The Impact of No Child Left Behind on Non-cognitive Skills

Sean Tanner
Goldman School of Public Policy
and
Department of Economics
University of California, Berkeley

VERSION: 10-28-2015
The most current version is available at
https://sites.google.com/site/patrickseantanner/job-market-paper

I would like to thank Rucker Johnson, Jesse Rothstein, Hilary Hoynes, Mark Wilson, David Kirp, Geno Smolensky, and seminar participants at the Goldman School of Public Policy, the Institute for the Study of Societal Issues, and APPAM's 2014 conference "The Decline of the Middle Class?" for their helpful comments. All errors are my own.

Abstract

Beginning in 2002, the No Child Left Behind Act (NCLB) mandated the monitoring of student performance and rewarded or sanctioned schools based on students' standardized "high-stakes" test scores. Non-cognitive skills, such as self-control, have been identified as key determinants of adult success in the labor market. Students' growth in non-cognitive skills are not well captured by standardized tests of literacy and numeracy. Consequently, this crucial dimension of child development may be neglected due to NCLB's focus on cognitive testing outcomes. I use the Early Childhood Longitudinal Study of Kindergarten, a panel data set of children attending kindergarten in 1998 that spans the onset of NCLB, in order to analyze the impact of the federal mandate on non-cognitive skills. In a departure from most of the prior economic and policy research on non-cognitive dimensions, this paper focuses on the growth of underlying non-cognitive skills (or "traits") rather than aberrant behaviors like truancy. I find little impact of NCLB on nine dimensions of non-cognitive growth for students overall. However, these precise null effects at the national level mask distinct heterogeneity across student subgroups and domains of behavior- classroom skills improved for African-American and Hispanic students but worsened for white and Asian students. Parents reported declines in home traits across most demographic subgroups.

Sean Tanner
Goldman School of Public Policy
University of California, Berkeley
2607 Hearst Avenue
Berkeley, CA 94720
stanner@berkeley.edu

Introduction

The quality of a child's school, beginning in the early years, can have persistent and marked effects on her adult well-being (Card & Krueger, 1992; Chetty et al., 2011; Hanushek, 2011). The inequitable distribution of high-quality schools by ethnicity and socioeconomic status has been well documented (Adamson & Darling-Hammond, 2012; Reardon, 2011) and has been a formidable, perennial educational policy challenge. For example, disadvantaged students are more likely to be taught by less qualified and experienced teachers (Clotfelter, Ladd, & Vigdor, 2005; Goldhaber, Lavery, & Theobald, 2015; Lankford, Loeb, & Wyckoff, 2002). By shaping both cognitive and non-cognitive skills, school quality influences the pattern of human capital accumulation and plays a role in the inequality of life outcomes (Chetty, Friedman, & Rockoff, 2013; Heckman, Pinto, & Savelyev, 2013). The frustrating aspect of these findings is that altering school quality has been difficult to achieve with existing policy levers (Hanushek, 2003), leading to cycles of major reforms followed by public disappointment at the perceived lack of progress (McGuinn, 2006; Vinovskis, 2009).

School accountability polices are among the more recent waves of school reform in the United States. Focusing on the lack of performance incentives in the public K-12 system, supporters of these policies believed that accountability would enhance student achievement if state governments illuminated disparities in achievement and levied rewards or sanctions on schools based upon explicit, mandated measures of student performance (Hanushek & Raymond, 2001, pp. 368–369). These "high-stakes" tests incentivized schools to focus narrowly on the tested subjects, and especially on the performance of groups of students whose test scores would have a meaningful impact on the schools' accountability rating (Manna, 2011, Chapters 2, 6). An abiding concern for critics of accountability is how the focus on standardized tests of literacy and numeracy impacts domains of child development that are known to be key determinants of adult success but are not included in the accountability testing regime (Duckworth, Quinn, & Tsukayama, 2012).[1]

By the time the No Child Left Behind Act (NCLB) mandated such systems for all public schools nation-wide, half of states had already implemented similar systems (Dee & Jacob, 2011, p. 423). Several studies have attempted to assess the impact of this federal mandate by using the states that had already adopted accountability ("early adopters") to construct a counterfactual for states that were newly compelled to do so ("late adopters") (Dee & Jacob, 2011; Lee & Reeves, 2012; Reardon, Greenberg, Kalogrides, Shores, & Valentino, 2013; Wong, Cook, & Steiner, 2011). This body of work has produced somewhat mixed results, with most research showing positive effects on math and little effect on reading.

My research contributes to this literature by providing the first national estimates of the impact of NCLB on non-cognitive skills. The study analyzes data from the Early Childhood Longitudinal Study of Kindergarten 1998 (ECLS-K), a nationally-representative panel data set of 21,000 children attending kindergarten in 1998. I compare the developmental trajectories of children residing in early adopter states to those living in late adopter states. These two groups of children are balanced with respect to race, socioeconomic status, early childhood education, and cognitive achievement scores at the entrance to kindergarten. The identification of NCLB effects comes from the differentially binding nature of the law's accountability mandate between these two groups of students. Only students in the late adopter states (those without prior accountability regimes) experienced their first exposure to accountability at the onset of NCLB- fourth grade for this sample. By tracking the ECLS-K students as they progress through elementary school, this design captures pre-treatment (NCLB) differences in student growth between groups of students who had the same cognitive achievement and

[1] For a representative example of such criticism in the popular press, see Gewertz (2003) and Darling-Hammond (2007).

socioeconomic profile at the onset of their public schooling career. The impact of NCLB is the change in non-cognitive skills, relative to pre-NCLB trends, between the early and late adopters. This comparative interrupted time series with student fixed-effects is robust to linear and quadratic pre-NCLB trends as well as the inclusion of contemporaneous changes in paternal employment, socioeconomic status, and household composition. As I elaborate upon in the estimation strategy section, this research design is similar to that used in prior work on NCLB, but with a key distinction: this analysis tracks divergent growth between two groups of students as they progress through elementary school where the prior literature analyses divergent trends in average test scores across successive cohorts of students.

In a departure from prior work, I examine measures of non-cognitive skills and traits rather than aberrant behaviors such as truancy. Specifically, the nine measures are teachers' assessment of students' self-control, approaches to learning, peer relationships, and externalizing and internalizing problem behaviors, and parents' assessment of independence, ability to solve problems, attention span, and overall behavior. Furthermore, I consider both measures of cognitive and non-cognitive development to provide a more comprehensive assessment of the impacts of school accountability policies on broader dimensions of student performance.

To preview the results, I find little effect of NCLB on any of these measures on average nationally. However, as the law specifically mandated that test scores be disaggregated by socioeconomic sub-groups, national averages obscure potentially meaningful heterogeneity in NCLB's effects. When analyzing the data by socioeconomic sub-groups, I find mixed results. Teachers' ratings of several skills increase for underrepresented minorities (African-American and Hispanic) and a decrease for white and Asian students. The parents' assessments are mostly negative or indistinguishable from zero. Aggregate indices of these measures support this general pattern. The data allows changes in teachers' child development training and emphasis on non-cognitive assessment to be ruled out as potential mechanisms.

The paper begins with a description of No Child Left Behind and the prior work analyzing its impacts on cognitive achievement. The next section briefly summarizes literature on non-cognitive skills and their relation to educational achievement and labor market performance. I then discuss the identification strategy in detail before moving to a description of the ECLS-K data. The results section provides estimates of the impact of NCLB on each of the nine measures overall and for white, Asian, African-American, Hispanic, free-lunch, and limited English proficiency students. Supplementary results from cognitive tests and children's own rating of their non-cognitive skills are included in addition to two tests of potential mechanisms. A discussion section follows that addresses the potential reasons for the heterogeneity in results and policy implications. I conclude with a concise summary of the paper's contribution and fruitful directions for future research.

No Child Left Behind

Accountability systems shifted the focus to short-run educational outcomes—particularly, math and reading standardized test scores -- rather than school inputs, fundamentally altering the metrics by which school quality is judged. The central motivation of these policies is the perceived lack of progress in American students' achievement since the early 1970s (Hanushek & Raymond, 2001, pp. 366–368). Since the mid-1950s, per-pupil spending has increased markedly as a percentage of GDP (Springer, Houck, & Guthrie, 2015, p. 4). Consequently, traditional markers of school quality had been increasing, such as teacher-pupil ratios, the percentage of teachers with master's degrees, and teacher tenure. From 1960 to 2000, teacher salaries more than tripled (in constant dollars) while the proportion of teachers with at least a master's degree doubled and average tenure increased 30% (Hanushek, 2003, p. F68). However, aggregate measures of student progress have failed to increase commensurately with these significant increases in spending, leading many to question the wisdom of increased resources without concurrent incentives to use resources wisely (McGuinn, 2006, pp. 134–135).

To address this concern, accountability policies measure student performance, publish school performance data, and subject schools to rewards and sanctions based on this data (Figlio & Ladd, 2015). These performance measures are standardized and disaggregated by demographic characteristics such as race/ethnicity and socioeconomic status (SES), so that the educational deficiencies of disadvantaged students are not masked by school-wide averages or inflated grades (Manna, 2011, pp. 121–122). Operationally, this means that schools are primarily judged on the basis of their students' performance on state-administered tests of literacy and numeracy.

Accountability began as state-level reforms, with Illinois being the first to adopt such a system in 1992. By the 2002-2003 school year, when NCLB required schools in all states to implement accountability plans, roughly half of states had already adopted such plans (Dee & Jacob, 2011, p. 423). Several key elements of the law bear directly on this analysis. First, tests were standardized by grade and state, so that all students were assessed by the same tests of literacy and numeracy within each grade and state. Second, each test had a score above which students were considered to have adequately mastered the subject by grade level- the "proficiency threshold." Third, school-level accountability scores were calculated for each demographic sub-group as the percentage of students in that group who scored above the proficiency threshold. These pass rates were then used to determine which schools were not making "adequate yearly progress" (AYP) toward 100% of students scoring above the proficiency threshold by 2014. An escalating series of punitive measures would be levied on schools that continued to fail to make AYP year after year, ranging from a requirement to offer public school choice to massive staff layoffs and outright closure (Manna, 2011, pp. 26–29). In order to underscore the severity of this new regime and distance the law from the ineffectual regulations of prior administrations, Secretary of Education Roderick Paige warned states that they were expected to heed the letter of the law and that the Bush administration would not tolerate noncompliance (McGuinn, 2006, p. 183).

A large body of research has been devoted to assessing the impact of accountability, and NCLB in particular, on student achievement and school resources (see Figlio and Ladd (2015, pp. 204–207) for a concise overview). The literature summarized below focuses on studies that analyze the impact of NCLB rather than the impact of accountability in early adopter states. As is made clear in the estimation strategy section and the following literature review, studies of the impact of NCLB are designed so that the precise treatment is the federal accountability *mandate*, not aspects of accountability systems themselves or various other features of NCLB. Due to jurisdictional and political tensions, NCLB allowed states substantial freedom in crafting their own accountability systems (McGuinn, 2006, pp. 177–179; Vinovskis, 2009, pp. 158–170). This state freedom led to consequential policy variation in how much pressure was put on schools across states, primarily through variation in proficiency thresholds and ability of schools to discard certain students' tests (Davidson, Reback, Rockoff, & Schwartz, 2013). Because this variation is endogenous to student achievement, valid causal inference is limited to the effect of the NCLB mandate rather than states' choices in crafting systems. Moreover, NCLB contained numerous provisions in addition to the accountability mandate, such as the requirement that teachers be "highly qualified"[2] (Manna, 2011, pp. 29–30). These provisions affected both states that had and had not adopted accountability prior to the law, rendering the identification strategy favored here and in the prior literature inappropriate for assessing those effects.

Prior research on NCLB has found mostly positive effects on math and weak to null effects on reading. Of the five separate research papers summarized here, four have used variants of a comparative interrupted times series (CITS) design on data from the National Assessment of Educational Progress (NAEP). The control group is typically those states that had adopted consequential accountability prior to the onset of the national mandate (the early adopters) and the treated group is the states that were

[2] Highly qualified meant, essentially, that teachers had formal training in and a manifest command of the subjects in which they taught.

compelled to enact new systems in the 2002-2003 school year (the late adopters). The CITS design leverages the variation in the timing at which children were exposed to accountability, using a causal framework similar to that of a difference-in-difference model, but allowing pre-treatment differences in the (usually) linear trends in outcomes across treatment and control groups. The ability to control for differences in pre-treatment trends rather than merely pre-treatment levels is crucial in analyses of NCLB, where pre-2002 trends in NAEP scores are demonstrably different in early and late adopter states (Dee & Jacob, 2011, pp. 432–433). In the CITS models, the impact of NCLB is the difference in the difference in achievement trends (and levels) between the early adopter and late adopter states before and after the onset of the law in 2002.[3] This design has been used on NAEP data, which contains nationally-representative tests of literacy and numeracy given to successive cohorts of fourth and eighth grade students from 1990 to the present. As these scores are not used in accountability regimes and are thus "low-stakes," they are plausibly immune to the "teaching to the test" phenomenon (Jacob & Levitt, 2003) and so provide credible measures of students' cognitive achievement.

Using this strategy, Dee & Jacob (2011) find strong increases on fourth grade math scores (between .15 to .47 standard deviations), fewer and more moderate increases on eighth grade math scores (around .22 standard deviations), and little to no effects on 4^{th} grade reading achievement. These results are stronger for underrepresented minority (URM) and lower-income students (Dee & Jacob, 2011, pp. 438–440). Wong, Cook, and Steiner (2011) further separate states into those that had relatively high vs. low proficiency thresholds. They find positive effects on fourth and eighth grade math overall (in the range of effects found by Dee and Jacob (2011)), with positive effects on reading limited to states that had high proficiency thresholds. Lee and Reeves (2012) use propensity score weights that place greater analytic weight on early adopter states that were similar to late adopters states on aggregate levels of

[3] A full treatment of this model is provided in the estimation strategy section

high school attainment, poverty rates, Caucasian proportion of the population, state education revenue, SAT scores, and a political culture variable. They also include data on fidelity of implementation (data tracking capacity, funding, and stringency of proficiency thresholds) as explanatory variables. They find an effect on eighth grade math scores, but none on fourth grade math and reading scores. Finally, Reardon et al. (2013) find that NCLB narrowed achievement gaps (a small effect size of roughly 3%[4] decrease in the magnitude of the gap per year) between white and URM students only in states with larger pre-NCLB gaps, a greater degree of between-school segregation, and more accountability pressure focused on URM sub-groups.

Reback, Rockoff, and Schwartz (2014) use an alternate identification strategy to assess the impact of NCLB pressure for students in schools at risk of failing to meet proficiency thresholds. Using the same ECLS-K data employed in the present study, they analyze the cross-sectional variation in schools' risk of failure conditional on observable school and student characteristics. Their logic is that similar students face differing levels of accountability pressure due to state policy choices (ex. proficiency thresholds) and school-level demographics. They find positive effects of being near the failure boundary on reading scores, with positive but statistically insignificant effects on math and science tests.

Several analyses of changes in school resource allocation and pedagogical activities induced by NCLB support this set of findings of effects on achievement gains. Using the Schools and Staffing Survey (SASS), Dee and Jacob report that schools increased teacher salaries and hired more teachers with master's degrees, while teachers re-allocated their time toward reading and (to a lesser extent) math (2010, pp. 184–187). Further evidence from the SASS suggests that the onset of NCLB and the resulting reallocation of school resources was not associated with a decrease in teacher morale or increase in

[4] Reardon et al. (2013) use the V-statistic outlined in Reardon and Ho (2015), a measure of between-group disparity more appropriate to ordinal, non-Gaussian data such as the NAEP. The baseline V for the gaps was between .7 and .6 and the effect of NCLB per year ranged from -.02 to .02, depending on state characteristics.

hours worked (at least as measured by survey responses in the SASS) (Grissom, Nicholson-Crotty, & Harrington, 2014).

On balance, these results provide evidence that NCLB shifted resources toward instruction in literacy and numeracy, increased performance on certain tests of those subjects, and did little to impact overall teacher morale.

School Accountability and Non-Cognitive Skills

In addition to the traditional focus on cognitive achievement, teachers also impact their students' development of non-cognitive skills, such as self-control, adaptability, and motivation. The category labeled "non-cognitive"[5] encompasses a broad range of attributes such as those related to concentration, social interactions, and judgement. In a school setting, they manifest themselves in children's ability to complete tasks, pay attention, and cooperate with teachers and peers, to name but a few. These skills are complementary to those traditionally labeled cognitive skills, such as numeracy and literacy, yet are not well captured by narrow aptitude or achievement tests. For example, in a study undertaken in direct response to NCLB, Duckworth, Quinn, and Tsukayama (2012) find that teacher and parent ratings of children's self-control in fourth grade are far better predictors of ninth grade G.P.A. than are fourth grade IQ tests. Conversely, fourth grade IQ is a far better predictor of ninth grade standardized achievement tests than is self-control. This suggests that a narrow focus on standardized testing will fail to account for the impact of schools and teachers on the elements of academic achievement more closely related to non-cognitive skills.

Non-cognitive skills are typically not in the large, state data sets that are used to monitor student achievement in the U.S. Nevertheless, the data that is available reveals that such non-cognitive skills predict acquisition of cognitive skills and are rewarded in the labor market independently of IQ and test

[5] These skills have also been referred to as character, grit, social and emotion learning competencies, disposition, and temperament (Duckworth & Yeager, 2015, p. 238).

scores (Almlund, Duckworth, Heckman, & Kautz, 2011). These skills are not stable over a lifetime, suggesting a role for the educational system in manipulating them. Heckman, Pinto, and Savelyev (2013) provide compelling evidence that the long-run gains from the Perry Preschool program are largely attributable to its impact on externalizing problems- aggressive, anti-social behavior. A similar pattern is found in children who participated in Project STAR, the canonical experiment of class size reduction on student achievement. The experimental impact of higher quality classrooms on test scores fades out in later grades, yet impacts on non-cognitive skills remain and are likely mechanisms through which the treatment affects long-run success (Chetty et al., 2011).

Adopting the methodology used in typical value-added models to control for observable characteristics and prior achievement, Jackson (2012) shows that teachers have meaningful impacts on their students' non-cognitive skills. Interestingly, teacher quality in regards to non-cognitive skills appears orthogonal to a measure of teacher quality based on gains in cognitive skills. In other words, a teacher that is above average in her effect on test scores is no more or less likely to improve her students' non-cognitive ability than an average (or below average) teacher. This has severe consequences for accountability policies whereby teachers and schools are rated only by their students' cognitive scores. There is little evidence of how teachers respond to those policies in terms of promoting cognitive vs. non-cognitive skills. Critics of accountability policies worry that the focus on preparation for high stakes testing hinders the acquisition of non-cognitive skills by crowding out pedagogical activities that encourage such skills in favor of those that increase scores on standardized achievement tests (Darling-Hammond, 2007; Duckworth et al., 2012).

Estimation Strategy

This paper identifies the impact of NCLB's school accountability requirements by exploiting the differentially binding nature of the law between states that had already adopted accountability policies by 2001 (the control or "early adopter" group) and those states that were newly compelled by NCLB to adopt such policies (the treatment group). The preferred estimation strategy, a propensity-score weighted comparative interrupted time series, is robust to endogenous selection into treatment groups and both linear and non-linear differences in outcome dynamics due to heterogeneity in observed and unobserved determinants of students' non-cognitive skills. The basic intuition is similar to that of a difference-in-difference (DID) approach, but does not require the assumption of parallel pre-treatment time trends in the outcome. Additionally, the panel structure of ECLS-K allows the analysis of student achievement growth within the same cohort of students rather than changes in averages across successive cohorts of students as are analyzed in the NAEP studies. Though the research designs in both this analysis and the prior NAEP studies can be appropriately labeled comparative interrupted times series designs, the student fixed effect panel vs. pooled cohort model is a key distinction. The panel study design employed here has the benefit of avoiding threats to validity from endogenous selection in and out of the tested sample as well as increased precision from using student fixed-effects.[6] Analyses of the public school NAEP data are reliant on an assumption that schooling decisions (public vs. private) are exogenous to the onset of NCLB conditional on the demographic data available from NAEP. This is a strong assumption, one that is not necessary when tracking the same cohort of children across grades as ECLS-K does. However, this validity and precision come at a price- the divergent trends being captured in the CITS model are within-student growth rather than changes in between-student averages. The validity of the CITS design rests on correctly modeling pre-treatment divergence between early adopter and late adopter states. Modeling student growth over time presents challenges that are likely not present in successive cohort models, where the assumption of stability or linearity in averages over time is easier to maintain. To deal with this challenge, I standardize the teacher ratings of non-cognitive outcomes by grade-level to net out any year to year changes in overall student growth. The raw

[6] The increase in precision over pooled cohort models is conditional on similar sample sizes and objects of measurement.

measures from ECLS-K suggest that teachers are already informally standardizing their ratings, as there is very little year to year change in the central tendencies of all five constructs.

A simple DID strategy would model the impact of NCLB as the change in non-cognitive skills from third to fifth grade in the treated states minus the change in the control states. This would be operationalized as b_3 in equation 1.

$$Y_{ist} = b_0 + b_1 NCLB_t + b_2 Treat_s + b_3(NCLB_t \times Treat_s) + \varepsilon_{ist} \qquad (1)$$

Where Y_{ist} is the outcome of individual i in state s at time t, NCLB is a dummy variable equal to one in fifth grade and zero in third, Treat is a dummy variable equal to one if that state had not implemented school accountability prior to the onset of NCLB in 2002, and ε is a stochastic error term. One of the strongest identifying assumptions necessary for b_3 to be the causal impact of NCLB is that the students in treated and control states would have had parallel growth if not for the policy change. This is a strong assumption that is likely to be violated if the treated and control states differ in ways that affect the growth and levels of students' non-cognitive skills. As Figure 1 reveals, this is almost certainly true in the case of NCLB, where the early adopters are highly clustered in the South. As can be seen in the third row of table Table 9, teachers' ratings of students' externalizing problem behaviors and approaches to learning diverge between the two groups of states in the pre-NCLB period (the coefficients on Year*Treat are statistically significant and negative). A DID would incorrectly include a continuation of this pre-treatment trend in b_3.

This problem of non-parallel trends can be attenuated with a comparative interrupted time series (CITS) design, which explicitly controls for pre-treatment trends across treated and control groups. The basic linear strategy models the impact of the treatment as the change in level *and* slope across the treatment boundary in the treated states minus the change in level *and* slope in the control states.

A linear CITS model specification with individual panel data is below:

$$Y_{ist} = b_0 + b_1 YEAR_t + b_2 NCLB_t + b_3(Year_t \times NCLB_t) + b_4(T_s \times YEAR_t) + b_5(T_s \times NCLB_t) \quad (2)$$
$$+ b_6(T_s \times Year_t \times NCLB_t) + \mu_i + \varepsilon_{ist}$$

Where Y_{ist} is the outcome of individual i in state s at time t, Year is the calendar year minus 1998, NCLB is a dummy variable equal to one at 2002 and after (when Year ≥ 4), T is a dummy variable equal to one if that state had not implemented school accountability prior to the onset of NCLB in 2002, μ is an individual fixed effect, and ε is a stochastic error term. The total effect of NCLB at any given year after its onset is $b_5 + b_6 \times t$, where t is the number of years since 2002. The total effect of NCLB at the final survey wave in 2007 is thus b_5 (the shift in level from 3rd to 5th grade) + $b_6 \times 3$ (the coefficient on the linear trend from 5th to 8th multiplied by 3 years). This strategy has the benefit of accounting for linear, non-parallel pre-treatment trends in the treated and control states. For instance, the states that adopted accountability policies may have also been more likely to attract and retain high quality teachers, who could have increased student test scores in addition to the gains from accountability. A basic CITS model allows for both time-invariant differences such as these as well as time-dependent differences so long as the functional form (linearity) of the time-dependency does not change from the pre-NCLB era to the post-NCLB era. Similarly, the onset of NCLB is allowed to co-vary with secular changes in the determinants of student achievement. However, this covariance must be similar across treatment and control groups. This design has been used extensively in education research, particularly to estimate the impact of NCLB and accountability generally on student achievement (Jacob, 2005; Somers, Zhu, Jacob, & Bloom, 2012). In a series of sensitivity analyses available in the online appendix[7], I relax the linearity assumption by adding a quadratic year term to equation 2.

Heterogeneous, non-linear trends in cognitive achievement are an empirical regularity in many social experiments such as the Perry Preschool project (Heckman et al., 2013). In these experiments, treated

[7] Available at the author's Open Science Framework site: https://osf.io/zq6fn/

children have initially higher cognitive achievement scores, but those scores decay and collapse to control group means around 3rd and 4th grades- the precise point at which the ECLS-K children experience the onset of NCLB. An estimation strategy that does not account for such differential early childhood education (ECE) would attribute the lack of fade out in those not exposed to ECE as an impact of a 3rd or 4th grade intervention rather than the fading out of ECE-related gains in those exposed to it. To account for potential differences such as these, I weight the early adopter group to fit the late adopter group by multiple baseline characteristics such as early childhood education, socioeconomic status, and cognitive assessments of literacy and numeracy. Specifically, I fit a saturated logistic regression of treatment status on binary indicators for Head Start attendance, non-Head Start preschool attendance, free/reduced price lunch and limited English proficiency status, ethnic category (African-American, Hispanic, White, other), and terciles of socioeconomic status (an index of parental education, income, and job categories) and cognitive scores on tests of literacy and numeracy taken upon entrance to kindergarten. These variables are completely interacted with one another to allow a saturated regression. The predicted treatment status ($\hat{t}$) is used to calculate a propensity score p wherein the late adopter group is assigned a value of 1 and the early adopter group is assigned $p = \hat{t}/(1-\hat{t})$. This weighted CITS design places more analytic weight on control group (early adopter) students with demographic and cognitive profiles that are most similar to the students in the treatment group (late adopter). Similar weighting strategies have been used for DID designs to overcome the problem of non-parallel trends (Abadie, 2005), Lee and Reeves' CITS analysis of NCLB (2012, pp. 214–215), and two CITS designs found to produce similar causal estimates to those from a regression discontinuity analysis (Somers et al., 2012) and a randomized field trial (St.Clair, Cook, & Hallberg, 2014) of two separate educational programs. As an addition guard against spurious causal inference, I control for contemporaneous changes in family structure, socioeconomic status, and paternal employment.

Specifically, I include indicator variables for a father's presence in the home and fulltime employment status as well as a continuous measure of family socioeconomic status provided by ECLS-K.

The actual years observed in the ECLS-K data when both teacher and parent data on non-cognitive skills are available are years 1 (spring Kindergarten), 2 (spring 1st grade), 4 (spring 3rd grade), and 6 (spring 5th grade), so only the initial level shift can be estimated (the impact of ≈2 years of accountability). The full model is presented in equation 3.

$$Y_{ist} = b_0 + b_1 YEAR_t + b_2 NCLB_t + b_3(Year_t \times Treat_t) + b_4(NCLB_s \times \text{Treat}_t) + \sum_{j=5}^{k} \beta_j X_{ist} + \mu_i + \varepsilon_{ist} \quad (3)$$

Where Y_{ist} is the outcome of individual i in state s at time t, Year is the calendar year minus 1998, NCLB is a dummy variable equal to one at 2002 and after (when Year ≥ 4), Treat is a dummy variable equal to one if that state had not implemented school accountability prior to the onset of NCLB in 2002, X_{ist} is a vector of socioeconomic variables for individual i in state s at time t, μ is an individual fixed effect, and ε is a stochastic error term. The effect of NCLB is b_4, which represents the level shift in non-cognitive skills after approximately two academic years under NCLB.

In a less parametric model, I create grade-level dummy variables and estimate saturated regressions of the outcomes on grade-level dummies and their interactions with the treatment identifier. This has the advantage of being completely transparent and non-parametric, however it does not readily yield an estimate of the impact of NCLB under observed differences in the pre-NLCB outcome trends.

ECLS-K 1998 Data

This analysis uses data from the Early Childhood Longitudinal Study of Kindergarten 1998 (ECLS-K), a nationally representative panel study that began surveying children in the fall of their kindergarten year in 1998 and concluded with a survey of the same cohort in the spring of their eighth grade year in 2007.

Seven rounds of data were collected, six of which are used in this analysis.[8] The ECLS-K data has several features that are advantageous for an analysis of the impact of NCLB. First, the timing of the study is such that the children's elementary school career spanned the onset of NCLB: K-3 prior to NCLB and 4-8 under NCLB. Second, unlike alternate data sets of student achievement such as NAEP or state-level achievement tests, ECLS-K contains non-cognitive assessment data on the same cohort of children as they develop through elementary school. Third, ECLS-K contains a rich set of data on students, teachers, schools, and parents- both upon entrance to kindergarten and throughout the students' elementary school career.

The analysis sample includes data on approximately 10,000 children sampled each round.[9] Table 1 contains descriptive data across early and late adopter states in the fall of 1998 – the children's first semester of kindergarten and the first survey period. The states compelled by NCLB to adopt accountability are considerably less ethnically diverse, have a larger share of children attending Head Start, and have fewer children who receive free or reduced price lunch than the states that adopted accountability prior to NCLB. Additionally, schools in the late adopter states are less likely to have kindergarten teachers who completed post-baccalaureate training and are far more likely to receive Title 1 funding. These differences indicate likely differences in additional, unobserved determinants of student achievement and non-cognitive growth, suggesting that the early adopters are a poor counterfactual for the late adopter states in a simple, cross-sectional comparison design.

This analysis primarily focuses on two sets of non-cognitive domains that are measured from kindergarten to fifth and eighth grade in the ECLS-K surveys: the Social Rating Scale (SRS) from the teacher questionnaires (K to 5th) and parents' rating of their own children's behavior (K to 8th).

[8] The fall 1st grade round was dropped, leaving fall and spring kindergarten, spring 1st, spring 3rd, spring 5th, and spring 8th in the analysis sample. The fall 1st grade survey round was conducted only on a small sub-sample.
[9] The complete dataset contains information on 21,409 children. Attrition, missing data, and private school attendance whittle the analysis sample down to roughly 10,000.

Additional analyses are conducted on the children's self-reported non-cognitive skills. Focusing on teachers' and parents' ratings of the children as they mature through elementary school is advantageous for two reasons. First, the parent and teacher surveys are capturing distinct elements of children's non-cognitive growth in school and home settings. This allows tests of whether changes in classroom behaviors spill over into the children's behavior elsewhere. The second advantage of the teacher and parent surveys is that they are more likely to be reliable in the early grades than children's self-report (Junttila, Voeten, Kaukiainen, & Vauras, 2006). As the children mature and their responses become more reliable, ECLS-K begins to collect self-reported socio-emotional data. In the final year, eighth grade for most of the sample, the SRS is no longer collected from teachers as the children are not with a single teacher throughout the school day as they are in earlier years. The parent surveys of non-cognitive skills are given throughout the entire panel periods- kindergarten through eighth grade. Table 4 contains the details of when each construct is measured and from whom.

The four measures from the parents survey, independence, ability to pay attention, problem solving skills, and overall behavior are each collected from a single survey item on a 1 (is better than peers) to 4 (is much worse than peers) scale. Table 2 contains the distribution of responses in each category for all four variables over the entire survey period (kindergarten through eighth grade). For each item, more than half of parents report their child as being average ("as well as peers") for their age and at least a quarter reported their child as being better than average. Few parents labeled their child as below average and fewer still labeled their child as far below average. Given this distribution of responses, these ordinal scales have been transformed into dummy variables for this analysis, with "better than peers" coded as 1 and all other responses coded as 0. Doing so focuses the analysis on a readily interpretable boundary (better/not better than average), one that is also likely to yield more informative results that apply to a larger share of children in the sample.

The SRS from the teacher survey is composed of five constructs measured on a continuous one to four scale that is the combination of multiple, ordinal items in the questionnaires. The constructs are self-control, interpersonal skills, approaches to learning, internalizing problem behaviors (ex. anxiety), and externalizing problem behaviors (ex. class disruptions). The first three capture positive aspects of child development (higher values are normatively better) and the latter two capture problematic behavior (higher values are normatively worse). The child surveys contain self-description questionnaires (SDQs) on externalizing and internalizing problem behaviors as well as interpersonal relationship skills in the third and fifth grades (on either side of the onset of NCLB). Like the SRS, these measures are on a continuous one to four scale that is the combination of multiple items in the questionnaires. For each item from which the five SRS variables and three SQD variables are constructed, the respondent is asked how often particular behavior is observed (teachers) or experienced (students): never, sometimes, often, very often. The scale scores for the SRS and SDQ are the mean for all items in each scale. Table 3 contains summary statistics from all five constructs in the SRS and all three in the SDQ across all survey periods. Overall, the teachers and children reported relatively higher levels of positive constructs and relatively lower levels of the problem behaviors. For the causal analysis, the problem behaviors in the SRS and SDQ have been reverse coded in a manner that preserves ordinal relationships and standard deviations but eases interpretation so higher values are normatively better across all non-cognitive constructs in the teacher and child surveys. The SRS and SDQ measures are also standardized by grade, so that each variable in each survey period has mean zero and unit variance.[10] Table 5 contains the non-cognitive skills, their labels for this analysis, the number of survey items from which they are constructed, and examples of each skill.

[10] The means of the non-standardized variables remain fairly stable across all survey periods, suggesting that teachers were already conceptually standardizing their responses by grade level. Standardizing merely formalizes this.

Table 6 contains the correlations amongst the non-cognitive skills in the teacher and parent surveys in the fall kindergarten period. As is apparent, there are higher correlations amongst skills within surveys than between them. Correlations of constructs range from .22 to .28 within the parent survey and from .24 to .78 within the teacher survey. However, many of the between-survey correlations are less than .1. This is further evidence that these two surveys are picking up distinct elements of children's behavior across two different environments: home and school. As is expected, the measures of non-cognitive skills from teachers are more highly correlated with children's cognitive assessments than are the skills measured by parents. Table 7 contains the correlations amongst the SRS and cognitive assessments of literacy, numeracy, and general knowledge in the fall kindergarten period. Approaches to learning correlates more strongly than the other four SRS constructs with the three cognitive assessments (.34 to .41 vs. .13 to .25). This is also commensurate with the prior research demonstrating that non-cognitive skills measure aspects of child development that are separate from and only modestly correlated with cognitive development (Almlund et al., 2011).[11] These non-cognitive skills also vary significantly over the course of a child's elementary school years. Table 8 contains the proportion of the variance in each construct due to variation within children over time, as well as the white-black, white-Hispanic, and treat-control gaps. Between one third and one half of the variation in non-cognitive skills is within a child over time rather than between children. Though this is less within-child variation than that of the ECLS-K literacy assessment, is it enough to suggest malleability and response to policy interventions. The ethnic achievement gaps are merely the percentage point difference for the parent survey measured and for the SRS are calculated using the non-parametric method preferred by Reardon and Ho (2015, p. 159), which is equivalent to Cohen's d for normally distributed variables. The SRS

[11] A natural question is how these skills correlate with students' academic grades. ECLS-K collected report cards from teachers in each survey period, but has not yet transferred the grades to the data. Such an analysis is therefore not possible at present.

ethnicity gaps are considerably smaller that the white-black and white-Hispanic literacy gaps, yet many are still troubling large.

Taken together, the data supports the prior evidence that non-cognitive skills vary over childhood, are distinct from measures of cognitive skills, and are thus likely malleable by education policy.

Results

Nationally, I find scant evidence that NCLB had any impact on students' non-cognitive growth, as measured by parents or teachers. Figure 2 contains the mean, standardized self-control ratings from the SRS by grade-level and treatment group. As is apparent, the control (early adopter) and treatment (late adopter) groups do not significantly diverge in either the pre- or post-NCLB periods. Figure 3 reveals similar patterns across the remaining four SRS skills- little change due to NCLB. The formal CITS analysis confirms this visual pattern. Table 9 contains the results from the full regressions, with the treatment effect captured by NCLB*Treat- the shift in levels due to approximately two years of treatment. In no case is the impact statistically significant even at the .1 level. Negative impacts of one fortieth to one tenth of a standard deviation can be rejected and positive impacts of one seventieth to one seventh of a standard deviation can be rejected.

The skills measured by parents display the same pattern, both visually and statistically, as those measured by teaches: very little impact from NCLB. Figure 4 contains the time series graphs for the four measures from the parent survey and Table 10 presents the statistical analysis. The impact of NCLB (NCLB*Treat) is not statistically significant for any of the four variables at the .1 level. Negative impacts of 1.6 to 4.1 percentage points can be rejected and positive impacts of 1.2 to 3.5 percentage points can be rejected.

To provide a sense of scale, the point estimates and confidence intervals are given in Figure 5 along with the standardized effect from an RCT of twelve months of exposure to a pedagogy focused on non-cognitive skills (Barnett et al., 2008), in the case of the SRS, and the female percentage point advantage in the case of the parent survey. As this figure reveals, NCLB had very little impact at the national level.

When the results are separated by socioeconomic categories, the results become more mixed- with suggestive evidence of some positive effects for African-American and Hispanic students from the teacher surveys and negative effects for white, Asian, and LEP students. Table 11 reveals small negative effects on self-control and peer relationships for white students, a modest negative impact on externalizing problem behaviors for Asian students, and a modest negative impact on approaches to learning for students with limited English proficiency. Conversely, African-American students displayed fewer internalizing problem behaviors and Hispanic students had improved peer relationships. Beyond statistical significance, all of the coefficients for African-American students and four out of five for Hispanic students are positive; four out of five for white and Asian students are negative. This is reflected in the final column of Table 11, which provides the mean effect of NCLB on all SRS variables with accompanying standard errors.[12] Though none of the mean effects can be distinguished from zero at the .1 level of significance, the directions suggest that NCLB had weak positive effects on the classroom behavior of certain students who were more likely to be targets of the policy, contrary to the concerns of its critics.

The sub-group specific results from the parent survey reveals mostly negative effects. Table 12 shows weak to modest negative effects for African-American, Asian, free lunch, and limited English proficiency students. Conversely, Hispanic students displayed greater independence due to NCLB. The two statically significant mean effects are both negative (African-American and LEP). Assuming that these results are not due to differential changes in psychometric validity of either the parent survey or SRS,

[12] The mean effect is the arithmetic mean of the five standardized beta coefficients for each subgroup. Standard errors were calculated by simulating the sampling distribution of the mean effects with 1000 samples (taken with replacement) clustered at the student level. For recent uses of mean effects in economic literature, see Kling, Liebman, and Katz (2007) and Casey, Glennerster, and Miguel (2012).

NCLB appears to have had divergent impacts on school and home behavior. In the classroom, the policy improved non-cognitive skills for ethnic groups that were explicitly targeted by it, and worsened those skills for others. At home, the impact of NCLB was a decline in certain skills for most sub-groups, with overall negative or statistically insignificant effects.

Tables Table 13 through Table 20 provide the results from the supplementary analysis of students' ratings of their own non-cognitive skills as well as standardized tests of literacy and numeracy. Table 13 contains the DID results for the entire sample and reveals little change due to NCLB. Table 14 through Table 16 uncover significant heterogeneity in the results. White and free lunch students rated themselves worse and Hispanic students rated themselves better for both sets of problem behaviors. LEP and Asian students reported no significant change, while African-American students reported a decline in peer relations. In tables Table 17 through Table 20, I find effects on student achievement that are qualitatively commensurate with the prior literature. Overall, I find a statistically significant increase in math scores of one tenth of a standard deviation and no change in reading scores. I find statistically significant increases in both scores for white and African-American students of roughly one tenth (math) to one twentieth (reading) of a standard deviation in each group. I find an increase in math scores and decrease in reading scores for free lunch students. No significant changes are present for Hispanic, Asian, and LEP students. Though the nature of the heterogeneity is somewhat different, taken as a whole these results support the conclusion from the prior literature that NCLB had an overall stronger effect on math scores and a weaker or non-existent effect on reading scores.

The ECLS-K data allows tests of several potential mechanisms through which NCLB's accountability requirements impacted non-cognitive skills. The stock of teachers might have changed with respect to pedagogical training in or emphasis of non-cognitive skills development. ECLS-K asks teachers how many child development courses they have taken in college as well as how much emphasis they place on non-cognitive skills and standardized testing when evaluating students. Figure 6 lists the coefficients and 90% confidence intervals for the impact of NCLB on the number of child development courses teachers report taking in college. Overall, the policy appears to have decreased teacher training in this area, though only two coefficients are negative and significant and one is positive and significant. A systematic increase can be ruled out.

Table 21 contains the impact of NCLB on how much importance teachers place on students' effort, completion of homework, participation, and behavior as well as students' achievement on standardized tests. The variables are ordinal scales (not important to extremely important) that have been dichotomized into extremely important (1) and not extremely important (0). The results are not suggestive of non-cognitive skills becoming less important due to NCLB. African-American students have teachers who place more emphasis on completion of homework and less on participation, Hispanic students have teachers who place more emphasis on standardized tests and less on behavior, Asian students have teachers who place more emphasis on completion of homework and less on standardized tests. Together, these results rule out increases in post-secondary child development courses and shifts in evaluative emphasis (as reported in the ECLS-K surveys) as mechanisms through which NCLB might have impacted non-cognitive skills.

Discussion

What to make of the divergence in effects on teacher vs. parent ratings is unclear. Table 22 provides suggestive evidence that no single domain should be privileged over the others. The table contains the adjusted R^2 statistic taken from regressions of standardized test scores and teachers' ratings of overall academic proficiency in reading and math on three sets of non-cognitive skills: those measured from surveys of teachers, parents, and the children themselves. The data comes from fifth grade tests and surveys. The only clear pattern is that teachers' ratings of non-cognitive skills are more closely correlated with their ratings of academic proficiency than are parents' and children's ratings of non-cognitive skills.

Teacher ratings have the benefit of being drawn from a greater number of survey items that address specific situations rather than the parent ratings which are each from a single question. Moreover, teachers are more likely than parents to have exposure to children along the entire distribution of each skill being measured. The potential impacts of exposure on ratings is manifest in the drop in parents' ratings of children between the fall of kindergarten and the spring of first grade (see figure Figure 4). No such change is observed on the teacher ratings, suggesting that parents think more highly of their children when they have less exposure to a school setting. A similar dynamic could be happening with the onset of NCLB if the law increased parents' interactions with schools. Table 23 reveals that NCLB increased parental attendance at open houses, parents' informal meetings with teachers, and parent-initiated contact with teachers. The law also marginally increased the number of other parents the survey child's parents spoke with regularly, but the effect is just under the .1 significance threshold (p = .11). Ultimately, the data is unable to fully account for the discrepancy of NCLB effects on parents and teacher ratings.

An important limitation of this research is that the structure of this data does not allow the estimation of NCLB effects at numerous ages and stages of child development. The weak and overall null effects of the law on non-cognitive skills might be local to the age group represented here: late elementary school. If teachers are sufficiently focused on non-cognitive skill development at these ages, then a pedagogical shift toward literacy and numeracy may not have reduced this focus appreciably. Alternately, teachers of this age group may specifically target non-cognitive skills as a means to enhancing literacy and numeracy. The impact of NCLB on the non-cognitive skills of older children cannot be ascertained from this analysis, yet is vital to understand.

Finally, it bears repeating that this is a test of the federal accountability mandate on students' non-cognitive skills rather than a test of any particular accountability system and accompanying endogenous implementation choices. Though limited, this focus on the federal mandate provides salient information on the type of policy the federal government is likely to be able to enact. As the federalist tensions in the drafting and implementation of NCLB attest to, the U.S. Department of Education has a limited amount of power to impact local schools (Manna, 2011). It is therefore essential to understand how this major federal policy impacted students overall, without limiting the analysis to only those states whose implementation was particularly faithful to the law's intent.

Conclusion

Despite well founded fears, No Child Left Behind does not appear to have seriously harmed the non-cognitive skills of students targeted by the policy- at least as measured by teachers at the end of elementary school. However, the onset of NCLB is associated with parents' reduction in claiming their children to be above average on four survey items. The policy implications of these findings are twofold. First, the heterogeneous impacts of NCLB on non-cognitive skills as measured by teachers suggests that numeracy and non-cognitive skills are not pedagogical substitutes and might be complements. Given that NCLB did not explicitly incentivize the latter, the results suggest that a policy that focuses on both might reap even greater results. Second, the divergence between parents' and teachers' reports suggests that gains in non-cognitive skills from education policy might not spill over into home behavior and might even negatively co-vary with it. This underscores the importance of assessing differences across multiple measures of non-cognitive skills (Renk & Phares, 2004), particularly if they are to be components accountability or value-added regimes. At least several measures of non-cognitive skills should be included in state administrative data sets so that these skills can be measured more reliably than serendipitous panel studies allow. Doing so would allow research on the permanency of policy-induced changes in these skills. Future research should also establish the stability of the psychometric properties of these measures in the face of accountability and performance pay pressure. At very least, these measures should be routinely collected in educational randomized field trials. Currently, the Institute of Educational Science's What Works Clearinghouse lists 404 interventions focused on literacy and numeracy and only 56 focused on non-cognitive skills, many of which are focused on children with clinical behavior disorders.[13] The emerging research on non-cognitive skills is increasingly suggesting that this imbalance in research priorities is unjustified. As this analysis reveals, a more nuanced picture of NCLB arises when non-cognitive skills are included- future research on education policy should do the same.

[13] See http://ies.ed.gov/ncee/wwc/, accessed 9-15-15

Works Cited

Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *Review of Economic Studies*, *72*, 1–19.

Adamson, F., & Darling-Hammond, L. (2012). Funding Disparities and the Inequitable Distribution of Teachers: Evaluating Sources and Solutions. *Education Policy Analysis Archives*, *20*(37).

Almlund, M., Duckworth, A. L., Heckman, J. J., & Kautz, T. D. (2011). *Personality Psychology and Economics* (No. 16822). Cambridge. Retrieved from http://www.nber.org/papers/w16822

Barnett, W. S., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk, R., & Burns, S. (2008). Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly*, *23*(3), 299–313. doi:10.1016/j.ecresq.2008.03.001

Card, D., & Krueger, A. B. (1992). Does School Quality Matter ? Returns to Education and the Characteristics of Public Schools in the United States. *The Journal of Political Economy*, *100*(1), 1–40.

Casey, K., Glennerster, R., & Miguel, E. (2012). RESHAPING INSTITUTIONS : EVIDENCE ON AID IMPACTS USING A PREANALYSIS PLAN. *Quarterly Journal of Economics*, *127*(4), 1755–1812.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from project star. *Quarterly Journal of Economics*, *126*(4), 1593–1660. doi:10.1093/qje/qjr041

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2013). *Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood* (No. 19424). Retrieved from http://www.nber.org/papers/w19424

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review*, *24*(4), 377–392. doi:10.1016/j.econedurev.2004.06.008

Darling-Hammond, L. (2007). Evaluating "No Child Left Behind." *The Nation*. Retrieved from http://www.thenation.com/article/evaluating-no-child-left-behind/

Davidson, E., Reback, R., Rockoff, J., & Schwartz, H. (2013). Fifty ways to leave a child behind: Idiosyncrasies and discrepancies in states' implementation of NCLB (No. w18988). *National Bureau of Economic Research*. Retrieved from http://www.nber.org/papers/w18988

Dee, T. S., & Jacob, B. (2010). *The Impact of No Child Left Behind on Students, Teachers, and Schools*. *Brookings Papers on Economic Activity* (Vol. 2010). Washington, D.C. doi:10.1353/eca.2010.0014

Dee, T. S., & Jacob, B. (2011). The Impact of No Child Left Behind on Student Achievement. *Journal of Policy Analysis and Management*, *30*(3), 418–446.

Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2012). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology*, *104*(2), 439–451. doi:10.1037/a0026280

Duckworth, A. L., & Yeager, D. S. (2015). Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes. *Educational Researcher*, *44*(4), 237–251.

Figlio, D. N., & Ladd, H. F. (2015). School Accountability and Student Achievement. In H. F. Ladd & M. E. Goertz (Eds.), *Handbook of Research in Education Finance and Policy* (Second., p. 676). New York: Routledge.

Gewertz, C. (2003, September 3). Hand in Hand. *Education Week*, pp. 38–41. Bethesda, MD. Retrieved from http://www.edweek.org/ew/articles/2003/09/03/01social.h23.html

Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven Playing Field? Assessing the Teacher Quality Gap Between Advantaged and Disadvantaged Students. *Educational Researcher*, *44*(5), 293–307. doi:10.3102/0013189X15592622

Grissom, J. a., Nicholson-Crotty, S., & Harrington, J. R. (2014). Estimating the Effects of No Child Left Behind on Teachers' Work Environments and Job Attitudes. *Educational Evaluation and Policy Analysis*, *36*(4), 417–436. doi:10.3102/0162373714533817

Hanushek, E. A. (2003). The Failure of Input-based Schooling Policies*. *The Economic Journal*, *113*(485), F64–F98.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *30*(3), 466–479.

Hanushek, E. A., & Raymond, M. E. (2001). The Confusing World of Educational Accountability. *National Tax Journal*, *54*(2), 365–384.

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, *103*(6), 2052–2086.

Jackson, C. K. (2012). *Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina* (No. 18624).

Jacob, B. A. (2005). Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, *89*(5-6), 761–796. doi:10.1016/j.jpubeco.2004.08.004

Jacob, B. A., & Levitt, S. D. (2003). Rotten Apples : An Investigation of the Prevalence and Predictors of Teacher Cheating. *Quarterly Journal of Economics*, *118*(3), 843–877.

Junttila, N., Voeten, M., Kaukiainen, A., & Vauras, M. (2006). Multisource Assessment of Children's Social Competence. *Educational and Psychological Measurement*, *66*(5), 874–895. doi:10.1177/0013164405285546

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). EXPERIMENTAL ANALYSIS OF NEIGHBORHOOD EFFECTS.

Econometrica, *75*(1), 83–119.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37–62. doi:10.3102/01623737024001037

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. doi:10.3102/0162373711431604

Manna, P. (2011). *Collision Course: Federal Education Policy Meets State and Local Realities*. Washington, D.C.: CQ Press.

McGuinn, P. J. (2006). *No Child Left Behind and the Transformation of Federal Education Policy, 1965-2005*. Lawrence, KS: University Press of Kansas.

Reardon, S. F. (2011). The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances* (pp. 91–116). New York: Russell Sage Foundation.

Reardon, S. F., Greenberg, E. H., Kalogrides, D., Shores, K. A., & Valentino, R. A. (2013). *Left Behind? The Effect of No Child Left Behind on Academic Achievement Gaps*. *working paper*. doi:10.1227/00006123-197907010-00030

Reardon, S. F., & Ho, a. D. (2015). Practical Issues in Estimating Achievement Gaps From Coarsened Data. *Journal of Educational and Behavioral Statistics*, *40*(2), 158–189. doi:10.3102/1076998615570944

Reback, R., Rockoff, J., & Schwartz, H. L. (2014). Under pressure: Job security, resource allocation, and productivity in schools under No Child Left Behind. *American Economic Journal: Economic Policy*, *6*(3), 207–241. doi:10.1257/pol.6.3.207

Renk, K., & Phares, V. (2004). Cross-informant ratings of social competence in children and adolescents. *Clinical Psychology Review*, *24*(2), 239–254. doi:10.1016/j.cpr.2004.01.004

Somers, M.-A., Zhu, P., Jacob, R., & Bloom, H. (2012). *The Validity and Precision of the Comparative Interrupted Time Series Design and the Difference-in-Difference Design in Educational Evaluation*. Retrieved from http://appam.confex.com/data/extendedabstract/appam/2012/Paper_1758_extendedabstract_156_0.pdf

Springer, M. G., Houck, E. A., & Guthrie, J. W. (2015). History and Scholarship Regarding U.S. Education Finance and Policy. In H. F. Ladd & M. E. Goertz (Eds.), *Handbook of Research in Education Finance and Policy* (2nd ed., p. 676). New York: Routledge.

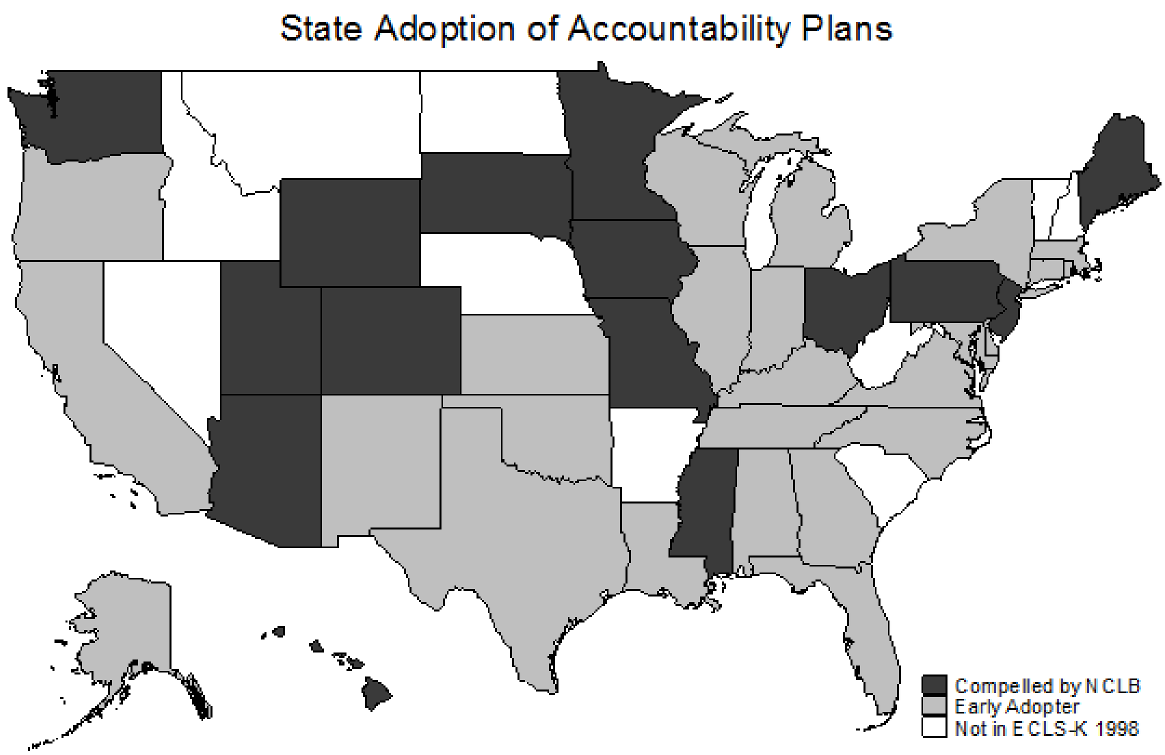
St.Clair, T., Cook, T. D., & Hallberg, K. (2014). Examining the Internal Validity and Statistical Precision of the Comparative Interrupted Time Series Design by Comparison With a Randomized Experiment. *American Journal of Evaluation*, *35*(3), 311–327. doi:10.1177/1098214014527337

Vinovskis, M. A. (2009). *From a Nation at Risk to No Child Left Behind: National Education Goals and the Creation of Federal Education Policy*. New York: Teachers College Press.

Wong, M., Cook, T. D., & Steiner, P. M. (2011). *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* (No. WP-09-11). *Institute for Policy Research*. Evanston, IL. Retrieved from http://www.ipr.northwestern.edu/publications/workingpapers/2009/wp0911.pdf

Appendix A

Figure 1: Early Adopter and Late Adopter States in the ECLS-K 1998



Source: Dee and Jacob (2011) *and the Early Childhood Longitudinal Study of Kindergarten, 1998, base year file*

Table 1: ECLS-K Summary Statistics at Fall Kindergarten Students Attending Public Schools

	Early Adopter	Late Adopter
	Students	
White	**49%**	**62%**
Black	**17%**	**13%**
Hispanic	**23%**	**11%**
Attended Head Start	**12%**	**17%**
Limited English Proficiency	**25%**	**16%**
Free/Reduced Price Lunch	**35%**	**27%**
	Teachers & Schools	
Teacher's with MA	**34%**	**27%**
Title 1 School	**52%**	**67%**

Source: Early Childhood Longitudinal Study of Kindergarten, 1998, base year file

Table 2: Distribution of Parental Assessments of Children's Non-cognitive Skills

Compared to children of the same age, the child can

	Independent	Pay Attention	Solve Problems	Behave/ Relate to others
Better	**33%**	**25%**	**34%**	**27%**
As well	**60%**	**61%**	**57%**	**66%**
Slightly less well	**5.6%**	**12%**	**7.9%**	**5.7%**
Much less well	**0.9%**	**1.8%**	**1%**	**0.8%**

Source: Early Childhood Longitudinal Study of Kindergarten, 1998, base year file

Table 3: Distributions of the Social Rating Scale (Teachers) and the Self-Description Questionnaire (Children)

Teacher SRS	Mean	Median	Standard Deviation
Self-Control	**3.18**	**3.25**	**.614**
Approaches to Learning	**3.07**	**3.14**	**.684**
Interpersonal Skills	**3.08**	**3**	**.64**
Externalizing Behaviors (Reversed)	**1.64 (3.36)**	**1.5 (3.5)**	**.612 (.612)**
Internalizing Behaviors (Reversed)	**1.58 (3.42)**	**1.5 (3.5)**	**.522 (.522)**
Child SDQ			
Interpersonal Skills	**3.01**	**3**	**.621**
Externalizing Behaviors (Reversed)	**1.83 (3.17)**	**1.67 (3.33)**	**.65 (.65**
Internalizing Behaviors (Reversed)	**2.1 (2.9)**	**2 (3)**	**.649 (.649)**

Source: Early Childhood Longitudinal Study of Kindergarten, 1998, base year file

Table 4: Data Sources in the ECLS-K

Outcome	**Questionnaire**	**Pre-NCLB**			**Post-NCLB**	
Outcomes for Primary Models		**Kindergarten**	**1st Grade**	**3rd Grade**	**5th Grade**	**8th Grade**
Independence	**Parent**	**X**	**X**	**X**	**X**	**X**
Pays Attention	**Parent**	**X**	**X**	**X**	**X**	**X**
Solve Problems	**Parent**	**X**	**X**	**X**	**X**	**X**
Behaves Well	**Parent**	**X**	**X**	**X**	**X**	**X**
Self-Control	**Teacher**	**X**	**X**	**X**	**X**	
Externalizing	**Teacher**	**X**	**X**	**X**	**X**	
Internalizing	**Teacher**	**X**	**X**	**X**	**X**	
Learning Approach	**Teacher**	**X**	**X**	**X**	**X**	
Interpersonal Skill	**Teacher**	**X**	**X**	**X**	**X**	
Outcomes for Secondary Models						
Interpersonal Skill	**Student**			**X**	**X**	
Externalizing	**Student**			**X**	**X**	
Internalizing	**Student**			**X**	**X**	**X**

Source: Early Childhood Longitudinal Study of Kindergarten, 1998

Table 5: Constructs from ECLS-K 1998

Construct	**Label**	**Items per Survey**	**Examples of Behavior**
Self-Control	CONTRO	4	controlling temper, accepting peer ideas
Approaches to Learning	LEARN	6	attentiveness, task persistence, eagerness to learn
Peer Relations	PEER	5	expressing feelings/ideas/opinions in positive ways
Externalizing Problems	EXTERN	5	argues, gets angry, acts impulsively, disruptive
Internalizing Problems	INTERN	4	anxiety, loneliness, low self-esteem, and sadness
Independence	INDEP	1	independent and takes care of herself
Pay Attention	ATTN	1	pays attentions
Solve Problems	SOLVE	1	learn, think, and solve problems
Overall Behavior	BEHAV	1	behaves /relates to children and adults

Source: Early Childhood Longitudinal Study of Kindergarten, 1998

Table 6: Correlations amongst Non-cognitive Skills in the Fall Kindergarten Survey (1998)

	INDEP	ATTN	SOLVE	BEHAV	EXTERN	INTERN	PEER	CONTRO	LEARN
INDEP	1								
ATTN	0.231	1							
SOLVE	0.254	0.280	1						
BEHAV	0.234	0.264	0.217	1					
EXTERN	0.059	0.195	0.051	0.117	1				
INTERN	0.061	0.060	0.093	0.065	0.244	1			
PEER	0.074	0.178	0.073	0.121	0.690	0.268	1		
CONTRO	0.096	0.175	0.103	0.141	0.553	0.341	0.779	1	
LEARN	0.134	0.241	0.192	0.140	0.507	0.352	0.662	0.696	1

Source: Early Childhood Longitudinal Study of Kindergarten, 1998

Table 7: Correlation amongst Social Rating Scale and Academic Assessments in the Fall Kindergarten Survey (1998)

	EXTERN	INTERN	PEER	CONTRO	LEARN	READ	MATH	GEN
EXTERN	1.00							
INTERN	0.28	1.00						
PEER	0.57	0.35	1.00					
CONTRO	0.70	0.29	0.79	1.00				
LEARN	0.51	0.37	0.70	0.67	1.00			
READ	0.13	0.13	0.20	0.17	0.34	1.00		
MATH	0.15	0.18	0.24	0.20	0.41	0.72	1.00	
GEN	0.14	0.15	0.25	0.20	0.34	0.50	0.61	1.00

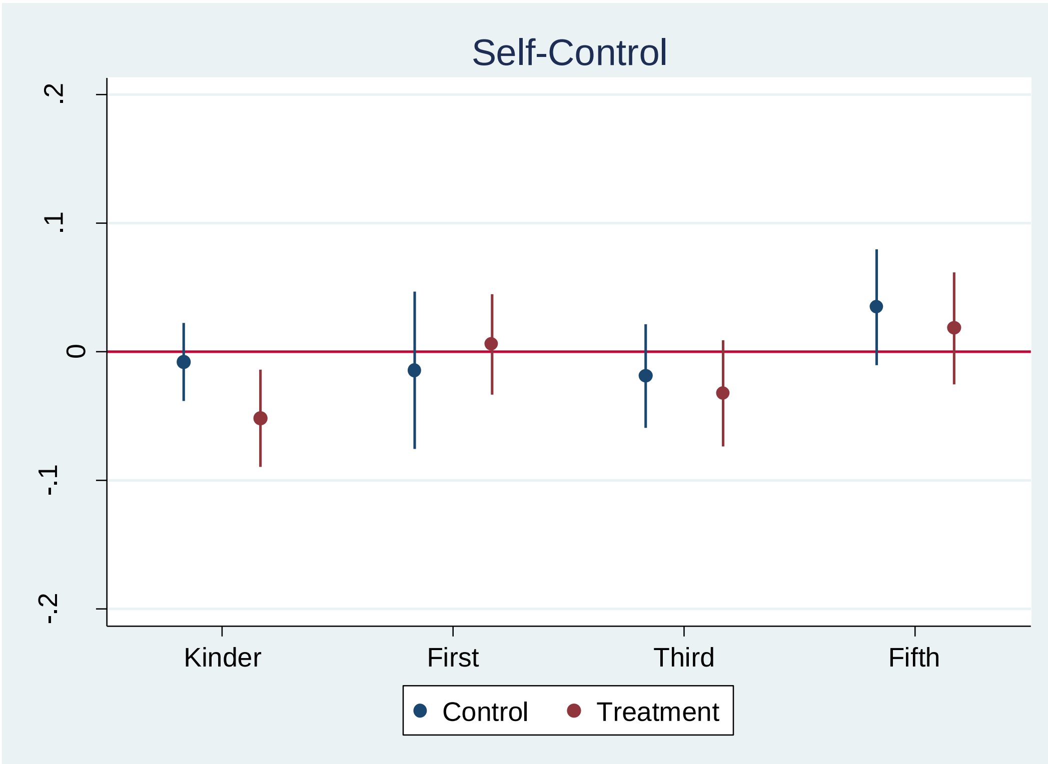
Source: Early Childhood Longitudinal Study of Kindergarten, 1998

Table 8: Within Child Variance and Skills Gaps

	% of Variance within Child	White-Black Gap	White-Hisp Gap	Treat-Control Gap
READ	0.82	0.56	0.55	-0.05
CONTRO	0.45	0.38	0.16	-0.04
EXTERN	0.36	0.32	0	-0.03
INTERN	0.55	0.1	0.12	0.02
PEER	0.46	0.34	0.2	-0.04
LEARN	0.37	0.38	0.26	0.01
BEHAV	0.52	0.04	0	-0.01
ATTN	0.5	0.08	-0.02	-0.02
SOLVE	0.48	0.08	0.06	0
INDEP	0.51	0.02	0	-0.01

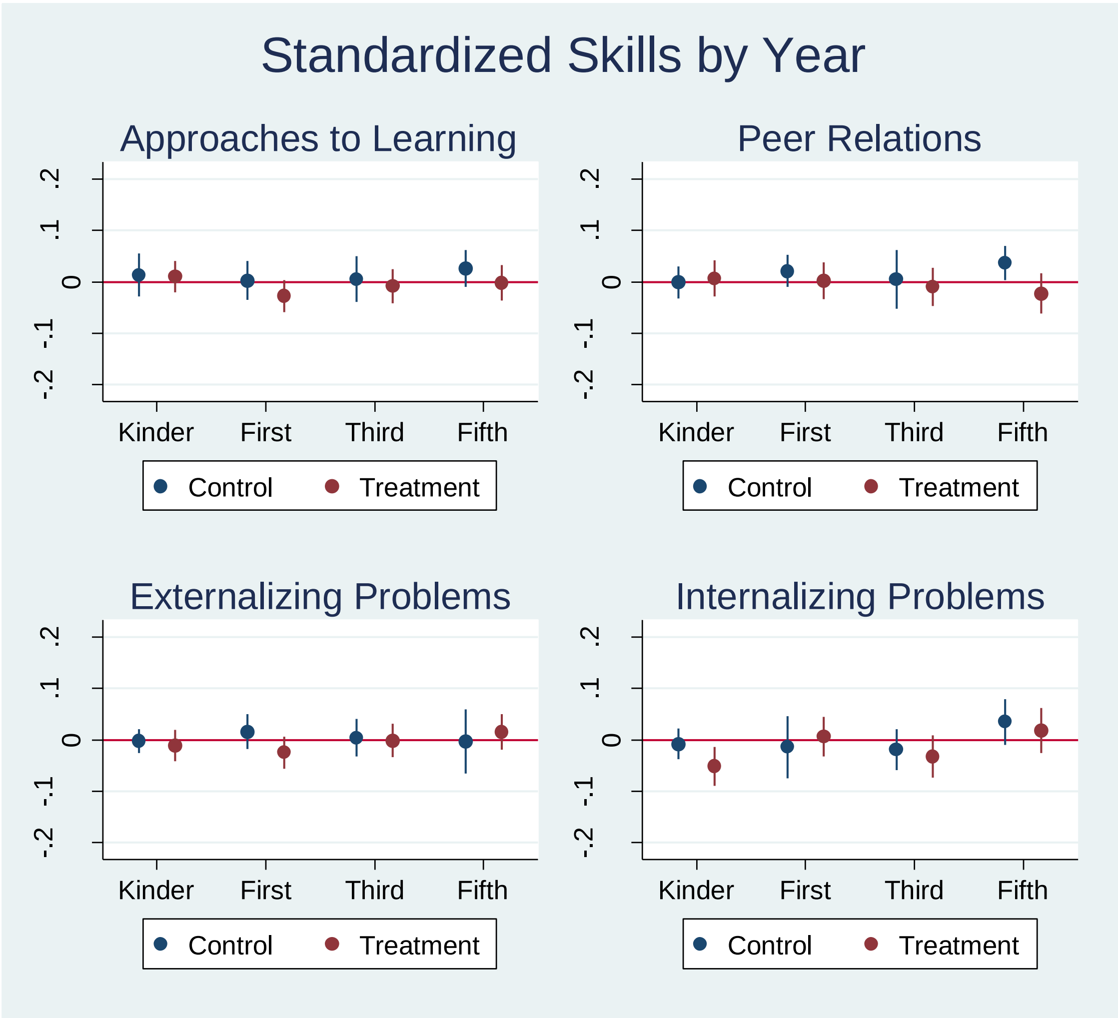
Source: Early Childhood Longitudinal Study of Kindergarten, 1998

Figure 2: Standardized Self-Control from the Social Rating Scale



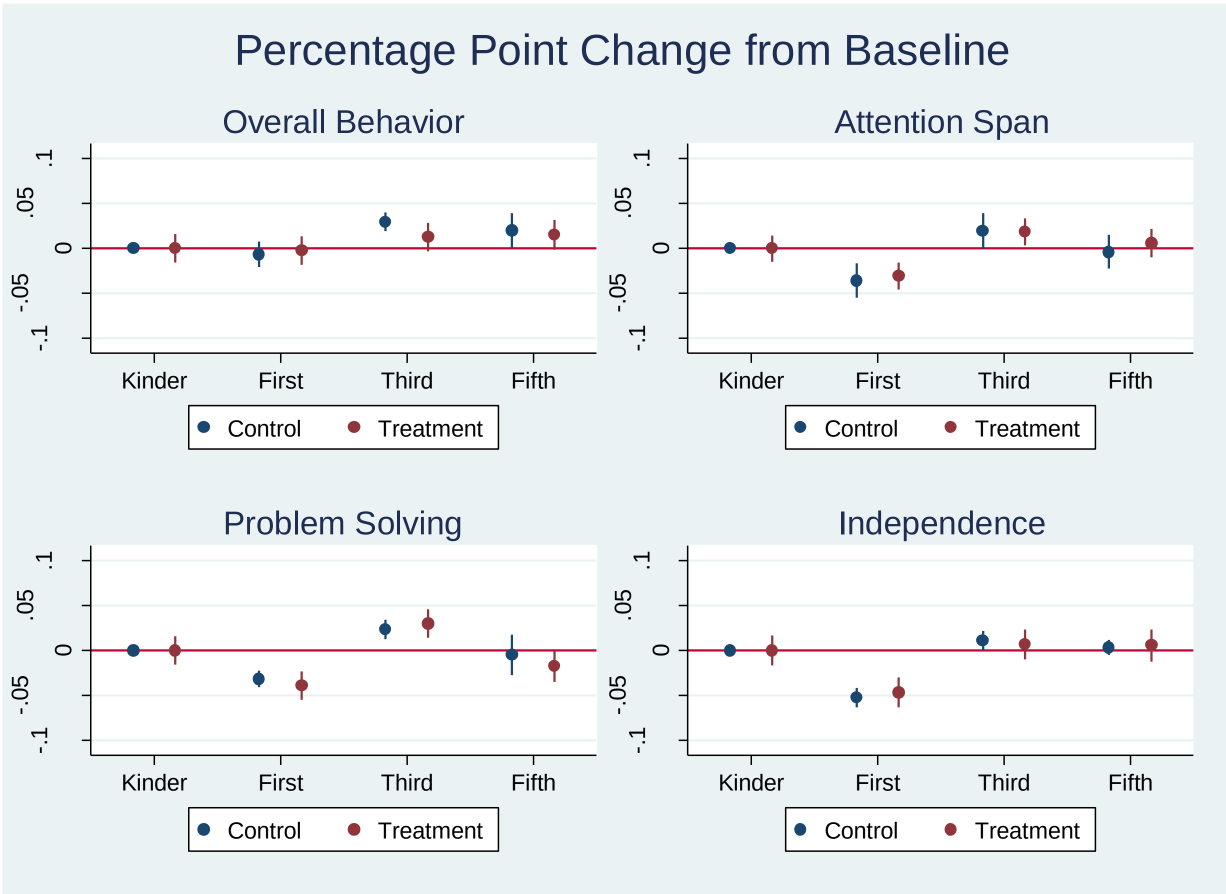
Source: Early Childhood Longitudinal Study of Kindergarten, 1998

Figure 3: Standardized Skills from the Social Rating Scale



Source: Early Childhood Longitudinal Study of Kindergarten, 1998

Figure 4: Parent Ratings of Students' Skills by Grade and Treatment Group



Source: Early Childhood Longitudinal Study of Kindergarten, 1998

Table 9: The Impact of NCLB on Teacher ratings of students' Non-cognitive Skills

	Self-Control	Externalizing Problems	Internalizing Problems	Peer Relations	Approaches to Learning
Year	0.009	0.005	-0.011	0.007	0.005
	(0.008)	(0.007)	(0.008)	(0.007)	(0.005)
NCLB	-0.001	-0.015	0.057	0.020	0.013
	(0.036)	(0.041)	(0.030)*	(0.028)	(0.026)
Year*Treat	-0.007	-0.015	-0.007	-0.009	-0.014
	(0.008)	(0.007)*	(0.008)	(0.007)	(0.005)**
NCLB*Treat	-0.037	0.057	-0.001	-0.044	0.003
	(0.037)	(0.041)	(0.030)	(0.029)	(0.026)
SES (Continuous)	-0.015	-0.031	0.036	-0.031	-0.036
	(0.023)	(0.017)*	(0.018)*	(0.025)	(0.024)
Father Present	0.014	0.018	0.111	0.045	0.062
	(0.021)	(0.015)	(0.027)***	(0.027)	(0.035)
Father Employed Fulltime	0.044	0.018	0.029	0.038	0.015
	(0.022)*	(0.015)	(0.030)	(0.019)*	(0.019)
Constant	-0.095	-0.069	-0.121	-0.123	-0.111
	(0.012)***	(0.009)***	(0.034)***	(0.015)***	(0.025)***
R^2	0.00	0.00	0.00	0.00	0.00
N	39,539	39,828	39,497	39,218	40,102

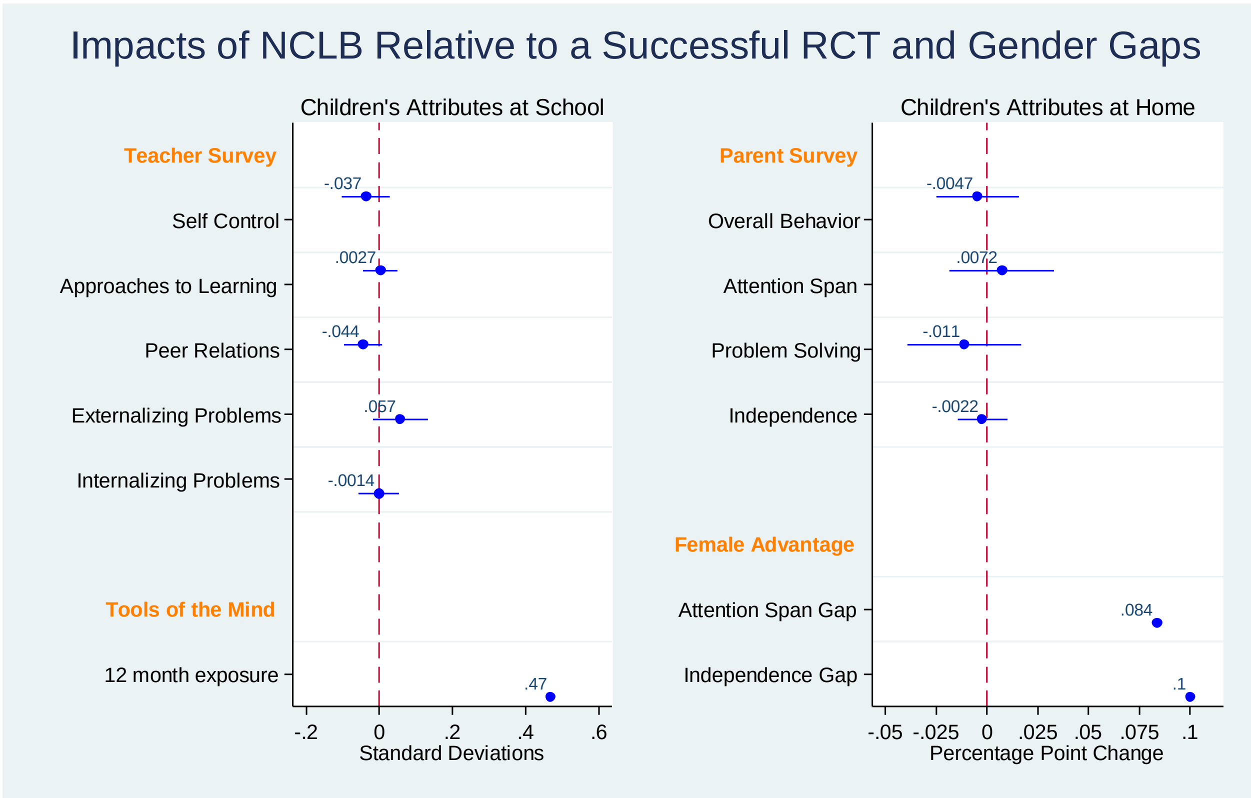
* p<0.1; ** p<0.05; *** p<0.01

Table 10: The Impact of NCLB on Parent ratings of students' Non-cognitive Skills

	Overall Behavior	Attention Span	Problem Solving	Independence
Year	0.003	-0.003	-0.001	-0.009
	(0.001)*	(0.003)	(0.002)	(0.001)***
NCLB	0.018	0.013	0.009	0.035
	(0.011)	(0.014)	(0.015)	(0.007)***
Year*Treat	-0.001	0.001	0.001	0.002
	(0.001)	(0.003)	(0.002)	(0.001)
NCLB*Treat	-0.005	0.007	-0.011	-0.002
	(0.011)	(0.014)	(0.015)	(0.007)
SES (Continuous)	0.019	0.016	0.007	0.005
	(0.006)**	(0.011)	(0.004)*	(0.006)
Father Present	0.002	-0.003	0.009	0.011
	(0.007)	(0.009)	(0.010)	(0.018)
Father Employed Fulltime	-0.024	0.007	-0.005	0.009
	(0.010)**	(0.009)	(0.007)	(0.014)
Constant	0.261	0.238	0.320	0.333
	(0.007)***	(0.011)***	(0.007)***	(0.009)***
R^2	0.00	0.00	0.00	0.00
N	47,878	47,884	47,817	47,904

* p<0.1; ** p<0.05; *** p<0.01

Figure 5: Null Results of NCLB in Comparison to a successful RCT and Female Advantage



Source: Early Childhood Longitudinal Study of Kindergarten, 1998

Table 11: Effects of NCLB on SRS Ratings

Sub-Group	Self-Control	Peer Relations	Approaches to Learning	Internalizing Problems	Externalizing Problems	Mean Effect
White	-0.071*	-0.080**	-0.030	-0.039	0.058	-.033
	(0.037)	(0.034)	(0.026)	(0.035)	(0.043)	(.032)
African-American	.092	.116	.200	.271**	.085	.153
	(.060)	(.103)	(.125)	(.100)	(.100)	(.098)
Hispanic	.035	.115**	.096	-.024	.081	.061
	(.060)	(.045)	(.060)	(.081)	(.096)	(.088)
Asian	.026	-.138	-.125	-.053	-.123**	-.083
	(.100)	(.090)	(.080)	(.093)	(.049)	(.103)
Free Lunch	-.031	-.017	.023	.118	.003	.019
	(.041)	(.067)	(.042)	(.075)	(.067)	(.059)
Limited English	-.063	-.091	-.135*	.090	.008	-.038
	(.080)	(.080)	(.073)	(.082)	(.062)	(.065)

*Standard Errors in Parentheses, * p<0.1; ** p<0.05; *** p<0.01, Source: ECLS-K 1998*

Table 12: The Impacts of NCLB on Parents' Ratings of Children's Non-cognitive Skills

Sub-Group	**Overall Behavior**	**Problem Solving**	**Attention Span**	**Independence**	**Mean Effect**
White	-0.006	0.004	0.011	0.008	.004
	(.014)	(.017)	(.012)	(.013)	(.011)
African-American	-.066*	-0.080**	-0.052	-0.112***	-.077**
	(.030)	(0.033)	(.053)	(.031)	(.036)
Hispanic	.006	-0.011	0.049	0.066**	.027
	(.020)	(.039)	(.046)	(.028)	(.033)
Asian	.072	-0.096***	-0.003	-0.059*	-.021
	(.041)	(.022)	(.027)	(.031)	(.053)
Free Lunch	.008	-0.031	-0.018	-0.075***	-.029
	(.026)	(.036)	(.039)	(.020)	(.022)
Limited English	-.034**	-0.017	-0.074***	-0.042**	-.042*
	(.012)	(.013)	(.016)	(.018)	(.026)

*Standard Errors in Parentheses, * p<0.1; ** p<0.05; *** p<0.01, Source: ECLS-K 1998*

Table 13: The Impact of NCLB on Students' Ratings of Non-cognitive Skills (Diff-in-Diff model)

	Peer Relations	Internalizing Problems	Externalizing Problems
NCLB	-0.008	0.007	-0.002
	(0.008)	(0.017)	(0.013)
NCLB*Treat	0.009	-0.027	-0.013
	(0.009)	(0.017)	(0.013)
SES (Continuous)	0.116	-0.060	-0.041
	(0.047)**	(0.030)*	(0.051)
Father Present	0.072	0.036	0.006
	(0.037)*	(0.039)	(0.044)
Father Employed Fulltime	-0.091	-0.046	0.010
	(0.032)**	(0.041)	(0.022)
Constant	0.014	0.011	0.014
	(0.021)	(0.028)	(0.038)
R^2	0.00	0.00	0.00
N	15,348	15,348	15,348

* *p*<0.1; ** *p*<0.05; *** *p*<0.01, *Source: ECLS-K 1998*

Table 14: The Impact of NCLB on Students' Ratings of Non-cognitive Skills by Race

	Peer Relations	Externalizing Problems	Internalizing Problems
NCLB	-0.022	-0.004	0.010
	(0.020)	(0.011)	(0.021)
NCLB*Treat	0.028	-0.022	-0.051
	(0.020)	(0.011)*	(0.021)**
SES (Continuous)	0.111	-0.042	-0.060
	(0.048)**	(0.050)	(0.031)*
Father Present	0.077	0.008	0.037
	(0.036)*	(0.042)	(0.041)
Father Employed Fulltime	-0.091	0.009	-0.045
	(0.032)**	(0.022)	(0.041)
African-American*NCLB	0.112	0.006	0.103
	(0.098)	(0.067)	(0.046)**
African-American*NCLB*Treat	-0.204	0.016	0.058
	(0.097)*	(0.067)	(0.046)
Hispanic*NCLB	0.099	-0.029	-0.037
	(0.052)*	(0.031)	(0.052)
Hispanic*NCLB*Treat	-0.014	0.126	0.162
	(0.051)	(0.031)***	(0.052)**
Asian*NCLB	-0.083	0.050	-0.090
	(0.063)	(0.085)	(0.046)*
Asian*NCLB*Treat	-0.013	-0.069	-0.074
	(0.063)	(0.086)	(0.046)
Constant	0.010	0.013	0.010
	(0.022)	(0.036)	(0.027)
R^2	0.00	0.00	0.00
N	15,348	15,348	15,348

* p<0.1; ** p<0.05; *** p<0.01, *Source: ECLS-K 1998*

Table 15: The Impact of NCLB on Students' Ratings of Non-cognitive Skills by Free/Reduced Price Lunch Status

	Peer Relations	Externalizing Problems	Internalizing Problems
NCLB	-0.040	-0.010	-0.016
	(0.016)**	(0.018)	(0.019)
NCLB*Treat	0.007	0.017	-0.005
	(0.017)	(0.018)	(0.019)
SES (Continuous)	0.112	-0.044	-0.064
	(0.049)**	(0.051)	(0.032)*
Father Present	0.072	0.004	0.035
	(0.035)*	(0.044)	(0.039)
Father Employed Fulltime	-0.090	0.012	-0.044
	(0.032)**	(0.022)	(0.041)
Free/Reduced Lunch*NCLB	0.126	0.032	0.091
	(0.035)***	(0.039)	(0.042)*
Free/Reduced Lunch*NCLB*Treat	0.013	-0.124	-0.088
	(0.036)	(0.039)***	(0.042)*
Constant	0.013	0.014	0.010
	(0.021)	(0.038)	(0.028)
R^2	0.00	0.00	0.00
N	15,348	15,348	15,348

* p<0.1; ** p<0.05; *** p<0.01, *Source: ECLS-K 1998*

Table 16: The Impact of NCLB on Students' Ratings of Non-cognitive Skills by LEP Status

	Peer Relations	Externalizing Problems	Internalizing Problems
NCLB	-0.011	-0.012	0.015
	(0.011)	(0.011)	(0.020)
NCLB*Treat	0.006	-0.015	-0.029
	(0.012)	(0.011)	(0.020)
SES (Continuous)	0.116	-0.039	-0.061
	(0.047)**	(0.050)	(0.030)*
Father Present	0.072	0.004	0.038
	(0.037)*	(0.043)	(0.039)
Father Employed Fulltime	-0.091	0.010	-0.046
	(0.032)**	(0.023)	(0.041)
Limited Eng Prof*NCLB	0.011	0.044	-0.035
	(0.025)	(0.038)	(0.042)
Limited Eng Prof*NCLB*Treat	0.022	0.033	-0.004
	(0.025)	(0.038)	(0.042)
Constant	0.015	0.016	0.010
	(0.021)	(0.038)	(0.028)
R^2	0.00	0.00	0.00
N	15,348	15,348	15,348

* p<0.1; ** p<0.05; *** p<0.01, Source: ECLS-K 1998

Table 17: The Impact of NCLB on Standardized Math and Reading Scores

	Std. Math Score	Std. Reading Score
Year	0.029	0.012
	(0.010)**	(0.014)
NCLB	-0.074	-0.045
	(0.036)*	(0.028)
Year*Treat	-0.039	-0.004
	(0.009)***	(0.014)
NCLB*Treat	0.104	0.044
	(0.036)**	(0.028)
SES (Continuous)	0.032	0.018
	(0.020)	(0.018)
Father Present	-0.015	-0.031
	(0.025)	(0.032)
Father Employed Fulltime	-0.007	-0.026
	(0.020)	(0.021)
Constant	-0.036	-0.064
	(0.032)	(0.028)**
R^2	0.00	0.00
N	21,063	19,628

* p<0.1; ** p<0.05; *** p<0.01, Source: ECLS-K 1998

Table 18: The Impact of NCLB on Standardized Tests of Math and Reading by Race

	Std. Math Score	Std. Reading Score
Year	0.034	0.036
	(0.011)**	(0.012)**
NCLB	-0.087	-0.087
	(0.040)*	(0.024)***
Year*Treat	-0.032	-0.002
	(0.011)**	(0.012)
NCLB*Treat	0.098	0.049
	(0.040)**	(0.024)*
SES (Continuous)	0.028	0.015
	(0.019)	(0.018)
Father Present	-0.011	-0.031
	(0.022)	(0.027)
Father Employed Fulltime	-0.000	-0.013
	(0.018)	(0.017)
African-American*Year	-0.077	-0.154
	(0.012)***	(0.016)***
African-American*NCLB	0.041	0.216
	(0.041)	(0.033)***
African-American*Year*Treat	-0.057	-0.035
	(0.012)***	(0.017)*
African-American*NCLB*Treat	0.095	0.065
	(0.041)**	(0.035)*
Hispanic*Year	0.011	-0.062
	(0.013)	(0.023)**
Hispanic*NCLB	0.025	0.128
	(0.039)	(0.085)
Hispanic*Year*Treat	-0.007	0.027
	(0.013)	(0.023)
Hispanic*NCLB*Treat	-0.013	-0.092
	(0.040)	(0.085)
Asian*Year	0.005	-0.089
	(0.018)	(0.037)**
Asian*NCLB	0.099	0.200
	(0.049)*	(0.096)*
Asian*Year*Treat	0.018	-0.002
	(0.018)	(0.037)
Asian*NCLB*Treat	-0.074	-0.078
	(0.050)	(0.096)
Constant	-0.045	-0.074
	(0.028)	(0.025)**
R^2	0.02	0.03
N	21,063	19,628

* p<0.1; ** p<0.05; *** p<0.01, Source: ECLS-K 1998

Table 19: The Impact of NCLB on Standardized Tests of Math and Reading by free/reduced price lunch status

	Std. Math Score	Std. Reading Score
Year	0.040	0.035
	(0.010)***	(0.013)**
NCLB	-0.091	-0.099
	(0.041)**	(0.022)***
Year*Treat	-0.026	0.001
	(0.009)**	(0.013)
NCLB*Treat	0.080	0.056
	(0.040)*	(0.022)**
SES (Continuous)	0.030	0.015
	(0.019)	(0.018)
Father Present	-0.012	-0.024
	(0.024)	(0.032)
Father Employed Fulltime	0.002	-0.013
	(0.018)	(0.018)
Free/Reduced Lunch*Year	-0.048	-0.106
	(0.010)***	(0.021)***
Free/Reduced Lunch*NCLB	0.072	0.252
	(0.039)*	(0.030)***
Free/Reduced Lunch*Year*Treat	-0.058	-0.026
	(0.010)***	(0.022)
Free/Reduced Lunch*NCLB*Treat	0.105	-0.058
	(0.038)**	(0.029)*
Constant	-0.045	-0.080
	(0.033)	(0.026)**
R^2	0.02	0.02
N	21,063	19,628

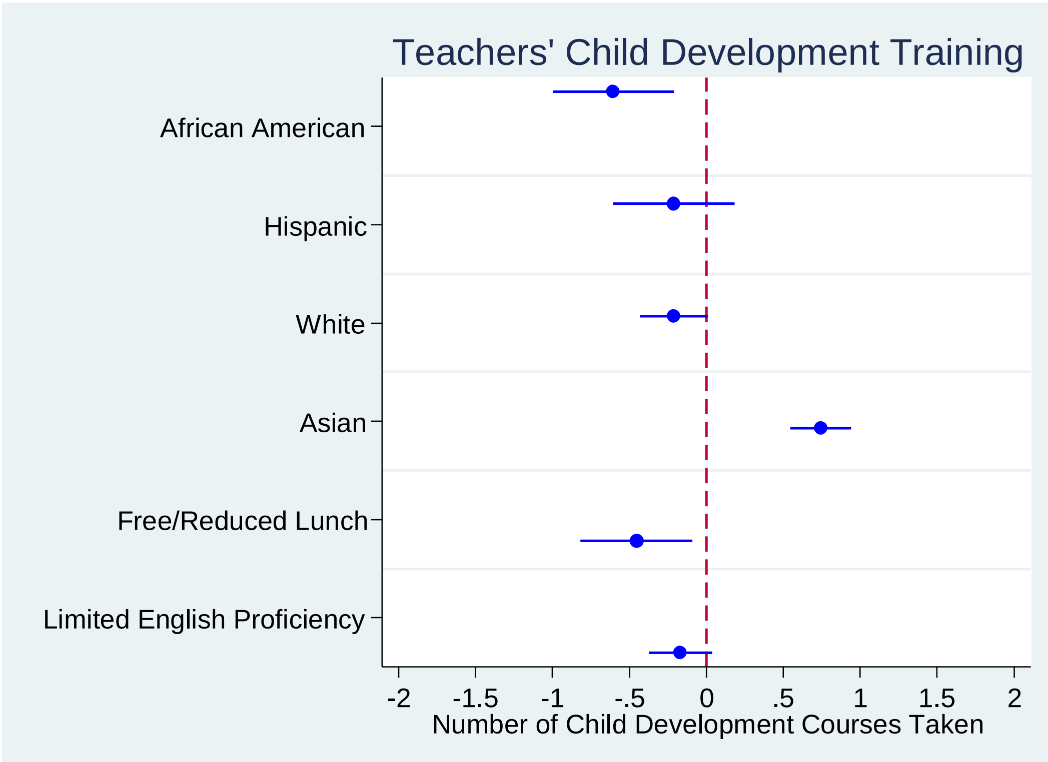
* p<0.1; ** p<0.05; *** p<0.01, Source: ECLS-K 1998

Table 20: The Impact of NCLB on Standardized Tests of Math and Reading by LEP status

	Std. Math Score	Std. Reading Score
Year	0.022	0.012
	(0.010)**	(0.018)
NCLB	-0.067	-0.060
	(0.042)	(0.034)
Year*Treat	-0.033	-0.003
	(0.009)***	(0.017)
NCLB*Treat	0.093	0.066
	(0.042)**	(0.034)*
SES (Continuous)	0.032	0.019
	(0.020)	(0.018)
Father Present	-0.015	-0.032
	(0.024)	(0.031)
Father Employed Fulltime	-0.007	-0.025
	(0.021)	(0.021)
Limited Eng Prof*Year	0.031	-0.001
	(0.010)***	(0.024)
Limited Eng Prof*NCLB	-0.034	0.073
	(0.042)	(0.070)
Limited Eng Prof*Year*Treat	-0.028	-0.005
	(0.010)**	(0.024)
Limited Eng Prof*NCLB*Treat	0.052	-0.118
	(0.042)	(0.070)
Constant	-0.036	-0.064
	(0.031)	(0.027)**
R^2	0.00	0.00
N	21,063	19,628

* p<0.1; ** p<0.05; *** p<0.01, Source: ECLS-K 1998

Figure 6: The Impact of NCLB on Pedagogical Training for Non-cognitive Skills, by Student Characteristic



Source: Early Childhood Longitudinal Study of Kindergarten, 1998

Table 21: The Impact of NCLB on Teacher Emphasis in Student Evaluation

	Effort	Homework	Participate	Behavior	Std. Tests
Year	-0.000	0.046	0.013	-0.031	0.023
	(0.010)	(0.011)***	(0.016)	(0.017)*	(0.012)*
NCLB	-0.008	-0.070	-0.074	-0.008	0.012
	(0.038)	(0.043)	(0.055)	(0.059)	(0.043)
Year*Treat	-0.029	0.007	-0.015	-0.040	-0.011
	(0.010)**	(0.011)	(0.016)	(0.017)**	(0.012)
NCLB*Treat	0.129	0.005	0.067	0.092	0.061
	(0.038)***	(0.043)	(0.055)	(0.059)	(0.043)
African-American*Year	0.025	0.010	-0.031	0.035	-0.019
	(0.023)	(0.022)	(0.019)	(0.024)	(0.017)
African-American*NCLB	-0.044	-0.042	0.148	0.007	0.052
	(0.070)	(0.054)	(0.087)	(0.047)	(0.049)
African-American*Year*Treat	0.048	-0.090	0.079	-0.004	0.005
	(0.023)*	(0.022)***	(0.020)***	(0.024)	(0.018)
African-American*NCLB* Treat	-0.033	0.268	-0.203	-0.019	0.009
	(0.070)	(0.054)***	(0.087)**	(0.047)	(0.049)
Hispanic*Year	-0.022	-0.021	-0.055	-0.037	0.014
	(0.021)	(0.030)	(0.025)*	(0.025)	(0.022)
Hispanic*NCLB	0.014	-0.020	0.104	0.106	-0.104
	(0.098)	(0.124)	(0.073)	(0.092)	(0.049)*
Hispanic*Year*Treat	0.040	-0.013	0.023	0.059	-0.036
	(0.021)*	(0.030)	(0.024)	(0.025)**	(0.023)
Hispanic*NCLB* Treat	-0.061	0.118	-0.019	-0.170	0.104
	(0.098)	(0.124)	(0.073)	(0.093)*	(0.050)*
Asian*Year	0.010	0.018	-0.016	0.008	0.001
	(0.023)	(0.023)	(0.011)	(0.019)	(0.013)
Asian*NCLB	-0.072	-0.142	0.043	-0.034	-0.012
	(0.051)	(0.057)**	(0.053)	(0.052)	(0.069)
Asian*Year*Treat	0.001	-0.037	-0.017	-0.011	0.057
	(0.023)	(0.023)	(0.011)	(0.019)	(0.013)***
Asian*NCLB*Treat	0.009	0.323	0.042	0.113	-0.180
	(0.051)	(0.058)***	(0.053)	(0.052)*	(0.068)**
Constant	0.697	0.251	0.417	0.717	0.148
	(0.032)***	(0.022)***	(0.030)***	(0.025)***	(0.012)***
R^2	0.01	0.02	0.00	0.03	0.02
N	21,208	20,892	21,196	21,150	20,987

* p<0.1; ** p<0.05; *** p<0.01

Table 22: Adjusted R^2 from Regressions of 5th grade Tests and Academic Rating Scales on Non-cognitive Skills

Student Skills	**Standardized Tests**		**Teachers' Academic Rating Scale**	
Measured by	Reading	Math	Reading	Math
Teacher	.153	.143	.368	.23
Parent	.154	.148	.147	.120
Child	.163	.154	.138	.112

Source: Early Childhood Longitudinal Study of Kindergarten, 1998, 5th grade file

Table 23: The Impact of NCLB on Parents' Interactions with School (Teachers and other Parents)

	# of Parents Contacted	Informal Meetings	Initiate Contact w/Teacher	Attend Open House
Year	0.279	-0.012	0.052	0.029
	(0.067)***	(0.005)**	(0.005)***	(0.002)***
NCLB	-0.791	-0.009	-0.137	-0.082
	(0.192)***	(0.027)	(0.011)***	(0.008)***
Year*Treat	-0.040	-0.010	-0.006	-0.006
	(0.068)	(0.005)*	(0.005)	(0.002)**
NCLB*Treat	0.338	0.056	0.021	0.019
	(0.195)	(0.027)*	(0.011)*	(0.008)**
SES (Continuous)	-0.009	-0.003	0.023	0.018
	(0.065)	(0.007)	(0.017)	(0.005)***
Father Present	0.313	0.003	0.006	0.005
	(0.082)***	(0.017)	(0.009)	(0.010)
Father Employed Fulltime	-0.107	-0.019	0.003	0.008
	(0.106)	(0.016)	(0.011)	(0.008)
Constant	1.676	0.695	0.493	0.719
	(0.116)***	(0.008)***	(0.008)***	(0.006)***
R^2	0.01	0.01	0.02	0.01
N	33,470	29,804	33,656	33,619

* p<0.1; ** p<0.05; *** p<0.01, *Source: Early Childhood Longitudinal Study of Kindergarten, 1998*