CITS Design for No Child Left Behind Evaluation

Manyee Wong, Thomas D. Cook & Peter M. Steiner

JREE, 2014
NCLB main Program Components

• Introduced 2002 and applied to all public schools nationally – national program
• Goal – all children proficient by 2014
• Testing Component – to decide school-level performance in 38 categories and decide if making adequate yearly progress (AYP)
• Graduated Sanctions Component for Repeated Failure -- if fail AYP one year, then...; If fail 5 consecutive years, then ...
NCLB Components

• State Variation in how much they had made similar “consequential accountability” reforms prior to 2002.
• State-level control over proficiency tests – easy or hard and consequences for AYP
• State-level control over rate of progress to 2014 – front-load or rear-load shift in “proficiency standards” – percentage of students in school who had to be proficient by a state-determined date
How would you do SITS

• Hypothesized Change would be in ....?
• Outcome would be measured how ?
• SITS Design would be?
• Internal Validity Threats?
Comparison: Contrasts of Type 1 & 2

• Contrast 1: Test for NCLB effect at national level
  – Compare student achievement in public schools with private schools (both Catholic and also non-Catholic)

• Contrast 2: Test for NCLB effects at the state level
  – Compare states varying in proficiency standards.
  • States with higher proficiency standards are likely to have more schools fail to make AYP and so more schools will need to “reform” to boost student achievement
  -- Compare states varying in timing of consequential accountability
Contrast 1: Public vs Private Schools

• Public schools got NCLB but private ones essentially did not
  – If NCLB is raising achievement in general, then public schools should do better than private ones after 2002

• Hypothesis is that changes in mean, slope or both after 2002 will favor public schools
Hypothetical NCLB effects on public (red) versus private schools (blue)
Contrast 1: Public vs Private Schools

- Two independent datasets: Main and Trend NAEP data can be used to test this
- Main NAEP four posttest points. Data available for both Catholic and other private schools
- Trend NAEP only one usable post-2002 point and then only for Catholic schools
Analytic Model

• NCLB Public vs. Catholic school contrast

• Model

\[ Y_{ij} = \beta_0 + \beta_1 (year)_{ij} + \beta_2 (group)_{j} + \beta_3 (policy)_{ij} + \beta_4 (year \times group)_{ij} + \beta_5 (policy \times year)_{ij} \]
\[ + \beta_6 (policy \times group)_{ij} + \beta_7 (policy \times year \times group)_{ij} + \varepsilon_{ij}, \]

• Low Power

  • Only 3 groups (public, Catholic, non-Catholic private) with 8 time points. So only 24 degrees of freedom

• Autocorrelation

  – Few solutions
  – Cannot use clustering algorithm because there are not enough groups
  – Robust s.e. used but results less conservative
Main NAEP Time Series Graphs

Public vs. Catholic
Public vs. Other Private
Main NAEP 4th grade math scores by year: Public and Catholic schools
Main NAEP 4th grade math scores by year: Public and Other Private schools
Difference in differences in Total change for $4^{th}$ Grade Math
Analyses based on Main NAEP data

<table>
<thead>
<tr>
<th></th>
<th>4th Grade Math (All Data)</th>
<th></th>
<th>4th Grade Math (Exclude 1990)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>S.E.</td>
<td>t</td>
</tr>
<tr>
<td>Public vs. Catholic</td>
<td>10.96</td>
<td>5.22</td>
<td>1.77+</td>
</tr>
<tr>
<td>Public vs. Other Private</td>
<td>6.46</td>
<td>8.39</td>
<td>0.77</td>
</tr>
</tbody>
</table>

+ p<0.10, * p<0.05
Main NAEP 8th grade math scores by year:
Public and Catholic schools
Main NAEP 8\textsuperscript{th} grade math scores by year: Public and Other Private schools
Difference in differences in Total change for 8th Grade Math Analyses based on Main NAEP data

<table>
<thead>
<tr>
<th>Diff in Total Δ (2009)</th>
<th>8th Grade Math (All Data)</th>
<th>8th Grade Math (Exclude 1990)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>S.E.</td>
</tr>
<tr>
<td>Public vs. Catholic</td>
<td>2.91</td>
<td>6.64</td>
</tr>
<tr>
<td>Public vs. Other Private</td>
<td>11.16</td>
<td>7.95</td>
</tr>
</tbody>
</table>

+ p<0.10, * p<0.05
Trend NAEP Time Series Graphs

Public vs. Catholic School Contrast

(Other private school data unavailable and only 1 post-intervention time point)
Trend NAEP 4th grade math scores by year: Public and Catholic schools
Trend NAEP 8th grade math scores by year: Public and Catholic schools
Difference in Differences in Mean Change in 2004 for Math Analyses based on *Trend NAEP* data

Public vs. Catholic Contrast

<table>
<thead>
<tr>
<th>4th Grade Math</th>
<th>Coef.</th>
<th>S.E.</th>
<th>t</th>
<th>8th Grade Math</th>
<th>Coef.</th>
<th>S.E.</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>4th Grade Math</td>
<td>10.93</td>
<td>5.53</td>
<td>1.97*</td>
<td>8th Grade Math</td>
<td>7.26</td>
<td>2.03</td>
<td>3.58*</td>
</tr>
</tbody>
</table>

* p<0.05
Public vs Private School Findings

• All effects on 4th and 8th grade math are in right direction, and some statistically significant

• No effect on 4th grade reading but all are in the right direction (results not shown here)

• This suggests that NCLB has a significant math effect nationally, but...
Concerns with Contrast 1

• Possible low power due to low number of groups (3: public, Catholic, non-Catholic private) with 8 time points
  – So only 24 degrees of freedom
• Catholic sex abuse scandals in 2002 result in parents taking their children out of Catholic schools
• Did this cause a drop in average student achievement in Catholic schools or an increase in public schools since transfers were higher scoring Catholic school attendees
• Did transfers from Catholic schools raise achievement in other private schools?
• So Contrast 1 by itself is not likely to be definitive
Contrast 2a: States and Consequential Accountability (CA)

• Dee & Jacob (2013)
• States with CA prior to 2002 versus after
• Main NAEP scores
• Hypothesis DiD
D & J Results: 4th Grade Math

[Graph showing NAEP scale score over years for two groups: one that adopted school accounting policy by 1998 and one that never adopted before NCLB.]
D & J Results: 8th Grade Math
Contrast 2b: Comparing States that vary in Proficiency Standards

• Some states set high standards for making AYP and so many schools fail and have to change their educational practices
  – (more serious NCLB implementers, higher dosage of treatment)

• Other states set low standards and so do not have to change much
  – (less serious NCLB implementers, lower dosage of treatment)
Define Proficiency Standards Based on the Percentage of Students Deemed Proficient

• To determine a state’s overall level of proficiency standard, we average the percentage of students deemed proficient across grades (4th and 8th grade) and across subjects (math and reading) using state assessment data from 2003

• States that deemed less than 50% of students proficient have **high proficiency standards**

• States that deemed 75% or more proficient are states have **low proficiency standards**

• States between 50% and 75% are **moderate**
## Evidence of Differing Standards

<table>
<thead>
<tr>
<th>State</th>
<th>State Test</th>
<th>NAEP Test</th>
<th>State</th>
<th>State Test</th>
<th>NAEP Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arizona</td>
<td>46</td>
<td>24</td>
<td>Colorado</td>
<td>83</td>
<td>35</td>
</tr>
<tr>
<td>Arkansas</td>
<td>46</td>
<td>25</td>
<td>Connecticut</td>
<td>76</td>
<td>39</td>
</tr>
<tr>
<td>California</td>
<td>36</td>
<td>22</td>
<td>Georgia</td>
<td>76</td>
<td>25</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>48</td>
<td>8</td>
<td>Minnesota</td>
<td>75</td>
<td>40</td>
</tr>
<tr>
<td>Hawaii</td>
<td>31</td>
<td>21</td>
<td>Nebraska</td>
<td>80</td>
<td>33</td>
</tr>
<tr>
<td>Kentucky</td>
<td>47</td>
<td>27</td>
<td>New Hampshire</td>
<td>76</td>
<td>39</td>
</tr>
<tr>
<td>Maine</td>
<td>35</td>
<td>34</td>
<td>North Carolina</td>
<td>85</td>
<td>34</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>50</td>
<td>41</td>
<td>Tennessee</td>
<td>80</td>
<td>24</td>
</tr>
<tr>
<td>Missouri</td>
<td>29</td>
<td>32</td>
<td>Texas</td>
<td>84</td>
<td>28</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>45</td>
<td>28</td>
<td>Virginia</td>
<td>75</td>
<td>35</td>
</tr>
<tr>
<td>South Carolina</td>
<td>26</td>
<td>27</td>
<td>Wisconsin</td>
<td>77</td>
<td>35</td>
</tr>
<tr>
<td>Washington</td>
<td>46</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyoming</td>
<td>39</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean: 40 | 27 | 79 | 33

1 Results are averaged across grades (4th and 8th grade), subjects (math and reading) in year 2003 for state and NAEP assessment.

Note: When state assessment data are missing in the grade examined, data from the next lower grade are used and if not available then data are from the next higher grade.

Source: Consolidated State Performance Report and Institute of Education Science
So Contrast 2 is...

- Compare states at three levels of standards - high, medium and low
  - Cut offs at 50% and 75% of students being proficient

- Hypothesis is that in states with higher standards there should be more of a post-2002 change in mean, slope or both

- Using Main NAEP data from 1990 to 2009 for Math and to 2007 for 4th grade Reading
Main NAEP Time Series Graphs

States with High vs. Medium vs. Low Proficiency Standards
Adjusting for Autocorrelation

• NCLB State Contrast

• Option 1:
  – HLM model can control for autocorrelation

Level 1:

\[ Y_{ti} = \gamma_{0i} + \gamma_{1i}(year)_{ti} + \gamma_{2i}(policy)_{ti} + \gamma_{3i}(year \times policy)_{ti} + \varepsilon_{ti} \]

Level 2:

\[ \gamma_{0i} = \beta_{00} + \beta_{01}(group)_{i} + \beta_{02}(percent\_free\_lunch) + \beta_{03}(pupil\_teacher\_ratio) + \tau_{0i} \]
\[ \gamma_{1i} = \beta_{10} + \beta_{11}(group)_{i} + \tau_{1i} \]
\[ \gamma_{2i} = \beta_{20} + \beta_{21}(group)_{i} + \tau_{2i} \]
\[ \gamma_{3i} = \beta_{30} + \beta_{31}(group)_{i} + \tau_{3i} \]
Analytic Model – Cont.

• Full Model

\[ Y_{ti} = \beta_0 + \beta_1((year)_{ti}) + \beta_2((policy)_{ti}) + \beta_3((policy \times year)_{ti}) + \beta_4((group \_h)_{ti}) + \beta_5((group \_m)_{ti}) \]
\[ + \beta_6((year \times group \_h)_{ti}) + \beta_7((year \times group \_m)_{ti}) + \beta_8((policy \times group \_h)_{ti}) + \beta_9((policy \times group \_m)_{ti}) \]
\[ + \beta_{10}((policy \times year \times group \_h)_{ti}) + \beta_{11}((policy \times year \times group \_m)_{ti}) \]
\[ + \beta_{12}((percent \_ free \_ lunch)_{i}) + \beta_{13}((pupil \_ teacher \_ ratio)_{i}) \]
\[ + \left[ \tau_0 + \tau_1((year)_{ti}) + \tau_2((policy)_{ti}) + \tau_3((policy \times year)_{ti}) + \varepsilon_{ti} \right], \]

• Main Variables of Interest

\[ \beta_8((policy \times group \_h)_{ti}) \]
\[ \beta_{10}((policy \times year \times group \_h)_{ti}) \]
Adjusting for Autocorrelation

• Option 2:

• Fixed effects model because we are looking at the entire population of states
  – Use cluster option in stata

\[ Y_{ti} = \beta_0 + \beta_1 (policy)_{ti} + \beta_2 (policy \times year)_{ti} + \beta_3 (year \times group \_h)_{ti} + \beta_4 (year \times group \_m)_{ti} \\
+ \beta_5 (policy \times group \_h)_{ti} + \beta_6 (policy \times group \_m)_{ti} \\
+ \beta_7 (policy \times year \times group \_h)_{ti} + \beta_8 (policy \times year \times group \_m)_{ti} \\
+ \beta_9 (percent \_ free \_ lunch)_{ti} + \beta_{10} (pupil \_ teacher \_ ratio)_{ti} + \mu_i + \tau_t + \varepsilon_{ti}, \]
Main NAEP 4th grade math scores by year and proficiency standards
Main NAEP 8th grade math scores by year and proficiency standards
Difference in differences in 2009 Total change for Math Analyses based on Main NAEP data

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
<th>S.E.</th>
<th>t</th>
<th>Coef.</th>
<th>S.E.</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>High vs. Low Proficiency Standards</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diff in Total $\Delta$ (2009)</td>
<td>7.38</td>
<td>3.33</td>
<td>2.22*</td>
<td>6.97</td>
<td>3.38</td>
<td>2.06*</td>
</tr>
</tbody>
</table>

+p<0.1, *p<0.05
Contrast 2b Study Conclusions

• NCLB increased
  – 4th Grade Math
    – 8 points.
    – 6 months of learning
    – .26 SD
    – .21 Pct
  – 8th grade math
    – 8 points
    – 12 months of learning
    – .19 SD
    – .20 pct

– 4th Grade Reading
  • No significant effect for either contrast but all are in the hypothesized direction
Overall Conclusions

• Similar results obtain from both strategies:
  – significant 4th and 8th grade math effects but not for reading
• Viable internal threats are factors independent of NCLB that changed in 2002 and are correlated with both the public/private and the high/low proficiency contrasts
• Most alternative interpretations do not apply to both strategies, and this should reduce concerns about internal validity
• A few are shared (e.g., changes in math standards in public schools in 2002). These are discussed in the paper and shown to highly unlikely.
The Use of Multiple Comparison Groups and Replications

Leaving aside Dee & Jacob,
39 tests presented of changes in mean, slope and mean plus slope – every single one has right coefficient
Some stat sig; most not – but few time points and low power
Great degree of replication of results pattern across levels of analysis (national vs states), across types of comparison groups nationally (Catholic vs non-Catholic) and across types of state level contrast (proficiency and CA), and even across Main and Trend NAEP), and across 4th and 8th grade math and 4th grade reading even by pattern.
Of the independent tests, many more are consistent with hypothesis than expect by chance in sign test.
Level of Aggregation of ITS DATA:
How Video Bail Hearings affects Bail Amounts
Log-Transformed CPI-adjusted Daily Averages
End 2015 Presentation here