattaneo, M., Idrobo, N. & Titiunik, R. (2018b). A Practical Introduction to Regression Discontinuity Designs: Volume II. *Cambridge Elements: Quantitative and Computational Methods for the Social Science*. Cambridge University Press.
- Cattaneo, M. D., Jansson, M. & Ma, X. (2018). Manipulation Testing Based on Density Discontinuity. *The Stata Journal*, 18(1), 234-261.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W. & Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics*, 126(4), 1593-1660.
- Connolly, M. & Haeck, C. (2022). Nonlinear Class Size Effects on Cognitive and Noncognitive Development of Young Children. *Journal of Labor Economics*, 40(51), 341-382.
- Danmarks Statistik. (2020). *Store klasser*. Juni 2020.
- Dobbelsteen, S., Levin, J. D. & Oosterbeek, H. (2002). The Causal Effect of Class Size on Scholastic Achievement: Distinguishing the Pure Class Size Effect from the Effect of Changes in Class Composition. *Oxford Bulletin of Economics and Statistics*, 64(1), 17-38.
- Dynarski, S., Hyman, J. & Schanzenbach, D. W. (2013). Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion. *Journal of Policy Analysis and Management*, 32(4), 692-717.
- Filges, T., Sonne-Schmidt, C. S. & Nielsen, B. C. V. (2018). Small class sizes for improving student achievement in primary and secondary schools: a systematic review. *Campbell Systematic Reviews*, 14(1), 1-107.
- Frandsen, B. (2017). Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design When the Running Variable is Discrete. *Regression Discontinuity Designs: Theory and Applications* (Advances in Econometrics, vol. 38), M. D. Cattaneo & J. C. Escanciano (red.). Emerald Group Publishing, 281-315.
- Fredriksson, P., Öckert, B. & Oosterbeek, H. (2013). Long-term effects of class size. *The Quarterly Journal of Economics*, 128(1), 249-285.
- Fredriksson, P., Öckert, B. & Oosterbeek, H. (2016). Parental Responses to Public Investments in Children: Evidence from a Maximum Class Size Rule. *The Journal of Human Resources*, 51(4), 832-868.

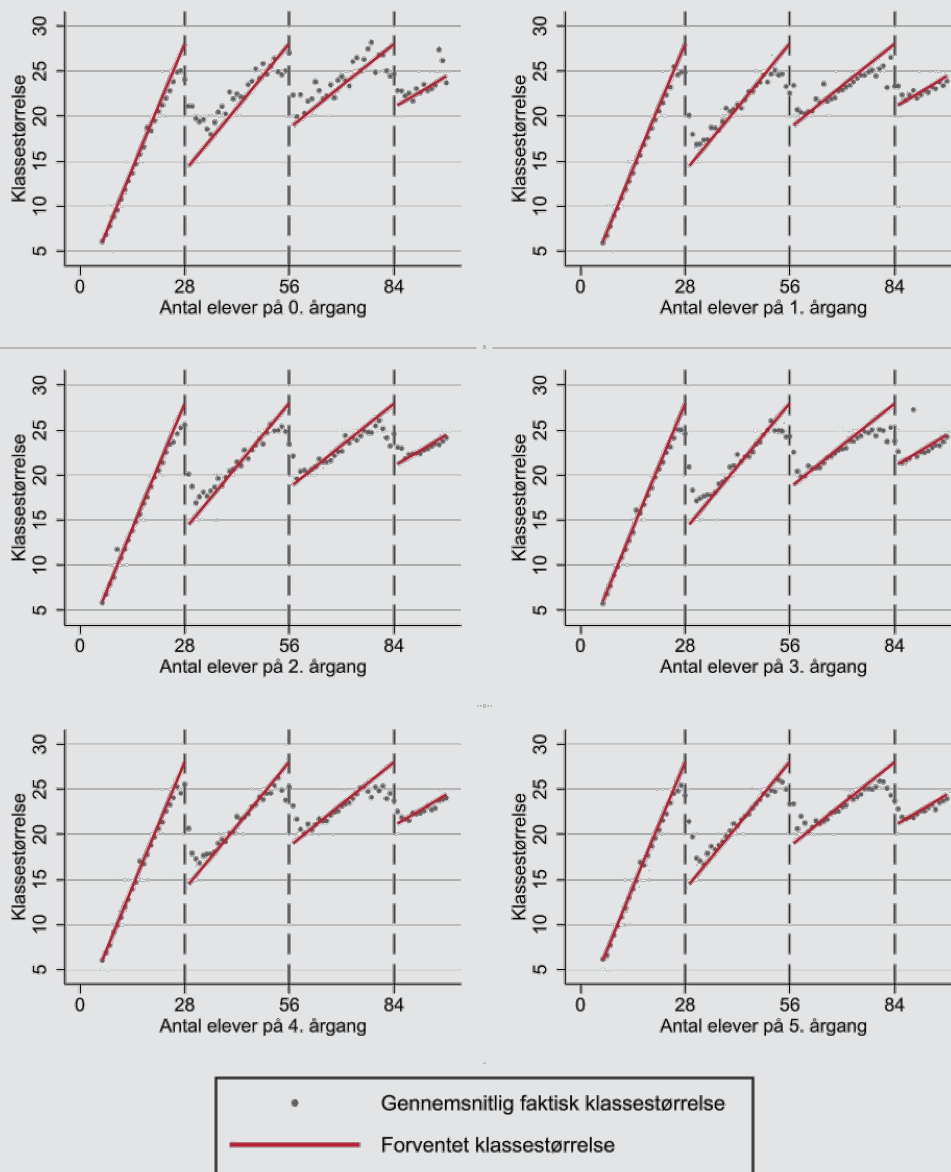
- Gerritsen, S., Plug, E. & Webbink, D. (2017). Teacher Quality and Student Achievement: Evidence from a Sample of Dutch Twins. *Journal of Applied Econometrics*, 32(3), 643-660.
- Heinesen, E. (2010). Estimating Class-size Effects using Within-school Variation in Subject-specific Classes. *The Economic Journal*, 120(545), 737-760.
- Hoxby, C. M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *The Quarterly Journal of Economics*, 115(4), 1239-1285.
- Iversen, J. & Bonesrønning, H. (2011). Disadvantaged students in the early grades: will smaller classes help them? *Education Economics*, 21(4), 1-20.
- Jepsen, C. & Rivkin, S. (2009). Class Size Reduction and Student Achievement. The Potential Tradeoff between Teacher Quality and Class Size. *The Journal of Human Resources*, 44(1), 223-250.
- Johansen, E. R., Mølbæk, L. B. & Nandrup, A. B. (2020). *A note on parental response to class size: Do parents shy away from large classes*. Ikke udgivet working paper. Aarhus Universitet.
- Kolesár, M. & Rothe, C. (2018). Inference in Regression Discontinuity Designs with a Discrete Running Variable. *American Economic Review*, 108(8), 2277-2304.
- Krassel, K. F., Lesner, R. V. & Nandrup, A. B. (2022a). *Udvikling af SØM på børne- og undervisningsområdet. Teknisk dokumentationsrapport*. VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.
- Krassel, K. F., Lesner, R. V., Nandrup, A. B. & Andersen, M. (2022b). *Nye indsatser og ny funktionalitet til prototypen til SØM på børne- og undervisningsområdet. Teknisk dokumentationsrapport, Del 1*. VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.
- Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, 114(2), 497-532.
- Lazear, E. P. (2001). Educational Production. *The Quarterly Journal of Economics*, 116(3), 777-803.
- Lee, D. S. & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48, 281-355.

- Leuven, E. & Løkken, S. A. (2020). Long Term Impacts of Class Size in Compulsory School. *Journal of Human Resources*, 55(1), 309-348.
- Leuven, E., Oosterbeek, H. & Rønning, M. (2008). Quasi-experimental Estimates of the Effect of Class Size on Achievement in Norway. *The Scandinavian Journal of Economics*, 110(4), 663-693.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2), 698-714.
- Mueller, S. (2013). Teacher experience and the class size effect — Experimental evidence. *Journal of Public Economics*, 98, 44-52.
- Nandrup, A. B. (2016). Do class size effects differ across grades? *Education Economics*, 24(1), 83-95.
- Pop-Eleches, C. & Urquiola, M. (2013). Going to a Better School: Effects and Behavioral Responses. *American Economic Review*, 103(4), 1289-1324.
- STIL (2020). *Metodenotat: Beregning af indikatorer i den nationale trivselsmåling i folkeskolen. 26. oktober 2020*. Ministeriet for Børn, Undervisning og Ligestilling, Styrelsen for IT og Læring.
- Stock J. & Yogo M. (2005). Testing for Weak Instruments in Linear IV Regression. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, D. W. K. Andrews & J. H. Stock (red.). Cambridge University Press, 80-108.

Bilag 1 Figurer

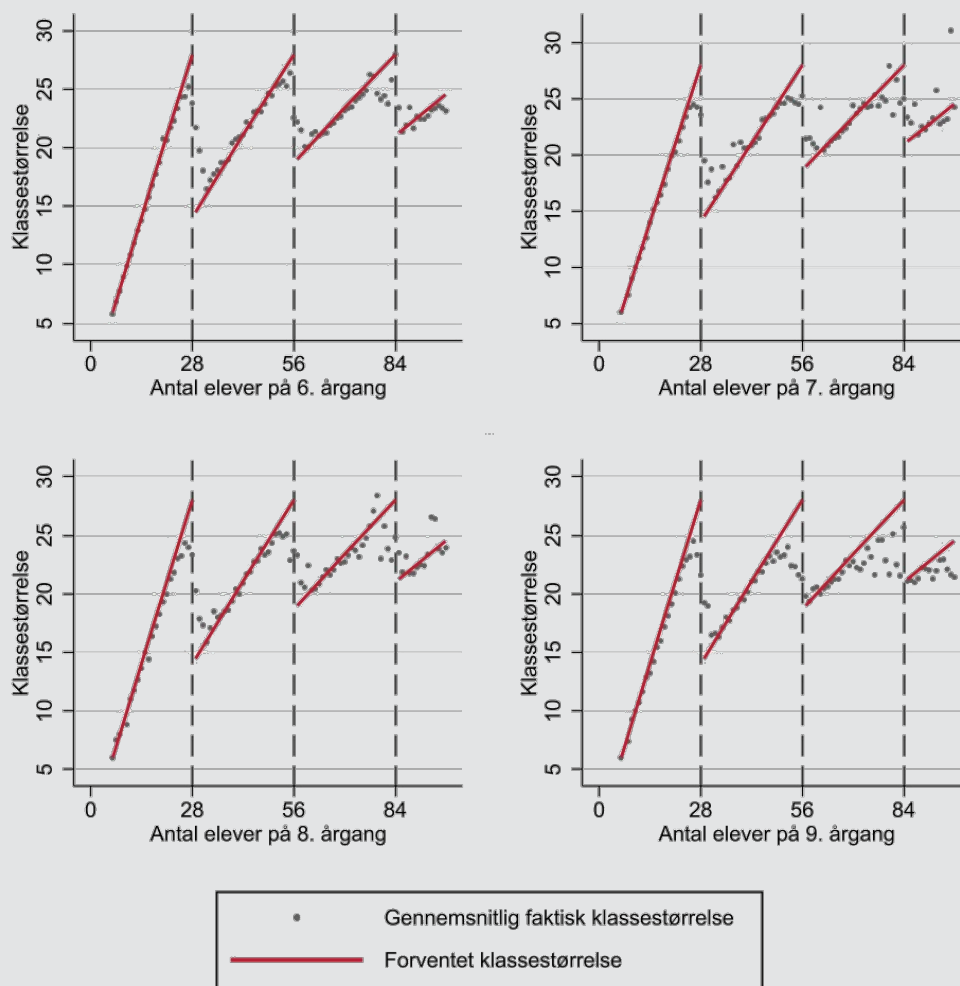
Bilagsfigur 1.1 Illustration af RD-designet, 0.-5. årgang

Figurerne viser, hvordan klasseloftsreglen danner tre tærskelværdier, hvor elever lige under tærsklen vil gå i relativt større klasser, end eleverne lige over tærsklen.



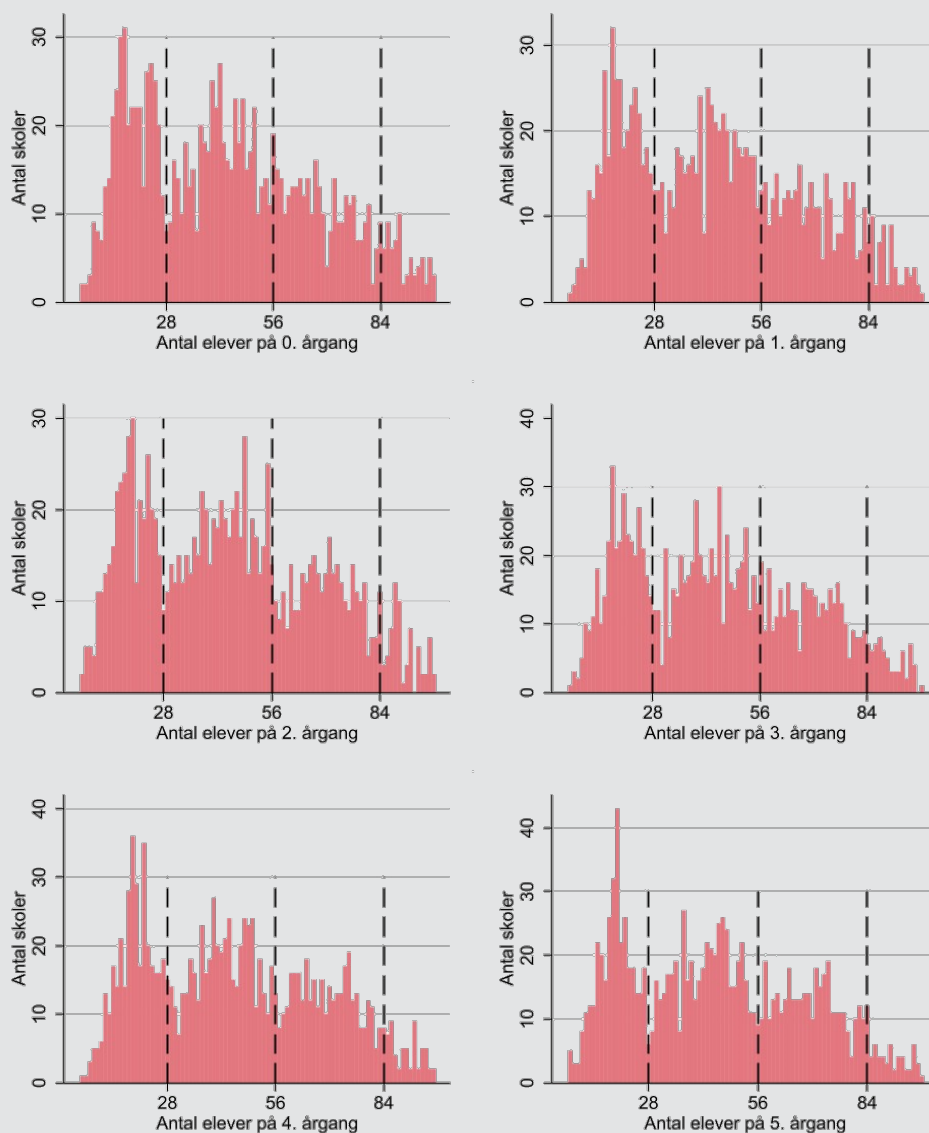
Bilagsfigur 1.2 Illustration af RD-designet, 6.-9. årgang

Figurerne viser, hvordan klasseloftsreglen danner tre tærskelværdier, hvor elever lige under tærsklen vil gå i relativt større klasser end eleverne lige over tærsklen.



Bilagsfigur 1.3 Årgangsstørrelse, 0.-5. årgang

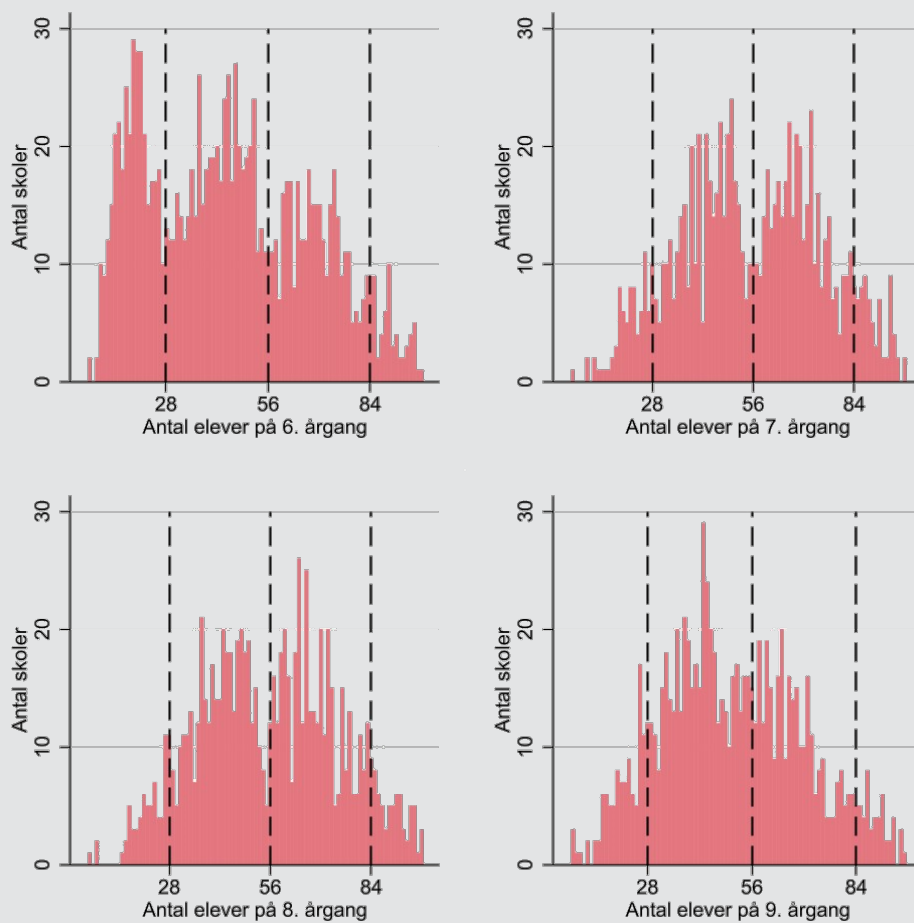
Histogrammerne angiver fordelingen af årgangsstørrelse for skoleårgangene i estimations-sampen.



Anm.: Den stiplede linje går igennem den bin, som er hhv. 28, 56 og 84 elever på årgangen. Årgangsstørrelser over 98 elever ekskluderes.

Bilagsfigur 1.4 Årgangsstørrelse, 6.-9. årgang

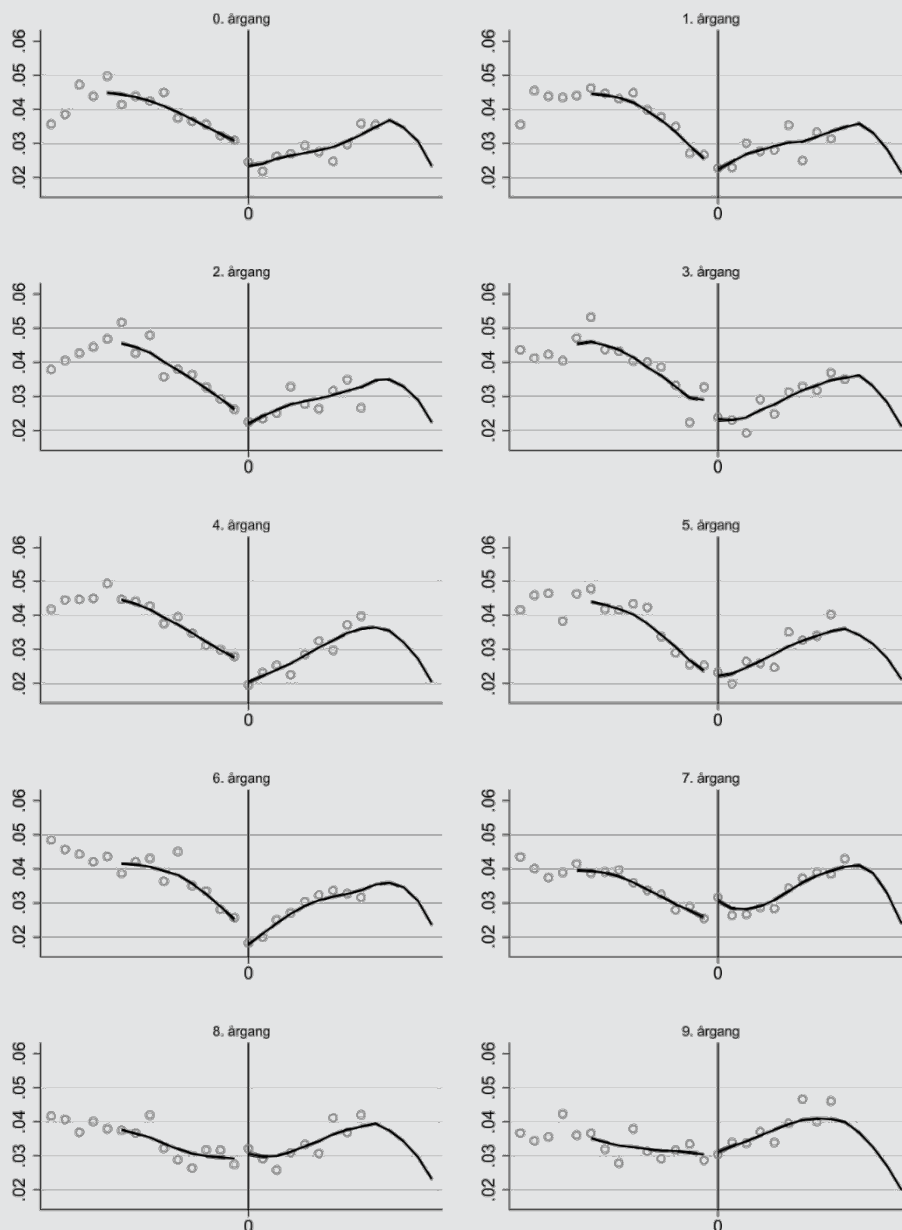
Histogrammerne angiver fordelingen af årgangsstørrelse for skoleårgangene i estimations-samplen.



Anm.: Den stiplede linje går igennem den bin, som er hhv. 28, 56 og 84 elever på årgangen. Årgangsstørrelser over 98 elever ekskluderes.

Bilagsfigur 1.5 McCrary density test af årgangsstørrelse: 0.-9. årgang

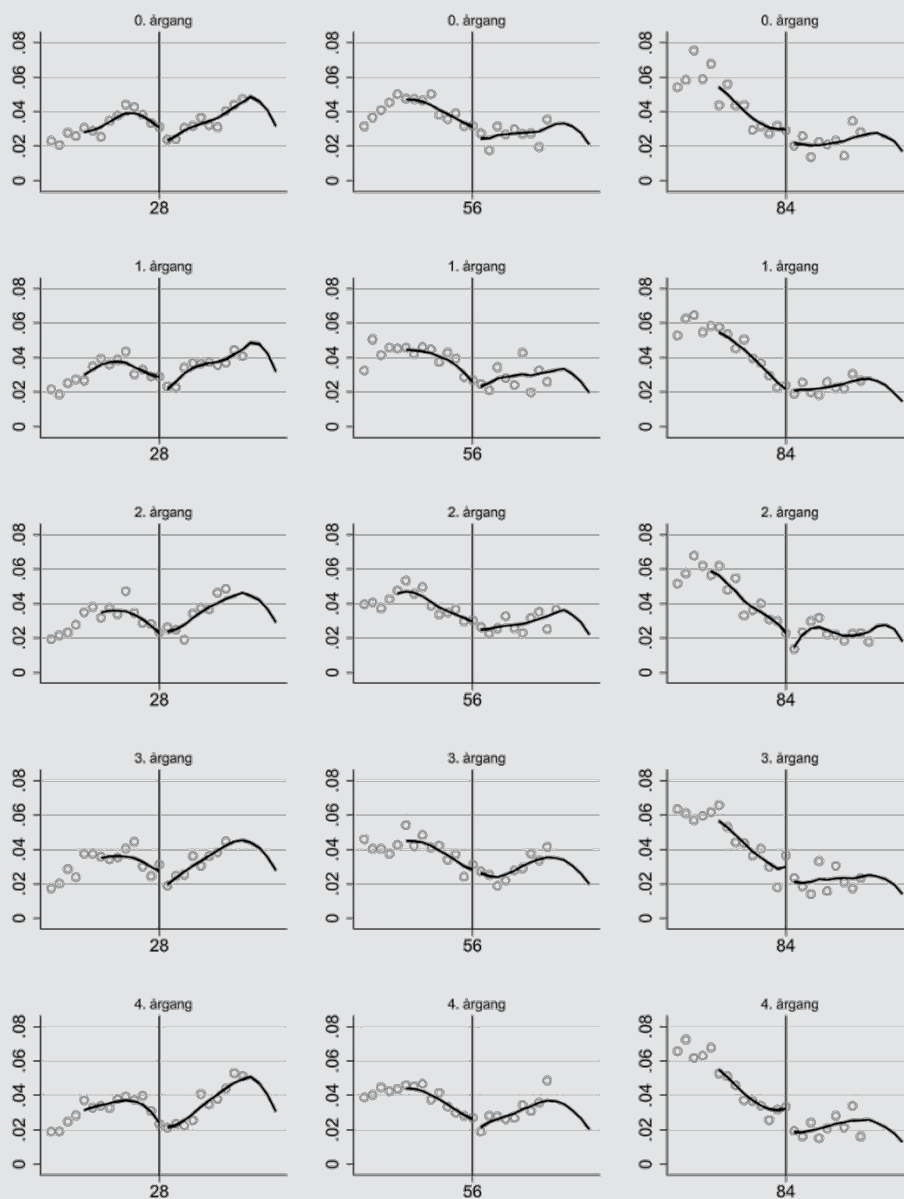
Figurerne viser densiteten af årgangsstørrelser over og under klasseloftstærsklerne for det samlede estimationsdata (med alle tre tærskelværdier centreret i nul).



Anm.: Figurerne er lavet ved hjælp af Statakommandoen `dcdensity`, hvor `binsize` er lig 1, og kommandoen vælger optimal bandwidth. Data indeholder alle år.

Bilagsfigur 1.6 McCrary density test af årgangsstørrelse, 0.-4. årgang, separat for klasseloftstærskler

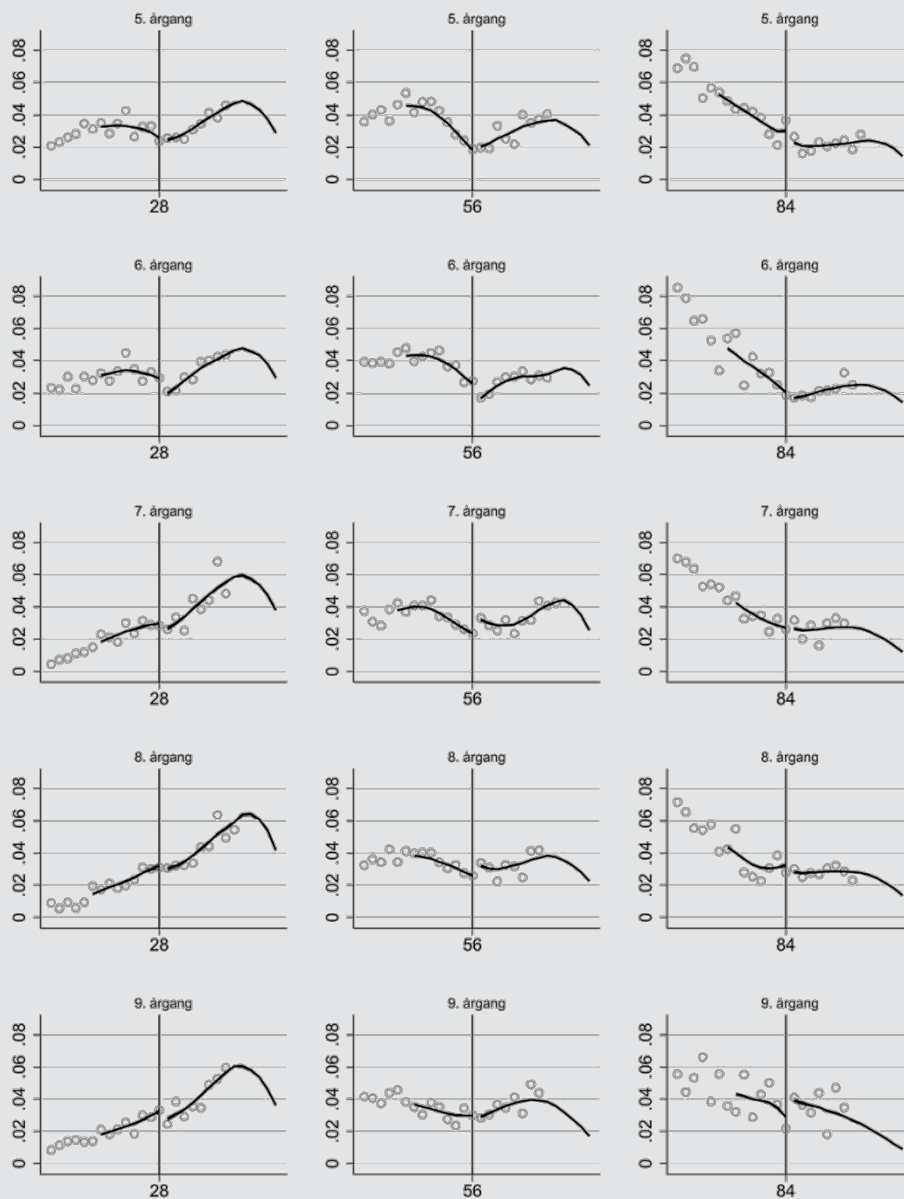
Figurerne viser densiteten af årgangsstørrelser over og under hver af de tre tærskelværdier: årgangsstørrelser på hhv. 28, 56 og 84.



Anm.: Figurerne er lavet ved hjælp af Statakommandoen `dcdensity`, hvor `binsize` er lig 1, og kommandoen vælger optimal bandwidth. Data indeholder alle år.

Bilagsfigur 1.7 McCrary density test af årgangsstørrelse, 5.-9. årgang, separat for klasseloftstærskler

Figurerne viser densiteten af årgangsstørrelser over og under hver af de tre tærskelværdier: årgangsstørrelser på hhv. 28, 56 og 84.



Anm.: Figurerne er lavet ved hjælp af Statakommandoen `dcdensity`, hvor `binsize` er lig 1, og kommandoen vælger optimal `bandwidth`. Indeholder alle år.

Bilag 2 Tabeller

Bilagstabel 2.1 Overblik signifikante (positive eller negative) estimater, default udfaldsmål

Udfaldsmål	År	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
Udfaldsmål 1: Specialskole	1							+	+		
	2								+		
	3										
	4										
Udfaldsmål 2: Kommunal grundskole, 0.-9. almenklasse	1		%			%				+	
	2					%				+	
	3				%	%					
	4			%							
Udfaldsmål 3: Kommunal grundskole, 10. almenklasse	1										
	2										
	3										
	4									%	
Udfaldsmål 4: Fri- og privat grundskole, 0.-9. almenklasse	1				+	+					
	2		+		+						
	3				+	+					
	4										
Udfaldsmål 5: Fri- og privat grundskole, 10. almenklasse	1										
	2										
	3										
	4										
Udfaldsmål 6: Specialundervisning i almenklasse	1								%		
	2	+		+							
	3										
	4										
Udfaldsmål 7: Specialklasse	1										
	2									+	
	3										
	4							+			
Udfaldsmål 8: Forberedende uddannelses tilbud (ekskl. STU)	1										
	2										
	3										
	4										
Udfaldsmål 9: Særligt tilrettelagt uddannelse (STU)	1										
	2										
	3										
	4										
Udfaldsmål 10: Ordinær ungdomsuddannelse	1										%
	2										%
	3										

Udfaldsmål	År	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
Udfaldsmål 20: Praktiserende læge	4										+
	1									%	
	2										
	3				%					%	
	4							%			%
Udfaldsmål 21: Øvrige behandlere under sygesikringen	1										
	2			+						%	
	3								%	%	
	4								%		%
Udfaldsmål 22: Somatisk indlæggelse	1						+				
	2					+	+				
	3	+					+				
	4										
Udfaldsmål 23: Somatisk ambulansbesøg	1										
	2			%			+				
	3										
	4										
Udfaldsmål 24: Somatisk skadestuebesøg	1	+					+				
	2	+			+				%		
	3		+				+		+		%
	4		+	%							
Udfaldsmål 25: Psykiatrisk indlæggelse	1										
	2										
	3										
	4										
Udfaldsmål 26: Psykiatrisk ambulansbesøg	1										
	2							+			
	3										
	4	+									
Udfaldsmål 27: Psykiatrisk skadestuebesøg	1										
	2										
	3							+			
	4										
Udfaldsmål 28: Fængselsdage	1										
	2										
	3										
	4										
Udfaldsmål 29: Domme for vold og sædelighedsforbrydelser	1										
	2										
	3										
	4										
Udfaldsmål 30: Domme for indbrud, tyveri og hærværk	1										
	2										
	3										
	4										
	1										%

Udfaldsmål	År	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
Udfaldsmål 31: Domme for overtrædelse af færdselsloven og øvrige særlove	2										
	3										
	4										
Udfaldsmål 37: Lønindkomst (elev)	1										
	2										
	3					+					
	4				+						
Udfaldsmål 47: Lønindkomst (mødre)	1		+								%
	2		+								%
	3		+								%
	4		+								%
Udfaldsmål 53: Lønindkomst (fædre)	1		+								%
	2										%
	3										%
	4					%					%
Udfaldsmål 54: Lærernes sygefravær	1		%	%	%	%	%	%	%	%	%
	2	%	%	%	%		%				
	3		%	%			%				
	4		%	%							

Anm.: Signifikante estimater samt retning på effekttestimat. Medtager kun signifikante estimater hvor $n \geq 100$. Alle udfaldsmål er standardiserede.

Note: + = signifikant positivt estimat på 10 %-signifikansniveau, hvis fed = 5 %-signifikansniveau.
% = signifikant negativt estimat på 10 %-signifikansniveau, hvis fed = 5 %-signifikansniveau.

Bilagstabel 2.2 Overblik over signifikante estimater på ikke-default udfaldsmål

Udfaldsmål	År	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
Udfaldsmål 12: Hjemmebase-rede foranstaltninger	1										
	2										
	3										
	4	%				%					
Udfaldsmål 13: Anbringelse i slægts- eller netværkspleje	1								+		
	2										
	3										
	4										
Udfaldsmål 14: Anbringelse i familiepleje	1										
	2										
	3						%				
	4			+		+					
Udfaldsmål 42: Kontanthjælp (mødre)	1										
	2								+	%	
	3	%						%	+	%	
	4	%									
Udfaldsmål 43: Førtidspension (mødre)	1				%				%		
	2			%					%		

Udfaldsmål	År	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
	3			%							
	4										
Udfaldsmål 44: Uddannelseshjælp (mødre)	1										
	2										
	3					%					
	4										
Udfaldsmål 45: Vejledning og opkvalificering (mødre)	1										
	2							%			
	3	%	+					%			
	4										
Udfaldsmål 46: Øvrig overførsler (ekskl. SU) (mødre)	1	%									
	2										
	3				%						
	4										
Udfaldsmål 48: Kontanthjælp (fædre)	1										
	2										
	3					%					
	4										
Udfaldsmål 49: Førtidspension (fædre)	1				%	%					
	2				%						
	3	%		%							
	4	%									
Udfaldsmål 50: Uddannelseshjælp (fædre)	1							%			
	2										
	3										
	4										
Udfaldsmål 51: Vejledning og opkvalificering (fædre)	1	%		%							
	2	%					+	%			
	3	%	+					%	+		
	4										
Udfaldsmål 52: Øvrig overførsler (ekskl. SU) (fædre)	1					%					
	2	%					%				
	3	%									
	4	%									

Anm.: Signifikante estimater samt retning på effektestimater. Medtager kun signifikante estimater hvor $n \geq 100$. Alle udfaldsmål er standardiserede.

Note: + = signifikant positivt estimat på 10 %-signifikansniveau, hvis fed = 5 %-signifikansniveau.
% = signifikant negativt estimat på 10 %-signifikansniveau, hvis fed = 5 %-signifikansniveau.

Der er ikke fundet signifikante estimater på følgende ikke-default udfaldsmål (og de er derfor udeladt fra oversigten i Bilagstabel 2.2):

- SØM-konsekvens 11 om modtagelse af forebyggende indsatser
- SØM-konsekvens 15 om anbringelse på institution
- SØM-konsekvens 16-19 om forbrug af efterværn
- SØM-konsekvens 32-36 om elevens senere forbrug af indkomstoverførsler (kontanthjælp, førtidspension, uddannelseshjælp og andet) og beskæftigelsesindsats (Vejledning og opkvalificering)
- SØM-konsekvens 38-41 om elevens senere forbrug af sociale services (herberg og stofmisbrugs- og alkoholbehandling).

Bilagstabel 2.3 Illustration af compliance med klasseloftsreglen

Faktisk klassestørrelse	Antal elever	Andel af elever i sample (%)	Antal klasser	Andel af klasser i sample (%)
0. årgang				
Mindre end 14 elever	7.526	3,4	1.074	10
14-28 elever	201.636	91	9.328	87,1
Mere end 28 elever	12.345	5,6	305	2,8
1. årgang				
Mindre end 14 elever	5.735	2,5	759	7
14-28 elever	215.301	95,5	9992	91,9
Mere end 28 elever	4.443	2	124	1,1
2. årgang				
Mindre end 14 elever	4.958	2,1	699	6,3
14-28 elever	222.102	96	10.258	92,6
Mere end 28 elever	4.416	1,9	124	1,1
3. årgang				
Mindre end 14 elever	5.101	2,1	717	6,1
14-28 elever	235.261	96,2	10.835	92,8
Mere end 28 elever	4.310	1,8	122	1
4. årgang				
Mindre end 14 elever	5.003	2	700	5,9
14-28 elever	237.984	96,2	10.957	93
Mere end 28 elever	4.454	1,8	128	1,1
5. årgang				
Mindre end 14 elever	5.098	2,1	714	6,1
14-28 elever	236.812	96	10.899	92,8
Mere end 28 elever	4.734	1,9	135	1,1
6. årgang				
Mindre end 14 elever	5.519	2,3	787	6,7

Faktisk klassestørrelse	Antal elever	Andel af elever i sample (%)	Antal klasser	Andel af klasser i sample (%)
14-28 elever	233.464	95,5	10.770	91,9
Mere end 28 elever	5.573	2,3	159	1,4
7. årgang				
Mindre end 14 elever	2.074	1	422	4,2
14-28 elever	208.246	96,3	9.499	94,3
Mere end 28 elever	5.984	2,8	153	1,5
8. årgang				
Mindre end 14 elever	2.301	1,1	500	5
14-28 elever	203.283	96,4	9.376	93,6
Mere end 28 elever	5.288	2,5	136	1,4
9. årgang				
Mindre end 14 elever	4.242	2,2	698	7
14-28 elever	188.672	96,2	9.129	92,1
Mere end 28 elever	3.173	1,6	82	0,8

Anm.: For 0. årgang er 5,6 % af eleverne (2,8 % af klasserne) i klasser med over 28 elever; kan opdeles på 1,2 % (0,8 %) går i en klasse med 29-30 elever og 4,4 % (2,0 %) går i en klasse med mere end 30 elever.

Bilagstabel 2.4 F-test for first stage

Konsekvens	År	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
Grundskole og specialundervisning 0.-9. klasse (elev)	1	46,2	104,9	107,6	120,7	147,6	125,7	142,0	100,5	79,1	73,0
	2	46,2	104,9	107,6	120,7	147,6	125,7	142,0	100,5	79,1	
	3	32,7	68,5	78,1	86,9	114,7	90,2	105,1	71,8		
	4	18,1	40,9	50,1	53,8	86,6	66,8	71,1			
Grundskole og specialundervisning 10. klasse (elev)	1										73,0
	2									79,1	73,0
	3								71,8	61,8	65,0
	4							71,1	52,2	40,1	44,7
Ungdomsuddannelse (elev)	1										73,0
	2									79,1	73,0
	3								71,8	61,8	65,0
	4							71,1	52,2	40,1	44,7
Sociale foranstaltninger til børn og unge under 18 år	1	46,2	104,9	107,6	120,7	147,6	125,7	142,0	100,5	79,1	73,0
	2	32,7	68,5	78,1	86,9	114,7	90,2	105,1	71,8	61,8	65,0
	3	18,1	40,9	50,1	53,8	86,6	66,8	71,1	52,2	40,1	44,7
	4	10,1	20,7	30,6	36,6	57,7	41,5	40,3	23,7	15,3	
Efterværn (til 18-22-årige)	1										
	2										65,0
	3									40,1	44,7
	4								23,7	15,3	26,6
Sundhedsydelse, egen læge (elev)	1	46,2	104,9	107,6	120,7	147,6	125,7	142,0	100,5	79,1	73,0
	2	46,2	104,9	107,6	120,7	147,6	125,7	142,0	100,5	79,1	73,0
	3	32,7	68,5	78,1	86,9	114,7	90,2	105,1	71,8	61,8	65,0
	4	18,1	40,9	50,1	53,8	86,6	66,8	71,1	52,2	40,1	44,7
Sundhedsydelse, hospital (elev)	1	32,7	68,5	78,1	86,9	114,7	90,2	105,1	71,8	61,8	65,0
	2	18,1	40,9	50,1	53,8	86,6	66,8	71,1	52,2	40,1	44,7
	3	10,1	20,7	30,6	36,6	57,7	41,5	40,3	23,7	15,3	26,6
	4	7,6	15,2	14,6	23,9	30,3	22,5	17,5	9,8	29,2	13,1

Konsekvens	År	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
Politi, retsvæsen og kriminalforsorg (elev)	1								100,5	79,1	73,0
	2							142,0	100,5	79,1	73,0
	3						90,2	105,1	71,8	61,8	65,0
	4					86,6	66,8	71,1	52,2	40,1	44,7
Indkomstoverførsler (elev)	1										
	2										73,0
	3									61,8	65,0
	4								52,2	40,1	44,7
Beskæftigelsesindsats (elev)	1										
	2										73,0
	3									61,8	65,0
	4								52,2	40,1	44,7
Beskæftigelse (elev)	1						125,7	142,0	100,5	79,1	73,0
	2					147,6	125,7	142,0	100,5	79,1	73,0
	3				86,9	114,7	90,2	105,1	71,8	61,8	65,0
	4			50,1	53,8	86,6	66,8	71,1	52,2	40,1	44,7
Herberg og forsorgshjem (§ 110) (elev)	1										
	2										73,0
	3									61,8	65,0
	4								52,2	40,1	44,7
Stofmisbrugsbehandling (elev)	1										
	2										44,9
	3									15,3	26,8
	4								9,9	29,4	13,2
Alkoholmisbrugsbehandling (elev)	1										
	2										44,7
	3									15,3	26,6
	4								9,8	29,2	13,1
Indkomstoverførsler, beskæftigelsesindsats og beskæftigelse (mødre)	1	46,2	105,4	107,6	120,6	147,4	126,1	142,0	100,8	79,9	73,2
	2	46,2	105,4	107,6	120,6	147,4	126,1	142,0	100,8	79,9	73,2
	3	32,7	68,8	78,3	86,8	114,7	90,6	105,2	71,8	62,4	65,2
	4	18,1	41,1	50,0	53,7	86,7	67,1	71,2	52,1	40,5	45,0
Indkomstoverførsler, beskæftigelsesindsats og beskæftigelse (fædre)	1	45,8	103,9	107,3	120,5	148,0	126,5	140,7	102,3	82,1	72,4
	2	45,8	103,9	107,3	120,5	148,0	126,5	140,7	102,3	82,1	72,4
	3	32,5	67,9	77,9	86,7	114,8	91,4	103,8	73,3	63,6	64,6
	4	17,9	40,6	50,0	54,0	86,9	68,1	69,8	53,7	41,7	44,5
Personalefravær	1	50,1	163,0	119,0	135,7	149,0	135,2	148,7	121,2	86,0	90,7
	2	35,1	70,7	81,9	90,2	123,8	100,8	108,0	77,6	63,5	65,0
	3	18,9	40,8	54,1	53,2	92,8	68,9	73,3	58,3	40,2	44,7
	4	9,6	20,9	31,0	39,0	57,6	44,6	40,5	23,8	15,3	26,6

Anm.: First stage F-tests for fælles signifikans af modellens instrumenter (indikatorer for årgangsstørrelse over klasseloftstærsker) i first stage-regressionerne tilhørende estimerne i Tabel 2.2-2.5.

Bilagstabel 2.5 Balancetest for hvert klasseøftstærskel: 1. klassetrin

Baggrundsvariabel	Data opdelt på nærmeste tærskelværdi												Data samlet			
	Omkring tærskel 28				Omkring tærskel 56				Omkring tærskel 84				Omkring nul			
	Hop	P-værdi	Bw, V	Bw, H	Hop	P-værdi	Bw, V	Bw, H	Hop	P-værdi	Bw, V	Bw, H	Hop	P-værdi	Bw, V	Bw, H
Alder	0,00	0,03	5,95	10,47	-0,01	0,29	8,66	8,86	0,00	0,70	19,34	10,17	-0,02	0,15	3,96	3,88
Kvadreret alder	0,00	0,02	6,04	10,47	0,00	0,31	8,41	8,84	0,00	0,79	16,85	10,12	0,00	0,15	3,98	3,89
Dreng	0,00	0,15	6,05	18,45	0,00	0,63	10,61	8,66	0,00	0,43	15,04	9,89	0,00	0,51	3,43	4,00
Ikkevestlig oprindelse	0,00	0,35	9,45	17,24	0,00	0,64	10,78	10,70	-0,01	0,11	20,79	11,70	-0,02	0,06	4,55	4,77
Placering i mors fødselsrækkefølge	-0,01	0,16	7,82	10,63	-0,01	0,33	9,51	9,13	-0,01	0,25	17,70	11,44	-0,05	0,17	3,86	3,91
Flerfødsfødsel	0,00	0,91	8,55	13,21	0,00	0,56	14,75	12,43	0,00	0,75	15,49	11,38	0,00	0,87	4,13	3,30
Mors uddannelse: gymnasial uddannelse eller lavere	0,00	0,29	9,68	19,19	0,00	0,63	9,18	10,75	0,00	0,34	22,25	11,64	0,00	0,33	4,41	4,94
Mors uddannelse: videregående uddannelse	0,00	0,25	9,49	14,14	0,02	0,28	9,54	7,98	0,01	0,40	16,10	11,28	0,02	0,36	4,19	4,08
Mors uddannelse: erhvervsuddannelse	0,00	0,68	7,68	14,37	-0,01	0,13	9,54	10,51	0,01	0,49	12,42	13,00	0,00	0,64	3,71	4,25
Mors alder ved barnets fødsel	0,02	0,59	7,62	14,84	0,67	0,49	6,28	7,85	0,13	0,21	14,01	10,52	0,23	0,32	3,85	3,84
Mors alder ved barnets fødsel kvadreret	0,01	0,60	7,80	15,86	0,31	0,42	6,73	7,91	0,08	0,19	14,90	10,55	0,13	0,34	3,92	3,82
Fars uddannelse: gymnasial uddannelse eller lavere	0,00	0,27	6,68	17,45	0,00	0,41	8,96	10,47	0,00	0,61	23,85	11,99	0,01	0,22	3,09	5,21
Fars uddannelse: videregående uddannelse	0,00	0,05	10,33	14,26	0,03	0,19	10,06	7,70	0,01	0,24	17,35	11,22	0,02	0,30	4,24	4,16
Fars uddannelse: erhvervsuddannelse	0,00	0,76	7,31	14,44	-0,01	0,43	7,35	10,82	0,00	0,46	22,58	12,40	0,00	0,86	4,61	4,69
Fars alder ved barnets fødsel	0,01	0,93	8,40	14,74	0,33	0,53	6,62	7,57	0,12	0,45	13,64	10,25	0,28	0,33	3,40	3,67
Fars alder ved barnets fødsel kvadreret	0,00	0,77	8,42	15,63	0,11	0,65	6,66	7,73	0,06	0,57	14,20	10,13	0,05	0,69	3,53	3,87
Mor diagnosticeret med opmærksomhedsforstyrrelse	0,00	0,33	10,76	15,37	0,00	0,09	11,15	14,41	0,00	0,41	20,51	11,44	0,00	0,11	4,22	4,14
Mor diagnosticeret med angst, depression eller anden affektiv lidelse	0,00	0,65	9,29	15,61	-0,01	0,03	11,64	11,05	0,00	0,83	14,88	11,93	-0,01	0,20	4,31	5,12
Far diagnosticeret med opmærksomhedsforstyrrelse	0,00	0,66	9,33	15,74	0,00	0,53	13,32	9,17	0,00	0,29	22,20	10,17	0,00	0,59	3,77	5,70
Far diagnosticeret med angst, depression eller anden affektiv lidelse	0,00	0,51	9,16	17,96	0,00	0,40	8,31	9,90	0,00	0,09	19,52	11,99	-0,01	0,11	3,90	5,70
Antal observationer	19.980-41.280				42.168-109.052				69.626-141.666				39.116-55.782			

Anm.: RD-model for den pooled model kontrollerer for års-dummies og segment dummies. Antal observationer varierer afhængigt af den estimerede optimale bandwidth. P-værdier, der er signifikante på et 5 %-niveau, er markeret med **fed**.

Bilagstabel 2.6 Balancetest for hvert klassetrin, pooled sample

Klassetrin	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.	9. kl.
BaggrundsvARIABLE	RDD	RDD	RDD	RDD	RDD	RDD	RDD	RDD	RDD	RDD
Alder	0,86	0,15	0,92	0,88	0,09	0,64	0,87	0,92	0,43	0,61
Kvadreret alder	0,74	0,15	0,91	0,93	0,10	0,64	0,88	0,90	0,44	0,64
Dreng	1,00	0,51	0,58	0,78	0,88	0,28	0,8	0,92	0,37	0,64
Ikkevestlig oprindelse	0,38	0,06	0,08	0,48	0,29	0,49	0,74	0,34	0,38	0,52
Placering i mors fødselsrækkefølge	0,36	0,17	0,73	0,20	0,84	0,87	0,71	0,27	0,65	0,60
Flerfødsfødsel	0,25	0,87	0,27	0,65	0,16	0,60	0,30	0,97	0,53	0,80
Mors uddannelse: gymnasial uddannelse eller lavere	0,63	0,33	0,07	0,23	0,59	0,09	0,28	0,77	0,89	0,37
Mors uddannelse: videregående uddannelse	0,84	0,36	0,75	0,21	0,04	0,57	0,54	0,55	0,38	0,28
Mors uddannelse: erhvervsuddannelse	0,35	0,64	0,92	0,56	0,04	0,72	0,57	0,20	0,62	0,30
Mors alder ved barnets fødsel	0,47	0,32	0,96	0,36	0,08	0,65	0,16	0,50	0,53	0,49
Mors alder ved barnets fødsel kvadreret	0,40	0,34	0,97	0,60	0,09	0,64	0,15	0,40	0,54	0,47
Fars uddannelse: gymnasial uddannelse eller lavere	0,33	0,22	0,36	0,87	0,87	0,67	0,52	0,61	0,37	0,42
Fars uddannelse: videregående uddannelse	0,69	0,30	0,51	0,20	0,07	0,31	0,35	0,16	0,68	0,30
Fars uddannelse: erhvervsuddannelse	0,13	0,86	0,39	0,67	0,14	0,66	0,32	0,29	0,45	0,35
Fars alder ved barnets fødsel	0,85	0,33	0,39	0,57	0,22	0,33	0,13	0,41	0,50	0,18
Fars alder ved barnets fødsel kvadreret	0,89	0,69	0,82	0,89	0,24	0,52	0,13	0,72	0,70	0,20
Mor diagnosticeret med opmærksomhedsforstyrrelse	0,73	0,11	1,00	0,09	0,18	0,25	0,96	0,57	0,61	0,84
Mor diagnosticeret med angst, depression eller anden affektiv lidelse	0,10	0,20	0,09	0,21	0,21	0,24	0,68	0,72	0,38	0,75
Far diagnosticeret med opmærksomhedsforstyrrelse	0,62	0,59	0,85	0,16	0,53	0,21	0,30	0,91	0,68	0,97
Far diagnosticeret med angst, depression eller anden affektiv lidelse	0,96	0,11	0,10	0,13	0,15	0,25	0,32	0,34	0,20	0,40

Anm.: Balancetesten er foretaget med Statapakken rdrobust. Der tillades forskellig optimal bandwidth på hver side af tærskelværdien. Estimer, der er signifikante på et 5 %-niveau, er markeret med **fed**.

Bilagstabel 2.7 Test for baggrundsvARIABLES joint signifikans

BaggrundsvARIABLE	0. kl	1. kl	2. kl	3. kl	4. kl	5. kl	6. kl	7. kl	8. kl	9. kl
Alder	0,098 (0,068)	0,182 (0,142)	-0,011 (0,130)	-0,056 (0,167)	0,076 (0,168)	0,126 (0,179)	0,414** (0,186)	0,227 (0,195)	0,337 (0,205)	-0,113 (0,097)
Kvadreret alder	-0,676 (0,509)	-1,099 (0,942)	0,143 (0,767)	0,387 (0,887)	-0,292 (0,810)	-0,542 (0,787)	-1,674** (0,751)	-0,812 (0,727)	-1,172* (0,712)	0,360 (0,317)
Dreng	-0,000 (0,004)	0,002 (0,004)	0,005 (0,004)	0,001 (0,004)	0,0023 (0,003)	-0,000 (0,004)	0,004 (0,004)	-0,001 (0,004)	-0,001 (0,005)	0,003 (0,005)
Ikkevestlig oprindelse	0,0049 (0,022)	0,033* (0,018)	0,042** (0,017)	0,021 (0,017)	0,054** (0,021)	0,050** (0,025)	0,045** (0,018)	0,001 (0,020)	-0,012 (0,019)	0,010 (0,019)
Placering i mors fødselsrækkefølge	-0,002 (0,004)	0,002 (0,004)	-0,001 (0,004)	0,003 (0,004)	0,008** (0,004)	0,005 (0,004)	0,001 (0,004)	0,005 (0,004)	0,003 (0,004)	-0,001 (0,004)
Flerfødsfødsel	-0,007 (0,013)	-0,012 (0,014)	0,000 (0,014)	-0,007 (0,013)	-0,004 (0,013)	-0,004 (0,013)	0,004 (0,013)	-0,010 (0,013)	0,008 (0,013)	-0,004 (0,013)
Mors uddannelse: gymnasial uddannelse	0,022** (0,011)	0,010 (0,012)	-0,016 (0,011)	-0,010 (0,012)	-0,011 (0,011)	-0,005 (0,011)	-0,007 (0,011)	0,008 (0,011)	0,012 (0,011)	0,035*** (0,010)
Mors uddannelse: videregående uddannelse	0,0053 (0,009)	0,010 (0,009)	-0,005 (0,009)	-0,005 (0,009)	-0,012 (0,009)	0,000 (0,009)	-0,017* (0,009)	0,006 (0,009)	0,004 (0,009)	0,020** (0,009)
Mors uddannelse: erhvervsuddannelse	0,000 (0,008)	0,010 (0,008)	-0,010 (0,008)	-0,007 (0,008)	-0,006 (0,008)	0,004 (0,007)	-0,007 (0,007)	-0,005 (0,007)	-0,000 (0,008)	0,001 (0,007)
Mors alder ved barnets fødsel	-0,003 (0,005)	0,004 (0,005)	0,0014 (0,005)	-0,003 (0,005)	-0,003 (0,004)	0,006 (0,004)	0,008* (0,005)	0,000 (0,005)	-0,001 (0,005)	0,000 (0,005)
Mors alder ved barnets fødsel kvadreret	0,005 (0,007)	-0,009 (0,008)	-0,003 (0,008)	0,003 (0,007)	0,004 (0,007)	-0,012* (0,007)	-0,013* (0,008)	-0,001 (0,007)	0,002 (0,008)	0,001 (0,007)
Mor diagnosticeret med opmærksomhedsforstyrrelse	0,019 (0,023)	0,006 (0,022)	-0,021 (0,022)	0,006 (0,023)	0,013 (0,024)	-0,047* (0,024)	0,013 (0,023)	0,057** (0,027)	0,021 (0,027)	0,048* (0,028)
Mor diagnosticeret med angst, depression mv.	-0,001 (0,007)	0,006 (0,007)	0,014* (0,007)	0,002 (0,007)	0,007 (0,007)	-0,001 (0,007)	0,011 (0,007)	0,0011 (0,007)	0,004 (0,007)	-0,0053 (0,007)
Fars uddannelse: gymnasial uddannelse	0,008 (0,012)	0,012 (0,012)	-0,005 (0,012)	0,004 (0,012)	-0,013 (0,011)	-0,024** (0,012)	-0,019 (0,012)	-0,009 (0,014)	-0,010 (0,013)	0,013 (0,012)
Fars uddannelse: videregående uddannelse	0,003 (0,010)	-0,006 (0,010)	0,013 (0,010)	0,009 (0,009)	-0,002 (0,010)	-0,015 (0,011)	-0,013 (0,010)	0,011 (0,011)	0,002 (0,010)	0,019* (0,010)
Fars uddannelse: erhvervsuddannelse	-0,014* (0,007)	0,000 (0,007)	0,001 (0,007)	0,004 (0,007)	-0,003 (0,007)	-0,007 (0,007)	-0,010 (0,007)	0,018*** (0,007)	0,014** (0,007)	0,001 (0,007)
Fars alder ved barnets fødsel	-0,000 (0,003)	-0,006* (0,003)	-0,000 (0,003)	-0,000 (0,003)	-0,003 (0,003)	-0,000 (0,003)	0,002 (0,003)	0,001 (0,003)	-0,000 (0,003)	-0,000 (0,003)
Fars alder ved barnets fødsel kvadreret	0,000 (0,004)	0,008* (0,004)	-0,001 (0,004)	-0,000 (0,004)	0,002 (0,004)	-0,001 (0,004)	-0,004 (0,004)	-0,002 (0,005)	-0,001 (0,004)	0,000 (0,004)
Far diagnosticeret med opmærksomhedsforstyrrelse	0,0114 (0,020)	-0,011 (0,021)	0,004 (0,023)	0,017 (0,021)	0,014 (0,023)	0,029 (0,024)	0,063*** (0,024)	0,023 (0,026)	-0,021 (0,029)	0,013 (0,032)
Far diagnosticeret med angst, depression mv.	0,007 (0,008)	0,007 (0,008)	0,018** (0,008)	-0,002 (0,008)	0,012 (0,008)	-0,002 (0,008)	-0,013 (0,008)	-0,006 (0,008)	-0,016* (0,008)	0,012 (0,009)
Antal elever	65.275	65.729	67.016	69.254	68.902	67.227	69.071	62.524	62.663	63.226
P-værdi fra F-test H0: Alle baggrundskontroller=0	0,417	0,121	0,025	0,419	0,072	0,107	0,002	0,098	0,507	0,104

Anm.: Tabellen rapporterer estimationskoefficient (og standardfejl i parentes) fra en OLS-model, der regresserer Above-indikatoren på alle baggrundskontroller og deres missing-indikatorer, segment-indikatorer og års-indikatorer. Standardfejl er robuste og klyngekorrigerede på skoliveau. R² er under 0.05 for alle klassetrin.

Bilagstabel 2.8 Persistens i klassestørrelse i efterfølgende skoleår

	0. kl.	1. kl.	2. kl.	3. kl.	4. kl.	5. kl.	6. kl.	7. kl.	8. kl.
Klassestørrelse	0,972***	0.882***	0.863***	0.858***	0.865***	0.792***	0.470***	0.679***	0.327***
	(0,079)	(0035)	(0.040)	(0.045)	(0.039)	(0.042)	(0.061)	(0.059)	(0.063)
Konstant	-1,410	0.890	2.002*	2.755**	3.046***	3.818***	11.40***	4.653***	11.49***
	(1,778)	(1.143)	(1.059)	(1.231)	(0.883)	(0.994)	(1.668)	(1.644)	(1.616)
Obs.	63.823	64,347	65,808	68,023	67,657	66,159	67,638	60,695	60,345
FS: F-test	47,77	142,0	109,3	121,6	146,9	123,9	140,9	103,6	79,0

Anm.: F-test for first stage angiver F-testet for fælles signifikans af modellens instrumenter (indikatorer for årgangsstørrelse over klasseloftstærskler) i first stage-regressionen, dvs. om instrumenterne har relevans i forhold til at forklare den observerede klassestørrelse.

Note: Outcome er klassestørrelsen i efterfølgende skoleår i fuldt årgangspdelte normalklasser. 2SLS-estimerne er baseret på populationen af folkeskoleelever i fuldt årgangspdelte normalklasser i skoleårgange med årgangsstørrelser +/- fem elever omkring de tre nederste klasseloftstærskler i skoleårene 2014/15-2018/19. 2SLS-regressionerne kontrollerer desuden for årgangsstørrelsessegment, alder og kvadreret alder, køn, ikkevestlig baggrund, antallet af ældre søskende, moderens og faderens uddannelse, alder ved fødsel og kvadreret alder ved fødsel, samt indikatorer for, om moderen og faderen er diagnosticeret med angst og depression hhv. opmærksomhedsforstyrrelser. Standardfejl er klyngekorrigeret på skoleniveau, ***p < 0,01, **p < 0,05, *p < 0,1. 9. klassetrin er udeladt, da kun få elever går 9. klasse om i folkeskoleregi.

Bilag 3 REFUD udfaldsmål

Bilagstabel 3.1 Overblik over REFUDs udfaldsmål

Hoveddimension	Antal udfaldsmål	Bemærkning
Grundskole og specialundervisning (elev)	7	0.-10. kl.
Sociale foranstaltninger til børn og unge under 18 år	5	
Sundhedsydelse (elev)	8	
Indkomstoverførsler (mødre)	4	
Beskæftigelsesindsats (mødre)	1	
Beskæftigelse (mødre)	1	
Indkomstoverførsler (fædre)	4	
Beskæftigelsesindsats (fædre)	1	
Beskæftigelse (fædre)	1	
Personalefravær (elevs lærer)	1	
Politi, retsvæsen og kriminalforsorg (elev)	4	Fra 15 år og frem
Ungdomsuddannelse (elev)	3	Fra 16 år og frem
Efterværn (til 18-22-årige)	4	Fra 18 år og frem
Indkomstoverførsler (elev)	4	Fra 18 år og frem
Beskæftigelsesindsats (elev)	1	Fra 18 år og frem
Beskæftigelse (elev)	1	Fra 18 år og frem
Sociale serviceydelser (elev)	4	Fra 18 år og frem
Trivsel 4.-9. klassetrin (elev)	5	
Total	59	

Bilagstabel 3.2 Modellens konsekvenser, opdelt på hoveddimension og aktivitet

Hoveddimension	Aktivitet (udfaldsmål)	Enhed	Kilde(r)	De-fault	Nr.
Grundskole og specialundervisning (elev)	Kommunal grundskole, 0.-9. almenklasse ¹	Antal uger indskrevet	Elevregister (KOTRE) og Institutionsregister (INST)	Ja	2
	Kommunal grundskole, 10. almenklasse ⁵	Antal uger indskrevet	Elevregister (KOTRE) og Institutionsregister (INST)	Ja	3
	Fri- og privat grundskole, 0.-9. almenklasse ¹	Antal uger indskrevet	Elevregister (KOTRE) og Institutionsregister (INST)	Ja	4
	Fri- og privat grundskole, 10. almenklasse ⁵	Antal uger indskrevet	Elevregister (KOTRE) og Institutionsregister (INST)	Ja	5
	Specialskele ¹	Antal uger indskrevet	Elevregister (KOTRE) og Institutionsregister (INST)	Ja	1
	Specialklasse ¹	Antal uger indskrevet	Elevregister (KOTRE) og Specialundervisning (UDSP)	Ja	7

Hoveddimension	Aktivitet (udfaldsmål)	Enhed	Kilde(r)	De-fault	Nr.
	Specialundervisning i almenklasse ¹	Antal uger indskrevet	Elevregister (KOTRE) og Specialundervisning (UDSP)	Ja	6
Ungdomsuddannelse (elev) ⁵	Forberedende uddannelsestilbud (ekskl. STU)	Antal uger indskrevet	Elevregister (KOTRE)	Ja	8
	Særligt tilrettelagt uddannelse (STU)	Antal uger indskrevet	Elevregister (KOTRE)	Ja	9
	Ordinær ungdomsuddannelse	Antal uger indskrevet	Elevregister (KOTRE)	Ja	10
Sociale foranstaltninger til børn og unge under 18 år	Forebyggende indsatser (§ 11, stk. 3)	Modtager foranstaltning	Børn og Unge Forebyggende Foranstaltninger (BUFO)		11
	Hjemmebaserede foranstaltninger	Modtager foranstaltning	Børn og Unge Forebyggende Foranstaltninger (BUFO) og Børn og Unge Anbragte Forløbsregister (BUAF)		12
	Anbringelse i slægts- eller netværkspleje	Antal dage	Børn og Unge Anbragte Forløbsregister (BUAF)		13
	Anbringelse i familiepleje	Antal dage	Børn og Unge Anbragte Forløbsregister (BUAF)		14
Efterværn (til 18-22-årige) ⁴	Efterværn med hjemmebaserede foranstaltninger	Modtager foranstaltning	Børn og Unge Forebyggende Foranstaltninger (BUFO) og Børn og Unge Anbragte Forløbsregister (BUAF)		16
	Efterværn med anbringelse i slægts- eller netværkspleje	Antal dage	Børn og Unge Anbragte Forløbsregister (BUAF)		17
	Efterværn med anbringelse i familiepleje	Antal dage	Børn og Unge Anbragte Forløbsregister (BUAF)		18
	Efterværn med anbringelse på institution	Antal dage	Børn og Unge Anbragte Forløbsregister (BUAF)		19
Sundhedsydelse (elev)	Praktiserende læge	Antal sygesikrings-kontakter	Sygesikringsregisteret (SSSY)	Ja	20
	Øvrige behandlere under sygesikringen	Antal sygesikrings-kontakter	Sygesikringsregisteret (SSSY)	Ja	21
	Somatisk indlæggelse	Antal sengedage	Landspatientregisteret (LPR)	Ja	22
	Somatisk ambulantbesøg	Antal besøg	Landspatientregisteret (LPR)	Ja	23
	Somatisk skadestuebesøg	Antal besøg	Landspatientregisteret (LPR)	Ja	24
	Psykiatrisk indlæggelse	Antal sengedage	Landspatientregisteret Psykiatri (PSYK)	Ja	25
	Psykiatrisk ambulantbesøg	Antal besøg	Landspatientregisteret Psykiatri (PSYK)	Ja	26
Politi, retsvæsen og kriminalforsorg (elev) ³	Fængselsdage	Antal dage	Kriminalstatistik Afgørelser (KRAF) og Kriminalstatistik Indsættelser	Ja	28
	Domme for vold og sædelighedsforbrydelser	Antal domme	Kriminalstatistik Afgørelser (KRAF)	Ja	29
	Domme for indbrud, tyveri og hærværk	Antal domme	Kriminalstatistik Afgørelser (KRAF)	Ja	30
	Domme for overtrædelse af færdselsloven og øvrige særlove	Antal domme	Kriminalstatistik Afgørelser (KRAF)	Ja	31
Indkomstoverførsler (elev) ⁴	Kontanthjælp	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		32
	Førtidspension	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		33
	Uddannelseshjælp	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		34
	Øvrig overførsler (ekskl. SU)	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		35
Beskæftigelsesindsats (elev) ⁴	Vejledning og opkvalificering	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		36

Hoveddimension	Aktivitet (udfaldsmål)	Enhed	Kilde(r)	De-fault	Nr.
Beskæftigelse (elev) ³	Lønindkomst	Kr. pr. år	Detaljeret lønmodtagerdata fra e-Indkomst (BFL)	Ja	37
Sociale service-ydelser (elev) ⁴	Herberg og forsorgshjem (§ 110)	Antal dage	Boformregisteret (BOFORM_F)		38
	Stofmisbrugsbehandling, dag	Antal dage	Register over Stofmisbrugere i Behandling (SIB)		39
	Stofmisbrugsbehandling, døgn	Antal dage	Register over Stofmisbrugere i Behandling (SIB)		40
	Alkoholmisbrugsbehandling	Antal dage	Det National Alkoholbehandlingsregister (NAB)		41
Indkomstoverførsler (mødre)	Kontanthjælp	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		42
	Førtidspension	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		43
	Uddannelseshjælp	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		44
	Øvrig overførsler (ekskl. SU)	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		46
Beskæftigelsesindsats (mødre)	Vejledning og opkvalificering	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		45
Beskæftigelse (mødre)	Lønindkomst	Kr. pr. år	Detaljeret lønmodtagerdata fra e-Indkomst (BFL)	Ja	47
Indkomstoverførsler (fædre)	Kontanthjælp	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		48
	Førtidspension	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		49
	Uddannelseshjælp	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		50
	Øvrig overførsler (ekskl. SU)	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		52
Beskæftigelsesindsats (fædre)	Vejledning og opkvalificering	Antal uger	Beskæftigelsesministeriets forløbsdatabase (DREAM)		51
Beskæftigelse (fædre)	Lønindkomst	Kr. pr. år	Detaljeret lønmodtagerdata fra e-Indkomst (BFL)	Ja	53
Personalefravær	Lærernes sygefravær	Antal fuldtidsdage	Fraværperioder (FRPE) og Lærernes Kompetencedata fra STIL	Ja	54

Anm.: Tabellen angiver hvilke konsekvenser og ydelser, der estimeres indsatseffekter for i, herunder måleenhed og registerkilde, samt om de i udgangspunktet er inkluderet i måling af indsatsers nettogevinst eller om de skal tilvælges af brugeren..

- 1) Konsekvensmål for 0.-9. klasse måles til og med 9. klassetrin.
- 2) Beskæftigelse på elevniveau måles tidligst fra klassetrin, hvor den øverste alderspercentil er 13 år.
- 3) Konsekvensmål måles tidligst fra klassetrin, hvor den øverste alderspercentil er 15 år.
- 4) Konsekvensmål måles tidligst fra klassetrin, hvor den øverste alderspercentil er 18 år.
- 5) Konsekvensmål måles tidligst fra 9. klassetrin.

VIVÉ