epaa ‖ aape

Evaluating the Impact of NCLB School Interventions in New York State: Does One Size Fit All?

Jaekyung Lee
Hyejin Shin
&
Laura Casey Amo
SUNY at Buffalo
United States

Abstract: This study examines the efficacy and moderators of New York State interventions for schools in need of improvement under NCLB, including: (1) school transfer, (2) supplementary education service (3) corrective action, (4) planning for restructuring, and (5) restructuring. Despite the fact that schools in increasingly aggressive treatment groups had higher performance gains relative to schools in good standing, propensity score matching analysis results reveal negative or null effects of the interventions. There are indications of treatment effect heterogeneity and the effects varied by the year of implementation and the propensity of treatment assignment (schooling conditions prior to interventions). The findings of our study have implications for both theory of action and program implementation.
Keywords: accountability, achievement, intervention.

Evaluación del impacto de las intervenciones escolares mandadas por la ley *NCLB* en el estado de Nueva York: ¿Sirven las mismas soluciones para todos?

Resumen: Este estudio examina la eficacia y los desafíos de las intervenciones estatales en Nueva York para escuelas que necesitan mejoramiento bajo el mandato de la ley *NCLB*, incluyendo: (1) la transferencia de la escuela, (2) servicios educativos suplementarios (SES) (3) acciones correctivas, (4) planificación de reestructuración y (5) reestructuración. Mientras que las escuelas en los grupos de tratamiento más agresivos tuvieron mayores ganancias de rendimiento en relación con las escuelas en buen estado, resultados de análisis de propensión revelan efectos negativos o nulos de las intervenciones. Hay indicios de heterogeneidad del efecto del tratamiento y los efectos varían según el año de puesta en práctica y la propensión de la asignación al tratamiento (condiciones de escolarización antes de intervenciones). Los resultados de nuestro estudio tienen consecuencias tanto para la teoría de la acción como de la implementación del programa.
Palabras-clave: rendición de cuentas; logros académicos; intervención.

Avaliar o impacto das intervenções escolares baseadas nos mandatos do *NCLB* no estado de New York: Será que servem as mesmas soluções para todos?
Resumo: O presente estudo analisa a eficácia e os desafios das intervenções do estado de Nova York em escolas que precisam ser melhoradas sob o mandato *NCLB*, incluindo: (1) a transferência da escola, (2) serviços educacionais suplementares (SES) (3) ações corretivas, (4) plano de reestruturação, e (5) reestruturação. Enquanto as escolas em grupos de tratamento mais agressivos tiveram ganhos de desempenho superiores em relação às escolas em bom estado, os resultados da análise de propensão revelam efeitos negativos ou nulos dessas intervenções. Há evidências de heterogeneidade do efeito do tratamento e os efeitos variam de acordo com o ano de aplicação e a propensão de atribuição do tratamento (antes das intervenções e condições de escolaridade). Os resultados do nosso estudo têm implicações para a teoria da ação e da implementação do programa.
Palavras-chave: responsabilidade, realização, de intervenção.

Research Problems

The No Child Left Behind Act of 2001 (NCLB) in the United States is aimed at ensuring 100 percent reading and math proficiency for all students across all states by 2014 ([Pub.L. 107-110](#), enacted January 8, 2002). While NCLB allows individual states to set Adequate Yearly Progress (AYP) targets based on states' own performance standards and assessments, states have to monitor and report the percentage of schools that did not make AYP as well as the percentage of schools identified as being in need of improvement. Initially, it was estimated that up to 80 percent of schools in some states could be identified as either needing improvement or corrective action in the first few years under NCLB (Marion et al., 2002; Olson, 2002). According to several national and state reports of actual AYP statistics in the past six years of NCLB implementation however, it turns out that the percentage of schools identified as being in need of improvement was much smaller than expected, despite variations among the 50 states and some signs of increasing rates (Council of Chief State School Officers [CCSSO], 2008; Padilla et al., 2006; LeFloch et al., 2007; Education Week, 2004, 2006).

Across states, the percentage of schools that did not meet performance targets and consequently received interventions depended not only on how well the schools performed but also on how rigorous the performance standards and targets were, relative to national standards (Lee, 2010). All between-state comparisons of assessment and accountability results are complicated by the fact that school accountability targets (Annual Measurable Objective, or AMO) and performance standards vary substantially among states. Recent policy changes by the U.S. Department of

Education have enabled states to abandon the unrealistic goal of meeting 100 percent proficiency targets by 2014 by allowing states to apply for the waiver of AYP requirements in exchange for adopting more rigorous performance standards for college/career readiness and linking teacher evaluation to student performance outcomes. Although this policy change has the potential to make the original NCLB school accountability rules obsolete, it is still important for educational policymakers and practitioners to reflect on past practices and results.

While the primary source of NCLB implementation was the law itself, another layer of interpretation has been added by the U.S. Department of Education who developed and negotiated regulations and operating guidelines for states to comply with NCLB (Mills, 2008). Previous studies showed problems with states' fidelity of NCLB implementation, particularly during the first several years after the law passed (Erpenbarch, Forte-Fast, & Potts, 2003; American Institutes for Research, 2006; Kim & Sunderman, 2004). Moreover, another study showed that the fidelity of states' implementation was not a significant predictor of state assessment proficiency gains or National Assessment of Educational Progress (NAEP) proficiency gains (Lee, 2010). Regardless of states' compliance with federal policy, the effectiveness of school-level participation, delivery, and receipt of services fall under more question. A study shows very limited participation in supplemental education services (24-28% in elementary and <5% in high school) and school choice (<1%) (Zimmer et al., 2007). As last resort interventions, corrective action and restructuring also appear to have been either underused or ineffective (Center on Education Policy [CEP], 2008; Mathis, 2009).[1] Although NCLB provides a federal mandate for states to develop statewide systems of support intended to build the capacity of underperforming districts and schools, this new expectation for an enhanced role of state education agencies in school improvement has faced serious challenges due to limitations associated with the state agencies' own fiscal, administrative, and technical capacities (Center on Education Policy, 2007c; McClure, 2005; Rhim, Hassel, & Redding, 2008). Moreover, the policy impact on student achievement depends on long-term statewide funding for school resources rather than short-term state agency support for data tracking and interventions (Lee & Reeves, 2012).

The working theory of test-driven school accountability policy postulates that such policy can bring about significant change in educational practice and academic improvement by holding schools accountable for test results with possible sanctions and interventions. However, evidence for the effects of pre-NCLB high-stakes testing and test-driven accountability on student achievement has been mixed (Lee, 2008). Similarly, recent studies on post-NCLB academic progress and the policy impact were mixed and inconclusive (see Dee & Jacob, 2009; Lee & Reeves, 2012; National Research Council, 2011; Wong, Cook, & Steiner, 2009**).** While these previous studies examined overall national policy impact across states, it is important to examine the effects of state-specific interventions at the school level as well. Further, it is critical to understand school-level contextual factors that facilitate or constrain the policy impact. This study examines the efficacy and moderators of New York State interventions for schools in need of improvement under NCLB, that is, schools that failed to meet the AYP target for at least two consecutive years and went through sequential stages of interventions, including: (1) school transfer, (2) supplementary education service (SES) (3) corrective action, (4) planning for restructuring, and (5) restructuring.

---

[1] Common interventions for corrective action involved changes in curriculum or the appointment of outside advisors. Reopening the school as a charter school, replacing all or most of the school staff, or turning over school operations either to the state or a private company with a demonstrated record of effectiveness were rarely used options (U.S. Department of Education, 2003).

Review of Prior Research on School Interventions under NCLB

Prior empirical studies that address school interventions under NCLB accountability have focused on either the implementation or the effects of specific intervention types (e.g., school transfer, supplemental educational services, or restructuring). A number of studies have examined the initial stages of accountability intervention, or the provision of Supplemental Educational Services (SES) and/or the option of school transfer. There was some heterogeneity of study results regarding the effect of SES on academic achievement. In a study on district implementation of SES at the early stage of NCLB, Sunderman, Kim, and Orfield (2005) found that the provision of these services was inconsistent for districts with large percentages of low-income and minority students. In another study that examined the effect of SES on student achievement, Ross et al. (2008) reported a small positive effect of SES on student achievement but noted that SES providers needed to offer direct tutoring services that targeted state standards and test content. In another study of the effect of SES and school choice, Zimmer et al. (2007) reported a modest, positive effect of SES on students' reading and math achievement across seven school districts. The researchers, however, reported an insignificant effect of school transfer/choice (i.e. the first tier of intervention) on students' achievement across six school districts (Zimmer et al., 2007). Using Milwaukee Public School data, Heinrich, Meyer, and Whitten (2010) reported no statistically significant effect of SES participation on students' math or reading achievements across the grade level.

There is evidence that SES was not sufficiently provided to all targeted students (see Heinrich et al., 2010; Office of Research, Evaluation, and Accountability as study of Chicago Public Schools, 2007; Potter et al., 2007; Rickles & White, 2006). Although federal regulations require that all states need to approve SES providers based on scientific review (including evidence that the programs are effective among students that are among targeted groups), many states are not using a specific form to monitor the quality of the SES providers. Furthermore, the monitoring activity at the state-level has focused mainly on compliance-related reporting based on occasional visits by external constituents or has been focused on school and district self-reports (Burch, Steinberg, & Donovan, 2007).

While more studies have examined interventions that occur earlier in the NCLB sequence, there has been at least one study that focused on more severe interventions such as restructuring. Based on an analysis of data from five states on schools that were restructuring in 2006-2007, Scott (2008) found no statistical evidence that one restructuring plan was more effective than another in terms of helping schools make AYP. Similar to problems associated with the implementation of other interventions, Scott (2008) also reported that state, district, and schools experienced difficulty in implementing the restructuring plan. Findings related to the prevalence of restructuring varied across states. While the number of schools requiring restructuring intervention has risen since 2004 in California, Ohio, and Maryland, the number of schools in the intervention of Georgia has declined during the same period of time (Duffrin, Scott, & Kober, 2008). At this time, it is still unclear whether the reduced number of schools in restructuring in Georgia can be attributed to the state's (or school's) efforts.

There is some evidence that state education agencies played an important role in offering direct guidance and technical assistance in order to help to implement district improvement plans (Crane et al., 2008; Hergert, Gleason, & Urbano, 2009). In a review of intervention plans for eight states (Maine, New York, Rhode Island, Connecticut, Massachusetts, New Hampshire, Puerto Rico, and Vermont), all of the eight states supported low performing schools or districts; states' plans generally fall into three categories: *support to launch the intervention*, *continuing consultant and communication*, and *topic-specific professional development* (Hergert et al., 2009). Each State provided templates, tools, and

consultations regarding assessment and improvement plan in order to assist schools or districts. New York State Education Department (NYSED) has established a Regional Network Strategy as a regional technical assistance system in order to provide technical assistance for schools and districts that required improvement under state and federal accountability systems (e.g., Regional School Support Centers, Special Education Training and Resource Centers, Bi-Lingual Education Technical Assistance Centers, Student Support Services Centers, and Regional Adult Education Network) (U.S. Department of Education, 2006b). However, Taylor et al. (2010) found that although states provided technical assistance for implementing NCLB interventions for school improvement, there was often insufficient technical assistance, especially for improving services intended for students with disabilities as well as for students with limited English proficiency.

Methods

Data Sources

We choose to focus on tracking the impact of school AYP interventions within a single state, New York State. New York State is one of the first generation accountability states that adopted high-stakes school accountability before NCLB. The state has also continued to strengthen its policy since NCLB; it was ranked very high on the measure of the fidelity of NCLB policy implementation (Lee, 2008). NYSED played an active role in the process by developing a rationale (theory of action) for interventions to assist in the implementation of NCLB; for schools and districts that fail to improve, the quantity and intensity of supports and monitoring would increase over time (Hergert et al., 2009). The state also raised student proficiency standards for its state assessment since NCLB. As a result, the percentage of schools that failed to meet AYP in New York State has fluctuated, changing from 25% in 2004, 16% in 2005, 29% in 2006, 20% in 2007, 16% in 2008, 12% in 2009, and 38% in 2010. This volatility is primarily attributable to changes in the level of performance standards during 2006 and 2010 when the failure rate increased substantially across the state.

Data include all New York State public schools with achievement outcome measures and time-varying AYP status and intervention history along with school and district profiles. There were two main sources of data used in this study: achievement data from the New York State Report Cards (NYSRC)[2] collected by the New York State Education Department (NYSED) and data from the Common Core of Data (CCD) collected by the National Center for Education Statistics (NCES). Since the NYSRC data include information on students' achievement in reading/English Language Arts (ELA) and math at the school level, data from 1999 – 2009 on schools' Mean Score, Proficiency Level, and Performance Index for both reading and math were recorded. This paper focuses on outcomes related to schools' Performance Index Gain. A school's Performance Index is a summative measure ranging from 0 – 200 that captures the percentage of students achieving at Level 2, 3, or 4 and the percentage of students achieving at Level 3 and 4. Schools' Performance Index Gain therefore represents changes in these percentages across a single academic year.

The information from NYSED on school districts in New York State was supplemented by data on school demographic information from the CCD. Two CCD data sets for New York State were used: the Local Education Agency (School District) Universe Survey Data[3] and Local

---

[2] https://reportcards.nysed.gov/
[3] http://nces.ed.gov/ccd/pubagency.asp

Education Agency (School District) Finance Survey (F-33) Data.[4] Certain information about students, such as percentage English Language Learner, percentage Students enrolled in Special Education, and Instructional Expenditure per Pupil was available only at the district-level; thus, this information in the data set is representative of the schools' respective district. The analytical sample was restricted to regular public schools[5] in New York State. Since certain schools changed school types between 1999 and 2009, only schools which were consistently categorized as regular for the 11 consecutive years were retained in the analytical sample.

   After data from multiple sources were combined, school accountability status was coded for each year. School accountability status was reported on the annual school report card (SRC), available through the NYSED. Among all school districts in New York State, school districts in New York City only provided the SRC information from 2004-2005 to 2009-2010 academic years. Further, due to significant changes in state accountability policy and performance standards in 2009-2010, the results were not directly comparable to previous years so that the data for 2009-2010 academic year were excluded from this study. As a result, this study used data on school accountability status for (regular) schools in New York State in the academic years 2004-2005 through 2008-2009.

   An artifact of the current state accountability system was that there were increasingly smaller numbers of schools in the upper category of the accountability intervention sequence; designation into the higher order of accountability groups with stronger intervention required school failure over several years so these data were more difficult to capture. In order to increase reliability of the analysis and facilitate comparison across years, the sets of annual school data for 2005 through 2009 were pooled to create a single merged data set. For example, information for School A in the year 2006 had prior year information from 2004-05 and present year information from 2005-06. All of the variables in the data set were transformed and were combined to produce repeated measures for each school. The final N in the 4th grade sample was 6,381 records from 1,605 schools for ELA assessment and 6,418 records from 1,620 schools for math assessment; the final N in the 8th grade sample was 2,564 records from 663 schools for ELA assessment and 2,572 records from 664 schools for math assessment.

Variables

   The dependent variable in this study was schools' *Performance Index (PI) Gain* for fourth and eighth graders' reading and math assessments. A school's PI determines the schools' Adequate Yearly Performance and is defined by the state education department as follows:

   (1) $PI_t = (\sum_2^4 PL_i + \sum_3^4 PL_i)_y$

   where PL_i = Percentage of Proficiency Level i under NY accountability where i =1 to 4[6]; t = years from 2004 to 2009.

   Therefore, the *PI Gain* for year t is defined as the gap between the PI for the current year and the PI for the prior year, as demonstrated in Equation (2):

   (2) $PI\ Gain_t = PI_t - PI_{t-1.}$

---

[4] http://nces.ed.gov/ccd/f33agency.asp

[5] Excludes 'special education schools', 'vocational schools', and 'other/alternative schools.'

[6] Student scores on the tests are converted to four performance levels, from Level 1 to Level 4. Each level represents student performance relative to learning standards: Level 1 = Not Meeting Learning Standards, Level 2 = Partially Meeting Learning Standards, Level 3 = Meeting Learning Standards, Level 4 = Meeting Learning Standards with Distinction. This PI formula gives full credit for performance at Level 3 and Level 4 and gives partial credit for Level 2.

The main independent variable in this study was accountability status. School Accountability status was coded in two different ways in order to estimate (a) an overall effect, as well as (b) the differential effects of accountability treatment on school achievement outcomes. In order to estimate an overall effect, schools in good standing were coded as 0 while schools in intervention (i.e. schools in need of improvement, in corrective action, planning for restructuring, and restructuring) were all coded as 1. In order to estimate differential effects of treatment on achievement relative to the achievement of schools in good standing, schools receiving interventions were coded as follows: schools in intervention Year One = 1; schools in intervention Year Two = 2; in corrective action = 3; planning for restructuring = 4; and, restructuring = 5. Due to a small number of cases, the interventions for schools at any stage of restructuring (e.g. restructuring year one, restructuring year two, etc.) were captured in the restructuring variable (i.e. at level 5).

The other variables used in the analysis were considered covariates. The following time-varying covariates were included: *%Free lunch*, *%White students*, *%Black students*, *%Hispanic students*, *%Asian students*, *School size*, *Pupil-teacher ratio*, *PI* (prior year only), *PI Gain* (prior year only) and *accountability status* (prior year only). Time-invariant covariates included type of the school as *Magnet* (compared to non-Magnet) and location of the school as *Urban* or *Rural* (compared to *Suburban*). District-level information on *%ELL*, *%IEP*, and *Expenditure per pupil* was only included for prior year in the analysis. These covariates include observable confounding factors that are associated with both treatment assignment and proficiency gain. More detailed description of the variables is provided in Appendix A.

Analytical Methods

The data analysis consists of two parts. First, the study compares schools in good standing (comparison group) with schools in need of improvement (treatment group) in terms of their underlying student/school characteristics (covariates) and PI gains. Further, the study differentiates treatment group schools at five different stages of intervention and compares each of five different treatment subgroups with the comparison group. Failure to observe more gains for treatment group after controlling for the covariates would indicate that the mandated interventions were ineffective. This logic is based on the NCLB theory of action that lagging schools with mandated interventions under threats of sanctions will make larger gains to reach proficiency.

Since schools' accountability status (treatment condition) changes over time and school characteristics that influence assignment to treatment conditions and outcomes also change as well, time-varying causal effect model estimation procedures are needed to examine the effects of school-level AYP interventions under NCLB. We employ propensity score matching with Inverse Probability Treatment Weighting (IPTW) and difference-in-differences methods to account for selection bias[7]. IPTW realizes this matching by assigning differential weights to subjects based on the

---

[7] Selection bias can be addressed through IPTW that weights subjects by the inverse probability of receiving a treatment at a given time conditional on prior treatment and outcome history as well as time-varying and time-invariant covariates (Hong & Raudenbush, 2008). The following formula was used for computing stabilized weight (*w*) for each school i:

$$w_i = \left[\frac{p(T=1)_i T_i}{p(T=1 \mid X)_i} \right] + \left[\frac{p(T=0)_i (1-T)_i}{p(T=0 \mid X)_i} \right]$$

For schools in need of improvement (*T =1*), the greater chance of treatment group assignment conditional on the covariates (*p(T=1|X)*), the smaller the weight it gets. For schools in good standing (*T =0*), the greater the chance of control group assignment given covariates (*p(T=0|X)*), the smaller the weight. The same logic

inverse probability of receiving a treatment at a given time, conditional on prior outcome history and other covariates (Hirano & Imbens, 2002; Rosenbaum & Rubin, 1984). Difference-in-differences method utilizes matched comparison of the two groups in their PI gain scores (differences between pretest and posttest scores).

Based on school identification rules and prior research, we identified covariates that were likely to be associated with both accountability status and achievement outcomes (see Appendix A). Then, the treatment variable (z) – either a dummy variable (for the binary analysis of schools in need of improvement) or a categorical variable (for the multinomial analysis of schools at different stages of intervention) for accountability status – was modeled as a function of the covariates (x) using binomial or multinomial logistic regression (respectively) to generate propensity scores for treatment group assignment. With the estimated propensity score (i.e. the predicted probability that each school was to receive accountability intervention under NCLB, conditional on all of the covariates[8]), schools were assigned weights in order to adjust for selection bias. Finally, we fit a regression model with IPTW weights at the school level to estimate NCLB accountability intervention effects (see the final model below). We included the propensity score in the final model in addition to weighting in order to reduce any remaining bias due to observed covariates, increase efficiency, and examine its interaction with treatment.

$$(Y_{it} - Y_{it-1}) = \beta_0 + \beta_1 (\text{Intervention})_{it} + \beta_2 (\text{Propensity})_{it} + \beta_3 (\text{Year})_t + \beta_4 (\text{Intervention} \times \text{Propensity})_{it} + \beta_5 (\text{Intervention} \times \text{Year})_{it} + e_{it}$$

Second, this study conducts an exploratory analysis of factors that could produce the heterogeneity of the treatment effect. At this stage of the analysis, we examine the question: What are the district-level or school-level characteristics and conditions that may account for different treatment outcomes under NCLB high-stakes accountability? This portion of the analysis differentiates and compares three matched groups of schools in the treatment group that went through an intervention in prior year but ended up with different accountability status a year later as a result of meeting or failing AYP targets: (1) the exit group, (2) the watch group, or (3) the fail group. The exit group includes schools that were identified as in need of improvement and had an intervention in the prior year, and exited the (intervention) status a year later by meeting AYP for two years in a row. The watch group includes schools that were identified as in need of improvement and had an intervention in prior year but remained in the same status a year later by not meeting AYP for two years in a row. The fail group includes schools that were identified as in need of improvement and had an intervention in prior year yet underwent progressive intervention in the subsequent year by failing to meet AYP. The Grade 4 ELA groups included 74 cases in the Exit group, 338 cases in the Watch group, and 232 cases in the Fail group; the Grade 4 math groups included 96 cases in the Exit group, 67 cases in the Watch group, and 42 cases in the Fail group. The Grade 8 ELA groups included 71 cases in the Exit group, 415 cases in the Watch group and 299

applies for the analysis of multiple treatment groups with differentiated interventions. Cases with extremely small or large weights (outside the value range of .10 to 10) have been excluded from analysis.

[8] The inclusion of current year covariates in propensity score matching may be questionable due to potential problems with post-treatment covariates. It is possible that change in school accountability status (treatment condition) may influence student demographic composition, pupil-teacher ratio and school size (through students' school transfer or changes in school staffing). However, such changes in school demographics and contexts were not only rare, but also they were not supposed to be part of the intended mechanism of high-stakes school accountability interventions (targeting changes in school practices and incentives). It turns out that the estimates of intervention effects with and without current year covariates are very similar.

cases in the Fail group; the Grade 8 math groups included 128 cases in the Exit group, 200 cases in the Watch group and 190 cases in the Fail group.

We conducted paired comparisons to explore the context of factors that might influence variations in the fidelity of implementation and the effectiveness of interventions. To enrich our understanding of policy implementation problems under specific school contexts, we selected two Title I schools (under restructuring stage) in a large urban school district for a supplementary case study, and conducted a content analysis of the Title I monitoring report documents as complied by NYSED. The monitoring reports give information on policy compliance and implementation in low-performing districts and schools in New York State. Through this document analysis along with the phone interview with the NYSED Title I office, we were able to gain a little more insight into how accountability works. Our study has several limitations, including test measures and statistical methods. The use of state assessment as a single measure of student achievement outcome may bring bias to results. As with many other states, the discrepancy between NAEP and state assessment results in New York State raises a concern about the transferability of test score gains from high-stakes tests. It is critical to track long-term results with multiple measures of outcomes and check if the interventions had lasting effects on school performance through real long-term changes as opposed to temporary test score gains through malpractices (e.g., narrowing curriculum and teaching to the test) irrespective of interventions. Our analytic method, propensity score matching, has inherent limitations in that only observable differences between schools are controlled. Finally, the lack of information on school-level fidelity of implementation does not allow us to examine "treatment on treated" (TOT) effects as opposed to "intent to treat" (ITT) effects.

Results

Matching

In order to identify initial group mean differences between schools in accountability treatment (i.e. treatment group) and schools in good standing (i.e. control group) in regard to covariates, the two groups were compared on all standardized covariates using *t*-tests for independent samples (see Tables 1 and 2). These results demonstrated that the differences between groups were both statistically and practically significant on a majority of the covariates. The two groups of schools were very different in terms of potentially confounding variables, across subjects and grade levels. Prior to matching, schools in accountability treatment had significantly (a) higher percentage of students eligible for free and reduced price lunch years, (b) higher percentage of Black students or Hispanic students, (d) higher percentage of students who had IEPs (i.e. special education students), (e) higher percentage of students who are English Language Learners, (f) higher number of enrolled students, and (g) higher pupil-teacher ratio. Similarly, the treatment group had significantly (a) lower percentages of White students and Asian students, and (b) lower Performance Index in the prior year. Schools in accountability treatment were also more likely to be in an urban setting and be a magnet school, and were less likely to be in a rural setting. The only advantage demonstrated by schools in accountability prior to matching was with regard to expenditure per pupil at the district level, and this is likely attributed to extra federal and state funding to such high-needs school districts.

Table 1

Results of Covariate Balance Checks: Standardized Mean Differences before and after Matching between Treatment Group (Schools in need of improvement) and Control Group (Schools in good standing) in Grade 4 ELA and Math

| | ELA | | Math | |
	Before Matching	After Matching	Before Matching	After Matching
Covariates Prior Year				
Accountability	2.536***	.018	4.605***	-.485***
% Free lunch	1.19***	.031	.686***	.131
% White students	-1.19***	-.062	-1.076***	-.525**
% Black students	.692***	.061	.878***	.084
% Hispanic students	1.286***	.037	.935***	.751***
% Asian students	-.119**	.021	-.223***	.098
School size	.976***	.086	.951***	-.213
Pupil teacher ratio	-.16***	-.029	-.276***	-.419*
Magnet school	.249***	-.067	.216*	-.232
Urban	1.06***	.075	1.221***	.390**
Rural	-.584***	-.126*	-.641***	-.416*
% ELL	1.196***	.059	..384	.420*
% IEP	-.698***	.077	.8124**	-.143
Expenditure per pupil	.552***	-.004	.236**	.345*
PI	-1.491***	-.049	-1.782***	-.199
PIGain	-.004	-.04	.206	1.009***
Covariates Present Year				
% Free lunch	1.274***	.048	1.296***	.504***
% White students	-1.189***	-.057	-1.064***	-.550**
% Black students	.688***	.058	.864***	.102
% Hispanic students	1.286***	.027	.925***	.734***
% Asian students	-.124**	.034	-.207***	.137
School size	.89***	.083	.819***	-.178
Pupil teacher ratio	-.199***	.197**	-.151***	.251

Note. All values are standardized. For after matching results, IPTW regression coefficients are reported.
*p < .05, **p < .01, ***p < .001.

Table 2
Results of Covariate Balance Checks: Standardized Mean Differences before and after Matching between Treatment Group (Schools in need of improvement) and Control Group (Schools in good standing) in Grade 8 ELA and Math

	ELA		Math	
	Before Matching	After Matching	Before Matching	After Matching
Covariates Prior Year				
Accountability	1.912***	-.142***	2.201***	-.145**
% Free lunch	1.235***	-.096	.899***	.260**
% White students	-1.4***	-.123*	-1.428***	-.310***
% Black students	1.037***	.018	1.165***	.216**
% Hispanic students	1.148***	.173**	1.163***	.260**
% Asian students	.414***	.098	.161***	.091
School size	.75***	.115	.714***	.096
Pupil teacher ratio	.336***	.01	.363***	.027
Magnet school	.431***	.123	.242**	.032
Urban	1.333***	.085	1.306***	.309***
Rural	-.885***	-.059	-.900***	-.166*
% ELL	1.171***	.142*	1.097***	.203*
% IEP	.48***	.161*	.398***	.119
Expenditure per pupil	.494***	-.135*	.341***	.244**
PI	-1.39***	-.117**	-1.578***	-.215***
PIGain	-.06	-.297***	-.048*	.029
Covariates Present Year				
% Free lunch	1.556***	.039	1.315***	.319***
% White students	-1.402***	-.123*	-1.425***	-.333***
% Black students	1.031***	.008	1.151***	.219**
% Hispanic students	1.149***	.175**	1.17***	.267**
% Asian students	.422***	.088	.169***	.087
School size	.669***	.121	.588***	.111
Pupil teacher ratio	.227***	-.014	.123**	.067

Note. All values are standardized. For after matching results, IPTW regression coefficients are reported.
*$p < .05$, **$p < .01$, ***$p < .001$.

After schools were matched, a balance check analysis for matched groups was performed by examining how the treatment and control groups differed on all covariates. The results from this balance check are shown in Tables 1 and 2, alongside findings from the unmatched analysis; the results show that the groups' differences on covariates were generally reduced after matching. On average, the percentage reduction in covariate imbalance was about 87% in Grade 4 ELA and about 63% reduction for Grade 4 math; the average percentage reductions was about 87% for Grade 8 ELA and only about 72% in Grade 8 math. The average reduction on the mean difference for all

covariates after matching was about 77%, and indicates that the full-matching efforts in the final analysis were overall successful, and resulted in fairly well-balanced groups[9].

Average Intervention Effects

Results from the unmatched group analysis are shown in Figure 1. There is a general upward trend, meaning that schools in increasingly aggressive treatment groups were experiencing higher mean PI gains relative to schools in good standing (i.e. the first category) in grade 4 ELA and Grade 8 ELA and math. In Grade 4 ELA, the mean PI gain for schools in good standing was 1.49, while the mean PI gain across all schools in accountability treatment groups was 5.56; in Grade 8 ELA, the mean PI gain for schools in good standing was 5.29, while the mean PI gain across all schools in accountability treatment groups was 7.44; and in Grade 8 math, the mean PI gain for schools in good standing was 7.58, while the mean PI gain across all schools in accountability treatment groups was 10.81. Results for Grade 4 math demonstrated that schools in good standing performed better than schools in intervention Year One, better than schools in intervention Year Two and better than schools in corrective action; schools planning for restructuring and in restructuring demonstrated higher PI Gains relative to schools in good standing in Grade 4 math. Thus, without consideration of any covariates (i.e. confounding variables), it would appear as though accountability interventions were overall successful.
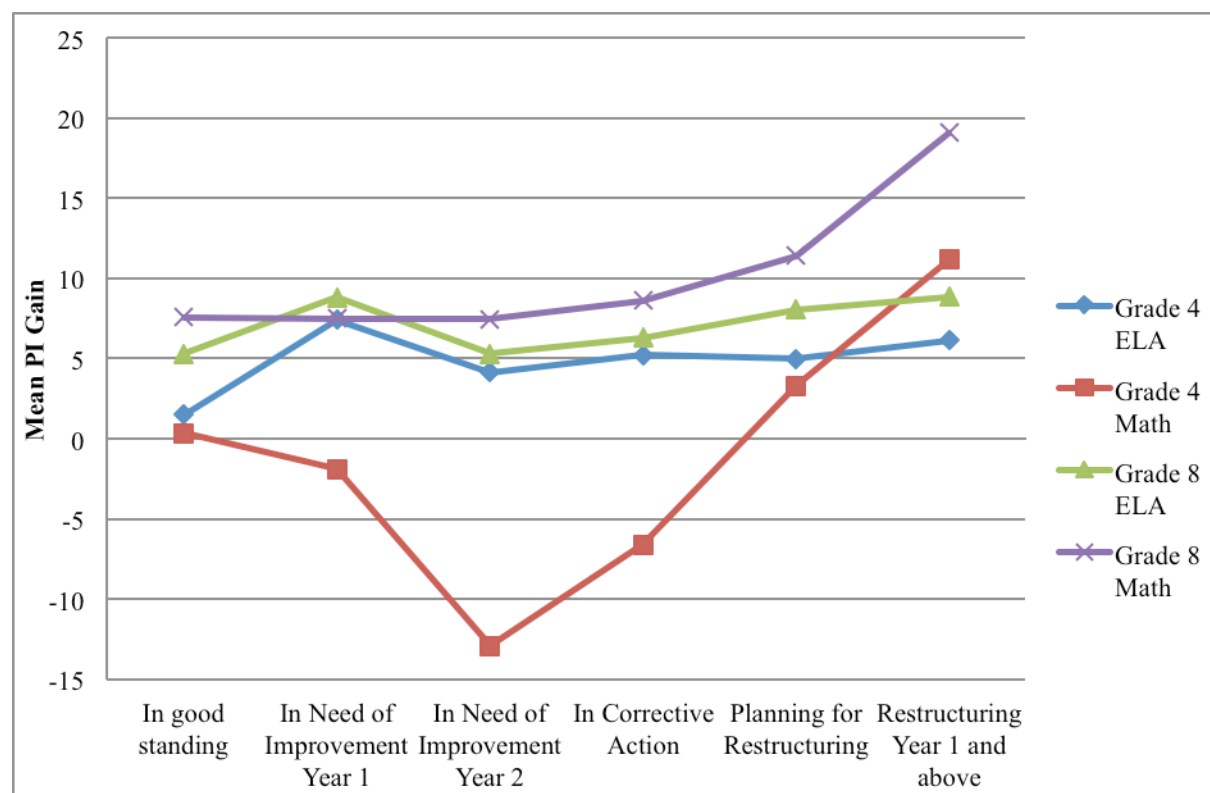
Figure 1. Mean PI gain in present year by accountability status: Unmatched group analysis

---

[9] There were some exceptional covariates which showed an opposite pattern in the percentage reduction; the covariates include Magnet_Prior, PELL_Prior, EXP_Prior, PIGain_prior year and pupil-teacher ratio.
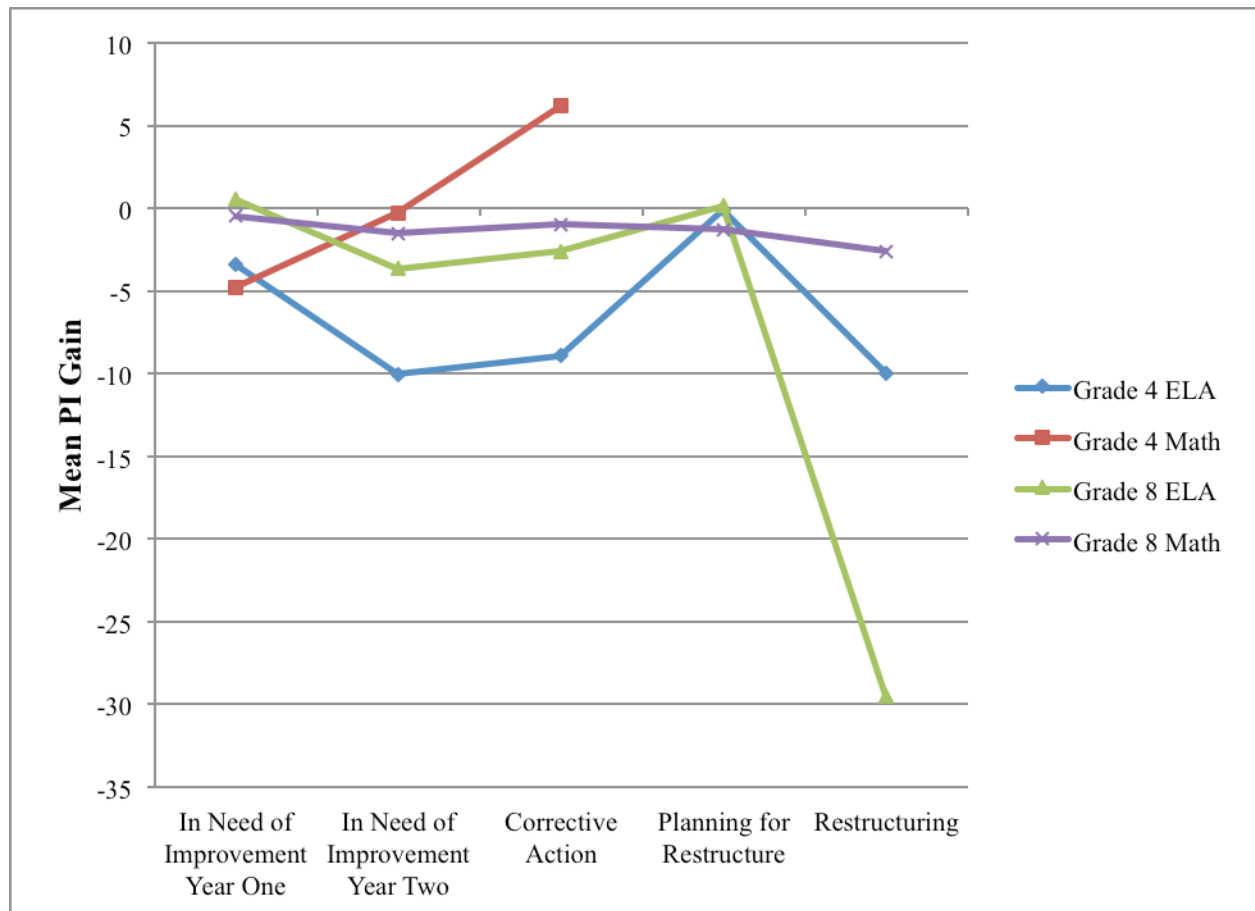
Figure 2. Differences across accountability treatment in PI gain relative to schools in good standing: Matched group analysis

The results from the IPTW-matched group analysis, displayed in Tables 3-4 and Figure 2, were very different from that of the unmatched group analysis. Generally, results from the IPTW-matched analysis demonstrated a null or mixed effect of treatment on achievement outcomes. In other words, the larger PI gains observed in the treatment group mostly get smaller or disappear after matching such that true effects of school interventions on ELA and math achievement outcomes are questionable.

Table 3
Results of IPTW Regression Analyses of Overall NCLB Intervention Effects on ELA and Math PI (Performance Index) Gain

	ELA		Math	
	Grade 4	Grade 8	Grade 4	Grade 8
Intervention	-3.238**	1.517	-10.567*	5.586*
Year	5.055***	4.746***	5.447***	8.046***
Intervention × Year	2.878**	.157	9.252***	5.524***
Propensity	1.068***	.223	.733***	1.501***
Intervention × Propensity	-.058	-.133	-.359	-1.850**
R^2	.16	.137	.174	.270

Note. Unstandardized regression coefficients are reported.
*p < .05, **p < .01, ***p < .001.

Table 4
Summary Results of IPTW Regression Analyses of Differentiated NCLB Intervention Effects on ELA and Math PI (Performance Index) Gain

	ELA		Math	
	Grade 4	Grade 8	Grade 4	Grade 8
Intervention 1 (Needing improvement year 1)	-3.417**	.562	-4.803	-.464
Intervention 2 (needing improvement year 2)	-10.042**	-3.634	-.275	-1.518
Intervention 3 (corrective action)	-8.934	-2.599	6.175	-.958
Intervention 4 (planning for restructuring)	-.071	.162	N/A	-1.285
Intervention 5 (restructuring)	-9.947	-29.484***	N/A	-2.629
Year	4.872***	4.553***	5.276***	7.696***
Propensity for Intervention 1	-.411*	-.822	1.232***	1.696***
Propensity for Intervention 2	1.885***	.452	.092***	1.784 ***
Propensity for Intervention 3	-.295**	-1.946**	-.286***	-.192
Propensity for Intervention 4	-.414**	.61	N/A	1.003***
Propensity for Intervention 5	-.073*	1.106***	N/A	-.410***
R^2	.18	.158	.172	.289

Note. Unstandardized regression coefficients are reported.
*p < .05, **p < .01, ***p < .001.

Table 3 summarizes the results of IPTW regression analysis of overall NCLB intervention effects on ELA and math PI gains. Regression coefficients for the overall treatment effect (i.e. the binary treatment analysis) demonstrated that after controlling for propensity of treatment assignment and year, schools in accountability treatment experienced significantly lower PI gains relative to schools in good standing in Grade 4 ELA (β = -3.238, p < .01) and Grade 4 math (β = -10.567, p < .05). The effect sizes[10] for these negative treatment effects in standard deviation units are small (d = -.14 in Grade 4 ELA) to moderate (d = -.69 in Grade 4 math). The treatment effect in Grade 8 ELA (β =1.517, p > .05) was insignificant. The only positive effect was found in Grade 8 math (β = 5.586, p < .05) and it was small (d = .18).

The effect of year was significantly positive across subjects and grades (Table 3). This may suggest that statewide average proficiency has improved regardless of school accountability status over the 2004-09 period. This could happen if high-stakes accountability policy gave system-wide incentives or pressure to schools in good standing as well (e.g., inducing school improvement efforts under the threat of potential sanctions and intervention in the future). Was this test score gain authentic and transferrable? Comparison of achievement gains on state assessment (high-stakes test) with gains on NAEP (low-stakes test) for the same cohort of students in New York State raises a question about the validity of gains on the state's own assessment. Except for grade 4 math, there were big discrepancies in proficiency rate gains during 2003-09 period: 13% (state) vs. 2% (NAEP) in grade 4 ELA/reading, 24% (state) vs. -2% (NAEP) in grade 8 ELA/reading, 9% (state) vs. 7% (NAEP) in grade 4 math, 29% (state) vs. 2% (NAEP) in grade 8 math.[11]

It is worth noting that propensity scores have significantly positive effects on PI gains across grades and subjects (Table 3). This suggests that the higher chance of schools being assigned to treatment group (i.e., identification of schools in need of improvement) may be associated with the higher PI gains. It seems to result partly from the tendency of regression to the mean that initially lower-performing schools (and thus higher chance of being assigned to treatment) gain more than their higher-performing counterparts. However, prior-year PI status accounts for only a small part of the variations in PI gain and treatment assignment. An alternative (equally plausible) explanation is that low-performing schools responded aggressively to the threat of being identified and assigned to interventions by making extra efforts to improve PI. These efforts might include many high-stakes test preparation strategies such as reallocation, coaching, etc., and indeed those well-documented practices (whether acceptable or unacceptable) have been observed more in low-performing schools at the higher risk of failure (Koretz & Hamilton, 2011). The underlying assumption is that schools want to avoid potential stigma attached to public identifications as well as potential threat of sanctions and interventions (e.g., staff replacement or school reconstitution). If this interpretation were true, it is ironic that those positive (positive in the narrow sense of test score gains) effects of "threats" occur in the absence of positive effects of real "interventions."

Results from the differential treatment analysis showed null or negative effects of interventions (see Table 4). After controlling for the propensity of treatment assignment, schools receiving the school choice intervention (i.e. in need of improvement, year 1) experienced lower PI

---

[10] The effect size was computed by diving unstandardized regression coefficient by standard deviation of prior year Performance Index.

[11] Clearly, a major limitation of this study is relying on data from states' own assessments that serves as a tool of both NCLB intervention and evaluation at the same time. Since NAEP test results are not available at the school level, we are unable to examine whether this seemingly statewide phenomenon of test score inflation under high-stakes accountability pressure occurred to the same or different extent between the treatment and comparison schools.

gains relative to schools in good standing in Grade 4 ELA (β = -3.417, p < .01). The mean PI gains for school choice intervention schools were not significantly different in Grade 4 math (β = -4.803, p > .05), Grade 8 ELA (β = .562, p > .05), and Grade 8 math (β = -.464, p > .05). Similarly, schools receiving SES intervention (i.e. in need of improvement, year 2) experienced significantly lower PI gains relative to schools in good standing in Grade 4 ELA (β = -10.042, p < .01). There were also insignificant gaps between schools receiving SES intervention and schools in good standing in Grade 4 math (β = -.275, p > .05), Grade 8 ELA (β = -3.634, p > .05) and Grade 8 math (β = -1.518, p > .05). The same pattern was found for schools receiving corrective action intervention in Grade 4 ELA (β = -8.934, p > .05), Grade 8 ELA (β = -2.599, p > .05). There was an insignificant positive effect of corrective action on Grade 8 math (β = 6.175, p > .05). Schools planning for restructuring experienced lower PI gains in Grade 8 math (β = -1.285, p < .05) and in Grade 8 ELA (β = .162, p > .05). Result of grade 4 ELA in planning for restructuring showed insignificant gap (β = -.071, p > .05). Results from Grade 4 math was not available due to the lack of schools available for full matching in this small accountability treatment group. Finally, after controlling for the propensity of treatment assignment, schools that were under restructuring experienced lower gains in Grade 8 ELA (β = -29.484, p < .001). There was no significant restructuring effect in Grade 4 ELA (β = -.9.947, p > .05) and in Grade 8 math (β =-2.629, p > .05).

Interaction between Treatment and Contextual Moderators

Despite the average negative or null effects of interventions, there are indications of treatment effect heterogeneity among subgroups, which were captured by interaction terms in Table 3. First, there was a positive interaction between intervention and year of implementation in Grade 4 ELA (β = 2.878, p < .01), Grade 4 math (β = 9.252, p < .001) and Grade 8 math (β = 5.524, p < .001). This general pattern (with the exception of Grade 8 ELA) suggests that schools that were subject to interventions in later years produced better achievement gains. Although it is uncertain whether this pattern is due to improvement in implementation fidelity and/or intervention design over time, it indicates that the intervention effect could become more positive in the longer term under NCLB.

Secondly, there were also some tendency of negative interactions between intervention and propensity (i.e., chance of treatment group assignment): Grade 4 ELA (β = -.058, p > .05), Grade 8 ELA (β = -.133, p > .05), Grade 4 math (β = -.359, p > .05) and Grade 8 math (β = -1.850, p < .01). This suggests that schools in the treatment group, ones that had relatively better pre-treatment conditions and thus were less likely to be assigned to treatment, produced more PI gains. Figure 3 illustrate this pattern with the example of Grade 8 ELA. Within the common area of support where matching was made possible with data available for both groups, the treatment group shows higher PI gain than the control group at the lower propensity score range (around the logit values of -4 to -2 on X-axis), but conversely lower gain at the higher propensity range (around the logit values of 1 to 3 on X-axis). This variation in the treatment effect might have occurred for two reasons. It is likely that those treatment group schools with relatively lower propensity score were already performing closer to AYP targets (higher PI status prior to the intervention) and thus better able to meet targets regardless of intervention itself. The other possibility is that those schools with lower propensity had more favorable educational conditions and capacity (such as smaller class sizes and fewer disadvantaged students) to implement the intervention.
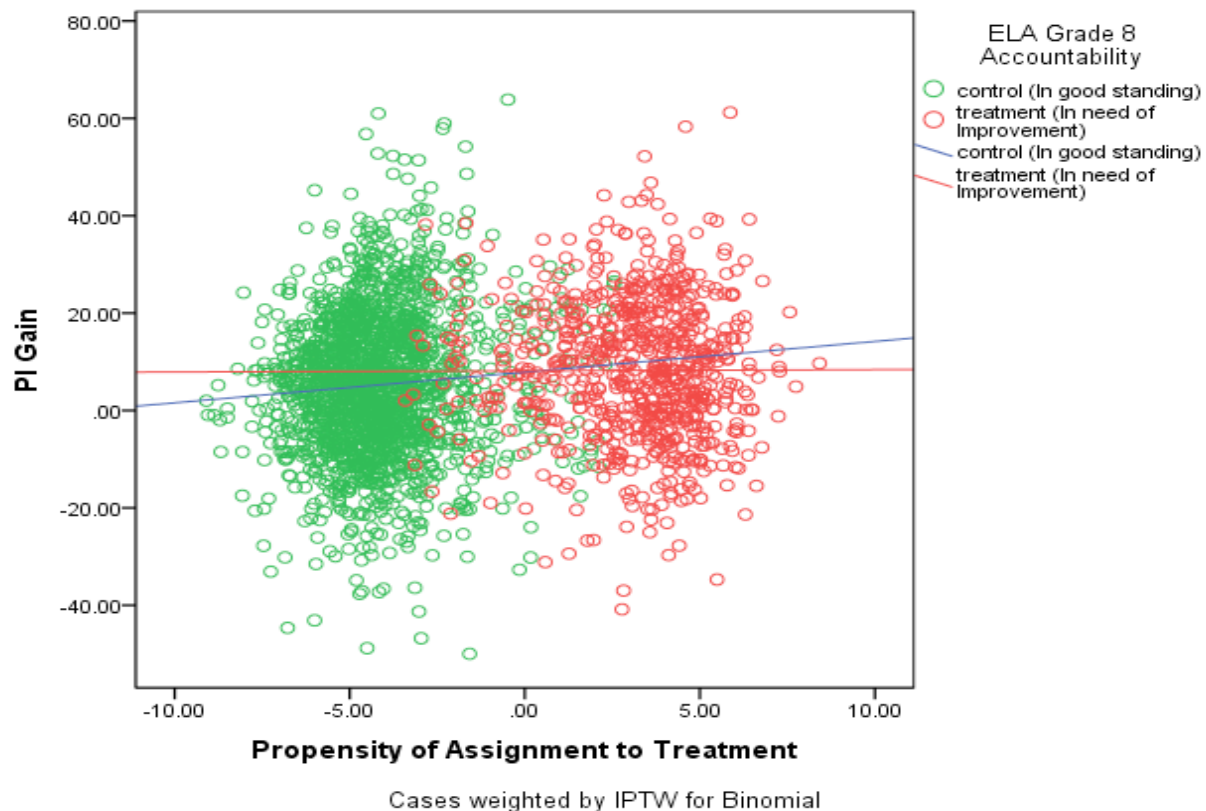
Figure 3. Scatterplot of PI gain vs. propensity score by school accountability status in Grade 8 ELA: Interaction between treatment and propensity of treatment assignment

To examine this issue further, we classified treatment schools into three categories and examined their schooling conditions (see Methods). The results from this analysis are summarized in Tables 5 and 6, and are displayed graphically in Figures 4 and 5; all values of the covariates were standardized (mean of zero and standard deviation of one). For this analysis, schools that received the same treatment yet produced different outcomes were compared against each other. From these results, schools in the Fail group demonstrated lower PI gains than schools in the Watch and Exit groups. Similarly, schools in the Watch group tended to demonstrate lower PI Gains relative to schools in the Exit group in Grade 4 ELA and math outcomes, however schools in the Watch group for Grade 8 ELA and math demonstrated higher PI Gains relative to more successful schools in the Exit group. Larger differences were found among these three groups in their prior year PI (i.e., performance status in the year right before an intervention). The Exit group performed relatively much better than the other groups even before the intervention. For the sake of space and because the patterns were similar across prior and present years, only the results from prior year covariates are displayed.

Table 5

Mean Score on Standardized Covariates among Treatment Subgroups: Results from Grade 4 ELA and Math for Prior Year

	ELA			Math		
	Exit	Watch	Fail	Exit	Watch	Fail
PIGain	0.346	0.306	-0.384 ***	.331	.464	-0.308
PI	-0.918	-1.472***	-1.747 ***	-1.272	-1.743*	-2.20***
%IEP	0.656	0.727	0.836	0.158	.668*	.891***
%ELL	1.036	1.352	1.051	1.264	1.235	0.342
Expenditure	0.529	0.589	0.456	0.285	0.264	0.149**
%Free lunch	1.097	1.405 **	1.384 **	0.576	0.902	0.728
School size	0.539	0.93 *	0.974 *	..897	1.08	1.07
Pupil-teacher ratio	-0.499	-0.279	-0.304	-0.262	-0.274	-0.337
Magnet	0.021	0.123	0.438	0.249	0.118	0.433
Urban	1.042	1.089	1.056	1.239	1.282	1.282
Rural	-0.481	-0.524	-0.476	-0.566	-0.566	-0.566
%White	-1.194	-1.257	-1.206	-1.346	-1.276	-1.095**
%Black	0.708	0.724	0.768	0.887	0.949	0.931
%Hispanic	1.269	1.395	1.266	1.366	1.210	0.943
%Asian	-0.111	-0.191	-0.211	-0.193	-0.239	-0.308

Note. Asterisks indicate the statistical significance of differences for exit group in comparison to watch group and fail group, respectively.

*$p < .05$, **$p < .01$, ***$p < .001$

Overall, results from the paired comparisons demonstrated that schools that performed better with an intervention (i.e. Exit schools) had greater access to resources and less exposure to variables associated with academic risk relative to schools in the Watch and Fail groups. For example, in Grade 4 ELA, schools in the Exit group tended to have lower pupil-teacher ratio (M = -0.499 for Exit group, M = -0.279 for Watch group, and M = -.304 for Fail group), lower percentages of Black students (M =.708 for Exit group, M = 0.724 for Watch group, and M = 0.768 for Fail group), and lower percentages of students eligible for free lunch (M = 1.097 for the Exit group, M = 1.405 for the Watch group, and M = 1.384 for the Fail group) relative to schools in the Watch group and schools in the Fail group. Schools in the Exit group were also less associated with urban location (M = 1.042 for Exit group, M = 1.089 for Watch group, and M = 1.056 for Fail group) and were less likely to be a magnet school (M = 0.021 for Exit group, M = 0.123 for Watch group, and M = 0.438 for Fail group), compared to schools in the Watch and Fail groups. Again, these are examples of patterns from Grade 4 ELA, however these types of patterns are seen across grades and subjects. The above results suggest that both prior year PI status and preexisting school conditions for the Exit group may have contributed to its chance of success in the year when their intervention was implemented.[12]

---

[12] The prior year achievement status and school condition variables are highly correlated with each other so that their relative effects cannot be disentangled.

Table 6

Mean Score on Standardized Covariates among Treatment Subgroups: Results from Grade 8 ELA and Math for Prior Year

	ELA			Math		
	Exit	Watch	Fail	Exit	Watch	Fail
PIGain	-0.07	0.137	-0.22	.521	.313	-0.387***
PI	-0.456	-1.163***	-1.289***	-.736	-1.355***	-1.693***
%IEP	0.535	0.494	0.418	.661	.522	.171***
%ELL	0.429	0.429***	0.8*	1.153	1.003	1.005
Expenditure	0.017	0.416**	0.314*	0.621	0.450*	0.081***
%Free lunch	0.529	1.1**	1.054*	1.006	1.186	0.658
School size	0.102	0.575**	0.426*	.601	.484	.664
Pupil-teacher ratio	-0.132	0.11	0.147*	.078	.142	.209
Magnet	-0.09	0.237	0.5*	0.152	0.243	0.265
Urban	0.314	1.085***	1.024***	1.08	1.129	1.124
Rural	-0.256	-0.618***	-0.573***	-0.634	-0.686	-0.724
%White	-0.496	-1.164**	-1.084**	-1.213	-1.132	-1.31***
%Black	0.477	0.808	0.936**	0.730	1.078**	1.078**
%Hispanic	0.427	1.025**	0.818*	1.207	1.093	1.100
%Asian	-0.137	0.316*	0.114	.286	.100	.023*

Note: *$p < .05$, **$p < .01$, ***$p < .001$. Asterisks indicate the statistical significance of differences for exit group in comparison to watch group and fail group respectively.
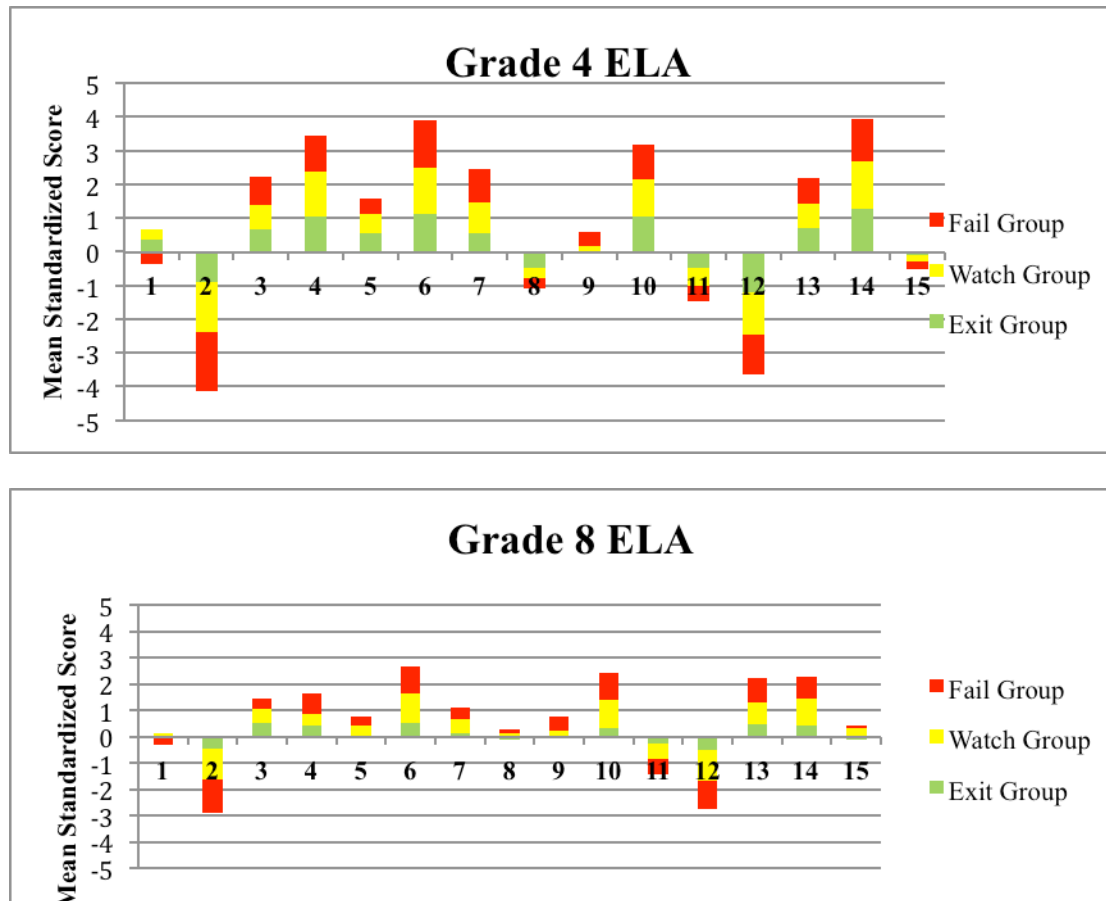
COVARIATE LEGEND

1	Performance Index Gain prior year
2	Performance Index prior year
3	Percentage of students with IEP
4	Percentage of students who are ELL
5	Instructional expenditure
6	Percentage of students w/Free Lunch
7	Total number of students in school
8	Pupil-teacher ratio in school
9	Magnet school
10	Urban location
11	Rural location
12	Percentage of White students
13	Percentage of Black students
14	Percentage of Hispanic students
15	Percentage of Asian students

Figure 4. Covariate profiles of three accountability groups in Grade 4 and Grade 8 ELA. The Y-axis represents standardized mean values of student demographics and schooling conditions, located in the legend.
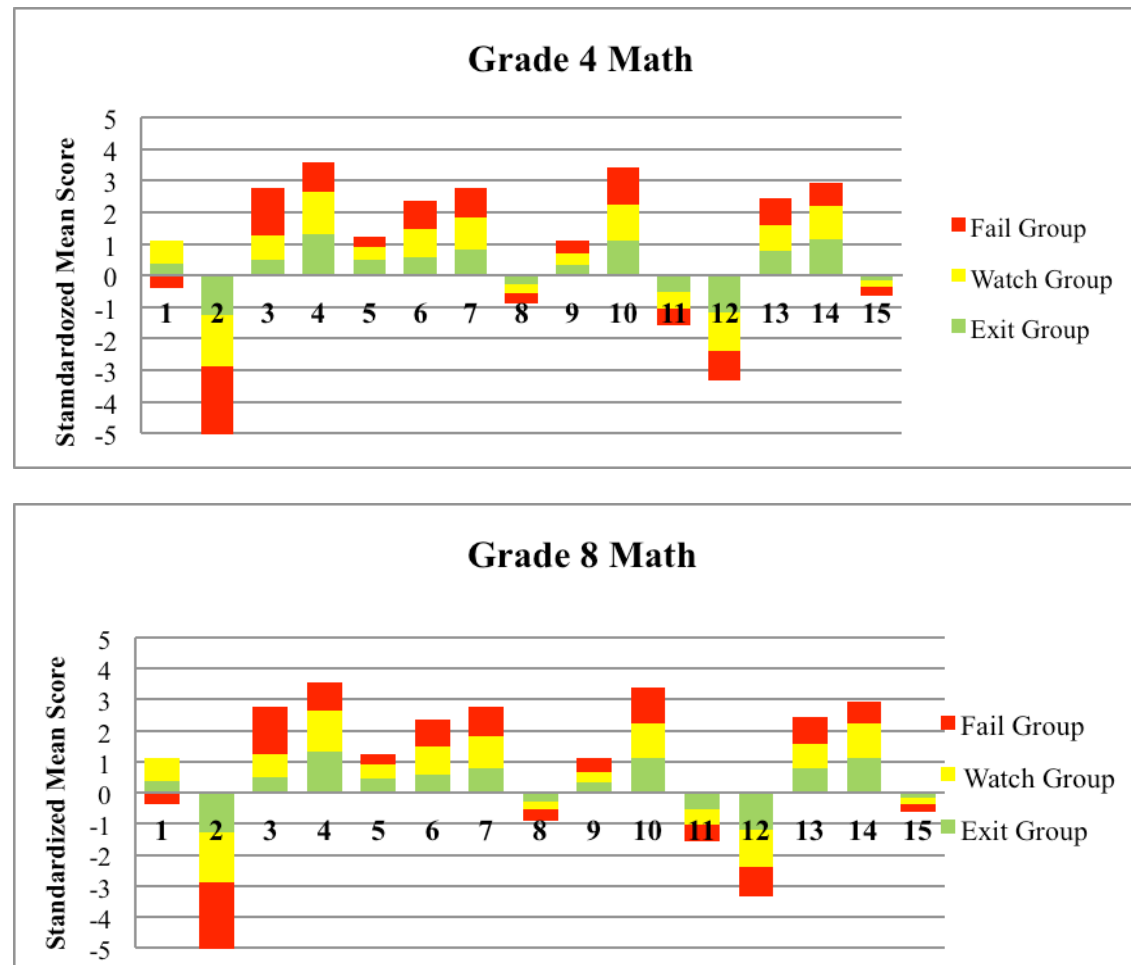
COVARIATE LEGEND
1   Performance Index Gain prior year
2   Performance Index prior year
3   Percentage students with IEP
4   Percentage students who are ELL
5   Instructional expenditure
6   Percentage students w/Free Lunch
7   Total number of students in school
8   Pupil-teacher ratio in school
9   Magnet school
10   Urban location
11   Rural location
12   Percentage of White students
13   Percentage of Black students
14   Percentage of Hispanic students
15   Percentage of Asian students

Figure 5. Covariate profiles of three accountability groups in Grade 4 and Grade 8 Math. The Y-axis represents standardized mean values of student demographics and schooling conditions, located in the legend

Findings from Case Study

   While there can be many possible reasons for the variability of observed effects for NCLB interventions, we pay attention to potential flaws in both theory of action and implementation practice. The results of the above quantitative analysis call for qualitative analysis that addresses why and how "one size fits all" approach to mandated interventions did not work at the school level. To help illustrate problems with school intervention processes and to enrich understanding of the results of statistical analysis, we examined the case of two selected Title I schools in a disadvantaged low-performing urban school district setting under restructuring stage (referred to as X and Y in this article). This case study is based on our analysis of the New York State Department of Education's (NYSED) building level monitoring review reports that incorporate information from a review of documents submitted by schools as well as a site visit to the school with interviews and observations.

   The review utilized a checklist form with 64 indicators (48 indicators in the area of instructional support; 15 indicators in the area of accountability; 1 indicator in the area of fiduciary responsibility). At the time of the NYSED review (Feb. 2009), elementary school X (with grades P-6) was under restructuring year 1 status for ELA, and the review shows that the school met state requirements for 43 indicators among 62 (except for 2 inapplicable indicators); the estimated rate of fidelity (as measured by the percentage of indicators that met requirements) was 69%. At the same time, middle-high school Y (with grades 7-12) was under restructuring year 4 for ELA and year 2 for math, and the review shows that the school met state requirements for 43 indicators among 64 and the rate of fidelity (the percentage of indicators that met requirements based on our calculation) was 67%. Although we do not know how typical or unique these implementation fidelity rates are across the state, the reviews reveal that not all schools were fully compliant.

   For school X, the monitoring report identifies several areas of implementation incompliance. One comment related to problems referred to School X's academic intervention services:

> The School Academic Intervention Services (AIS) Plan is not being implemented appropriately for students who are at risk of not meeting State Standards academic performance. Classroom teachers…reported that they frequently have difficulty finding additional time during the school day to provide AIS to all identified students.

Other problem-related comments on School X dealt with parental involvement. One comment was that "reviewers did not find evidence that parents are involved in an organized, on-going and timely way in the planning, review and implementation of the Title I program." This was reiterated in another comment from the report:

> The school has not developed a written Title I Parent Involvement Policy. They use the District developed Title I Parent Involvement Policy but that policy does not include all required components.

For School Y, the monitoring report identifies several areas of implementation incompliance. The report states that "(t)he school does not aggregate the data for subgroups." Furthermore, the report stated that there was an "inconsistent use of research-based strategies that targeted varied needs of students" and recommended that the school be "consistent with District plans, move to an RTI (Response to Intervention) model and ensure that services are varied to meet student needs." The report also stated that for School Y, "(t)here are no professional development activities within school to teach staff strategies to build partnerships between parents and the school. The District offers multiple opportunities but participation appears to be voluntary."

   These types of reviews identify the status of school compliance in terms of whether or not the school met requirements. For schools that did not meet the requirements, there are required

corrective actions or recommendations provided in the report. However, a key limitation of these types of reviews are the narrow focus on compliance with federal or state policy mandates without addressing school- and classroom-level instructional changes. In an interview with the director of Title I school and community services office, it was acknowledged that the site review does not involve experts with subject matter knowledge and thus classroom observations/reviews are superficial without substantial details. The interview also suggested that despite the seemingly high fidelity of implementation based on the checklist, many schools remain low-performing and the review does not address this deficiency. In fact, the state department discontinued building-level review due to the lack of staff and capacity; as conveyed in the interview, the way that this review is conducted does not really capture instructional dynamics. The structure of the review focuses on procedural compliance for provision of required services on the side of school administration/staff rather than focused on the effective delivery or receipt of services from the perspective of parents or students. The report does not have any information or evidence on how effective or ineffective current practices are, nor did it convey expected consequences of full or partial implementation of all required actions.

Although we don't have access to all other individual Title I schools' implementation fidelity information, the statewide report on policy implementation suggests that program delivery and service receipt were not highly consistent or effective (U.S. Department of education, 2006a; 2006b; 2007; 2008; 2009; 2010). The use of services under NCLB has been extremely low for school transfer and modest for SES (see Appendix C). For school transfer, application rates ranged from 2.2% (in 2005) to 3.3% (in 2009), and the percent of students who actually transferred to another school ranged from only 0.2 % (in 2005) to 1.3% (in 2009). For SES, both application and usage rates were around 32 % to 37%. For schools under correction actions, the most frequent choices were implementation of a new research-based curriculum or instructional program (49% in 2007) and extension of the school year or school day (20% in 2008). For schools under restructuring (implementation year 2), relatively few of them used the option of replacing school staff (21% in 2007 and 13% in 2008) and none went through extreme actions including charter school conversion, outsourcing to private management, and state take over (see Appendix C).

Discussion

The findings of this study on school intervention effects in New York State may add further insight (albeit perplexing) to our existing knowledge base on the efficacy of NCLB. First of all, the average "intent to treat" (ITT) effects of interventions were null or negative once those treatment schools have been matched to their counterparts without interventions. The treatment effects were sometimes worse for sequentially higher levels of interventions. The results do not support that "one size fits all" approach under NCLB school accountability system nor do findings support the underlying theory of action that the more chronically low-performing schools going through prescribed regimen of progressively intensive interventions, the greater academic improvement for turnaround and exit.

However, the average treatment effects obscure substantial variations in intervention effects in relation to the context of schools and the timing of implementation. Some positive effects were observed among low-performing schools with relatively more favorable conditions (and thus lower propensity of assignment to treatment) as well as among schools that were identified relatively later for interventions (and thus higher chance of policy adaptation and school organizational learning). It remains to be examined further what specific context and time factors influenced the effects of intervention and how.

Our findings underscore the importance of considering school context and background when developing a theory of action and evaluating the effect of accountability treatment on students' achievement gains. When designing or evaluating treatments, overlooking aspects of a school's social and racial composition and failing to account for district capacity and other characteristics may become fatal. In our study, ineffective schools in intervention (i.e. the fail group) had inferior environmental conditions (e.g., bigger school size, larger pupil-teacher ratio, or higher percentages of students eligible for free lunch) relative to effective schools in intervention (i.e. the exit group). School intervention plans should therefore be applied with consideration of each school's specific environmental context and specific school needs, rather than using a universal intervention plan to suit the needs of all schools. Schools with high percentage of at-risk students would need additional supports such as extra funding, technical assistance, or human resources from state education agencies.

Lastly, findings from the supplementary content analysis and interview provide insight into reasons why intervention may not be as effective as it could be. Notably, the way in which compliance information was being captured by NYSED seemingly provided very little useful information or guidance on schools' changes in instruction and its shifts in educational culture. The lack of technical and human resource capacity is a major barrier for state education agency to play a bigger role in those aspects. Moving forward, the gap in these crucial areas need to be addressed in monitoring reports with information for specific subject areas and student subgroups. For more contextually rich understanding of problems and data-driven accountability decision-making, quantitative information based on the analysis of state test results needs to be supplemented by information from qualitative analysis of instructional practices and needs. The well-known phenomenon of large gains in student proficiency based on high-stakes state test (vis-à-vis little or no gains based on low-stakes NAEP test) repeats in New York state data, and it may renew old debates. One may argue that it reinforces the theory of action for high-stakes testing in that the possible sanctions for schools that are at risk of failing to meet the standard motivates actions. Others may dismiss the results in that the standard itself as measured by such high-stakes tests is not a valid indicator of meaningful learning. Both sides of the debate have flaws and do not inform policy decisions. The theory of action for NCLB combines threats with mandated series of interventions that cost lots of taxpayer money, and thus key litmus test should focus on the efficacy of interventions. While we shared concerns about the test score inflations for high-stakes testing, we assumed that those effects would equally apply to schools under similar conditions and risk of failure (matched treatment and comparison groups) and thus do not prohibit comparing relative school performance against the state's own standards. We found that interventions per se did not produce systematically larger gains for schools in the treatment group when they were matched to the comparison group. This suggests that it may be threats rather than mandated interventions that induced those across-the-board gains observed on the state test only.

NCLB is at a crossroads. The law is up for reauthorization, while its test-driven school accountability policy has been under intense controversy with highly mixed findings on the implementation and efficacy of school interventions under NCLB. Top-down prescription of school interventions under NCLB may not be an exception to the fate of many previous education interventions that often failed to show scalable and sustainable effects on student achievement. When a policy fails to have the intended effect, it is often due to one of two types of failure: theory failure, or program failure. As the findings of our study imply, school policy interventions under NCLB might have had limited and heterogeneous impact on student outcomes because of problems with both theory of action and program.

As with many other states, New York State recently obtained NCLB waivers thorough its reform proposal that aligns with federal "Race to the Top Program" (RTTP), including raising the rigor of performance standards, adopting teacher evaluation based on student achievement, and implementing differentiated school interventions (NYSED, 2012). This move may indicate an acknowledgement that the past policy approach under NCLB did not work well. In light of our finding on the heterogeneity of treatment effects, it is more meaningful to allow for developing differentiated improvement plans that consider each school's unique context. It is also crucial to set more realistic targets based on growth trends and with sufficient time for policy adaptation and capacity building. The new intervention approach under waivers appears to follow these desired directions and thus may have the potential to be more flexible and cohesive than the old approach. However, results will be sadly predictable, if any state's initiatives under waivers remain to operate by the old theory of action for school turnaround and retain the same old interventions in new policy designs that under NCLB have shown to play a dubious role.[13]

References

American Institutes for Research. (2006). No Child Left Behind in 2004: Results from the National Longitudinal Study of NCLB and Study of State Implementation of NCLB. Symposium papers presented at the annual meeting of American Educational Research Association, San Francisco, CA.

Burch, P., Steinberg, M., & Donovan, J. (2007). Supplemental Educational Services and NCLB: Policy Assumptions, Market Practices, Emerging Issues. Educational Evaluation and Policy analysis. 29 (2), p. 115-133.

Center on Education Policy. (2007). *Choices, changes, and challenges: Curriculum and instruction in the NCLB era.* Washington, DC: Center on Education Policy.

Center on Education Policy (2008). *A call to restructure restructuring: Lessons from the No Child Left Behind Act in five states.* Washington, DC: Center on Education Policy.

Crane, E. W., Huang, C.W., Derby, K., Makkonen, R., & Goel, A. M. (2008). Characteristics of Arizona school districts in improvement. (Issues & Answers Report, REL 2008–No. 054). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West.

Council of Chief State School Officers [CCSSO] (2008). *Schools identified for improvement, or not making AYP, by state: 2007-08*. Retrieved on September 7, 2009 from http://www.ccsso.org.

---

[13] New York State Education Department claims that its policy approach under NCLB waivers attempts to bridge their existing differentiated accountability system with a new approach to school and district accountability while being consistent with the Secretary of Education's turnaround principles (NYSED, 2012). It involves enhanced turnaround efforts for lowest-performing schools newly labeled as "priority" schools: provide interventions via targeted technical assistance for subgroups; increase the quality of instruction by way of professional development opportunities; and develop partnerships with organizations with demonstrated success in helping implement proven interventions. However, the proposal does not address issues with the efficacy of the past interventions and the continuity between NCLB and interventions advocated under 'waivers'.

Dee, T. & Jacob, B. (2009). The impact of No Child Left Behind on student achievement. National Bureau of Economic Research working paper 15531, Retrieved March 1, 2010, from http://www.nber.org.

Duffrin, E., Scott, C., & Kober, N. (2008). Uncharted Territory: An Examination of Restructuring under NCLB in Georgia, Washington, DC: Center of Education Policy.

Education Week (2004). *AYP status.* Retrieved on September 1, 2007 from www.edweek.org.

Education Week (2006). *Preliminary NCLB results show slippage in 2006.* Retrieved on September 1, 2007 from www.edweek.org.

Erpenbarch, W. J., Forte-Fast, E., & Potts, A. (2003). *Statewide educational accountability under NCLB: Central issues arising from an examination of state accountability workbooks and U.S. Department of Education reviews under the No Child Left Behind Act of 2001.* Washington, DC: Council of Chief State School Officers. Available from CCSSO website: http://www.ccsso.org

Hergert, L. F., Gleason, S. C., & Urbano, C. (2009). How Eight State Education Agencies in the Northeast and Islands Region Identify and Support Low-Performing Schools and Districts. (Issues & Answers Report, REL 2009–No. 068). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education, Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands.

Heinrich, C. J., Meyer, R. H., & Whitten, Greg (2010). Supplemental Education Services Under No Child Left Behind: Who Signs Up, and What Do They Gain?, *Educational Evaluation and Policy Analysis*, 32(2), p. 273–298.

Hirano, K. & Imbens, G. W. (2002). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes Research Methodology*, 2, 259-278.

Hong, G. & Raudenbush, S. W. (2008). Causal inference for time-varying instructional treatments. Journal of Educational and Behavioral Statistics. 33 (3), 333-362.

Kim, J., & Sunderman, G. L. (2004). *Large mandates and limited resources: State response to the No Child Left Behind Act and implications for accountability.* Cambridge, MA: The Civil Rights Project at Harvard.

Koretz, D. M., & Hamilton, L. S. (2011). Testing for accountability in K-12. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 531–578). Westport, CT: Praeger Publishers.

LeFloch, K. C., Martinez, F., O'Day, J., Stecher, B., Taylor, J., & Cook, A. (2007). State and Local Implementation of the No Child Left Behind Act Volume III – Accountability under NCLB: Interim Report. Washington, DC: US DOE.

Lee, J. (2008). Is test-driven external accountability effective? Synthesizing the evidence from cross-state causal-comparative and correlational studies. *Review of Educational Research*, *78*(3), 608-644.

Lee, J. (2010). Trick or treat: New ecology of education accountability system in the USA. *Journal of Education Policy, 25*(1), 73-93.

Lee, J., & Reeves, T. (2012). Revisiting the Impact of NCLB High-stakes School Accountability, Capacity and Resources: State NAEP 1990-2009 Reading and Math Achievement Gaps and Trends. *Educational Evaluation and Policy Analysis, 34*(2), 209-231.

Marion, S. F., White, C., Carlson, D., Erpenbach, W. J., Rabinowitz, S., & Sheinker, J. (2002). *Making valid and reliable decisions in the determination of adequate yearly progress.* A Paper in the Series: Implementing The State Accountability System Requirements Under the No Child Left Behind Act of 2001. Washington, DC: Council of Chief State School Officers.

Mathis, W. J. (2009). NCLB's ultimate restructuring alternatives: Do they improve the quality of education? EPIC/EPRU Policy Brief. Retrieved on March 1, 2010 from www.greatlakescenter.org.

McClure, P. (2005). School Improvement Under No Child Left Behind. Report prepared for The Center for American Progress.

Mills, J. I. (2008). A legislative overview of No Child Left Behind. In T. Berry & R. M. Eddy (Eds.). *Consequences of No Child Left Behind for educational evaluation. New Directions for Evaluation, 117*, 9-20.

National Research Council (2011). *Incentives and Test-Based Accountability in Public Education.* Committee on Incentives and Test-Based Accountability in Public Education, M. Hout and S. W. Elliott (Editors). Washington, DC: National Academies Press.

New York State Education Department (May 21, 2012). *ESEA flexibility request.* Retrieved September 14, 2012, from http://www2.ed.gov/policy/eseaflex/approved-requests/ny.pdf.

Office of Research, Evaluation, and Accountability and Office of Extended Learning Opportunities. (2007). SES Tutoring Program: An evaluation of year 3 in the Chicago Public Schools. Chicago: Chicago Public Schools.

Olson, L. (2002). 'Inadequate' yearly gains are predicted. *Education Week.* Retrieved April 30, 2002, from http://www.edweek.org.

Padilla, C., et al. (2006). *Title I accountability and school improvement from 2001 to 2004.* Washington, DC: U.S. Department of Education.

Potter, A., Ross. S. M., Paek, J., & Mckay, D. (2007). Supplemental Educational Services in the State of Tennessee: 2005 – 2006 (2004 – 2005 Student Achievement Results). Center for Research in Educational Policy: Memphis, Tennessee.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*(387), 516-524.

Ross, S., Potter, A., Paek, J., McKay, D. Sanders, W., & Ashton, J. (2008). Implementation and Outcomes of Supplemental Educational Services: The Tennessee State-Wide Evaluation Study. Journal of Education for Students Placed at Risk, 13 (1), 26-58.

Rhim, L. M., Hassel, B., & Redding, S. (2008). State role in supporting school improvement. In S. Redding & H. Walberg (eds.) *Handbook on Statewide Systems of Support.* Charlotte, NC: Academic Development Institute and Information Age Publishing.

Rickles, J. H., & White, J. A. (2006). The impact of supplemental educational services participation on student achievement. Los Angeles, CA: Los Angeles Unified School District Planning, Assessment, and Research Division.

Scott, C. (2008). A Call to Restructure Restructuring: Lessons from the No Child Left Behind Act in Five States. Washington, DC: Center of Education Policy.

Sunderman, G. L., Kim, J. S., & Orfield, G. (2005). *NCLB meets school realities: Lessons from the field.* Thousand Oaks, CA: Corwin Press.

Taylor, J., Stecher, B., O'Day, J., Naftel, S., & Le Floch, K. C. (2010). State and Local Implementation of the "No Child Left Behind Act". Volume IX--Accountability under "NCLB": Final Report. U. S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.

U. S. Department of Education (2003). No child left behind: a parent's guide. Washington, DC.

U.S. Department of education (2006a). Consolidated State Performance Report: Part I and II, for State Formula Grant Programs under the elementary and Secondary Education Act as amended by the No Child Left Behind Act of 2001 for reporting on School Year 2004-2005

Washington DC: U.S. Department of education,
http://www.pbs.org/newshour/bb/education/nclb/map/perfreport/2004/ny.pdf

U.S. Department of education (2006b). Consolidated State Performance Report: Part I and II, for
reporting on School Year 2005-2006 New York State, Washington DC: U.S. Department of
education, http://www2.ed.gov/admins/lead/account/consolidated/sy05-06/ny.pdf

U.S. Department of education (2007). Consolidated State Performance Report: Part I and II, for
reporting on School Year 2006-2007 New York State, Washington DC: U.S. Department of
education 2006-07: http://www2.ed.gov/admins/lead/account/consolidated/sy06-
07part1/ny.pdf

U.S. Department of education (2008). Consolidated State Performance Report: Part I and II, for
reporting on School Year 2007-2008 New York State, Washington DC: U.S. Department of
education http://www2.ed.gov/admins/lead/account/consolidated/sy07-08part1/ny.pdf

U.S. Department of education (2009). Consolidated State Performance Report: Part I and II, for
reporting on School Year 2008-2009 New York State, Washington DC: U.S. Department of
education http://www2.ed.gov/admins/lead/account/consolidated/sy08-09part1/ny.pdf

U.S. Department of education (2010). Consolidated State Performance Report: Part I and II, for
reporting on School Year 2009-2010 New York State, Washington DC: U.S. Department of
education http://www2.ed.gov/admins/lead/account/consolidated/sy09-10part1/ny.pdf

Wong, M., Cook, T. D., & Steiner, P. (2009). No child left behind: An interim evaluation of its
effects on learning using two interrupted time series each with its own non-equivalent
comparison series. Northwestern University Institute for Policy Research Working Paper ,
Retrieved April 5, 2010, from http://www.northwestern.edu.

Zimmer, R., Gill, B., Razquin, P., Booker, K., & Lockwood III, J.R. (2007). State and local
implementation of the No Child Left Behind Act Volume I -- Title I school choice,
supplemental educational services, and student achievement. Washington, DC: US DOE.

Appendix A

Table A1
Variables description

Variable Name	Description
Dependent Variables	
PI Gain	Gain of Performance Index of ELA and Math assessment
Independent Variables	
Overall Effect models	
Intervention	Dummy variable indicating NCLB intervention under school accountability (In good standing (control group)=0, the other accountability status (intervention treatment group)=1)
Year	Academic year (2005~2009)
Intervention X Year	Interaction effect of intervention and year
Differentiated Effects models	
Intervention 1	Dummy variable indicating intervention of "Needing improvement year 1" under NCLB accountability (e.g., school choice)
Intervention 2	Dummy variable indicating intervention of "Needing improvement year 2" under NCLB accountability (e.g., supplemental educational service)
Intervention 3	Dummy variable indicating intervention of "Corrective action" under NCLB accountability
Intervention 4	Dummy variable indicating intervention of "Planning for restructuring" under NCLB accountability
Intervention 5	Dummy variable indicating intervention of "Restructuring" under NCLB accountability (e.g., Replacing principal, Reopen to charter school)
Year	Academic year (2005~2009)
Covariates	
Accountability	School Accountability status under NCLB prior year (Interventions under NCLB/accountability)
%Free lunch	Percentage of students who are eligible in free lunch program in school
%White students	Percentage of White students in school
%Black students	Percentage of Black students in school
%Hispanic students	Percentage of Hispanic students in school
%Asian students	Percentage of Asian students in school
School size	Total number of students in school
Pupil-teacher ratio	Pupil teacher ratio in school
Magnet school	A dummy variable indicating magnet school
Urban	A dummy variable indicating school location as urban
Rural	A dummy variable indicating school location as rural
% ELL	Percentage of ELL students in district
%IEP	Percentage of students in special education in district
Expenditure per pupil	Instructional expenditure per pupil in district
PI	Performance Index prior year
PIGain	Gain of performance index prior year

Appendix B

Table B1
Descriptive statistics

	Reading Mean (SD)		Math Mean (SD)	
Variables	4th	8th	4th	8th
Dependent Variable				
PIGain	1.9 (14.73)	5.82 (14.05)	.14 (14.44)	7.93 (16.94)
Covariates: Prior Year				
Accountability	.09 (.29)	.27 (.44)	.03 (.18)	.18 (.39)
%Free lunch	.33 (.295)	.30 (.27)	.33 (.30)	.30 (.27)
%White students	.58 (.381)	.64 (.378)	.58 (.38)	.64 (.38)
%Black students	.18 (.255)	.16 (.236)	.18 (.25)	.16 (.24)
%Hispanic students	.18 (.227)	.15 (.218)	.17 (.23)	.15 (.22)
%Asian students	.06 (.115)	.05 (.089)	.06 (.12)	.05 (.09)
School size	525.54 (238.252)	674.13 (355.662)	527.30 (237.54)	673.86 (355.80)
Pupil-teacher ratio	13.55 (2.286)	12.95 (2.59)	13.54 (2.26)	12.80 (4.6)
Magnet school	.05 (.21)	0.03 (.17)	.0491 (.22)	.03 (.17)
Urban	.38 (.49)	0.32 (.47)	.38 (.49)	.32 (.47)
Rural	.24 (.43)	0.38 (.48)	.24 (.43)	.38 (.48)
% ELL	.06 (.074)	.05 (.072)	.06 (.07)	.05 (.07)
% IEP	.14 (.037)	.14 (.038)	.14 (.04)	.14 (.05)
Expenditure per pupil	10940.72 (2504.73)	10845.13 (2700.457)	10942.89 (2504.74)	10689.15 (2760.77)
PI	163.38 (23.48)	148.38 (24.87)	177.64 (19.73)	151.60 (31.25)
PIGain	1.65 (14.65)	2.4 (12.94)	.437 (14.54)	3.68 (17.10)
Covariates: Present Year				
%Free lunch	.38 (.299)	.34 (.281)	.38 (.30)	.30 (.27)
%White students	.57 (.379)	.64 (.378)	.57 (.38)	.64 (.38)
%Black students	.18 (.253)	.15 (.234)	.17 (.25)	.15 (.23)
%Hispanic students	.18 (.228)	.15 (.219)	.18 (.23)	.15 (.22)
%Asian students	.06 (.117)	.05 (.092)	.06 (.12)	.05 (.09)
School size	520.05 (235.779)	656.53 (348.92)	521.57 (234.92)	656.16 (349.03)
Pupil teacher ratio	13.51 (5.574)	12.80 (4.61)	13.49 (5.59)	12.80 (4.6)

Note: Mean values are listed with standard deviations underneath

Appendix C

Description of School Identification Rules and Intervention Practices

1. School Identification and Exit Rules
   - To be identified for improvement status, a school must fail to make adequate yearly progress (AYP) for two consecutive years in the same grade and subject.
   - If a previously identified school fails to make AYP in the grade and subject in which it was identified, it moves to the next highest status on the continuum.
   - If an identified school makes AYP, it remains in the same status on the continuum.
   - To be removed from improvement status in a subject and grade, the school must make AYP in that subject and grade for two consecutive years. The school may remain or be placed in improvement status in another subject and/or grade for which it has not made AYP.
   - (Source: School and District Accountability Reports Implementing No Child Left Behind (NCLB), March 2004, New York State Department of Education website) http://www.p12.nysed.gov/repcrdfall2003/accountability/html/understanding-accountability-reports_frame.htm, slide 31p.)
   - For schools to make AYP, they must1. meet the 95 percent participation requirement, *and* 2. *either* meet its Effective AMO *or* make safe harbor.

Despite these seemingly common rules for identification, there are exceptions and variations that further complicate the results. New York State allows for adjustment of performance targets based on group size and applies "confidence interval" to the determination of effective AMO. The accountability rule says "Recognizing that the annual performance data for relatively small groups of students are not statistically reliable, the State has established Effective AMOs based on the number of students in a measured group. The Effective AMO is the lowest PI that an accountability group of a given size can achieve in a subject for the group's PI not to be considered significantly different from the AMO." The range of acceptable performance index varies by school size in the same subject and grade. Therefore, depending on accountability subgroup size even within the same school, different AMOs are applied (e.g, Grade 4 ELA AMO =116 for White group with N= 40 vs. AMO = 121 for Black group with N = 100). The majority of school districts were not required to make AYP for the students with disabilities accountability subgroup because they did not have a minimum enrollment of 30 students with disabilities.

These complicated rules that involve multiple assignment variables with group size-dependent cutscores make it impractical to apply regression discontinuity design due to concerns about the loss of power and generalizability. Therefore, we chose to employ propensity score matching strategy as an alternative. It matches two groups (schools in good standing vs. schools in need of improvement) based on the broader range of covariates that may influence both treatment assignment and performance outcomes. They include demographic compositions and school size as well as prior year performance and accountability status. Other option is to use fixed effects regression that capitalizes on temporal variations of intervention status within the same schools only.

Specifically, the first analysis addressed the overall treatment effect while the second addressed differential treatment effects.

Analysis of Overall Treatment Effect. In order to investigate the overall treatment effect, school accountability status was dichotomized such that 0 = *in good standing*, and 1 = *receiving accountability treatment*. Then, this dichotomous variable was regressed on all of the covariates to produce a probability, or propensity score. Since time-varying covariates affect the propensity score of

intervention variously by time period, covariates from prior and present years were used to create propensity scores. This propensity score was used in the weighting and estimation of overall effects. The binary treatment analysis used the predicted probability of being assigned into any one of the intervention condition; schools in good standing were coded as 0 and schools in any phase of intervention were coded as 1. This analysis allowed us to evaluate the effectiveness of overall treatment after controlling for all covariates, year, and interactions between treatment and year.

Analysis of Differential Treatment Effect. The second analysis included an evaluation of a differential treatment effects model. In this analysis, a multinomial logistic regression model was used to predict six accountability intervention conditions that may be considered progressively more aggressive. With schools in good standing as the reference group (or 0), the other treatment conditions were coded as follows: *Intervention Year One* = 1, *Intervention Year Two* = 2, *In Corrective Action* = 3, *Planning for Restructuring* = 4, *Restructuring* = 5. Schools that identified (in the primary data sources) as being *In Restructuring Year One*, *In Restructuring Year Two*, *In Restructuring Year Three* or *In Restructuring Year Four or above* were collapsed into a single category for this study and analysis. All of the schools that were in good standing or in different stage of interventions were assigned weights according to the estimated conditional probability of their group assignment.

2. School Intervention Practices

The following tables summarize the number of students who were eligible and received Public School Choice and Supplemental Educational Services in Title I schools identified for improvement, corrective action or restructuring in New York State.

Public School Choice					
	2004-05	2005-06	2006-07	2007-08	2008-09
-Number of students who were eligible for services	228,450	429,683	315,326	314,436	298,999
-Number of students who applied for services	4,996	N/A	11,469	9,884	10,010
-Number of students who received services	551	4,102	3,692	4,090	3,900
Supplemental Educational Services					
	2004-05	2005-06	2006-07	2007-08	2008-09
-Number of students who were eligible for services	268,768	272,164	274,114	251,240	289,647
-Number of students who applied for services	N/A	88,934	81,073	92,378	98,011
-Number of students who received services	105,035	87,814	64,906	87,579	92,651

The following tables summarize the frequency of specific intervention strategies adopted by Title I schools under *Corrective action* and *Restructuring year 2* in New York State.

Corrective Action	# of Title I Schools in Corrective Action	
	2006-07	2007-08
-Required implementation of a new research-based curriculum or instructional program	60	32
-Extension of the school year or school day	28	72
-Replacement of staff members relevant to the school's low performance	3	4
-Significant decrease in management authority at the school level	8	18
-Replacement of the principal	12	5
-Restructuring the internal organization of the school	11	23
-Appointment of an outside expert to advise the school	2	8

Restructuring Action	# of Title I Schools in Restructuring Year 2	
	2006-07	2007-08
-Replacement of all or most of the school staff (which may include the principal)	10	6
-Reopening the school as a public charter school	0	0
-Entering into a contract with a private entity to operate the school	0	0
-Take over the school by the State	0	0
-Other major restructuring of the school governance	37	40**

Note:
** NYC Schools Only: Of the 24 NYC schools in Restructuring Year 2 during the 2007-2008 school year, 22 implemented activities that supported other major restructuring of the school governance. The specific "other major restructuring of school governance" actions that were implemented include:
A. School Organization
Creation of "houses" or "academies" Smaller Learning Communities Change in grade configurations Change in student programming (block scheduling, self-contained, departmentalized, etc)
B. Zoning
Change in feeder patterns, Change in zoning
C. Targeted Interventions for specific identified subgroups
Multi-faceted and drastic changes in the curriculum and/or delivery of the educational program for the specific subgroup(s) of students that caused the school to be designated as Restructuring Year 2
D. Professional Development
To support the educational program of the restructured school (professional development before the start of the implementation year; differentiation of professional development appropriate to the

assignment of needs of staff due to the new organizational structure of new grade configurations, etc; professional development for supervisory/administrative staff due to the new structure)
E. Changes in Resource Allocations
F. School Support Organization

Sources:

U.S. Department of education (2006b). Consolidated State Performance Report: Part I and II, for reporting on School Year 2005-2006 New York State, Washington DC: U.S. Department of education, http://www2.ed.gov/admins/lead/account/consolidated/sy05-06/ny.pdf

U.S. Department of education (2007). Consolidated State Performance Report: Part I and II, for reporting on School Year 2006-2007 New York State, Washington DC: U.S. Department of education 2006-07: http://www2.ed.gov/admins/lead/account/consolidated/sy06-07part1/ny.pdf

U.S. Department of education (2008). Consolidated State Performance Report: Part I and II, for reporting on School Year 2007-2008 New York State, Washington DC: U.S. Department of education http://www2.ed.gov/admins/lead/account/consolidated/sy07-08part1/ny.pdf

U.S. Department of education (2009). Consolidated State Performance Report: Part I and II, for reporting on School Year 2008-2009 New York State, Washington DC: U.S. Department of education http://www2.ed.gov/admins/lead/account/consolidated/sy08-09part1/ny.pdf

About the Author

Jaekyung Lee
University at Buffalo, SUNY
jl224@buffalo.edu
Jaekyung Lee is the Dean and Professor of the Graduate School of Education at the University at Buffalo, SUNY. He specializes in educational policy research and evaluation.

Hyejin Shin
University at Buffalo, SUNY
hyeshin@buffalo.edu
Hyejin Shin is a Ph.D. Candidate in Educational Administration, University at Buffalo-SUNY. Her research interests include accountability, K-12 policy evaluation, and educational equity.

Laura C. Amo
University at Buffalo, SUNY
lccasey@buffalo.edu
Laura C. Amo is a Ph.D. Candidate in Educational Psychology and Quantitative Methods at the University at Buffalo-SUNY. She currently holds an appointment as Visiting Assistant Professor in the UB School of Management in the Management Science and Systems Department. Her research interests include school accountability, policy evaluation, and organizational outcomes among teachers and principals.

Join EPAA's Facebook community at https://www.facebook.com/EPAAAAPE and **Twitter feed** @epaa_aape.

education policy analysis archives
editorial board

Editor **Gustavo E. Fischman** (Arizona State University)
Associate Editors: **Audrey Amrein-Beardsley** (Arizona State University) **Rick Mintrop,**
(University of California, Berkeley) **Jeanne M. Powers** (Arizona State University)