

Til
**Undervisningsministeriet
(Kvalitets- og Tilsynsstyrelsen)**

Dokumenttype
Rapport

Dato
August 2014

SUPPLEMENT TIL EVALUERING AF DE NATIONALE TEST **RAPPORT**



NATIONALE TEST RAPPORT

INDHOLD

1. Indledning og resumé	1
2. Hvordan undersøges effekten af de nationale test i matematik	2
2.1 Empiriske Analyser	2
2.2 Datagrundlag	2
3. Effekten af de nationale test i matematik	3
3.1 Er elevernes faglige niveau styrket efter indførelse af de nationale test i matematik?	3
3.2 Sammenligning af effekten af de nationale test i matematik og i dansk, læsning	4
4. Konklusion	6
5. Perspektivering	6
6. Metodebilag	7

1. INDLEDNING OG RESUMÉ

Rambøll Management Consulting gennemførte i perioden november 2012 til september 2013 en evaluering af de nationale test i samarbejde med blandt andre professor Helena Skyt Nielsen og professor Simon Calmar Andersen, Aarhus Universitet. Undersøgelsen af effekterne af de nationale test på elevernes læring baseredes primært på effekten af at gennemføre de nationale test i dansk, læsning, da det var hvad datagrundlaget på det tidspunkt gav mulighed for.

I mellemtiden er nye data fra de nationale test i 2013 blevet tilgængelige. Det betyder, at det nu for første gang har været muligt at analysere betydningen af at gennemføre de nationale test i matematik. Resultaterne af disse analyser præsenteres i nærværende rapport, der derfor skal ses som et supplement til hovedrapporten. Analyserne er gennemført af Helena Skyt Nielsen og Simon Calmar Andersen med Rambøll i en mindre rolle. For en generel præsentation og evaluering af de nationale test henvises til hovedrapporten, Rambøll Management Consulting: "Evaluering af de nationale test i folkeskolen", 2013.

Resultaterne i denne supplerende rapport viser, at der er nogle positive sammenhænge mellem at gennemføre testene i matematik og elevernes senere resultater, men ingen af sammenhængerne er statistisk signifikante. Sammenhængerne er undersøgt både for den samlede matematik score og for dens tre profilområder for sig. Resultaterne er også sammenlignet mellem drenge og piger, for forskellige klassetrin og sammenlignet med resultaterne i dansk, læsning.

I det næste afsnit præsenteres de metoder og data, der er anvendt i rapporten. Derefter præsenteres resultaterne, og der afsluttes med en konklusion. Bagerst i rapporten findes et metodebilag.

2. HVORDAN UNDERSØGES EFFEKTEN AF DE NATIONALE TEST I MATEMATIK

Som udgangspunkt er det meget vanskeligt at undersøge effekten af de nationale test, fordi de blev indført på én gang for alle elever. Når man efterfølgende ser, hvordan eleverne klarer sig fagligt, mangler man derfor noget at sammenligne med. Måske ville eleverne have oplevet den samme fremgang (eller tilbagegang), hvis de nationale test ikke var blevet indført. Nedenfor beskrives, hvordan denne udfordring løses, og hvilket datamateriale der anvendes.

2.1 Empiriske analyser

For at analysere effekten af de nationale test på de enkelte elever udnyttes det, at it-systemet havde et nedbrud¹ i en periode på ni dage fra den 2. til 10. marts 2010. Det betød, at halvdelen af de elever, der skulle have gennemført en test på det tidspunkt, ikke fik nogen test². Ved at sammenligne disse elevers resultater ved efterfølgende nationale test med resultaterne for tilsvarende elever, der gennemførte testen i 2010, kan effekten af de nationale test undersøges. Den anvendte metode kaldes "two stage least squares" og foregår i to trin: I det første trin beregnes sandsynligheden for, at en elev er blevet testet. Her udnyttes det, at den gruppe, der blev ramt af nedbrud, har en lavere sandsynlighed for at blive testet, fordi kun nogle af dem blev genbooket til en ny test. I andet trin estimeres sammenhængen mellem denne sandsynlighed for at blive testet og testresultater to og tre år senere. På den måde beregnes hvor meget bedre, elever som gennemførte en test i 2010, klarer sig i forhold til lignende elever, som ikke gennemførte testen i 2010 på grund af nedbruddet.

Denne analyse hviler på den antagelse, at det var tilfældigt, hvilke elever der blev ramt af det midlertidige nedbrud. Analyser, der sammenligner de "nedbrudsramte" elever med de øvrige elever på en lang række baggrundsfaktorer fra Danmarks Statistiks registre (køn, alder, familieforhold, forældres uddannelse, indkomst, indvandrerstatus mv.), bestyrker, at den antagelse er korrekt. Der er kun få statistisk signifikante forskelle mellem de to grupper. Resultaterne af denne analyse ses i metodebilaget, tabel A1 og A2.

Evaluators har undersøgt muligheden for at lave lignende analyser af sammenhængen mellem nedbrud i 2010 og resultater ved folkeskolens afgangsprøve i 2011, 2012 og 2013. Det har dog vist sig, at informationerne i datamaterialet vedrørende bookinger i nedbrudsperioden ikke er intakt for dansk læsning, i 6. og 8. klasse og for matematik i 6. klasse, hvorfor der ikke er arbejdet videre med sådanne analyser.

2.2 Datagrundlag

Til analyserne benyttes registerdata fra UNI-C og Danmarks Statistik for elever, der gik i folkeskole fra skoleåret 2009/10 og til skoleåret 2012/13. Der benyttes data både om deltagelse i de nationale test, resultater fra de nationale test samt en lang række baggrundsplysninger om elevernes køn, alder og forældre. Alle individoplysningerne er anonymiserede. Fordelene ved registerdata er, at de er indsamlet på konsistent vis over tid, de kan kobles til andre datakilder via cpr-nummer og giver mulighed for at følge individer over tid.

¹ Begrebet "nedbrud" (eller lignende begreber) bruges som betegnelse for et midlertidigt teknisk nedbrud i testsystemet, der fandt sted i en kortere periode primo marts måned 2010. Det skal understreges, at driften af testsystemet efter evaluators oplysninger generelt er velfungerende.

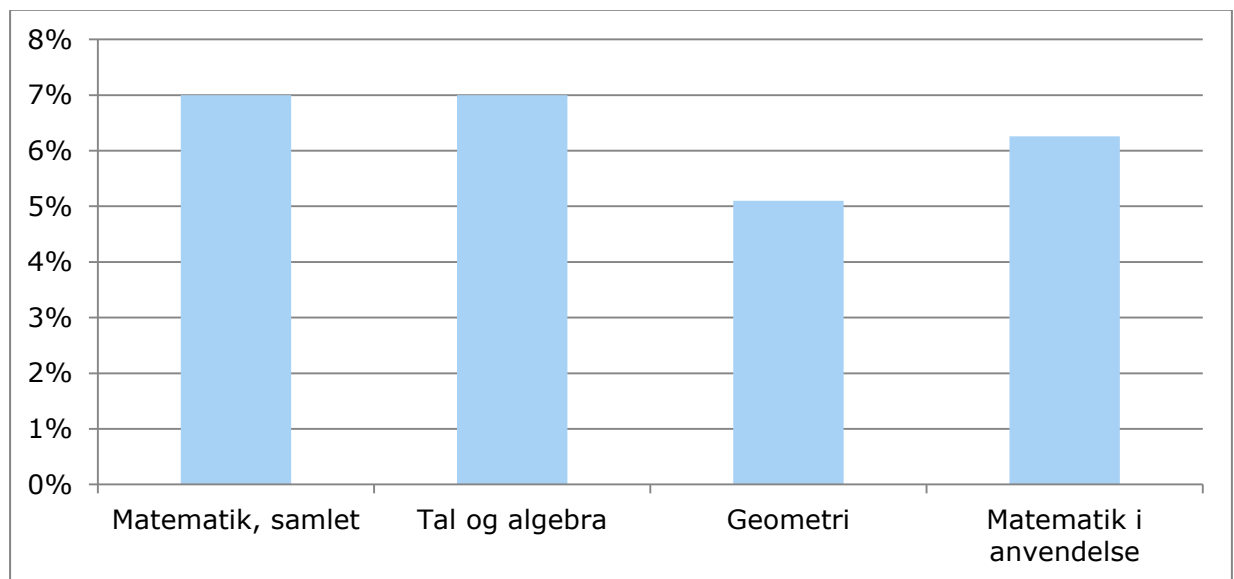
² Skolerne blev tilbudt mulighed for at booke en ny test, hvilket skete for mere end halvdelen af de 6.143 elever, som blev udsat for et nedbrud i den nationale test i matematik i 3.kl; 3.487 bookede og gennemførte en ny test, mens 2.656 elever ikke fik taget testen.

3. EFFEKTEN AF DE NATIONALE TEST I MATEMATIK

3.1 Er elevernes faglige niveau styrket efter indførelse af de nationale test i matematik?

Analysen af effekten af de nationale test i matematik viser, at der er en sammenhæng mellem, om eleverne gennemførte en test i matematik i 3. klasse i 2010 og elevernes præstationer tre år efter (2013) i test i 6. klasse, men at denne sammenhæng *ikke* er statistisk signifikant. Resultatet af denne analyse er illustreret i figur 3.1. Den samlede testscore i matematik er udregnet som et standardiseret gennemsnit af de standardiserede testscores på de tre profilområder.³ Når man standardiserer den samlede testscore, kan der sammenlignes med andre test på tværs af fag og på tværs af klassetrin. Figuren viser, at de elever, der gennemførte testen i matematik i 3. klasse, klarede sig 5 til 7 pct. (af en standardafvigelse⁴) bedre end elever, der ikke blev testet. Denne forskel er imidlertid ikke statistisk signifikant. Det vil sige, at der er en betydelig sandsynlighed for, at disse forskelle kan skyldes tilfældigheder – for eksempel små tilfældige forskelle i hvilke elever, der blev ramt af nedbrud, og hvilke der ikke gjorde. Mere specifikt viser analyserne, at der er mere end 10 pct. sandsynlighed for at finde disse forskelle, hvis der i realiteten ikke er nogen effekt af testene. Normalt kræves højst 5 pct. sandsynlighed, for at der er tale om statistisk signifikans. De præcise resultater af den *two-stage least squares* analyse, som ligger til grund for figuren, kan ses i metodebilaget, tabel A3.

Figur 3.1: Effekten af at gennemføre en test i matematik i 3. klasse i 2010 på præstation tre år efter ved test i 6. klasse



Note: Samlet testscore og opdelt på de tre profilområder målt på standardiseret skala. Effekten er *ikke* statistisk signifikant.

I hovedrapporten fandt evaluatoren signifikante effekter af at deltage i de nationale test i dansk, læsning for de enkelte elever. Disse effekter blev sidenhen opdelt på hvor meget, der lå mellem skoler, og hvor meget der kunne tilskrives forskelle mellem elever på samme skoler (en såkaldt skole "fixed-effects"-model). Denne forskel blev analyseret for at kunne sige noget om, hvorvidt der var tale om teaching-to-the-test. Da der i dette tilfælde med matematik ikke er påvist hovedeffekter, er der ikke grundlag for at kigge isoleret på sammenhænge indenfor skoler.

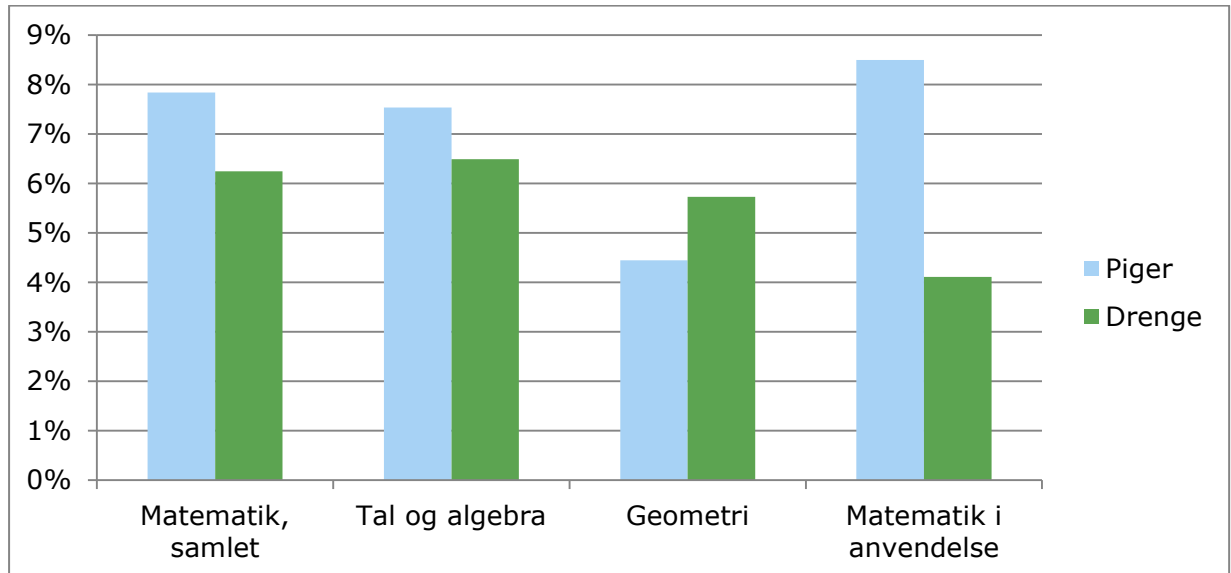
Evaluatoren har derimod undersøgt, om effekten af de nationale test i matematik er forskellig for drenge og piger. Det er der nogle små tendenser til som illustreret i figur 3.2, men generelt ikke

³ Standardisering af testscoren foregår ved at fratække gennemsnittet af testscores for årgangen og derefter dividere med standardafvigelsen af testscores for årgangen. Dette gøres først for hvert af de tre profilområder for sig, og derefter for den gennemsnitlige standardiserede score.

⁴ De nationale test måles som udgangspunkt på en såkaldt Rasch-skala. Vi har standardiseret skalaen, så den har gennemsnit på 0 og standardafvigelse på 1. Effekten måles derfor i, hvor meget af en standardafvigelse (der er mål for den gennemsnitlige spredning i resultaterne) de nationale test flytter eleverne.

større end at det kan skyldes tilfældigheder i målingerne. De præcise resultater af denne analyse kan ses i metodebilaget, tabel A4.

Figur 3.2: Effekten for piger og drenge af at gennemføre en test i matematik i 3. klasse i 2010 på præstation tre år efter ved test i 6. klasse



Note: Samlet testscore og opdelt på de tre profilområder målt på standardiseret skala. Forskellene mellem piger og drenge er ikke statistisk signifikante.

3.2 Sammenligning af effekten af de nationale test i matematik og i dansk, læsning

I nedenstående figur 3.3 har evaluator sammenlignet effekterne af at blive testet i henholdsvis dansk, læsning og matematik.

Den første søjle viser effekten af at blive testet i dansk, læsning i 2. klasse på resultatet i den nationale test i dansk, læsning i 4. klasse (to år efter). Denne søjle viser, at en elev, der blev testet i 2. klasse, klarer sig næsten 10 % bedre i dansk, læsning, i 4. klasse sammenlignet med en elev, som ikke blev testet i 2. klasse på grund af nedbrud i it-systemet. Denne effekt er statistisk signifikant og er en af de effekter, der også blev fundet i hovedrapporten.

Den anden søjle viser, at der er en lille (4 %) og ikke-signifikant effekt af at blive testet i dansk, læsning i 2. klasse på resultatet i den nationale test i matematik året efter i 3. klasse. Signifikansen er målt på det traditionelle 5 pct-niveau. Det vil sige, at der er mere end 5 pct. sandsynlighed for at effekterne kan skyldes tilfældigheder.

Den tredje søjle viser en lille (3 %) og ikke-signifikant effekt af at blive testet i dansk, læsning i 4. klasse på resultatet af den nationale test i dansk, læsning i 6. klasse.

Den fjerde søjle viser en negativ og ikke signifikant effekt af at blive testet i dansk, læsning i 4. klasse på resultatet af den nationale test i matematik i 6. klasse.

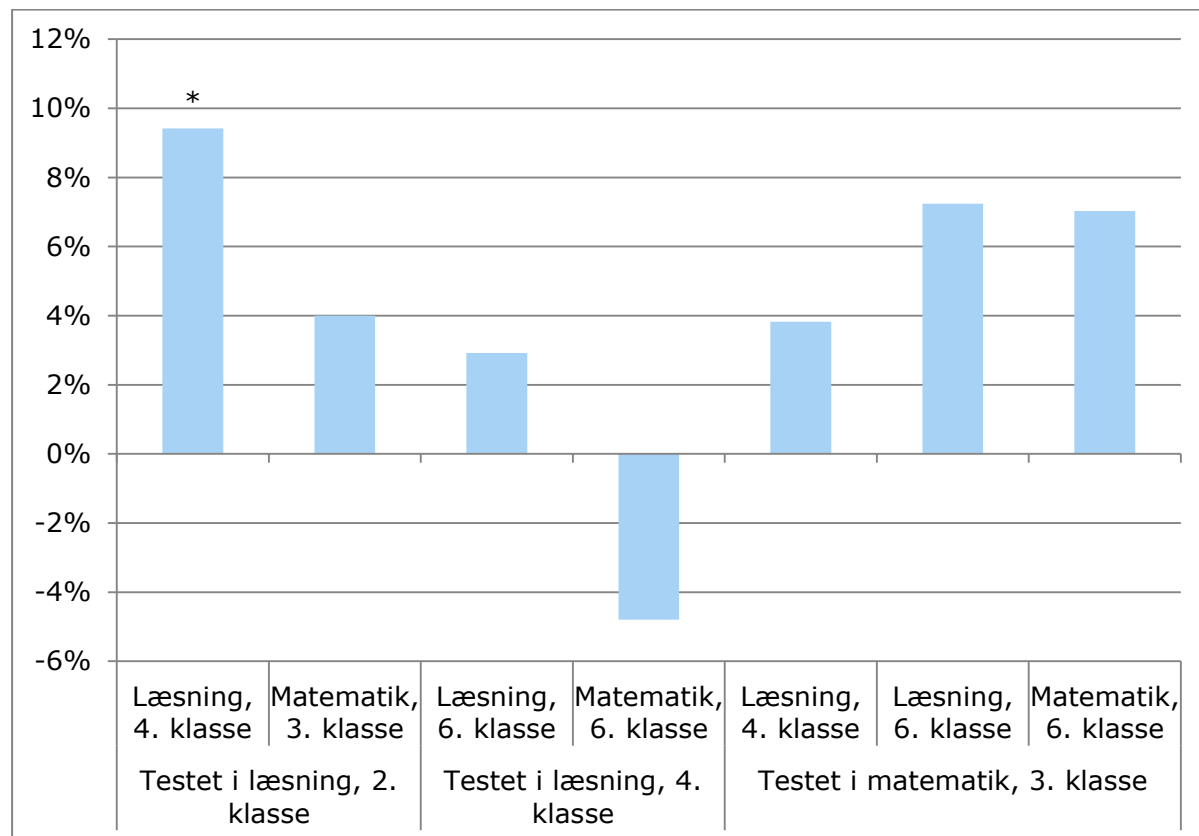
Den femte søjle viser en lille (næsten 4 %) og ikke-signifikant effekt af at blive testet i matematik i 3. klasse på resultatet af den nationale test i dansk, læsning i 4. klasse.

Den sjette søjle viser en lille (7 %) og ikke-signifikant effekt af at blive testet i matematik i 3. klasse på resultatet af den nationale test i dansk, læsning i 6. klasse.

Og endelig viser den syvende søjle en lille (7 %) og ikke-signifikant effekt af at blive testet i matematik i 3. klasse på resultatet af den nationale test i matematik i 6. klasse.

Som nævnt er signifikans målt på det traditionelle 5 pct-niveau. Det vil sige, at der er mere end 5 pct. sandsynlighed for at effekterne kan skyldes tilfældigheder. De præcise resultater af denne analyse kan findes i metodebilaget, tabel A5.

Figur 3.3: Effekten af at gennemføre test i dansk, læsning og i matematik på præstation i test på efterfølgende klassetrin



Note: Sammenhæng med samlet testscore målt på standardiseret skala. Figuren viser hvor meget bedre elever, som blev testet i 2010, klarer sig ved efterfølgende test i 2011-13 sammenlignet med elever, som blev ramt af nedbrud i 2010. * indikerer, at effekten er signifikant på 5 pct.-niveauet.

4. KONKLUSION

Hovedrapporten "Evaluering af de nationale test i folkeskolen" viste, at gennemførelse af de nationale test i dansk, læsning ser ud til at have en positiv effekt på elevernes efterfølgende resultater i dansk, læsning. Effekterne sås primært i udviklingen fra 2. til 4. klasse og fra 6. til 8. klasse. Nærværende supplement til hovedrapporten har undersøgt, om der er tilsvarende effekter af at gennemføre de nationale test i matematik, og der er benyttet de nye data til at anvende en alternativ analyse af effekterne i både dansk, læsning og matematik. Der er nogle positive sammenhænge mellem at gennemføre testene og de senere resultater i matematik, men sammenhænge er ikke statistisk signifikante.

5. PERSPEKTIVERING

Det kan ikke ses ud af disse tal, hvorfor der findes forskellige resultater for matematik og dansk. I hovedrapporten⁵⁵ sås det, at dansklærere og matematiklærere er nogle af de faglærere, som anvender resultaterne fra de nationale test mest, når det drejer sig om at få et overblik over klassens og den enkelte elevs faglige niveau. Når det drejer sig om at anvende testresultaterne som input i planlægningen af undervisningen kom der dog nogle tydelige forskelle frem i casestudierne.

Dansklærerne fremhævede, at det var nemt at bruge testene som input til planlægning af undervisningen. Dels måler testen færdigheder indenfor et afgrænset felt af danskfaget. Dels bliver eleverne testet hyppigt, hvilket muliggør systematisk og kontinuerlig opfølgning. Dels sætter lærerne pris på profilberegneren og det tilhørende undervisningsmateriale, som hjælper med at oversætte testresultaterne til pædagogisk praksis.

Matematiklærerne fremhævede på den anden side, at det var svært at bruge testene som input. Dels fordi timingen er uheldig, eftersom eleverne typisk skifter til mellemtrinnet og får en ny matematiklærer efter 3. klassetrin og skifter til udskolingen og får en ny matematiklærer efter 6. klassetrin. Dels fordi testene ikke spiller godt sammen med fagets trinmål og undervisningsmaterialer.

En anden fortolkning kan være, at dansklærere er bedre klædt på til at omsætte den viden, som de nationale test giver om de enkelte elevers styrker og svagheder indenfor de forskellige profilområder af faget. Man har lang tradition for at arbejde målrettet med læsetest og man gør stor brug af læsevejledere. Dette kan betyde, at dansklærerne i højere grad har redskaber til at arbejde med elevernes specifikke faglige udfordringer og dermed medvirke til at skabe en effekt ved næste test.

Man kunne også forestille sig, at selve det faktum, at vi undersøger effekten af testen i 2010 på præstationer i 2013 betyder, at en eventuel testeffekt udvaskes. Matematiklærerne har i denne treårige periode testet masser af andre elever og har formodentlig brugt andre individuelle test, hvorfor den nationale test for tre år siden overskygges af disse andre effekter. Disse er imidlertid kun nogle få ud af mange mulige fortolkninger.

⁵⁵ Se kapitel 4 i hovedrapporten.

6. METODEBILAG

I tabel A1 og A2 sammenlignes de grupper af elever, der henholdsvis gennemførte og blev ramt af nedbrud i de nationale test i matematik i 2010. Som det fremgår, er der kun få forskelle i gruppernes socioøkonomiske baggrund. Det bestyrker antagelsen om, at grupperne er sammenlignelige i udgangspunktet. Tabel A1 angiver gennemsnittet af baggrundsvariablene for elever, som var booket i nedbrudsperioden, men som ikke gennemførte testen i denne periode ("Nedbrud") samt elever, som var booket udenfor nedbrudsperioden og gennemførte testen som planlagt ("Gennemførte"). Tabel A2 viser resultatet af en lineær regression. Her ses, at kun ganske få variable er signifikant korrelerede med oplevet nedbrud, f.eks. etnisk oprindelse og antal børn i husholdningen.

Tabel A1: Gennemsnit og standardafvigelse for baggrundskarakteristika

	(1)		(2)	
	Intet nedbrud		Nedbrud	
	Gns	Std.afv.	Gns	Std.afv.
<i>Barns karakteristika:</i>				
Dreng	0,505	0,500	0,505	0,500
Tidligt i skole	0,013	0,115	0,016	0,125
Sent i skole	0,173	0,379	0,177	0,381
Etnisk dansker	0,909	0,288	0,913	0,282
Etnisk europæer	0,030	0,169	0,034	0,181
Enebarn	0,091	0,288	0,100	0,301
Antal børn	2,370	0,939	2,332	0,939
Mangler, antal børn	0,017	0,128	0,017	0,131
<i>Forældres uddannelse:</i>				
Mor, gymnasial uddannelse eller mindre	0,072	0,258	0,075	0,264
Mor, erhvervsuddannet	0,368	0,482	0,373	0,484
Mor, kort videregående uddannelse	0,047	0,211	0,048	0,213
Mor, mellemlang videregående uddannelse	0,231	0,421	0,222	0,416
Mor, lang videregående uddannelse	0,082	0,274	0,079	0,270
Mangler, mors uddannelse	0,018	0,133	0,018	0,131
Far, gymnasial uddannelse eller mindre	0,049	0,217	0,053	0,223
Far, erhvervsuddannet	0,403	0,491	0,387	0,487
Far, kort videregående uddannelse	0,067	0,250	0,070	0,255
Far, mellemlang videregående uddannelse	0,122	0,327	0,122	0,328
Far, lang videregående uddannelse	0,099	0,299	0,101	0,302
Mangler, fars uddannelse	0,075	0,264	0,077	0,266
<i>Arbejdsmarkedsstatus og forældres gns. indkomst:</i>				
Mor, selvstændig	0,031	0,174	0,037	0,188
Mor, under uddannelse	0,009	0,096	0,013	0,115
Mor, ledig	0,024	0,153	0,026	0,159
Mor, øvrig aktivitet	0,127	0,333	0,122	0,327
Mangler, mors beskæftigelse	0,016	0,124	0,015	0,121
Far, selvstændig	0,080	0,271	0,081	0,272
Far, under uddannelse	0,001	0,036	0,002	0,049
Far, ledig	0,014	0,036	0,013	0,113
Far, øvrig aktivitet	0,060	0,238	0,063	0,242
Mangler, fars beskæftigelse	0,075	0,263	0,076	0,265
Gennemsnitsindkomst, m. 125.000-220.000	0,209	0,406	0,219	0,413
Gennemsnitsindkomst, m. 220.000-345.000	0,422	0,494	0,418	0,493
Gennemsnitsindkomst, o. 345.000	0,296	0,457	0,293	0,455
Mangler, gennemsnitsindkomst	0,012	0,109	0,011	0,105
Mor, ledig m. 1-50%	0,188	0,391	0,196	0,397
Mor, ledig o. 50%	0,025	0,157	0,027	0,161
Far, ledig m. 1-50%	0,076	0,265	0,082	0,274
Far, ledig o. 50%	0,013	0,115	0,012	0,108
Antal observationer	40195		6143	

Tabel A2: Regression af nedbrud i matematik på baggrundskaraktetika

	(1)
	Nedbrud, matematik 2010
<i>Barns karakteristika:</i>	
Dreng	-0.000259 (0.00320)
Tidligt i skole	0.0216 (0.0136)
Sent i skole	0.00236 (0.00434)
Etnisk dansker	0.0191** (0.00746)
Etnisk europæer	0.0312*** (0.0113)
Enebarn	0.00539 (0.00643)
Antal børn	-0.00408* (0.00215)
<i>Forældres uddannelse:</i>	
Mor, gymnasial uddannelse eller mindre	0.00371 (0.00712)
Mor, erhvervsuddannelse	0.000683 (0.00489)
Mor, kort videregående uddannelse	0.000619 (0.00852)
Mor, mellemlang videregående uddannelse	-0.00617 (0.00568)
Mor, lang videregående uddannelse	-0.00981 (0.00796)
Far, gymnasial uddannelse eller mindre	0.00563 (0.00809)

(fortsættes.)

(fortsat)	(1)
	Nedbrud, dansk, læsning 2010
Far, erhvervsuddannelse	-0.00786* (0.00460)
Far, kort videregående uddannelse	0.00405 (0.00729)
Far, mellemlang videregående uddannelse	0.00121 (0.00624)
Far, lang videregående uddannelse	0.00550 (0.00703)
<i>Arbejdsmarkedsstatus og forældres gns. indkomst:</i>	
Mor, selvstændig	0.0204** (0.00923)
Mor, under uddannelse	0.0491*** (0.0170)
Mor, Ledig	0.00330 (0.0117)
Mor, øvrig aktivitet	-0.00525 (0.00596)
Far, selvstændig	0.00308 (0.00598)
Far, under uddannelse	0.0892** (0.0419)
Far, ledig	-0.0130 (0.0155)
Far, øvrig aktivitet	0.00799 (0.00718)
Gennemsnitsindkomst, m. 125.000-220.000	0.0117 (0.00778)
Gennemsnitsindkomst, m. 220.000-345.000	0.00544 (0.00817)
Gennemsnitsindkomst, o. 345.000	0.00618 (0.00858)
Mor, ledig m. 1-50%	0.00529 (0.00425)
Mor, ledig o. 50%	0.00605 (0.0113)
Far, ledig m. 1-50%	0.0101 (0.00625)
Far, ledig o. 50%	-0.0101 (0.0158)
Antal observationer	46,338

Note: Et konstant led og indikator variabler for manglende variabler er inkluderet. Std fejl i parenteser. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

I tabel A3 vises hovedresultaterne af analysen – sammenhængen mellem om eleverne har gennemført de nationale test i matematik i 3. klasse og deres resultater ved de efterfølgende nationale test i matematik i 6. klasse. Analysen er gennemført som en *two stage least squares* regression. I det første trin beregnes sandsynligheden for, at en elev er blevet testet. Her udnyttes det, at den gruppe der blev ramt af nedbrud, har en lavere sandsynlighed for at blive testet, fordi kun nogle af dem blev genbooket til en ny test. Som det ses, er der som forventet i 1. trin en stærkt signifikant sammenhæng mellem om en elev er ramt af nedbrud, og om eleven er endt med at have gennemført en test. I andet trin estimeres sammenhængen mellem denne sandsynlighed for at blive testet og testresultater to år senere. På den måde beregnes hvor meget bedre elever, som gennemførte en test i 2010, klarer sig i forhold til lignende elever, som ikke gennemførte testen i 2010 på grund af nedbruddet. Som det fremgår, er der i 2. trin ikke nogen signifikant sammenhæng mellem at være blevet testet og de efterfølgende resultater.

Model 1 viser hovedresultater uden brug af kontrolvariabler. I model 2-5 kontrolleres for en lang række baggrundsvARIABLE (jf. tabel A1). Model 3-5 viser resultater for de tre profilområder under matematik.

Tabel A3: Effekt af nedbrud i matematik på efterfølgende resultat i matematik. Two stage least squares.

	(1)	(2)	(3)	(4)	(5)
	Total Score	Total Score	Tal og Algebra	Geometri	Anvendelse
<i>Test Score (2.trin)</i>					
Testet	0.0794 (0.0610)	0.0757 (0.0501)	0.0743 (0.0517)	0.0563 (0.0423)	0.0672 (0.0481)
<i>Testet (1. trin)</i>					
Nedbrud	-0.436*** (0.0306)	-0.436*** (0.0306)	-0.436*** (0.0306)	-0.436*** (0.0306)	-0.436*** (0.0306)
Nobs	46,338	46,338	46,338	46,338	46,338
R-squared	0.001	0.145	0.113	0.107	0.108
Kontrolvariable:					
Køn	NEJ	JA	JA	JA	JA
+ ung/gammel-for-klasse, immig.status	NEJ	JA	JA	JA	JA
+ forældres uddannelse	NEJ	JA	JA	JA	JA
+ familie struktur og søskende status	NEJ	JA	JA	JA	JA
+ forældreindkomst og arbejdsmarkeds historik	NEJ	JA	JA	JA	JA

Note: Robuste standard fejl i parenteser. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Tabel A4 viser sammenhængene opdelt på drenge og piger. Til sammenligning vises i model 1 resultatet for begge køn samlet (modellen er den samme som model 2 i tabel A3). Som det fremgår af interaktionen mellem at være testet og at være dreng (Testet * dreng), er der ikke signifikant forskel i effekten mellem piger og drenge (model 2) – heller ikke for nogen af profilområderne (model 3-5).

Tabel A4: Resultater opdelt på piger og drenge. Two stage least squares.

	(1)	(2)	(3)	(4)	(5)
	Total score	Total score	Tal og Algebra	Geometri	Anvendelse
<i>Test Score (2.trin)</i>					
Testet	0.0757 (0.0501)	0.0814 (0.0596)	0.0791 (0.0619)	0.0445 (0.0541)	0.0850 (0.0574)
Testet * dreng		-0.0112 (0.0644)	-0.00934 (0.0677)	0.0176 (0.0612)	-0.0374 (0.0666)
<i>Testet (1.trin)</i>					
Nedbrud	-0.436*** (0.0306)	-0.432*** (0.0312)	-0.432*** (0.0312)	-0.432*** (0.0312)	-0.432*** (0.0312)
Nedbrud * dreng		-0.441*** (0.0312)	-0.441*** (0.0312)	-0.441*** (0.0312)	-0.441*** (0.0312)
Nobs	46,338	46,338	46,338	46,338	46,338
R-squared	0.145	0.145	0.107	0.108	0.113

Note: Alle kontrolvariable er medtaget, inklusiv de relevante main effects. Robuste std.fejl i parenteser.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

I tabel A5 sammenlignes effekten af at have gennemført testen i henholdsvis dansk og matematik på de efterfølgende resultater i de to fag. Som det fremgår, er den eneste sammenhæng (i 2. trin), der er signifikant på det konventionelle 5 pct.-signifikansniveau, sammenhængen mellem at have gennemført testen i dansk, læsning i 2. klasse og så resultatet i testen i dansk, læsning i 4. klasse.

Tabel A5: Sammenligning af resultater for dansk og matematik. Two stage least squares.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Matematik 3. klasse	Læsning 4. klasse	Læsning 4. klasse	Læsning 6. klasse	Matematik 6. klasse	Læsning 6. klasse	Matematik 6. klasse
<i>Test Score (2.trin)</i>							
Testet Læsning, 2.kl.	0.0411 (0.0434)	0.0958*** (0.0333)					
Testet Læsning, 4.kl.				0.0295 (0.0370)	-0.0468 (0.0489)		
Testet Matematik, 3.kl.			0.0319 (0.0461)			0.0787** (0.0398)	0.0757 (0.0498)
<i>Testet (1.trin)</i>							
Nedbrud Læsning, 2.kl.	-0.532*** (0.0285)	-0.532*** (0.0285)					
Nedbrud Læsning, 4.kl.				-0.451*** (0.0285)	-0.451*** (0.0285)		
Nedbrud Matematik, 3.kl.			-0.436*** (0.0306)			-0.436*** (0.0306)	-0.436*** (0.0306)
Nobs	46,890	46,294	46,153	46,250	46,178	46,448	46,338
R-squared	0.112	0.145	0.143	0.161	0.142	0.163	0.145

Note: Alle kontrolvariable er medtaget. Robuste standard fejl i parenteser. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$