

Comments Prepared for Urban NCLB Research Conference August 12, 2009

Steven Rivkin
Amherst College and UTD Dallas





Four papers in NCLB Effects

- **Dee and Jacob examine impacts on average achievement**
- **Three examine distributional effects**
 - Krieg: Does NCLB cause schools to focus on failing subgroups?
 - Figlio, Rouse, and Schlosser; and Ladd and Lauen: Does the impact of accountability vary by program type?

Order of Comments



- **Discuss some general issues relevant to all four papers**
- **Briefly discuss results of each paper**

School Incentives



- **No formal accountability system**
 - Focus of school efforts is likely to vary substantially among districts and schools
 - More likely to focus on subjects other than math and reading and on non-academic areas
 - No incentive to teach to a test
 - Less of an incentive to adopt best practices in terms of program and curricular choices?



- **Accountability system such as NCLB**
 - Focus on tested subjects and grades
 - Focus curricula and efforts in ways that have highest payout in terms of rating, constrained by laws, school boards, etc.
 - Details of accountability structure very important
 - Pass rate criteria as opposed to continuous measure of performance or multiple thresholds
 - subgroups

Impacts on Educational Outcomes



- Strength and distribution of impacts depend in large part on school responses
- Adopt better curriculum and instructional methods in tested subject areas (e.g. use better text books; make better use of Title 1 funds)
- Support use of evidence in decision-making
 - May benefit students across achievement distribution
 - Potentially positive spillovers in other subjects from organizational improvement



- **Devote additional resources and student time to tested subjects and grades**
 - May benefit all students in these areas
 - Likely to have adverse impact on learning in other subject areas
- **Focus attention on particular students for whom accountability payoff is higher**
 - Distributional effects on outcomes



Measurement of effects

- **Need to identify a counterfactual (comparison group) – very difficult**
 - Estimate of what achievement would have been in the absence of accountability system
 - Same school in previous years (Figlio, Ladd)
 - Different school (Krieg, Ladd)
 - Different state in different years (Dee, Wong)
 - Concern about selection out of or into public schools following NCLB



Academic Outcome

- **State Test**
 - Administered widely except for specific exceptions
 - Longitudinal data (prior test score info)
 - High stakes (teaching to test?)
- **NAEP**
 - Selected sample (raises concerns)
 - Low stakes and common across states



Impediments to Estimation

- **Contamination from confounding factors invalidates comparison group**
 - Unobserved changes in other school policies, unobserved differences in students
- **Test scores are measured with error**
 - Negative errors contribute to failing AYP
 - Expect 'recovery' even in the absence of any intervention

Findings



- **Dee and Jacob**

- Compare state NAEP trends by timing of accountability system adoption (how long pre-NCLB?)
- Accountability improves NAEP score in 4th grade math but only small effect on 8th grade math
 - Might observe delay in effect showing up if it takes years for curricular changes to have an impact on 8th grade scores
- Results might be driven in part by increase in spending and test exclusion
 - Not consistent with reduction of inefficiency
- Why does timing of accountability adoption vary among states?



- **Krieg**

- Estimate effect of being in a school in which another subgroup failed AYP on subgroup achievement. Does effect increase following adoption of NCLB?
- Race is a blunt signal of probability of failing
- Errors for subgroup more positive in schools in which others fail, so regression to mean could explain post-NCLB results BUT not pre-NCLB results which are different
- Should examine performance of failing subgroup as well



- **Figlio, Rouse, and Schlosser**
 - Find that Florida incentives based on average test score gain more powerful for poor-performing subgroups than explicit subgroup sanctions of NCLB
 - Consistent with overall curricular and system change being important for improving schools
 - Analysis focus is on achievement level rather than pass rate which favors FL system
 - Look at effect of failing FL system but of being in a subgroup post NCLB – Why not effect of being in a subgroup that failed to make AYP?



- **Ladd and Lauen**

- Compares NC system focusing on test score growth with NCLB
- Find distributional effects of NC system and NCLB differ in ways consistent with incentives of the two systems
- Little or no evidence that incentives cause schools to ignore very low achievers
- Analysis uses both between and within school variation. Might focus on variation within a school, grade, and year
- Value added specification with interactions imposes some strong assumptions on structure of learning

Summary



- Research consistently finds that incentives matter
- Methodological issues are complicated, and details are very important in terms of specific studies and program structure
- Using pass rate weakens incentives to work with students who are above grade level or doing very poorly
- Rating systems based on pass rate alone unfair to schools serving educationally disadvantaged students, because students may show substantial improvement but remain below passing threshold
- System with multiple thresholds (e.g. low pass, pass, high pass) would provide broader based incentives for schools to improve the quality of instruction across the curriculum and student body