Economics of Education Review 44 (2014) 1–16

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Accountability incentives and academic achievement: Distributional impacts of accountability when standards are set low

J. T. Richardson*

University of California, Davis, 1 Shields Avenue, Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 1 November 2013
Revised 16 October 2014
Accepted 19 October 2014
Available online 28 October 2014

JEL classification:
I21
I24
I28

Keywords:
Educational economics
Resource allocation

ABSTRACT

This paper examines the effects of a compositional shift in a school's testing population brought about by the elimination of special education testing exemptions. The policy change forced schools to add varying levels of generally low-achieving students to their testing pools, altering accountability incentives. I provide evidence that the elimination of exemptions caused significant test score increases for initially low-achieving students and narrowed the black-white test gap. I show that the measured effects were not caused by changes in classroom composition. Rather, benefits flowed to low-achieving students because Texas' accountability standard was low relative to the skills of its students.

1. Introduction

Public school accountability systems are one of the most important education policy interventions to occur in the last 20 years. In one common form of accountability, a school's rating is determined by the fraction of its students who achieve passing scores on standardized exams. The primary goal of these policies is to ensure that student achievement at each school meets specific, minimum standards.[1] Schools that routinely fail to meet their states' accountability standards can face a series of punishments, including decreased funding, increased state monitoring, or even state takeover of the school's operation. Therefore, school administrators and teachers have strong incentives to maximize their school's fraction passing.

There is substantial evidence that student performance on state standardized exams has increased as a result of accountability.[2] However, researchers have also shown that school accountability systems create incentives for strategic behavior that undercut these policies' intended effects.[3] Therefore, it is important for policy makers to understand how to best design accountability systems to deliver efficient and equitable results for students across the ability distribution.

One strand of literature examines how accountability-driven achievement gains differ across the initial

* Present address: Value-Added Research Center, University of Wisconsin, 1025 West Johnson Street Madison, WI 53706, USA. Tel.: +1 6124189583.
E-mail address: jtrichardson@wisc.edu, jtrichardson@edanalytics.org

[1] Carnoy and Loeb (2002).

[2] For evidence of test score increases on state exams, see Dee and Jacob (2011), Figlio and Rouse (2006), Hanushek and Raymond (2005), and Carnoy and Loeb (2002).

[3] Jacob (2007) compares gains on Texas state exams to gains on the national, low-stakes National Assessment for Educational Progress (NAEP). He finds that while discrepancies in fourth grade gains can be explained by skill and format differences across tests, divergence in eighth grade state and NAEP scores cannot. Jacob (2005) concludes that math and reading gains in Chicago were absent from a contemporaneous, low-stakes exam, and thus were attributable to student effort and test-specific skills. Chiang (2009) finds that in Florida, 70 percent of math gains persist at least two years, a result he attributes to "real" learning gains.

http://dx.doi.org/10.1016/j.econedurev.2014.10.006
0272-7757/

achievement distribution. Many have hypothesized that proficiency-based accountability systems, those in which schools are graded according to the fraction of students who attain some minimum-proficiency level of achievement, motivate schools to concentrate on students whose previous test scores place them closest to the passing margin. There is ample empirical evidence in support of this hypothesis. Using discrete cutoffs in the formulae for both schools' accountability ratings and whether a student attains a passing score, Reback (2008) establishes that achievement in Texas increased most for students whose test scores had the greatest probability of affecting their schools' accountability ratings. Neal and Schanzenbach (2010) demonstrate that higher achievement standards may have harmed low-achieving students in Chicago. They compare two cohorts in Chicago, one that tested in third and fifth grades before Illinois' accountability began, and one cohort that tested in third grade pre-accountability but took its fifth grade exams after accountability was introduced. Their results show the greatest gains for students in the middle of the initial achievement distribution, those closest to the passing margin, with mixed evidence of smaller gains or even losses for students in the tails.[4] They argue that because Illinois had chosen a high standard (approximately half of the nation's fifth graders would be expected to attain a passing score in Illinois), schools expended relatively few resources trying to increase scores for students at the very bottom of the achievement distribution who had little probability of passing even with additional intervention.

My aim is to further study the incentives associated with accountability testing using an exogenous change in the ability composition of a school's testing population while holding exogenous peer effects constant.[5] I do so using Texas' elimination of testing exemptions for special education students in the state's pre-NCLB accountability system. The elimination of exemptions created a natural experiment that allows me to test how the ability composition of a single school-cohort influences the incentive effects of proficiency-based accountability. When Texas eliminated testing exemptions, it caused a varying number of primarily low-achieving students to be added to each school's accountability testing population. This meant that for schools to maintain their accountability ratings from the prior year, they needed to increase their numbers passing. I take advantage of this plausibly exogenous increase in accountability pressure to illustrate the distributional impacts of Texas' accountability system. I argue that these distributional impacts are related to Texas' low student passing standards, but those standards remain unchanged throughout the sample period.

This unique research design offers several advantages. First, it is necessary for studying Texas' accountability system. Most of the literature rightly exploits the inception of accountability to estimate treatment effects. However, Texas only sporadically tested students in the lead-up to its accountability policy, making this type of analysis difficult. Taking advantage of exogenous changes in the testing population also enables me to study the distributional effects of Texas' particular policies, which, as the precursor to NCLB, have far-reaching policy implications.[6] Furthermore, identifying effects downstream from the program's inception allows me to study the effects of accountability after the initial shock of introduction. This more accurately reflects the current policy situation; every state already has an accountability program in place. I am also able to study how accountability incentives affect traditional racial and socioeconomic achievement gaps because of the extensive demographic variables available in the Texas data. Finally, unlike the Chicago system used in Neal and Schanzenbach, Texas had a low proficiency standard relative to the skills of its testing population, allowing me to test whether a low standard has different effects on the distribution of achievement gains given a plausibly exogenous increase in accountability pressure.

Using panel data of student characteristics and achievement assembled by the Texas Schools Project, I am able to construct a measure of the fraction of each school-cohort's students receiving special education exemptions in third grade. After the 1999–2000 academic year, schools were no longer permitted to grant exemptions based on special education status. I provide evidence that this did not qualitatively change which students were present in the classroom; special education students were no more prevalent in mainstream classes after the policy change than they were before. Instead, what changed was that students could no longer be omitted from a school's accountability calculation. The school's (and its teachers') response to these changes is what I am interested in investigating. Students who received exemptions would have been affected by both the policy's direct effect, the fact that they became eligible to test, and the indirect effect of the teacher and school response. To that end, I limit my analysis to those students who were never exempt in order to isolate the teacher and school response to the change in the testing population.

In order to credibly estimate the school response, I employ a difference-in-differences framework that controls for school-by-grade and grade-by-year fixed effects as well as individual- and cohort-specific demographic controls. Consequently, my estimation strategy identifies the school response to accountability incentives through time-varying differences across cohorts within a specific school. I test for selection using a number of falsification tests showing that school-grade demographics and the probability of missing a test are not meaningfully related to a cohort's fraction of exemptions. I also find that my results are robust to school time trends and provide evidence that common shocks do not explain my results.

I find that the elimination of special education exemptions in Texas had little impact on overall achievement but caused statistically and economically significant increases in test gains for initially low-achieving students relative to their higher-scoring peers. As such, the policy change acted

[4] Donovan, Figlio, and Rush (2006) and Burgess, Propper, Slater, and Wilson (2005) find adverse effects on initially high- and low-achieving students, respectively. Booher-Jennings (2005) and White and Rosenbaum (2007) present case studies describing how individual schools react to accountability pressure by focusing resources on students whose passing could be most influenced.

[5] For a discussion of the negative effects of adding low-achieving students to classrooms, see Carrell and Hoekstra (2010), Lavy, Paserman, and Schlosser (2012), and Figlio (2007).

[6] McNeil, Coppola, Radigan, and Heilig (2008).