Effets de la réforme scolaire au Québec sur les compétences mathématiques des enfants du primaire et du secondaire:
LES DÉFIS DE L'ÉVALUATION D'IMPACT

Catherine Haeck[†], Pierre Lefebvre[††] and Philip Merrigan[††]

[†]Katholieke Universiteit Leuven et [††]Université du Québec à Montréal

CIQSS, Février 2012

Impact evaluation, what do you need?

1. A natural experiment (controlled experiments are not always possible from an ethical point of view)
2. The experiment needs to be of interest (in my case, it needs to be economically relevant and rarely studied and/or new methodology)
3. A group that is treated, and one that is not... and they need to be convincing!
4. A good data set, with data on both groups and interesting outcome variables
5. A strong empirical strategy combined with a number of robustness checks and falsification tests

In our case, where did it all started?

General question: How does schooling matters?

- A consensus seems to have emerged from international surveys suggesting that a sizeable proportion of young people around age 15 in many countries do not appear to possess all of the skills required to meet the challenges of today's knowledge societies.
- Empirical research has shown that measures of schooling attainment alone may not be sufficient to capture the extent to which human capital triggers economic growth and impacts individual labour market outcomes (e.g. Currie and Thomas, 2001; Hanushek and Woessmann, 2008)
- Research shows that concrete measures of academic achievement and cognitive skills, along with educational attainment, are strongly correlated with labour market outcomes, such as earnings and unemployment.
- A number of studies have documented the specific importance of mathematical abilities in adulthood socioeconomic success (e.g. Murnane et al., 1995; Rose and Betts, 2004; Ingram and Neumann, 2006).

Knowing that math skills were important, we wondered...

What influences the development of these skills...

- Teacher's quality has been shown to be of great importance in predicting the success of student (e.g. Hanushek and Rivkin (2010)).
- Other types of resources (e.g. per pupil expenditure, school facilities, class size) have generally been shown to be poor predictors of student performance (e.g. Hanushek (2003), Angrist and Levy (1999), Hoxby (2000), and Rivkin et al. (2001). In educational research, see Teddlie (2000)).
- But there are few research studies in economics on the impact of different teaching approaches...

Specific question: How did the Quebec school reform, implemented in the early 2000s, impact students' ability in procedural mathematics on average and across the distribution throughout primary (K-6) and secondary (7-11) school?

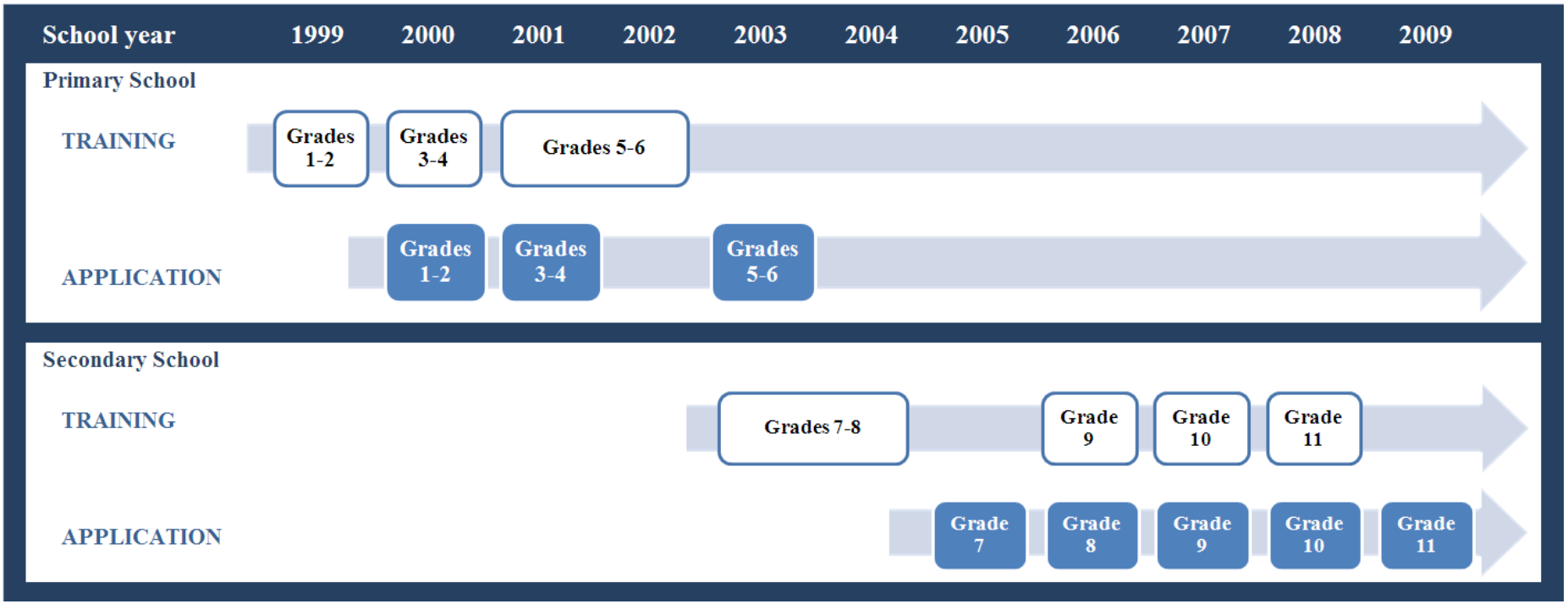
The Reform

The Quebec reform aimed at making schools more responsive to the changing needs of children in order to improve the **chances of success for all**.

It moved teaching away from the traditional/academic approaches of memorization, drills and activity books, to a much more comprehensive approach focused on *learning in a contextual setting* in which children are expected to *find the answers for themselves*.
$\longrightarrow$ **change of teaching approach**

It was implemented in the early 2000s following a precise calendar over a 10 year period.

Reform schedule and implementation



Is this reform interesting to study...

Provincial interest: did the reform ensure the success for all?
$\longrightarrow$ Need to measure not only the mean effect, but also the distributional effects...

International interest: the teaching approach promoted by the Quebec reform was supported by a number of countries
$\longrightarrow$ e.g. reform-oriented teaching approach (United States) supported by leading organizations such as

- the National Council of Teachers of Mathematics
- the National Research Council
- the American Association for the Advancement of Science

Is this reform interesting to study...

Studies on similar reforms rely on

- targeted reforms on more disavantage students
- small sample size
- short period of treatment and observation (max. 3 years)
- non experimental methods

Methodological advantages of the reform

The Québec experiment/reform provides some advantages for the purpose of evaluation and cuts across some of the methodological difficulties mentioned above.

1. Québec's Department of Education implemented the reform and all schools (public and private) were forced to apply the new education program.
2. at the same time teaching in the rest of Canada continued to be delivered in the same way.
3. the reform and data set used provides an observation and treatment period that is longer than typically seen in the literature.

Empirical Strategy

2 groups, 2 periods
$G_i \in \{0, 1\}$: group 0 is the RofC and group 1 is Québec
$T_i \in \{0, 1\}$: period 0 is prior to the school reform and period 1 is after

Y is the outcome measured

- Y_{gt}^I is the outcome if treated
- Y_{gt}^N is the outcome if *not* treated

We observe Y_{11}^I, Y_{10}^N, Y_{01}^N, Y_{00}^N

Empirical Strategy

The obvious candidate:

- the difference-in-differences (DID) estimator...

$\longrightarrow$ but DID has raised a number of concerns (e.g. Bertrand, Duflo, and Mullainathan, 2004; Donald and Lang, 2007; and Besley and Case, 2000).

As a result, we also implement

- the changes-in-changes (CIC) model (Athey and Imbens, 2006)
- the matching difference-in-differences (MDID) estimator (Heckman, Ishimura and Todd, 1997)

The DID approach

We first implement the following difference-in-differences estimator:

$$\tau^{DID} \equiv E[Y_{11}^{I}] - E[Y_{11}^{N}] = (E[Y_{11}] - E[Y_{10}]) - (E[Y_{01}] - E[Y_{00}]).$$

Assumptions:

- outcomes are additive in time period, group and unobservable characteristics of the individual (linearity)
- the treatment effect is constant across individuals or the effect differs across individuals but the distribution of outcomes without treatment is common across groups

The CIC approach

We implement the following changes-in-changes estimator:

$$\tau^{CIC} \equiv E\left[Y_{11}^{I}\right] - E\left[Y_{11}^{N}\right] = E\left[Y_{11}^{I}\right] - E\left[F_{Y,01}^{-1}\left(F_{Y,00}\left(Y_{10}\right)\right)\right].$$

Assumption:

- the underlying production functions for treated individuals and non-treated individuals, mapping the relationship between the outcomes and the unobservables at a given point in time, do not vary across groups

This model relaxes some of the assumptions of the standard DID:

- nonparametric identification, estimation, and inference for the ATE
- the time and treatment effect may vary across individuals
- estimates the entire counterfactual distribution of effects of not receiving the treatment for the treatment group
- it accommodates the possibility – but does not assume – that the treatment group adopted the policy because it expected greater benefits than in the control group

The MDID approach

In our approach to CIC, we control for X through a linear specification. To address the possibility of non linearity of response with respect to X, we implement the following matching difference-in-differences estimator:

$$\tau^{MDID} = \sum_{i \in S_{11}} \left\{ \left[y_{it_1} - \sum_{j \in S_{10}} \tilde{w}_{ijt_0} y_{jt_0} \right] - \left[\sum_{j \in S_{01}} \tilde{w}_{ijt_1} y_{jt_1} - \sum_{j \in S_{00}} \tilde{w}_{ijt_0} y_{jt_0} \right] \right\} w_i .$$

In contrast to standard DID, MDID allows the possibility of

- non linearity of response with respect to X
- selection into treatment

*** We implement kernel matching, local linear regression matching and nearest neighbor matching. Bootstrap standard errors are calculated for local linear regression and kernel matching to account for the underlying matching procedure (not consistent for nearest neighbor).

The Data set

Choose wisely... check response rate, outcome and control variables, etc.

We decided to use Statistics Canada's National Longitudinal Survey of Children and Youth (NLSCY) which provides three cohorts of children of primary and secondary school age:

1. students in grades 1 to 6 in academic year 1994-95 up to grades 9 to 12 in academic year 2008-09
2. students in grades 1 to 4 in academic year 2006-07
3. students in grades 1 and 2 in academic year 2008-09

The NLSCY provides **one** measure of cognitive development for school age children: the CAT/2 mathematics test.

Characteristics of the math test (CAT/2)

- developed after careful consideration of the differences across the main school curricula across Canada
- designed to measure procedural skills in mathematics (addition, subtraction, multiplication, division on integers, etc.)
- administered to students enrolled in grades 2 to 10, aged 7 to 15 years old
- difficulty of the test varies with the school grade of the child
- the standardized scores are designed to numerically represent the relative level of mathematics a child has attained
- the test score is positively correlated with
  1. the probability of not dropping out of high school
  2. the total personal income at age 24-25
  3. the highest level of education attained

Is this outcome variable relevant?

	High school dropout	Highest level of education	Total personal income
CAT/2 score			
Coef.	-0.015120	0.009723	25.367860
dy/dx	-0.000665	0.001226	25.367860
z	-4.74	4.18	2.16
Regression	logit	ord. logit	linear

The CAT/2 test (grades 5-6) is related to adulthood labour market outcomes.

Characteristics of the test

Points of caution...

1. The response rates for waves 1 to 3 were uncharacteristically low: 51% in wave 1, 74% in wave 2, and 54% in wave 3.
2. The difficulty level of the test for comparable students is different in wave 1 (compared to all other waves).
3. The difficulty level of the test for grades 9 and 10 is only comparable from wave 5 onwards.

Students' observations in wave 1 are dropped, and so are students' observations in grades 9-10 prior to wave 5.

School grades observed pre and post reform

Grade 1 entry year	Academic Year 1996	1998	2000	2002	2004	2006	2008
1989-90	7 - 8	.					
1991-92	5 - 6	7 - 8	.				
1993-94	3 - 4	5 - 6	7 - 8	.			
1995-96	2	3 - 4	5 - 6	7 - 8	9 - 10		
1997-98		2	3 - 4	**5*** 6	**7*** 8	**9*** 10	
1999-00			2	3 - 4	5 - 6	7 - 8	9 - 10
2001-02				.	.	.	.
2003-04					.	3 - 4	.
2005						2	.
2007							2

Are my groups comparable and do we observe selection into treatment?

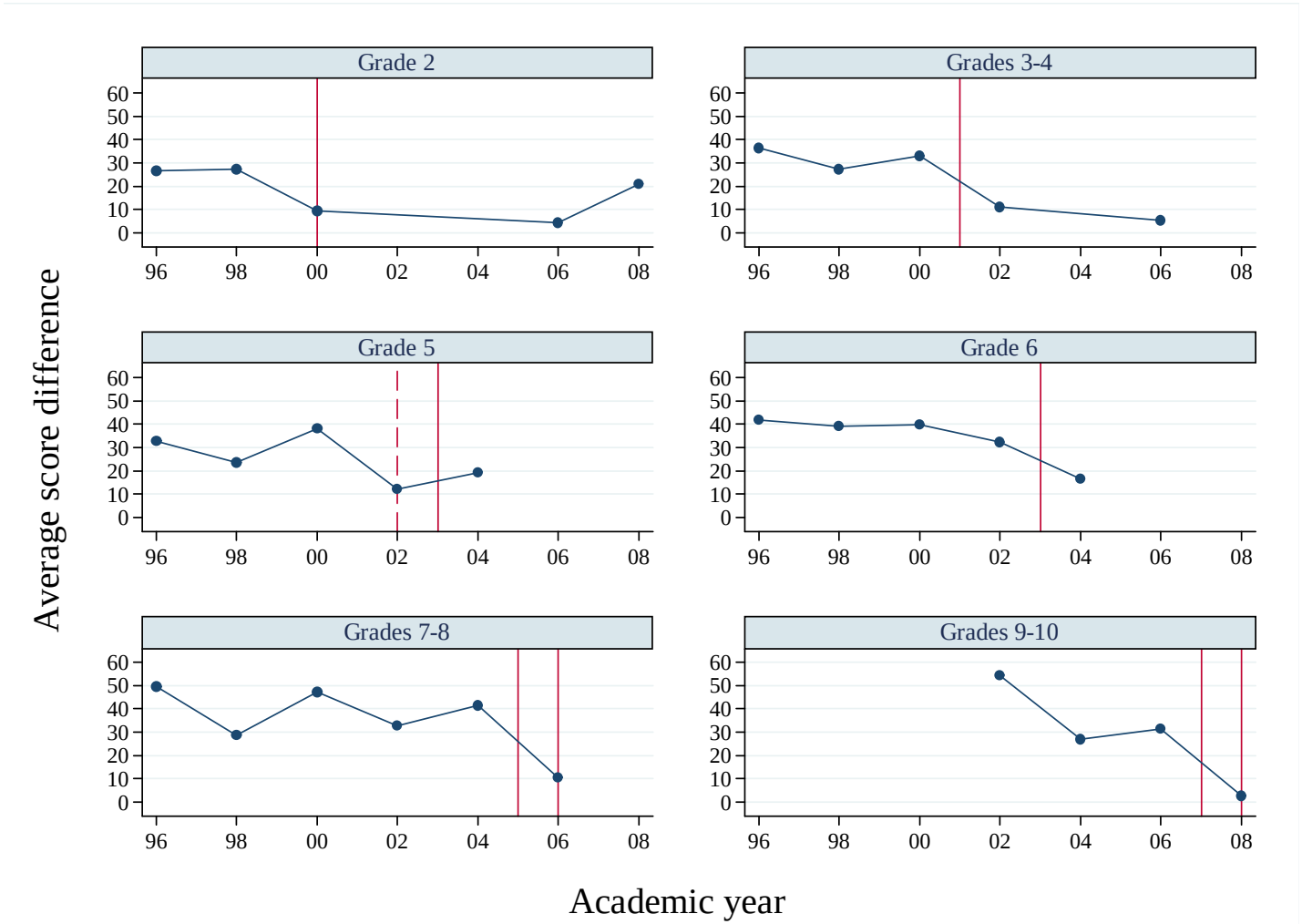
	Québec (treated)		Rest of Canada (control)	
	Mean	Std. dev.	Mean	Std. dev
Student characteristics				
male	0.49	(0.50)	0.51	(0.50)
ppvt (age 4-5)	99.64	(15.11)	100.16	(14.81)
school grade				
2	0.10	(0.30)	0.11	(0.31)
3 and 4	0.22	(0.42)	0.22	(0.41)
5 and 6	0.24	(0.43)	0.23	(0.42)
7 and 8	0.28	(0.45)	0.27	(0.44)
9 and 10	0.16	(0.36)	0.18	(0.38)
Family characteristics				
family structure				
one parent	0.21	(0.41)	0.18	(0.38)
two parents	0.79	(0.41)	0.82	(0.39)
maternal education				
less than secondary	0.16	(0.37)	0.09	(0.29)
secondary	0.24	(0.42)	0.23	(0.42)
some post-secondary	0.18	(0.38)	0.20	(0.40)
college or university	0.41	(0.49)	0.47	(0.50)
mother works (dummy)	0.81	(0.39)	0.84	(0.37)
household income ('000s)	66.92	(47.26)	77.56	(60.44)
area of residence				
rural	0.14	(0.34)	0.13	(0.34)
urban, $\leqslant$30,000	0.15	(0.36)	0.19	(0.39)
urban, 30,000 to 99,999	0.09	(0.29)	0.09	(0.29)
urban, 100,000 to 499,999	0.06	(0.24)	0.21	(0.41)

CAT/2 summary statistics

		Québec (treated)							Rest of Canada (control)						
		1996	1998	2000	2002	2004	2006	2008	1996	1998	2000	2002	2004	2006	2008
GRADE 2															
Years in reform		0	0	1	.	.	2	2	0	0	0	.	.	0	0
CAT/2	Mean	340	320	298	.	.	284	301	313	293	289	.	.	279	280
	Std. dev.	(48)	(40)	(39)	.	.	(38)	(39)	(47)	(38)	(42)	.	.	(40)	(39)
	% reponse	65	49	91	.	.	94	94	77	52	75	.	.	88	86
.	Nbr. obs	135	114	285	.	.	242	163	683	506	929	.	.	1,065	906
GRADES 3-4															
Years in reform		0	0	0	3	.	3-4	.	0	0	0	0	.	0	.
CAT/2	Mean	431	402	391	377	.	353	.	394	375	358	366	.	347	.
	Std. dev.	(60)	(48)	(53)	(44)	.	(53)	.	(58)	(55)	(52)	(49)	.	(52)	.
	% reponse	70	40	90	91	.	95	.	74	51	81	84	.	88	.
	Nbr. obs.	281	185	396	572	.	787	.	1,377	1,076	1,429	2,022	.	3,862	.
GRADES 5-6															
Years in reform		0	0	0	1-0	5	.	.	0	0	0	0	0	.	.
CAT/2	Mean	507	484	469	465	418	.	.	474	451	431	440	436	.	.
	Std. dev.	(48)	(50)	(53)	(53)	(55)	.	.	(60)	(59)	(56)	(51)	(53)	.	.
	% reponse	73	47	90	95	94	.	.	73	53	77	88	89	.	.
	Nbr. obs.	292	184	307	388	528	.	.	1,322	998	1,275	1,469	1,956	.	.
GRADES 7-8															
Years in reform		0	0	0	0	2-0	7	.	0	0	0	0	0	0	.
CAT/2	Mean	590	533	540	524	517	495	.	540	504	492	492	473	485	.
	Std. dev.	(73)	(57)	(71)	(58)	(60)	(67)	.	(78)	(69)	(73)	(68)	(60)	(70)	.
	% reponse	73	41	89	93	94	91	.	76	45	76	77	89	87	.
	Nbr. obs.	225	175	269	336	354	518	.	1,241	878	1,147	1,236	1,402	1,834	.
GRADES 9-10															
Years in reform		.	.	.	0	0	2-0	9	.	.	.	0	0	0	0
CAT/2	Mean	.	.	.	637	606	605	599	.	.	.	583	579	567	596
	Std. dev.	.	.	.	(90)	(82)	(87)	(86)	.	.	.	(85)	(91)	(81)	(88)
	% reponse	.	.	.	87	92	91	84	.	.	.	66	84	86	82
	Nbr. obs.	.	.	.	184	221	261	343	.	.	.	864	1,166	1,209	1,538

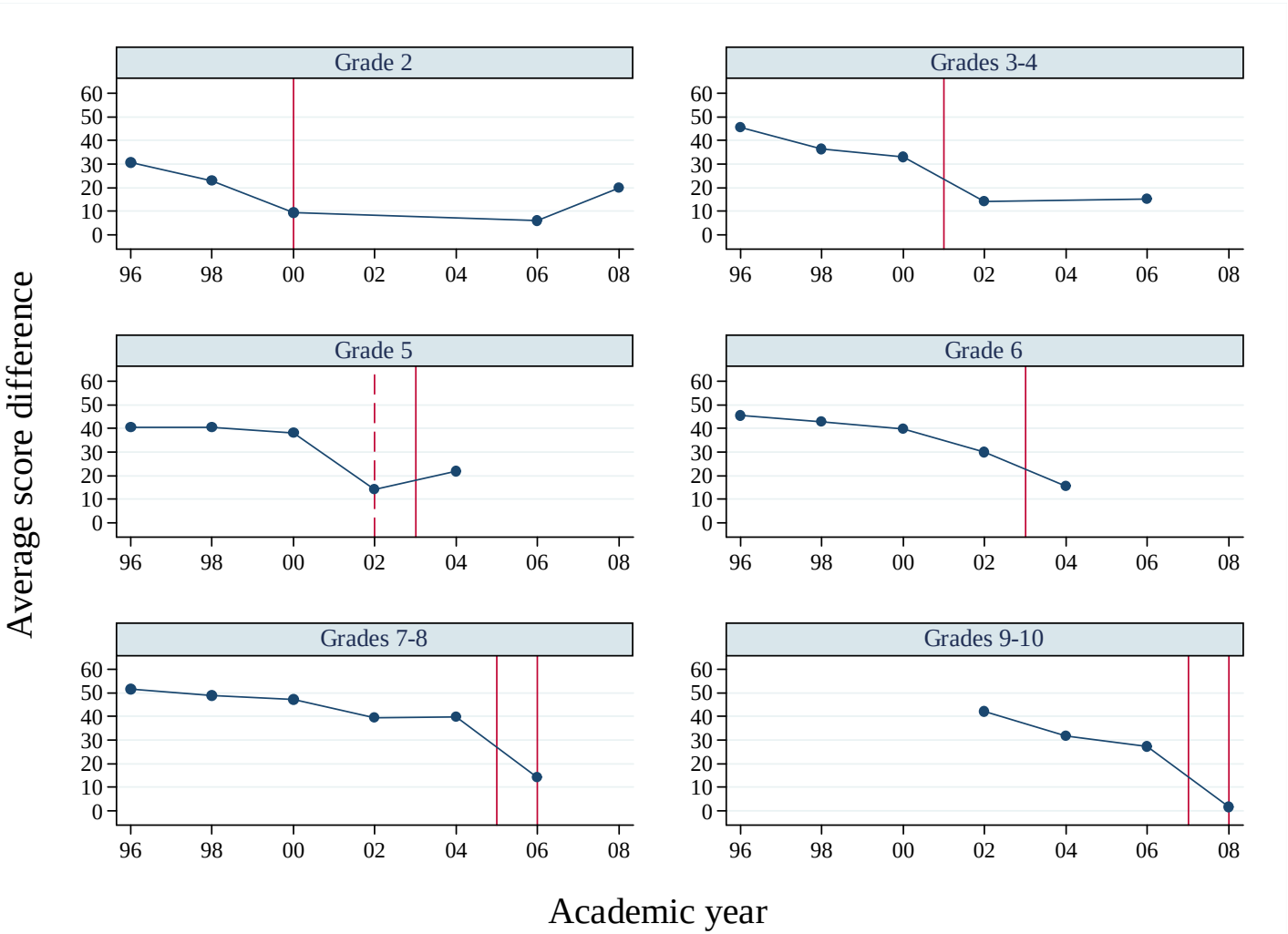
Trends in average score differences... are they stable pre reform?

No matching



Trends in average score differences... are they stable pre reform?

Matching



DID and MDID Estimates - Grades 2 to 6

	Base specifications		Alternative specifications					
	DID		DID			MDID		
	without covariates	with covariates	with PPVT	impute m(10)	excl. Ontario	kernel	llr	nn (5)
GRADE 2								
Years 1996, 2000								
Coef.	-17.32**	-21.18***	-14.23	-15.64**	-21.04**	-15.68***	-14.13***	-13.75***
Std.err.	(8.38)	(8.17)	(9.27)	(7.88)	(9.01)	(3.49)	(4.44)	(4.81)
Nbr.obs	2,032		1,668	2,620	1,502	1,139		
Years 1996, 2006								
Coef.	-22.29***	-24.31***	-19.89**	-18.51**	-32.70***	-22.39***	-25.40***	-15.62***
Std.err.	(8.37)	(8.09)	(9.33)	(7.97)	(8.57)	(3.01)	(3.58)	(3.79)
Nbr.obs	2,125		1,809	2,543	1,510	968		
Years 1996, 2008								
Coef.	-6.06	-11.07	-5.13	-6.05	-11.51	-9.37**	-10.15*	-6.18
Std.err.	(8.55)	(8.37)	(10.04)	(8.09)	(8.86)	(4.30)	(5.22)	(6.54)
Nbr.obs	1,887		1,565	2,317	1,312	652		
GRADES 3-4								
Years 2000, 2002								
Coef.	-21.97***	-19.64***	-16.80***	-17.40***	-16.21***	-11.92***	-7.08**	-6.57**
Std.err.	(5.03)	(4.43)	(4.50)	(4.56)	(4.73)	(2.46)	(2.85)	(2.97)
Nbr.obs	4,419		4,088	5,194	3,303	2,278		
Years 2000, 2006								
Coef.	-27.60***	-16.66***	-14.12***	-15.13***	-18.03***	-11.05***	-3.38	-3.88
Std.err.	(5.27)	(4.87)	(4.91)	(4.94)	(5.30)	(3.22)	(3.19)	(3.55)
Nbr.obs	6,474		6,115	7,818	4,943	3,133		
GRADES 5-6								
Years 2000, 2002								
Coef.	-13.40**	-13.43**	-12.54*	-12.69**	-9.41	-13.77***	-9.19**	-18.04***
Std.err.	(6.23)	(5.97)	(6.63)	(6.09)	(6.13)	(3.13)	(3.78)	(4.06)
Nbr.obs	3,439		3,139	4,053	2,608	1,550		
Years 2000, 2004								
Coef.	-20.30***	-20.13***	-17.79***	-18.33***	-19.41***	-19.10***	-20.81***	-21.35***
Std.err.	(6.12)	(5.86)	(6.50)	(6.25)	(6.12)	(3.02)	(3.45)	(3.84)
Nbr.obs	4 066		3,632	4,698	3,054	2,103		

DID and MDID Estimates - Grades 7 to 10

	Base specifications DID		Alternative specifications DID			MDID		
	without covariates	with covariates	with PPVT	impute m(10)	excl. Ontario	kernel	llr	nn (5)
GRADES 7-8								
Years 2000, 2006								
Coef.	-36.86***	-33.43***	.	-29.85***	-31.07***	-31.00***	-31.97***	-28.53***
Std.err.	(7.81)	(7.46)	.	(7.70)	(7.71)	(4.47)	(4.97)	(5.84)
Nbr.obs	3,768		.	4,446	2,817	2,059		
Years 2002, 2006								
Coef.	-22.47***	-23.08***	-26.44***	-21.86***	-27.70***	-26.29***	-32.47***	-22.32***
Std.err.	(7.52)	(6.70)	(6.49)	(7.10)	(7.25)	(4.45)	(5.40)	(6.12)
Nbr.obs	3,924		3,471	4,599	2,944	2,067		
Years 2004, 2006								
Coef.	-31.25***	-33.89***	-32.40***	-30.66***	-38.44***	-22.11***	-23.84***	-19.81***
Std.err.	(7.49)	(6.74)	(6.94)	(7.14)	(7.40)	(4.40)	(4.64)	(5.42)
Nbr.obs	4,108		3,803	4,627	3,059	2,065		
GRADES 9-10								
Years 2002, 2008								
Coef.	-51.53***	-45.12***	.	-40.49***	-31.58***	-23.93***	-34.61***	-31.44***
Std.err.	(12.41)	(10.75)	.	(10.76)	(11.27)	(6.24)	(8.29)	(7.39)
Nbr.obs.	2,929		.	3,719	2,102	1,364		
Years 2004, 2008								
Coef.	-23.98**	-28.58***	-29.30***	-24.04**	-25.30**	-26.39***	-29.17***	-22.79***
Std.err.	(11.47)	(10.73)	(11.24)	(10.42)	(11.56)	(6.19)	(7.08)	(6.85)
Nbr.obs	3,268		2,948	3,832	2,347	1,360		
Years 2006, 2008								
Coef.	-28.63**	-33.75***	-39.44***	-28.14**	-18.72	-9.00	-7.12	-6.67
Std.err.	(12.08)	(11.23)	(11.58)	(11.21)	(11.71)	(6.10)	(6.63)	(6.65)
Nbr.obs	3,351		3,093	3,938	2,404	1,363		

DID and MDID Estimates

Mean effects with and without covariates are comparable and generally increasing with grade.

Results are robust to:

1. controlling for the PPVT at age 4-5 (heterogeneity in ability).
2. imputation of missing CAT/2 scores (non response).
3. exclusion of Ontario students.
4. controlling linearly for X (DID vs MDID comparable).

CIC Estimates - Grades 2 to 6

	Mean	(Std.err.)	25th Perc.	(Std.err.)	50th Perc.	(Std.err.)	75th Perc.	(Std.err.)	90th Perc.	(Std.err.)
GRADE 2										
Years 1996, 2000										
CIC	-17.04**	(7.51)	-14.00	(8.97)	-10.00	(12.60)	-28.00**	(12.44)	-13.00	(13.36)
CIC lower	-17.81**	(7.53)	-14.00	(9.08)	-10.00	(12.65)	-28.00**	(12.93)	-13.00	(14.12)
CIC upper	-16.48**	(7.50)	-14.00	(8.92)	-10.00	(12.60)	-25.00**	(12.34)	-13.00	(13.36)
Years 1996, 2006										
CIC	-21.22***	(8.00)	-8.00	(6.84)	-14.00	(14.84)	-37.00**	(14.74)	-31.00**	(13.25)
CIC lower	-22.02***	(8.00)	-9.00	(6.96)	-14.00	(14.81)	-37.00**	(14.71)	-39.00***	(13.31)
CIC upper	-20.58**	(7.99)	-8.00	(6.82)	-14.00	(14.89)	-37.00**	(14.85)	-28.00**	(13.55)
Years 1996, 2008										
CIC	-5.43	(7.92)	2.00	(7.31)	-1.00	(13.36)	-18.00	(13.48)	0.00	(18.14)
CIC lower	-6.23	(7.97)	-1.00	(7.49)	-1.00	(13.52)	-18.00	(13.51)	-5.00	(18.34)
CIC upper	-4.83	(7.89)	2.00	(7.27)	-1.00	(13.31)	-18.00	(13.63)	0.00	(18.49)
GRADES 3-4										
Years 2000, 2002										
CIC	-15.17***	(4.18)	-16.00***	(5.91)	-17.00***	(4.77)	-11.00**	(4.76)	-15.00*	(8.10)
CIC lower	-15.59***	(4.19)	-16.00***	(6.04)	-18.00***	(4.90)	-11.00**	(4.91)	-15.00*	(8.19)
CIC upper	-14.71***	(4.18)	-16.00***	(5.92)	-17.00***	(4.82)	-11.00**	(4.74)	-15.00*	(8.08)
Years 2000, 2006										
CIC	-16.72***	(5.01)	-21.00***	(6.29)	-16.00**	(7.25)	-9.00	(5.84)	-11.00	(7.87)
CIC lower	-17.22***	(5.01)	-22.00***	(6.30)	-16.00**	(7.35)	-9.00	(5.98)	-11.00	(7.94)
CIC upper	-16.14***	(5.01)	-21.00***	(6.30)	-16.00**	(7.20)	-9.00	(5.78)	-9.00	(7.87)
GRADES 5-6										
Years 2000, 2002										
CIC	-9.47	(6.05)	-10.00*	(5.62)	0.00	(7.80)	-17.00	(10.63)	-16.00	(11.44)
CIC lower	-10.03*	(6.07)	-10.00*	(5.65)	0.00	(7.89)	-17.00	(10.68)	-16.00	(11.34)
CIC upper	-8.84	(6.04)	-10.00*	(5.65)	0.00	(7.78)	-17.00	(10.79)	-16.00	(11.55)
Years 2000, 2004										
CIC	-19.46***	(6.23)	-13.00*	(6.83)	-14.00	(9.03)	-35.00***	(10.32)	-25.00**	(11.80)
CIC lower	-20.09***	(6.23)	-14.00**	(6.98)	-14.00	(9.09)	-35.00***	(10.45)	-25.00**	(11.89)
CIC upper	-18.85***	(6.22)	-13.00*	(6.77)	-14.00	(9.02)	-32.00***	(10.24)	-24.00**	(11.77)

CIC Estimates - Grades 7 to 10

	Mean	(Std.err.)	25th Perc.	(Std.err.)	50th Perc.	(Std.err.)	75th Perc.	(Std.err.)	90th Perc.	(Std.err.)
GRADES 7-8										
Years 2000, 2006										
CIC	-29.82***	(7.36)	-28.00**	(11.24)	-39.00***	(10.28)	-22.00**	(10.74)	-6.00	(14.65)
CIC lower	-30.28***	(7.38)	-28.00**	(11.25)	-39.00***	(10.30)	-22.00**	(10.79)	-19.00	(14.73)
CIC upper	-29.24***	(7.34)	-24.00**	(11.26)	-39.00***	(10.30)	-21.00**	(10.70)	-6.00	(14.69)
Years 2002, 2006										
CIC	-23.74***	(7.34)	-25.00***	(7.58)	-24.00***	(8.75)	-12.00	(11.20)	-1.00	(12.78)
CIC lower	-24.35***	(7.36)	-26.00***	(7.55)	-26.00***	(8.85)	-12.00	(11.20)	-5.00	(13.13)
CIC upper	-23.13***	(7.33)	-25.00***	(7.61)	-23.00***	(8.75)	-12.00	(11.26)	-1.00	(12.68)
Years 2004, 2006										
CIC	-34.47***	(7.35)	-34.00***	(8.81)	-35.00***	(10.26)	-26.00**	(13.25)	-32.00*	(17.12)
CIC lower	-35.01***	(7.36)	-35.00***	(8.79)	-35.00***	(10.30)	-26.00*	(13.58)	-32.00*	(17.19)
CIC upper	-33.94***	(7.35)	-33.00***	(8.85)	-35.00***	(10.27)	-26.00**	(13.20)	-32.00*	(17.10)
GRADES 9-10										
Years 2002 2008										
CIC disc ci	-43.51***	(11.11)	-66.00***	(23.37)	-46.00***	(16.78)	-47.00***	(15.85)	-24.00	(17.09)
CIC disc lower	-43.91***	(11.11)	-66.00***	(23.38)	-48.00***	(16.88)	-47.00***	(15.84)	-24.00	(17.11)
CIC disc upper	-43.13***	(11.10)	-65.00***	(23.46)	-46.00***	(16.78)	-47.00***	(15.86)	-24.00	(17.13)
Years 2004, 2008										
CIC	-26.87***	(10.29)	-37.00***	(14.14)	-24.00	(15.12)	-27.00*	(14.22)	-16.00	(15.67)
CIC lower	-27.45***	(10.28)	-37.00***	(14.17)	-24.00	(15.15)	-27.00*	(14.14)	-17.00	(15.56)
CIC upper	-26.37**	(10.30)	-36.00**	(14.18)	-24.00	(15.23)	-27.00*	(14.40)	-16.00	(15.85)
Years 2006, 2008										
CIC	-33.57***	(11.27)	-37.00***	(14.17)	-40.00***	(15.02)	-36.00**	(15.97)	-42.00*	(23.07)
CIC lower	-34.13***	(11.27)	-38.00***	(14.22)	-41.00***	(14.86)	-36.00**	(16.08)	-42.00*	(23.23)
CIC upper	-32.95***	(11.27)	-37.00***	(14.13)	-40.00***	(15.14)	-34.00**	(15.98)	-42.00*	(23.07)

Impact of exposure

Focusing on students entering grade 1 in 1999 or 2000, we find that the magnitude of the effect **increases sigmoidally with exposure** to the reform.

- Grade 2: mean CIC is 17.0 (5.0% of mean score)
- Grades 3-4: mean CIC is 15.2 (3.9% of mean score)
- Grades 5-6: mean CIC is 19.5 (4.1% of mean score)
- Grades 7-8: mean CIC is 23.7 to 34.5 (4.5% to 6.6% of mean score)
- Grades 9-10: mean CIC is 26.9 to 43.5 (4.4% to 6.7% of mean score)

Short term versus long term effects

We find that grade 2 students, **8 years after** the implementation of the reform, no longer seem to experience a significant negative effect (the CIC estimator for academic years 2008 is small and not different from zero). Possible explanations:

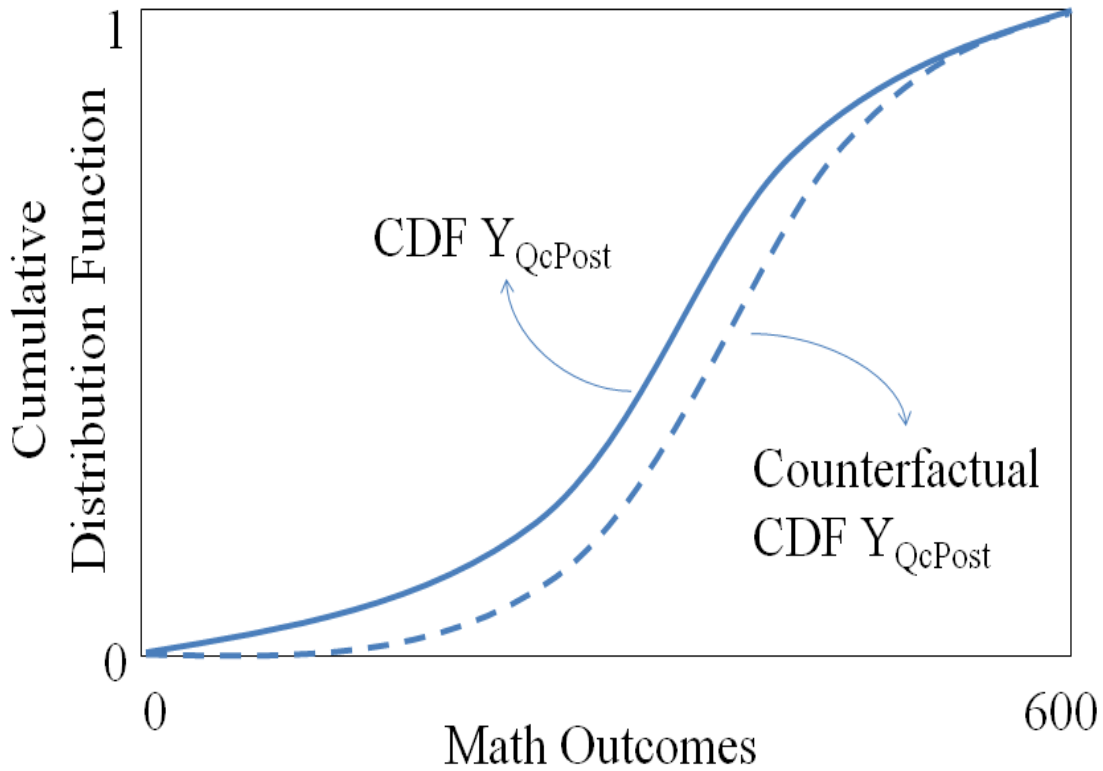
- took a fair number of years for teachers to develop the necessary skills to fully deploy all aspects of the reform.
- observing the decline in students' academic performance, teachers informally decided to reintroduce some of their pre-reform teaching approaches, and set aside in part or in totality the reform approach.

***This conclusion is derived from one set of grade 2 students at one point in time. Further research is needed to fully understand the long term effects of the reform.

Across the skill distribution

Results hold true for

- lower and middle performing students
- top performing students (but often not significant)



The entire distribution was impacted negatively.

Is this impact material?

We know the correlation between the grades 5-6 score and adulthood outcomes... we found that the mean effects on grades 5-6 scores (years 2000, 2004) was about -19.46...

Assuming the marginal rate is constant, this decrease in math score would correspond to a

- 1.29% decrease in the probability of dropping out of high school.
- 2.39% increase in the probability of having a University degree.
- 494$ increase in total personal income at age 24-25.

Falsification tests: were the observed trends really flat?

	Coef..	Std.err.	Nbr.obs.	All groups
GRADE 2				
2000-2006	-11.02*	(5.65)	1,788	after
2000-2008	9.49	(6.19)	1,590	after
2006-2008	21.22***	(5.93)	1,598	after
GRADES 3-4				
1996-2000	-3.50	(5.74)	2,642	before
2002-2006	-2.02	(4.61)	5,480	after
GRADES 5-6				
1996-2000	1.41	(6.55)	2,443	before
2002-2004	-8.75*	(5.32)	3,248	after*
GRADES 7-8				
2000-2002	-1.61	(7.39)	2,281	before
2002-2004	9.73	(7.91)	2,523	before
2000-2004	6.09	(8.64)	2,396	before
GRADES 9-10				
2002-2004	-8.93	(12.28)	1,811	before
2004-2006	-5.75	(13.28)	2,113	before
2002-2006	-10.86	(12.16)	1,868	before

Are these results specific to the NLSCY? Further evidence from TIMSS.

	Mathematics Achievement Grade 4				Mathematics Achievement Grade 8			
Year	1995	1999	2003	2007	1995	1999	2003	2007
International	500	500	500	500	500	500	500	500
Québec	550	-	506	519	556	566	543	528
Ontario	489	-	511	512	501	517	521	517
Alberta	523	-	-	505	-	-	-	-
British Columbia	-	-	-	505	-	522	-	509
	Coef.	Std.err.	t-stat	N	Coef.	Std.err.	t-stat	N
DID estimate	-40.83	20.08	-2.03	8	-31.00	18.07	-1.72	8

	Science Achievement Grade 4				Science Achievement Grade 8			
Year	1995	1999	2003	2007	1995	1999	2003	2007
International	500	500	500	500	500	500	500	500
Québec	529	-	500	517	510	540	531	507
Ontario	516	-	540	536	496	518	533	536
Alberta	555	-	-	543	-	-	-	-
British Columbia	-	-	-	537	-	542	-	526
	Coef.	Std.err.	t-stat	N	Coef.	Std.err.	t-stat	N
DID estimate	-24.67	23.29	-1.06	8	-28.75	24.56	-1.17	10

Conclusion

Main findings:

- Negative impact across the entire distribution
- The longer the exposure the stronger the impact
- Long run effects may be nulle

Since the least performing students were impacted negatively, it appears that on procedural mathematics skills the reform failed to meet its main objective which was to raise the proportion of students who were successful in school.

$\rightarrow$This also holds true more generally, since the expected dropout rate for the first completely treated reform cohort is expected to be flat versus the past 10 years (MELS).

Limits:

- Other outcomes may be important: reading, science, behaviour.
- Long term effects may be different... only one grade 2 cohort

MERCI