

**U. S. Department of Education
Policy and Program Studies Service
Report Highlights**

An Exploratory Analysis of Adequate Yearly Progress, Identification for Improvement, and Student Achievement in Two States and Three Cities

Background

The *No Child Left Behind Act of 2001 (NCLB)* requires states to hold schools and districts accountable for making “adequate yearly progress” (AYP) toward the goal of every child achieving proficiency in reading and mathematics by the year 2014. In schools that repeatedly fail to make AYP toward meeting state proficiency standards, *NCLB* provides for a progressive series of increasingly intensive interventions. Schools that do not make AYP targets for two consecutive years are “identified for improvement,” and states and districts are expected to provide assistance and interventions to help these schools improve student achievement as well as providing parents with additional educational options. Schools that continue to miss AYP targets for two additional years are identified for “corrective action,” at which point the district must implement at least one of a set of specific interventions that include replacing staff, replacing the curriculum, reducing the school’s management authority, bringing in an outside expert, adding time to the school calendar, or reorganizing the school internally. A fifth year of missing AYP leads to “restructuring,” which requires major changes to the governance of the school.

This report presents the results of exploratory quasi-experimental analyses that were conducted as part of the National Longitudinal Study of *No Child Left Behind (NLS-NCLB)*, in two states and three school districts, to examine the relationships between certain features of *NCLB* accountability and subsequent student achievement in Title I schools. The most rigorous method for examining the effectiveness of educational interventions is a randomized controlled trial, which randomly assigns students or schools to “treatment” and “control” groups. However, this approach would not be legal in the context of Title I accountability provisions, which under the law must be applied equally to all Title I schools. Instead, this report examines the relationship between the first two stages of the *NCLB* accountability system and student achievement using a quasi-experimental regression discontinuity (RD) design, which can provide causal inferences that approach the validity of randomized controlled trials.

The analyses discussed in this report do not answer the question of whether the *NCLB* accountability system *as a whole* was effective in raising student achievement in the two states and three school districts that were studied. Rather, these analyses were intended to explore the usefulness of the regression discontinuity method for examining the effects of certain aspects of the *NCLB* accountability system, specifically, the effects of not making AYP or of being identified for the first year of school improvement status (after missing AYP for two consecutive years), which are far narrower questions than the effects of the entire *NCLB* accountability system.

Methods

The analyses in this report rely on student-level state assessment data in three districts and school-level assessment data in two states, based on tests administered in 2003-04 and 2004-05; however, some of the analyses could only be conducted in one district or state because of sample size limitations.

The RD analyses conducted for this report used an assignment rule that accounts for the fact that to make AYP, each Title I school must reach the relevant state proficiency standards in reading and mathematics overall and for all relevant subgroups. Similarly, schools that did not make AYP for the first time must achieve AYP for the school overall and for all relevant subgroups to avoid becoming identified for improvement. This means that a school's assignment to treatment status (either not making AYP or being identified for improvement) depends on having its lowest-scoring subgroup-subject combination fall short of the state standard.¹ The RD analyses examine the relationship between the minimum AYP score for which a school is accountable and the school's subsequent achievement, assessing whether schools with minimum scores below AYP cutoffs (the treatment group) show achievement bumps in the subsequent year that distinguish them from schools with minimum scores that exceed AYP cutoffs (the comparison group).

The study does not assess the total impact of NCLB on all schools, including those that are making AYP. It is possible that *NCLB* affects schools that are currently making AYP as well as schools that have not made AYP and those that have been identified for improvement. Schools that are currently meeting AYP may perceive a threat of missing AYP and becoming identified for improvement in the future and take action to avoid being identified. Thus, no school can be viewed as entirely unaffected by the *NCLB*. In addition, this study examined only the effects of missing AYP or being identified for improvement for the first time, and did not examine the effects of assignment to later stages of school improvement status, such as corrective action or restructuring. Consequently, the schools included in this analysis may have experienced a relatively weak intervention relative to the full set of progressively more intensive interventions prescribed by NCLB. Although missing AYP once provides a warning of potential interventions that may lie ahead if the school does not make AYP again, and although this warning could potentially have an effect, the warning itself is not the primary treatment that NCLB is intended to provide. The RD analysis also examined schools that were identified for improvement for the first time in 2004–05 (based on 2003–04 testing), but we do not know whether these schools experienced substantial external assistance or undertook serious improvement efforts by the time the study's outcome measure was collected about 6–8 months later (i.e., spring 2005 testing).

In this context, readers should bear in mind that *all the estimates produced by the analyses in this report may understate the full, systemic effect of NCLB on student achievement in the two states and three districts that were studied*. The analyses conducted for this report should be viewed as estimating the *marginal* effect on student achievement of a school having not made AYP or being identified for the first year of school improvement status in these states and districts. Assessing the larger systemic effects of *NCLB* on all schools (including those that made AYP and those identified for later stages of school improvement status) would require a different approach, such as one that examines differences in achievement trajectories across states.

Key Findings and Implications

Utility of RD method for assessing effect of missing AYP and first-time identification for improvement. Findings from two states and three cities cannot be generalized to produce a national estimate of the effect of missing AYP or first-time identification for improvement on subsequent student achievement, but the RD method could be applied to each state across the country. Our analysis using school-level data in two large states suggests that the RD method applied to aggregated, school-level data would lack sufficient statistical power to produce a useful state-level estimate of these effects. But the state-level estimates could nonetheless be used in a 50-state meta-analysis that could produce a valid

¹ This description is slightly oversimplified, because it ignores complications associated with safe harbor gains, confidence intervals, and standards for non-test outcomes, such as test participation rates, attendance, and graduation.

estimate of the average effect across the country. Student-level data, where available, can substantially increase the precision of the RD analysis, making it possible to produce useful state-level estimates of the effect of missing AYP or first-time identification for improvement.

Utility of RD method for assessing effect of later stages of improvement status. Regardless of whether school-level or student-level data are available, it is not clear that the RD method can produce useful estimates of the marginal impacts of later phases of improvement status (i.e., School Improvement II, Corrective Action, and Restructuring), when the most intensive interventions of NCLB are triggered. The RD method estimates impacts only for schools entering a particular stage of improvement in a particular year, and the comparison schools are only those that were at risk of entering the same stage of improvement in the same year but made AYP and avoided entering the stage. These numbers are quite small in any individual state, severely limiting the power of the analysis. A 50-state, multiyear meta-analysis might have sufficient statistical power to produce useful national average estimates of the effects of some of the later stages of improvement status; further investigation of the number of schools across the country moving into each stage (and at risk for moving into each stage) would be necessary to assess this prospect.

Effect in three cities and two states of not making AYP on school performance. In two cities, RD analyses using longitudinal, student-level data found that schools that did not make AYP showed positive impacts for some student achievement outcomes in 2003–04 or 2004–05, but the effects were not consistent across years and outcomes. In the third city, we found no significant student achievement effects. Statewide RD analyses conducted using aggregate, school-level data found a significant positive impact on the achievement of the lowest-achieving subgroup from the preceding year, but only in one of two states and only in one of two years examined.

Effect in three cities of not making AYP on subgroups' performance. In two cities where RD analyses could be conducted, the analyses showed no evidence that gains in schools that did not make AYP were concentrated among bubble students (students who had prior achievement scores just below the proficient level). Similarly, in three cities where RD analyses could be conducted, the analyses produced no evidence that not making AYP had specific effects for particular racial or ethnic groups (white, Hispanic, black), special education students, students with limited English proficiency, or economically disadvantaged students.

Effect in one city and one state of first-time identification for improvement on school performance. The study found no statistically significant achievement effects in schools identified for improvement in the year subsequent to identification, in the one state and one district where RD analysis was possible.

This report is available on the U.S. Department of Education's Web site at www.ed.gov/about/offices/list/opepd/ppss/reports.html#title.