

Unintended Consequences of School Accountability Reforms: Evidence from Australia

Michael Coelli¹ and Gigi Foster²

¹Department of Economics, University of Melbourne

²Department of Economics, University of NSW

ACE Online – 14 July, 2021

Research Questions

- 1 Do schools respond to the public provision of test score information by attempting to influence the pool of tested students?
- 2 Do private schools respond to such public provision differently to government (public) schools?

Motivation

- Objectives of public provision of school performance data:
 - (1) to lower information barriers for parents, and
 - (2) to increase public accountability of schools.
- Schools may respond to accountability pressures via:
 - ① improving the overall quality of instruction,
 - ② concentrating on tested subjects / “teaching to the test”,
 - ③ focus efforts on students with the largest “payoff”,
 - ④ outright cheating (changing students’ answers), or
 - ⑤ altering the pool of test-takers.
- Responses may be stronger among private schools,
–if “market” pressures are greater.

The Accountability Intervention under Investigation

- 1 The **My School** website was launched on 28 January 2010.
- 2 It provided searchable school-level information on “ALL” Australian schools (but no tables of rankings).
- 3 School-average literacy and numeracy test scores are provided, – along with clear comparisons to all other schools and to “similar” schools.

Students may not sit these NAPLAN tests for three reasons:

- (1) **exemption** – learning disability / recent NESB (CALD) migrant,
- (2) **absence** on testing days, or simply
- (3) **withdrawal** by parental request (in writing).

Other School Responses

Coelli, Foster and Leigh (2018) conducted surveys of school principals pre (2009) and post (2012) the launch of **My School**.

We asked about a number of potential school attributes:

- length of school day,
- time spent on literacy and numeracy,
- teacher pay and incentives,
- extra assistance to poor performing students or teachers,
- additional classroom resources,
- PLUS many others.

There was little evidence of consistent responses to **My School**,
– even among poor-performing schools.

Literature – Unintended Responses to Accountability

- Jacob and Levitt (2003) – teachers completed student exams.
- Figlio and Winicki (2005) – changed meals program.
- Figlio (2006) – changed discipline / suspension patterns.
- Cullen and Reback (2006); Figlio and Getzler (2007)
– selective exclusion of students from testing
(assign low-performing students to “special education”).
- Bokhari and Schneider (2009) – use of psychostimulants.

Existing Literature – Accountability and Private Schools

Most accountability interventions have only affected public schools.

– see Figlio and Loeb (2010) for a survey; also Rouse et al. (2013).

A similar school accountability intervention in Brazil also affected private schools.

– see Firpo et al. (2014) and Camargo et al. (2015).

Public identification of “effective” schools in Chile also included private schools.

– See Mizala and Urquiola (2013).

None of these studies investigated responses via changes in the pool of tested students.

Estimation Strategy

- We focus on within-school changes from before to after the introduction of **My School**.
- We compare changes within schools that were revealed to have poor **initial** performance relative to changes within other schools that had performed better.
- We investigate whether responses differ by school sector:
 - public (Government) vs. private (Catholic / Independent).
- We focus on non-assessment due to withdrawal:
 - potentially the least costly path.

“Withdrawals are intended to address issues such as religious beliefs and philosophical objections to testing.”

Australian School System

- 1 One third of Australian school students attend private schools, which are mostly religious-based:
 - half Catholic system, half “Independent”.
- 2 These private (non-profit) schools receive public funding:
 - approx. 70% of the per-student funding of public schools.
- 3 Some private schools are high fee, selective and high prestige,
 - but many are not, particularly Catholic system schools.
- 4 The Australian school year runs from late January to mid-December.

Literacy and Numeracy Testing

- 1 National testing via NAPLAN was introduced in 2008.
- 2 Five domains are tested:
reading, persuasive writing, spelling, grammar and punctuation,
and numeracy.
- 3 Students are tested in grades 3, 5, 7 and 9 in May each year.
- 4 Prior to 2008, testing conducted separately by the states
(states have responsibility for public schools in Australia).

Accountability Intervention Details

- 1 **My School** website – launched on 28 January 2010.
- 2 Provides searchable school-level information on “ALL” schools.
- 3 NAPLAN mean scores for each domain and grade are provided, with comparisons to (a) all schools and to (b) “similar” schools.
“Similar” schools identified using:
 - parental background of students (education and occupation),
 - percentage of indigenous students, and
 - remoteness.
- 4 Red / green flags highlight performance below / above comparison schools.
- 5 Information on enrollment, school finances, parental background of students and general school profiles are also provided.

St Bernard's Primary School, Coburg, VIC

Results in numbers

The National Assessment Program – Literacy and Numeracy (NAPLAN) assesses all students in Australian schools in Years 3, 5, 7 and 9. For more information visit the [NAPLAN website](#).

The chart below displays average NAPLAN scores for each [domain](#). The selected school's scores are displayed in blue. Also displayed are average scores for statistically similar schools (SIM) and all Australian schools (ALL). The coloured bars indicate whether the selected school's scores are above, close to, or below the other scores.

	2008	2009	2010	2011	2012	2013	2014			
Colour Scheme	Red & Green		<input type="button" value="Submit"/>		Alternate view: Results in graphs					
	Reading		Narrative Writing		Spelling		Grammar and Punctuation		Numeracy	
Year 3	401 379 - 423		435 417 - 453		415 395 - 435		398 375 - 421		370 352 - 388	
	SIM 428 419 - 437	ALL 411	SIM 427 419 - 435	ALL 414	SIM 418 409 - 427	ALL 405	SIM 432 422 - 442	ALL 420	SIM 408 400 - 416	ALL 394
	502 482 - 522		512 493 - 531		489 471 - 507		525 503 - 547		465 448 - 482	
Year 5	509 500 - 518		497 489 - 505		497 489 - 505		514 505 - 523		499 491 - 507	
	SIM 509 500 - 518	ALL 494	SIM 497 489 - 505	ALL 485	SIM 497 489 - 505	ALL 487	SIM 514 505 - 523	ALL 500	SIM 499 491 - 507	ALL 487
	502 482 - 522		512 493 - 531		489 471 - 507		525 503 - 547		465 448 - 482	

Expected Timing of Responses

- **My School** launched essentially on first day of school in 2010.
- Parents (and thus school enrollments) are unlikely to respond until 2011.
- Schools were aware of website launch in late 2009,
 - thus any responses are not likely until 2010 at the earliest.

NOTES:

Any school closures or openings relate to demographic changes rather than poor test performance.

My School was controversial:

– opposed by teachers, principals and others.

Table 1a - School Summary Statistics by Sector, 2008

Statistic	Govt.	Indep.	Catholic	Total
Total school enrolment	422.7 (324.8)	610.8*** (489.8)	418.8 (312.3)	442.8 (349.6)
Combined school (%)	4.9	76.2***	7.3***	13.3
Students per tested grade	66.1 (61.3)	56.8*** (45.8)	63.8 (54.6)	64.6 (58.5)
Student / faculty ratio	15.4 (3.0)	13.1*** (3.9)	16.3*** (2.9)	15.3 (3.2)
Total funding (\$A, 2009) per student	10,954 (3,053)	13,602*** (4,459)	9,954*** (2,495)	11,040 (3,289)
Fees / contributions (\$A, 2009) per student	357 (255)	6,130*** (5,152)	1,696*** (1,433)	1,274 (2,574)
Parental socio-economic status (standardised, 2010)	-0.239 (0.981)	0.901*** (0.872)	0.390*** (0.710)	0.011 (0.998)
Schools	4,905	798	1,492	7,195

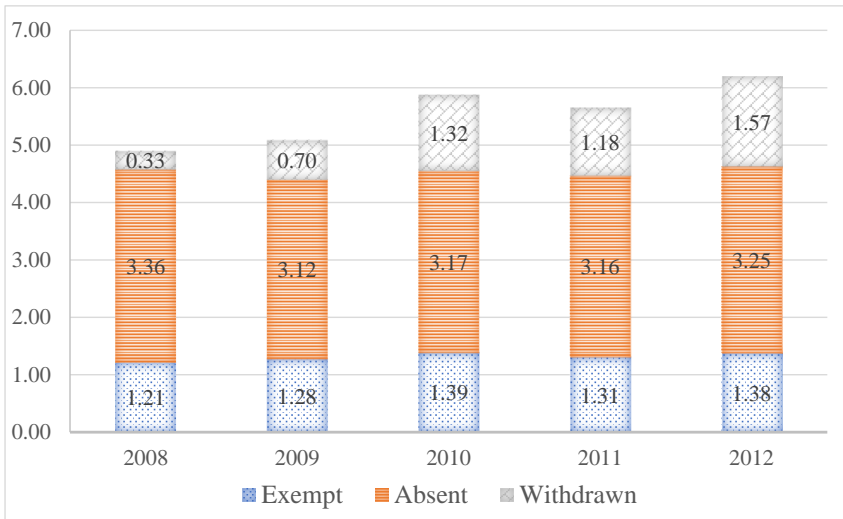
Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, difference versus Government schools.

Table 1b - School Summary Statistics by Sector, 2008

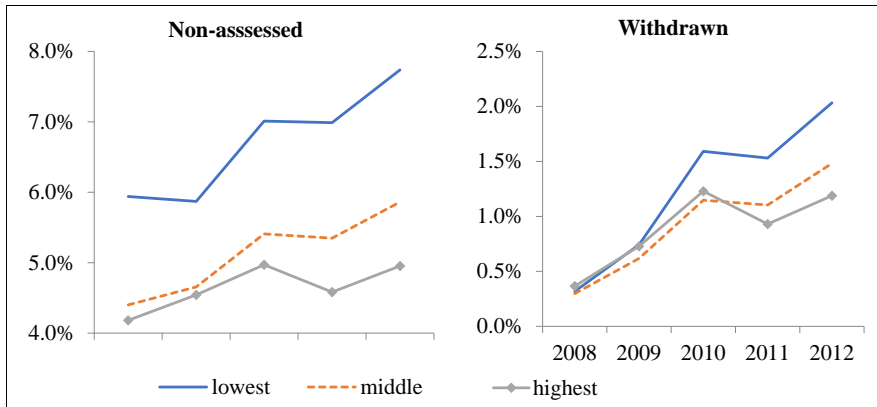
Statistic	Govt.	Indep.	Catholic	Total
Average test scores (2008/09) (standardised, all schools)	-0.095 (0.415)	0.263 (0.388)	0.153 (0.283)	-0.004 (0.412)
Average test scores (2008/09) (standardised, similar schools)	-0.034 (0.256)	0.110 (0.235)	0.087 (0.210)	0.007 (0.252)
Non-assessed percentage	5.50 (5.44)	3.92 (6.29)	3.10 (3.27)	4.83 (5.27)
Absent percentage	3.86 (4.23)	2.81 (4.55)	2.07 (2.40)	3.37 (4.03)
Exempt percentage	1.41 (2.98)	0.56 (1.48)	0.88 (2.01)	1.21 (2.69)
Withdrawn percentage	0.31 (1.34)	0.61 (3.55)	0.22 (0.86)	0.33 (1.67)
Schools	4,905	798	1,492	7,195

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, difference versus Government schools.

Non-assessed Percentages by Type



Non-assessed Percentages by Initial Performance



Distribution of Withdrawal over Time

Withdrawal	2008	2009	2010	2011	2012
Mean	0.33	0.70	1.32	1.18	1.57
Standard deviation	1.67	3.04	3.60	3.40	3.94
Proportion zero	0.84	0.78	0.62	0.63	0.55
75th percentile	0.00	0.00	1.00	1.00	1.67
90th percentile	0.60	1.67	4.00	3.50	4.50

Estimation Preliminaries

We compare within-school changes in outcomes of schools with poor versus good initial performance.

- Initial performance is averaged over 2008 and 2009,
 - the initial information revelation.
- We investigate responses to performance relative to both:
 - (1) all schools and
 - (2) “similar” schools.
- We combine normalised scores into summary measures:
 - across all 5 testing domains, and
 - across all grades tested.

Estimating Equation

$$y_{it} = \gamma_i + \sum_{t=2009}^{2012} [\alpha_t(\mu_t) + \alpha_t^I(\mu_t \cdot I_i) + \alpha_t^C(\mu_t \cdot C_i) + \beta_t(s_i \cdot \mu_t) + \beta_t^I(s_i \cdot \mu_t \cdot I_i) + \beta_t^C(s_i \cdot \mu_t \cdot C_i)] + \varepsilon_{it}$$

y_{it} = outcome (e.g. withdrawn percentage) in school i in year t

γ_i = school i fixed effect

μ_t = year t indicator

$I_i = 1$ if school i is an Independent private school

$C_i = 1$ if school i is a Catholic private school

s_i = test score performance for school i averaged over 2008 and 2009

ε_{it} = i.i.d. error term

Estimated Equation - Differential Effects by Sector

We are interested in the estimated β_t coefficients.

They denote responses to “initial” performance relative to the base year of 2008.

The β_t coefficients denote responses by government (public) schools.

The β_t^I and β_t^C coefficients denote whether Independent and Catholic school responses differ from government school responses.

The Betas and Alphas for Public Schools

	Non-assessed		Withdrawn vs Similar
	vs All	vs Similar	
Score × 2009	0.556 (0.400)	0.530** (0.217)	0.103 (0.152)
Score × 2010	-0.411 (0.430)	-0.054 (0.207)	-0.332 (0.207)
Score × 2011	-1.591*** (0.410)	-0.548** (0.225)	-0.988*** (0.194)
Score × 2012	-1.673*** (0.412)	-1.278*** (0.229)	-1.114*** (0.199)
2009	0.157** (0.0726)	0.189*** (0.0707)	0.399*** (0.0411)
2010	1.015*** (0.0781)	1.024*** (0.0755)	1.103*** (0.0522)
2011	0.826*** (0.0777)	0.829*** (0.0755)	0.904*** (0.0453)
2012	1.481*** (0.0814)	1.417*** (0.0790)	1.273*** (0.0486)

The Betas – Sectoral Differences in Score Responses

	Non-assessed		Withdrawn vs Similar
	vs All	vs Similar	
Score × 2009 × Independent	-0.592 (1.082)	-0.333 (0.686)	-1.550 (1.068)
Score × 2010 × Independent	-1.697 (1.288)	-0.870 (0.873)	-2.859*** (1.023)
Score × 2011 × Independent	-0.249 (1.765)	-0.179 (0.821)	-2.945** (1.287)
Score × 2012 × Independent	-1.003 (1.884)	-0.082 (0.963)	-4.135*** (1.570)
Score × 2009 × Catholic	0.648 (0.878)	0.009 (0.553)	0.229 (0.296)
Score × 2010 × Catholic	-0.246 (0.687)	-0.501 (0.441)	0.289 (0.366)
Score × 2011 × Catholic	1.831*** (0.647)	0.270 (0.445)	0.485 (0.539)
Score × 2012 × Catholic	1.258* (0.652)	0.232 (0.455)	-0.0782 (0.576)

The Alphas – Sectoral Differences in Time Effects

	Non-assessed		Withdrawn vs Similar
	vs All	vs Similar	
2009 × Independent	0.039 (0.287)	-0.048 (0.324)	0.321 (0.271)
2010 × Independent	0.139 (0.312)	0.140 (0.382)	0.300 (0.275)
2011 × Independent	-0.071 (0.372)	-0.086 (0.389)	0.488* (0.291)
2012 × Independent	-0.128 (0.390)	-0.002 (0.435)	0.902** (0.351)
2009 × Catholic	0.071 (0.141)	0.062 (0.156)	-0.253*** (0.0552)
2010 × Catholic	-0.226* (0.133)	-0.207 (0.146)	-0.539*** (0.0755)
2011 × Catholic	-0.190 (0.136)	-0.129 (0.147)	-0.317*** (0.0849)
2012 × Catholic	-0.598*** (0.141)	-0.409*** (0.158)	-0.369*** (0.0952)

Which Students Are Being Withdrawn?

Can we say for certain that it was the low-performing students being withdrawn?

- We only have individual student data from 2013:
 - withdrawn students performed poorly in prior (2011) tests,
 - but we cannot separate Independents from Catholic.
- We also look at the school-level relationship between:
 - (1) the proportion of students scoring in the lowest band in 2012,
 - (2) with the change in non-assessed (withdrawn) percentage, controlling for initial (2008) proportion scoring in lowest band.
- We think of both as “smoking gun” evidence.

2013 Year 5 Participation – Scores in Prior (year 3) Tests

	Standardised score	Proportion lowest band	Any in lowest band	Class rank (0 to 1 = top)
Absent	-0.442*** (0.013)	0.0537*** (0.0028)	0.126*** (0.0061)	-0.113*** (0.0038)
Withdrawn	-1.095*** (0.0249)	0.203*** (0.0079)	0.430*** (0.0138)	-0.308*** (0.0065)
Exempt	-1.403*** (0.0340)	0.305*** (0.0138)	0.579*** (0.0211)	-0.371*** (0.0079)
Private	0.226*** (0.0119)	-0.0232*** (0.00087)	-0.0614*** (0.0024)	-0.00459*** (0.0004)
Absent * Private	0.150*** (0.0246)	-0.0286*** (0.00439)	-0.0609*** (0.0101)	0.0375*** (0.0078)
Withdrawn * Private	0.0511 (0.0570)	-0.0646*** (0.0149)	-0.0823** (0.0321)	0.00155 (0.0165)
Exempt * Private	-0.0163 (0.0732)	-0.0644** (0.0287)	-0.0767 (0.0477)	-0.0117 (0.0172)
Constant	-0.0373*** (0.0079)	0.0403*** (0.00072)	0.121*** (0.0018)	0.522*** (0.00028)

Proportion in Lowest Band

	(1)	(2)	(3)	(4)
Δ non-assessed %	-0.056*** (0.014)	-0.057*** (0.018)		
Δ non-assessed % \times Indep.		-0.013 (0.032)		
Δ non-assessed % \times Catholic		0.026 (0.032)		
Δ withdrawn %			-0.106*** (0.020)	-0.117*** (0.029)
Δ withdrawn % \times Indep.				0.015 (0.041)
Δ withdrawn % \times Catholic				0.088 (0.074)
Percent lowest band, 2008	0.809*** (0.014)	0.790*** (0.015)	0.807*** (0.014)	0.788*** (0.015)
Sector indicators	No	Yes	No	Yes
Schools	7,195	7,195	7,195	7,195
R-squared	0.613	0.617	0.615	0.618

Other Findings

- **Enrollment** growth was stronger among better-performing schools,
 - but this trend was evident prior to *My School*.
- **Exemption** effects only marginally negative among public schools,
 - marginally positive among Catholic and Independent schools.
- **Absence** effects not evident among public schools,
 - marginally positive among Independent schools.

Effect Sizes and Potential Impact

- Number of students affected:
 - counterfactual exercise implies 1,070 fewer students tested,
 - only 0.1% of tested students, but 8% of withdrawal.
- Effect of removing **one** marginal “lowest band” student from testing:
 - raises school average standardised score by 0.045 on average,
 - greater than a 0.09 increase in 10% of schools,
 - 23% of schools with a dark red flag would “lose” it,
 - 20% of schools with a light red flag would “lose” it.

Concluding Remarks

- There is evidence of larger increases in withdrawal of students from testing in schools revealed on **My School** to be low-performing.
- This is consistent with schools “gaming” the testing regime, – particularly independent private schools.
- It is also consistent with individual responses to increased pressures to “perform” on tests.
- Either way, the potential benefits of testing and information provision are being undermined.