



The influence of an NCLB accountability plan on the distribution of student test score gains

Matthew G. Springer*

*Department of Leadership, Policy, and Organizations, Peabody College of Vanderbilt University,
Peabody #43, 230 Appleton Place, Nashville, TN 37203, USA*

Received 6 October 2006; accepted 20 June 2007

Abstract

Previous research on the effect of accountability programs on the distribution of student test score gains is decidedly mixed. This study examines the issue by estimating an educational production function in which test score gains are a function of the incentives schools have to focus instruction on below-proficient students. NCLB's threat of sanctions are positively correlated with test score gains by below-proficient students in failing schools; greater than expected test score gains by below-proficient students do not occur at the expense of high-performing students in failing schools. This pattern of results tends to suggest that failing schools were able to benefit low-performing students in ways that were consistent with having operational slack, and that the threat of sanctions may stimulate greater productivity within failing schools. © 2007 Elsevier Ltd. All rights reserved.

JEL classification: I21; I22; J18

Keywords: School accountability; No child left behind; Distributional effects; Incentives

1. Introduction

No Child Left Behind Act of 2001 (NCLB) is the reauthorization of the nation's omnibus *Elementary and Secondary Education Act of 1965* (ESEA). The central outcome of NCLB is that all traditional public school students, and defined student subgroups thereof, reach academic proficiency by the 2013–2014 academic year. NCLB monitors progress toward meeting academic proficiency through Adequate Yearly Progress (AYP) calculations, a series of minimum competency performance targets defined by state education agencies that must be met

by schools and school districts to avoid sanctions of increasing severity.

Although there exists an increasing amount of scholarly evidence regarding educational and non-educational responses of schools to accountability pressures, studies have not examined the influence of NCLB accountability policy on the distribution of student test score gains.¹ Studies of pre-NCLB

¹An accountability system implemented by the United Kingdom government in 1988 was found to reduce the educational gains and exam performance of traditionally low-performing students (Burgess, Proper, Slater, & Wilson, 2005). Other studies have exploited a break in trend of Florida's accountability programs to compare differences in the effects of pressures associated with the A+ Program versus NCLB (West & Peterson, 2006), and have examined the impact of Florida's

*Tel.: +1 615 322 5524; fax: +1 615 322 6018.

E-mail address: matthew.g.springer@vanderbilt.edu

state accountability programs in Florida, North Carolina and Texas that analyze student-level longitudinal data reported the presence of achievement trade-offs whereby traditionally low-performing students demonstrated greater than expected gains to the detriment of high-performing students (Deere & Strayer, 2001; Figlio & Rouse, 2006; Holmes, 2003; Reback, 2006). On the other hand, individual state studies using subgroup achievement data from Florida and Tennessee concluded that increased proficiency rates of low-performing students did not come at the expense of high-performing peers' achievement (Ballou, Liu, & Rolle, 2006; Chakrabarti, 2006).

This study contributes to this slender body of research by using a longitudinal student-level data collection to estimate an educational production function in which test score gains are a function of the incentives schools have to focus instruction on below-proficient students. The presence of an achievement trade-off is a function both of: (1) whether a school failed to make AYP in the prior year; and (2) how much effort might need be expended on a particular student to reach proficiency, as defined and measured by whether a student is expected to fail a high-stakes test and by how far that student is from the state-defined performance threshold in a particular year. The primary data for this study are drawn from the Northwest Evaluation Association's Growth Research Database (GRD), a data system that contains high-stakes test score information for over 90% of students in the state under consideration. Findings indicate that NCLB's threat of sanctions are positively correlated with test score gains by low-performing students in failing schools, and that greater than expected test score gains by low-performing students did not occur at the expense of high-performing students in failing schools. This pattern of results tends to suggest that failing schools in the state were able to benefit low-performing students in ways that were consistent with having operational slack,² and that the threat

of sanctions may stimulate greater productivity within failing schools.

This study makes several contributions to understanding the influence of NCLB accountability programs on the distribution of student test score gains. Previous research on the topic mostly comes from pre-NCLB accountability programs. Although select pre-NCLB accountability programs were products of the Improving America's Schools Act of 1994, and these programs served in many respects as a model for NCLB, idiosyncratic features of state-defined accountability programs might differentially impact teacher and school behavior (Swanson & Stevenson, 2002). This study also benefits from the availability of fall and spring student test scores, eliminating the confounding influence of summer months present in prior research.

While this study offers the first statewide analysis of the influence of an NCLB accountability plan on the distribution of student test score gains, it is important to acknowledge some limitations. Demographic characteristics of the state under study are homogenous and unrepresentative of other states and their respective school systems. NCLB provides states considerable autonomy in the design of their respective accountability programs. Thus, in response to reform, interstate variations attributable to variation in the standards that each state requires for a student to be considered "proficient" can be expected. Distributional effects may also be different across time since NCLB's long-term goal is for all students and student subgroups to reach academic "proficiency" by the 2013-14 school year. Finally, data limitations preclude me from precisely predicting the impact of the state accountability system on the distribution of student test score gains in the absence of a counterfactual condition.

2. Model

A general linear model is used to analyze the influence of the state's accountability program on the distribution of student test score gains in failing schools. The test score gain measure and the determinants of student test score gains are described later.

2.1. Dependent variable

The dependent variable in this study is a student's fall-to-spring test score gain in mathematics. This state's testing regime is unique in that the

(footnote continued)

A+ accountability program on high-achieving college-bound students (Donovan, Figlio, & Rush, 2007). Springer (2007) examined whether schools intent on meeting minimum competency benchmarks practice education triage.

²Operational slack in this context refers to a school's ability to benefit some students without creating losses for others. If schools were at the production frontier, presumably these gains would not be possible.

state-designed accountability assessment is administered to public school students twice per year, allowing for construction of a fall-to-spring test score gain for each individual student in the 2002–2003, 2003–2004, and 2004–2005 school years. Using a fall-to-spring test score gain as the dependent variable is advantageous because past research using spring-to-spring test score gains are subject to the confounding influence of the summer months, meaning that a school's effect on any gain (or potential loss) in a student's test score cannot be disentangled easily from how much gain (or loss) occurred as a result of summer activities (Alexander, Entwisle, & Olson, 2001). In addition, a fall-to-spring test score gain as the dependent variable increases the sample size by approximately one-third of the total observations when compared to a spring-to-spring test score gain.

2.2. Indicator variables

A central issue in testing for the influence of accountability policy on the distribution of student test score gains is to take into consideration a school's short-run incentive to target instruction. Holmes (2003) was the first researcher to construct such an indicator, albeit an approach that restricted the estimated amount of targeting predicted at the school-level. Reback (2006) developed a more comprehensive technique that estimated the marginal effect of a hypothetical improvement in the expected performance of a particular student on the probability a school would obtain a certain accountability rating.

This study takes a different approach to capitalize on the availability of fall and spring test score data. Two indicator variables are created, $GAP \cdot PASS_{ij(t-1)}$ and $GAP \cdot FAIL_{ij(t-1)}$, that quantify the distance of a student's score from the state-defined passing threshold after adjusting for the expected fall-to-spring test score gain in a particular grade and year. Using both the $GAP \cdot PASS_{ij(t-1)}$ and $GAP \cdot FAIL_{ij(t-1)}$ variables is preferred over a single indicator that would impose a strong linear assumption on the model specification; that being, increased gains of a student who starts 5 points below the cut-off equals the amount by which the gain will diminish when a student starts 5 points above the cut-off. These measures were also modeled using a quadratic and cubic function and found not to be statistically significant.

To capture a school's incentive to target instruction, this study relies on a single binary indicator

denoting whether a school met the state's minimum proficiency standard. The measure replicates the state's AYP calculations by determining whether each public school, and all defined subgroups therein, met the state-prescribed proficiency standard in each of the 3 years represented in the panel. A student subgroup must have at least 34 students for that subgroup to be included in a school's AYP determination. Schools where the size of certain failing subgroups fluctuates around the minimum "n-size" of 34 will result in that subgroup's continued exemption from AYP designation an uncertainty, thus possibly preserving the school's incentive to elevate the performance of particular students. To account for this feature of NCLB policy, the AYP variable was also constructed using minimum subgroup "n" sizes of 32, 28, 24 and zero students. Each version of this variable yielded similar results in cross-wise comparisons.³

A limitation of this approach is the inability to predict precisely the influence of the state's accountability program on the distribution of student test score gains in the absence of a counterfactual condition. This western state implemented their accountability program in response to NCLB legislation, and the policy applies to all public schools in the state, there is no readily available comparison group with which to examine the relative performance of student test score gains. Consequently, the relative performance identified in this study may be the result of customary school behavior irrespective of NCLB's threat of sanctions should schools fail to make AYP. Generalizability of this study is limited by the fact that the state's education system is disproportionately white and rural and has much smaller than typical schools and districts.

2.3. Control variables

This study includes controls for school effects, peer effects, and testing effects. A school fixed

³This research also explored whether other programmatic features of NCLB, such as severity of sanction further explained student test score gains. No evidence was found of the severity of sanction impacting student test score gains. The least squares mean difference using the Tukey–Kramer adjustment for multiple comparisons between schools that failed to meet the state's proficiency standard for no years, 1 year, and 2 consecutive years was used to determine that the mean difference between schools failing to meet the proficiency standard for 1 and 2 years, respectively, was not statistically different.

effects estimator controls for inter-campus differences that are time invariant and likely correlated with student achievement gains. A grade by year interaction is used to control for testing effects. A vector of demographic variables is used to control for the effect of peer composition, including poverty status, as measured by eligibility for free or reduced price lunch, and school-level student background characteristics including percent of students by race/ethnicity.

2.4. Summary

The basic general linear model is specified as follows. Let $FSGAIN_{ijt} = \Delta Y_{ijt} = Y_{ijt} - Y_{ij(t-1)}$ be the math test-score gain for student i in school j from fall-to-spring administration of the high-stakes test, where t denotes spring administration of the test.⁴ Then

$$\begin{aligned} FSGAIN_{ijt} = & \alpha_0 + \alpha_1 FAIL \cdot AYP_{j(t-2)} + \alpha_2 HISPANIC_{ijt} + \alpha_3 WHITE_{ijt} \\ & + \alpha_4 GAP \cdot PASS_{ij(t-1)} + \alpha_5 GAP \cdot FAIL_{ij(t-1)} + \alpha_6 FAIL \cdot AYP_{j(t-2)} \times GAP \cdot PASS_{ij(t-1)} \\ & + \alpha_7 FAIL \cdot AYP_{j(t-2)} \times GAP \cdot FAIL_{ij(t-1)} + X_{jt} + \mu_j + \eta_g \gamma_t + e_{ijt}. \end{aligned} \quad (1)$$

X_{jt} is a vector of student characteristics, μ_j is a school fixed-effects estimator, η_g is a grade fixed-effects estimator, and γ_t is a year fixed-effects estimator. Other variable definitions are given in Table 1.

3. Data

The primary data for this study are drawn from the Northwest Evaluation Association's Growth Research Database (GRD), a data system that collects longitudinal student-level achievement results from approximately 2200 school districts in 45 states. Starting with the 2002–2003 school year, NWEA's GRD contains high-stakes test score information and a unique identifier for over 90% of this state's students in mathematics. The high-stakes test is scored on a single cross-grade and equal-interval scale ranging from 150 to 300 using a one parameter Rasch model (Kingsbury, 2003).

⁴The school failure variable has a $t-2$ subscript and the prior achievement variable has a $t-1$ subscript since points in time are based on fall and spring observations where t is the current spring observation.

Eligible cases were restricted to students enrolled in public schools in grades three through eight. Ten reconfigured schools were deleted from the sample since the state's accountability program does not hold reconfigured schools accountable until three full years of achievement data is collected. Students without both fall and spring test scores in a given school year were excluded from analysis.⁵ In total, these restrictions eliminated less than 17,500 student observations, or an average of 5.5% of student observations per year. Results were qualitatively similar when analyses were run without these restrictions.

Table 1 includes select sample descriptive statistics for schools that made AYP, schools that failed to meet AYP, and all schools in the sample. There are modest demographic and test score gain differences among failing and non-failing schools. Differences are unlikely to bias findings

since the school fixed-effects estimator compares schools to themselves over time and all model specifications include school-level controls for race/ethnicity and free and reduced priced lunch status.⁶

4. Results

Table 2 displays results from models that estimate differences in student test score gains and losses across the achievement distribution while allowing for different outcomes when a student is expected to pass or fail the spring high-stakes test and/or whether a school met the state-defined AYP

⁵This restriction was also placed on data in light of the fact that state administrative code requires that only students who are continuously enrolled in the same public school from the end of the first 8 weeks, or 56 calendar days, of the school year through the spring test administration period be included in AYP calculations. While coding procedures employed are not based on student attendance patterns as defined in Section 112.03, this restriction is believed to produce a closer approximation to actual AYP calculations.

⁶To check data reliability following cleaning and development procedures, resultant student achievement descriptive statistics were compared to reported values published in a biannual series of statewide results brochures maintained by the state board of education. Estimates were virtually identical.

Table 1
Variables and descriptive statistics

Variable	Description	School made AYP	School failed AYP	All schools
Dependent variable				
FSGAIN	Math test score gain from fall to spring test administration	9.25 (7.57)	8.28 (7.38)	9.01 (7.53)
Indicator variables				
GAP · PASS	Distance of a passing student's score from the state-defined passing threshold after adjusting for the average fall-to-spring test score gain in a particular year and grade	12.88 (8.25)	12.75 (8.49)	12.84 (8.31)
GAP · FAIL	Distance of a failing student's score from the state-defined passing threshold after adjusting for the average fall-to-spring test score gain in a particular year and grade	8.76 (8.02)	9.72 (8.19)	9.46 (8.66)
Student characteristics				
AI.AN	American Indian/Alaska Native students as a proportion of the total student population	0.01 (0.12)	0.02 (0.14)	0.02 (0.13)
A.PI	Asian/Pacific Islander students as a proportion of the total student population	0.02 (0.12)	0.01 (0.12)	0.02 (0.12)
BLACK	African American students as a proportion of the total student population	0.01 (0.10)	0.01 (0.08)	0.01 (0.09)
HISPANIC	Hispanic students as a proportion of the total student population	0.10 (0.30)	0.18 (0.39)	0.12 (0.32)
WHITE	White, Non-Hispanic students as a proportion of the total student population	0.86 (0.35)	0.77 (0.42)	0.84 (0.37)
OTHER	Students categorized as "other" as a proportion of the total student population	0.004 (0.06)	0.002 (0.04)	0.003 (0.06)
FRPL	Free and reduced price lunch status students as a proportion of the total student population	0.39 (0.49)	0.48 (0.50)	0.41 (0.49)

Standard deviation in parentheses.

standard under NCLB. Models 2.1–2.3 report a positive, significant association between AYP status ($FAIL \cdot AYP_{j(t-2)}$) and a student's fall-to-spring test score gain. Estimates suggest that the average student enrolled in a failing school gained, on average, 0.24 points more than the average student enrolled in a non-failing school. Although the practical significance of this gain is negligible, the positive, statistically significant failing coefficient on $FAIL \cdot AYP_{j(t-2)} \times GAP \cdot FAIL_{ij(t-1)}$ indicates that failing students in failing schools have larger fall-to-spring test score gains relative to failing students in non-failing schools. Such results may imply that failing schools have concentrated more efforts on below proficient students.

In substantive terms, model 2.2 predicts that the average failing, student enrolled in a failing school gains 3.93 points more than the average passing, student also enrolled in a failing school. Moreover, when comparing the predicted fall-to-spring test score gain value for the average failing student enrolled in a failing school to that of the average failing student enrolled in a non-failing school, the

former scores 1.22 points greater, or the equivalent of 0.15 standard deviations in a single year.

An important question still remains; did test score gains realized by failing students in failing schools come at the expense of test score gains by high-performing students? Evidence indicates that the state's accountability system appears to be helping low-performing students in failing schools and not at the expense of high-performing student test score gains. There is a positive, significant coefficient on $FAIL \cdot AYP_{j(t-2)} \times GAP \cdot PASS_{ij(t-1)}$ and a student's fall-to-spring test score gain. The average student expected to pass the spring state assessment who attended a school that failed to make AYP the year prior will have a larger fall-to-spring test score gain than a student that had the same $GAP \cdot PASS_{ij(t-1)}$ score but was enrolled in a non-failing school.⁷ However, the predicted test score

⁷As displayed in Table 1, estimates are robust to model specifications that include a school by grade fixed effects or a school by grade by year fixed effects, respectively, to test for the interdependence of the errors within the same grade at the same

Table 2
Influence of distance from state accountability program performance threshold on student test score gains

Regression type	General linear model					
	Student math gain (fall-to-spring)		Standardized student math gain (fall-to-spring)		School failed to make AYP (FAIL · AYP), distance from performance threshold (GAP · FAIL and GAP · PASS)	
Dependent variable	School grade × year (2.1)	School × grade Grade × year (2.2)	School × grade × year (2.3)	School Grade × year (2.4)	School × grade Grade × year (2.5)	School × grade × year (2.6)
Indicator variables	School failed to make AYP (FAIL · AYP), distance from performance threshold (GAP · FAIL and GAP · PASS)					
Fixed effects (model)	School grade × year (2.1)	School × grade Grade × year (2.2)	School × grade × year (2.3)	School Grade × year (2.4)	School × grade Grade × year (2.5)	School × grade × year (2.6)
FAIL · AYP (α_1)	0.2543 (0.0592) ^{***}	0.2378 (0.0590) ^{***}		0.0564 (0.0089) ^{***}	0.0534 (0.0090) ^{***}	
HISPANIC (α_2)	-0.7405 (0.0679) ^{***}	0.7505 (0.0673) ^{***}	-0.7373 (0.0668) ^{***}	-0.1087 (0.0103) ^{***}	-0.1099 (0.0102) ^{***}	-0.1083 (0.0101) ^{***}
WHITE (α_3)	0.3804 (0.0587) ^{***}	0.3769 (0.0582) ^{***}	0.3818 (0.0577) ^{***}	0.0558 (0.0089) ^{***}	0.0557 (0.0088) ^{***}	0.0560 (0.0088) ^{***}
GAP · FAIL (α_4)	0.1872 (0.0002) ^{***}	0.1865 (0.0026) ^{***}	0.1859 (0.0026) ^{***}	0.0034 (0.0004) ^{***}	0.0035 (0.0004) ^{***}	0.0036 (0.0004) ^{***}
GAP · PASS (α_5)	-0.1141 (0.0016) ^{***}	-0.1144 (0.0016) ^{***}	-0.1123 (0.0016) ^{***}	-0.0024 (0.0002) ^{***}	-0.0024 (0.0002) ^{***}	-0.0021 (0.0002) ^{***}
FAIL · AYP × GAP · FAIL (α_6)	0.0811 (0.0047) ^{***}	0.0825 (0.0047) ^{***}	0.0846 (0.0047) ^{***}	0.0016 (0.0007) ^{***}	0.0018 (0.0007) ^{***}	0.0021 (0.0007) ^{***}
FAIL · AYP × GAP · PASS (α_7)	0.0107 (0.0047) ^{***}	0.0116 (0.0032) ^{***}	0.0092 (0.0032) ^{***}	0.0014 (0.0005) ^{***}	0.0016 (0.0005) ^{***}	0.0012 (0.0005) ^{***}
R^2	0.1994	0.2187	0.2451	0.0410	0.0644	0.0960
Observations	307758	307758	307758	307758	307758	307758

***, ***, Estimate statistically significant from zero at the 10%, 5%, and 1% levels, respectively. School-level controls for race/ethnicity and free and reduced price lunch status included. Estimates are robust when suspicious values are removed; suspicious defined by values 1.5 IQR above Q3 and 1.5 IQR below Q1.

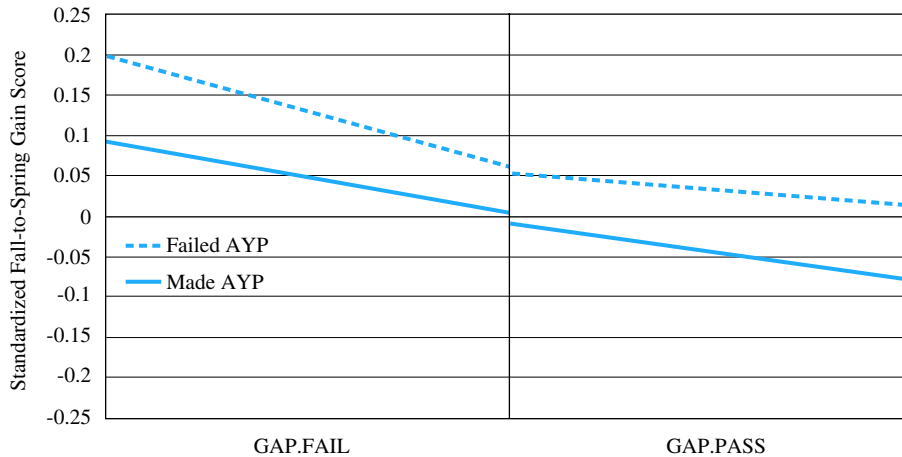


Fig. 1. Influence of accountability policy on the distribution of student test score gains.

gains of passing students are below average and decrease with high $GAP.PASS_{ij(t-1)}$ values.

Although models 2.1–2.3 give the appearance of elevated learning opportunities for traditionally low-performing students are coming at the expense of traditionally high-performing peers, Chay, McEwan, and Urquiola (2005) warn that mean reversion can lead conventional estimation approaches to overstate the impact of policy when using test scores as a left-hand side variable. To reduce the potential for mean reverting bias, the state's initial distribution of students' fall assessment scores is divided into 20 equal intervals for each year and grade combination, and the mean and standard deviation score gain computed for all students starting in a particular interval for each of those combinations. A student's test score gain is standardized by taking the difference between that student's nominal gain and the mean gain of all students in the interval over the standard deviation of all student gains in the interval. Gains in each interval are distributed with a mean of zero and standard deviation of one.

Results are displayed in the right panel of Table 2. Models yield qualitatively similar results to those parameters reported above; both the signs on the coefficients of interest and the attendant significance levels remain stable. There is a notable change in the

(footnote continued)

school. School by grade fixed effects essentially treat each grade within a school as a separate school, thus taking into consideration any observables common to all students in the same grade within a school. School by grade by year fixed effects control for similarity of outcomes among students in the same grade within a school by removing all the variation in cases over time.

predicted fall-to-spring test score gains of passing students in failing schools. Passing student test score gains are no longer below expectation. While this modest change in the right hand tail of the distribution is consistent with mean reversion, Fig. 1 illustrates that the effects at the other end of the distribution persist particularly for failing schools. No longer do gains of failing students in failing schools appear to be occurring at the expense of non-failing students.

5. Summary and conclusion

This study provides evidence that failing students in failing schools have greater than expected test score gains. The further a failing student is from this state's performance threshold, the greater their predicted fall-to-spring test score gain relative to that of similarly situated non-failing students. Yet, greater than expected test score gains by low-performing students do not appear to occur at the expense of high-performing students in failing schools.

These findings are robust across model specifications and when the dependent variable was standardized to gauge mean reverting measurement error. This pattern of results tends to suggest that failing schools in the state were able to benefit low-performing students in ways that were consistent with having operational slack, and that the threat of sanctions may stimulate greater productivity within failing schools.

Means by which schools become more productive, and why similar gains for traditionally low-performing students were not realized in failing schools prior to NCLB sanction, remain salient

questions and directions for future inquiry. Researchers need to continue to monitor potential achievement tradeoffs within the context of NCLB state accountability programs to better understand if the law's minimum competency standard produces perverse incentives and requires modification, perhaps to reward improvements across the entire achievement distribution. Indeed, responses by schools with historically low-performance relative to the average performance level in that state, and/or schools governed by a high-stakes accountability program with particularly rigorous standards, might differentially influence school behavior when compared to a school with above average performance in a state that has relatively weak standards. Finally, it is important to note that the finding that high-scoring students do not appear to suffer under NCLB has a particular meaning in this context because, as was discussed earlier, this study cannot know how these students would have fared in the complete absence of NCLB.

Acknowledgments

The author is grateful to Dale Ballou for helpful comments and insight in developing this work. He also wishes to thank James W. Guthrie, Warren Langevin, Derek Neal, Michael J. Podgursky, Randall Reback, Diane Schanzenbach, Jeffrey Springer, Martin West, and Kenneth Wong for providing useful feedback on this project. The author would also like to acknowledge the North-west Evaluation Association for the data used in this analysis. Any errors remain the sole responsibility of the author.

References

- Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis*, 23(2), 171–191.
- Ballou, D., Liu, K., & Rolle, E. (2006). *Response to no child left behind among tennessee schools*. Unpublished working paper. Peabody College of Vanderbilt University.
- Burgess, S., Propper, C., Slater, H., & Wilson, D. (2005). *Who wins and who loses from school accountability? The distribution of educational gain in English secondary schools*. Working paper no. 05/128. Bristol, England: Centre for Market and Public Organisation.
- Chakrabarti, R. (2006). *Do public schools facing vouchers behave strategically? Evidence from Florida*. Nashville, TN: National Research and Development Center on School Choice.
- Chay, K., McEwan, P. J., & Urquiola, M. (2005). The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review*, 95(4), 1237–1258.
- Deere, D., & Strayer, W. (2001). *Putting schools to the test: School accountability, incentives, and behavior*. Unpublished manuscript. Texas A&M University.
- Donovan, C., Figlio, D. N., & Rush, M. (2007). *Cramming: The effects of school accountability on college-bound students*. Working paper 12628. National Bureau for Economic Research.
- Figlio, D. N., & Rouse, C. E. (2006). Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics*, 89(2–3), 381–394.
- Holmes, G. M. (2003). *On teacher incentives and student achievement*. Unpublished working paper. East Carolina University, Department of Economics.
- Kingsbury, G. G. (2003). *A long-term study of the stability of item parameter estimates*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Reback, R. (2006). *Teaching to the rating: School accountability and the distribution of student achievement*. Unpublished manuscript. Barnard College, Columbia University.
- Springer, M. G. (2007). Accountability incentives: Do failing schools practice educational triage? *Education Next*, 8(1), 74–79.
- Swanson, C., & Stevenson, D. L. (2002). Standards-based reform in practice: Evidence on state policy and classroom instruction from the NAEP state assessments. *Educational Evaluation and Policy Analysis*, 24(1), 1–27.
- West, M. R., & Peterson, P. E. (2006). The efficacy of choice threats within accountability systems: Results from legislatively induced experiments. *The Economic Journal*, 116(510), c46–c62.