Accountability Incentives and Special Education

Sarah Prenovitz[1]

October 2017

Latest version available at www.sarahprenovitz.com

Abstract

Approximately 13 percent of students in US public schools receive special education services for a wide array of disabilities. Despite the size of the program, little is known about how special education affects students. Placement decisions are in theory based solely on students' needs, but prior literature suggests that schools alter their special education populations in response to other factors. Recent accountability policies put in place since the enactment of No Child Left Behind (NCLB) in 2002 have presented schools with a new set of incentives to alter the special education population. This paper is the first to investigate these incentives, which are similar to those under current accountability programs. I use administrative data from the universe of North Carolina Public Schools and a difference-in-difference framework in which incentives are determined by the interactions between schools' expectations about subgroup performance on the one hand and student performance and subgroup membership on the other. I find that schools responded to incentives to change the composition of the SWD subgroup to be higher-performing. Schools also used special education placement to target resources to students who were close to the passing threshold in reading, but not in math. I then use variation in incentives across schools and students as instruments to examine the effect of special education placement on achievement. For students whose special education placement was affected by incentives to select the SWD group to be relatively high-performing in math, special education decreased math scores. This suggests that special education decreases the achievement of some students.

Keywords: Accountability, Special Education

JEL codes: I24, I26, J14

---

[1] Department of Economics, Cornell University. Email: sjp268@cornell.edu. I am grateful to Ronald Ehrenberg, Michael Lovenheim, and Nicolas Ziebarth for their guidance and support. Thanks also go Maria Fitzpatrick, Douglas Miller, and to participants in the Cornell Institute for Health Behaviors and Disparities workshop, the Daniel Patrick Moynihan Syracuse/Cornell Summer Workshop in Education and Social Policy, and the Association for Education Finance and Policy Annual Conference for their comments. This work was supported by a Pre-Doctoral Fellowship in Disability Research from SSA via NBER and a Disability Research Consortium Fellowship from SSA via Mathematica Policy Research.

1. Introduction

About 13 percent of students in US public schools receive some form of special education, with specific learning disabilities and speech or language impairments the most common conditions (U.S. Department of Education, National Center for Education Statistics 2016). In 2012-2013, about 19 percent of public school spending was on students who received special education, amounting to approximately $118 billion (Federal Education Budget Project 2014, U.S. Department of Education, National Center for Education Statistics 2013). However, the effect of special education on student outcomes is not well understood. In theory, special education could either improve or hurt student achievement. Students who receive special education can benefit from extra attention and individualization, better understanding their own learning needs, or accommodations such as extra time for testing. On the other hand, they may suffer from stigma, respond to low expectations with low effort, or miss out on opportunities available to their peers in regular education (Bear, Clever, & Proctor, 1991, Lackaye & Margalit 2006).

Public schools are legally required to provide a free and appropriate public education to students with disabilities. Decisions about whether a student is disabled and what services are needed are made through a complicated interaction of many stakeholders – teachers, administrators, school-based specialists, doctors, parents, lawyers, and sometimes the students themselves. Although students' eligibility for special education is based entirely on their impairments and educational needs, prior work has shown that the size and composition of the special education population responds to school incentives. These incentives have included those created by funding formulae and by early accountability policies (Cullen 2003, Kwak, 2010, Cullen and Reback 2006, Figlio and Getzler 2006, Cohen 2007, Bokhari and Schneider 2011, Chakrabarti 2013, Jacob 2005, Mahitivanichcha and Parrish 2005, Winters and Greene 2011, Hanusheck and Raymond 2005, Morrill, 2016).

Under early accountability programs students enrolled in special education were not included in the accountability population and were in many cases exempt from testing. This presented schools facing accountability pressure with an incentive to encourage low-performing students to enter special education. No Child Left Behind, a federal law enacted in 2002, required all states to introduce accountability programs that held schools responsible for the performance of all students as well as student subgroups defined based on demographics. One of these subgroups was the group of students with disabilities (SWD). A school made adequate yearly progress (AYP) only if all of its subgroups, including students with disabilities, met targets for participation, proficiency, and either attendance or graduation rate.[2]

These new accountability programs presented schools with at least two incentives to alter the assignment of students to special education. First, schools could use special education placement to target services, individualization, and testing accommodations to “bubble” students expected to be near the passing threshold, particularly those in subgroups that were expected to fail to make AYP. Second, schools may have tried to select their special education population to be higher-performing in order to make it more likely that the SWD subgroup made AYP. In this scenario, schools that had previously failed to make AYP for the SWD group would be more likely to assign a student to special education if they expected that student to achieve a passing rather than a failing score, especially if the SWD group was close to the AYP threshold. It is unknown to what extent schools respond to accountability policies put into place since the enactment of No Child Left Behind (NCLB), which were crafted in part to eliminate the incentives in early accountability programs.

I use student-level administrative data from all public school students in grades 4-8 in North Carolina from 2007-2011 to examine how schools responded to accountability incentives under NCLB to

---

[2] Beginning in the 2011-2012 school year states were granted waivers from major requirements of NCLB, and it was replaced in 2015 by the Every Student Succeeds Act (ESSA). However, the accountability systems currently in place maintain the features of NCLB that underpinned its incentives for schools to alter the special education population. They judge performance at least in part based on the percentage of students who pass a given cut-score, include nearly every student in testing and accountability, and count the special education population as a separate subgroup that must meet standards.

classify particular students into special education. I do so by looking for evidence that schools use special education to target services to "bubble" students near the passing threshold when the school's AYP performance could be improved by their passing, or that schools select the special education population to be relatively high-performing when the school expects that the SWD subgroup will fail to make AYP. I use a difference-in-difference framework in which incentives are determined by the interactions between schools' expectations about subgroup performance and students' performance and subgroup membership. My analysis is focused on students with relatively malleable diagnoses, such as learning disabilities, speech and language impairments, and ADHD. I find evidence that schools used special education to target supports and services to bubble students in reading when the school would benefit from that student passing, with students who previously scored in the achievement level just below passing about 1 percentage point more likely to be in special education.[3] Students just below the passing threshold in math are less likely to be in special education, while those just above the passing threshold in math are unaffected. This pattern is probably shaped by the fact that math scores are generally easier to alter through instruction than are reading scores, as well as the fact that all schools had incentives to increase their overall passing rate. Also, schools face funding incentives to limit the overall size of the special education population, so may discourage some students from receiving special education in order to make room for others.

Responses to the second incentive are clearer. Compared with students who had previously failed, schools were relatively more likely to place previously passing students in special education when they were trying to improve the likelihood that the students with disabilities group achieved AYP, particularly in reading. A student who had previously passed their reading test would be about 2 percentage points more likely to be in special education, relative to a prior-failing student, if their school

[3] I consider schools able to benefit from a student passing if that student was a member of a subgroup that had previously failed to reach the AYP passing threshold.

faced the highest incentive to select the SWD group based on reading versus the lowest incentive. This finding is fairly robust across specifications, but the pattern in math is weaker and not consistently significant. Responses to the second incentive appear to be driven primarily by fewer prior-failing students being in special education when their school faces accountability pressure, rather than more prior-passing students.

I then use differences in the second set of incentives across schools and students as a source of plausibly-exogenous variation in the likelihood a student was assigned to special education to estimate the effect of special education on test scores in an instrumental variables framework. Special education decreased math scores by more than a standard deviation for students whose placements were driven by accountability incentives in math. Estimates for the reading scores for these students are similar in magnitude but not significantly different from zero, while those for students whose placements were altered by incentives to improve the performance of the SWD subgroup in reading are not statistically significant. I investigate several mechanisms and fail to find evidence of changes to grade retention or school switching, but do find changes in student effort as measured by attendance. This suggests that the lower engagement among special education students that has been documented previously is at least partially caused by special education placement, and has consequences for achievement.

This paper makes two main contributions to the literature. First, I extend our knowledge of how schools alter the assignment of students to special education in response to incentives. I do so by offering the first estimates of school responses to the special education incentives presented by NCLB, which are similar to those in current accountability policies. Special education placement should depend only on a student's impairments and needs, so any response to these incentives is important to understand. Prior work on this topic has focused on pre-NCLB policies in which schools had a straightforward incentive to place low-performing students into special education (Jacob 2005, Cullen and Reback 2006, Figlio and Getzler 2006, Cohen 2007). No Child Left Behind and current accountability policies were designed in part to eliminate this incentive, so the opportunities for schools to strategically change special education placement are more nuanced and targeted different groups of students.

Second, my estimates of the effect of special education placement on marginal students adds to a very sparse literature on the subject that has relied on strong assumptions for identification. Hanushek, Kain, and Rivkin (2002) examine students who move in and out of special education programs using student fixed effects and find that special education improves math scores. Their identification rests on two assumptions that I can relax: that any omitted variables that are correlated with both achievement and placement are static over time, and that changes in achievement do not cause changes in placement.[4] The paper most similar to this analysis is an unpublished working paper by Cohen (2007), who constructed instruments for special education placement based on Chicago's accountability program in the 1990s.[5] Her results are too noisy to draw conclusions about what effects, if any, placement has on student achievement. Cohen's analysis also rests on the assumption that schools that faced pressure to increase the percentage of students who performed at grade level did not undertake other measures, aside from encouraging special education placement, that would have improved the performance of low-achieving students.[6]

My findings suggest that near-universal testing requirements and an emphasis on the performance of malleable subgroups eliminates one set of incentives – to put low-achieving students in special education – but creates another – to target resources to bubble students and select the SWD group to be high-performing. This is a useful lesson for accountability design, particularly as these

[4] They are also missing data on the substantial portion of special education students who did not take standardized tests during this period. This could potentially induce selection bias, in either direction, depending on which students did not take tests.
[5] Chicago's accountability policy placed elementary schools on probation if less than 15 percent of students performed at grade level, defined as scoring least at the 50th percentile on the Iowa Test of Basic Skills (ITBS).
[6] A similar assumption would be needed in order to use the first incentive I consider – that to target resources to "bubble" students when the school would benefit from their passing – as an instrument for special education placement. Because it seems unlikely that this exclusion restriction would hold, my IV analyses use only the second incentive – that to select the SWD group to be higher performing.

features have continued beyond the end of NCLB. I also find evidence that special education can harm achievement for some of the marginal students whose placement is altered in response to accountability pressure. My results suggest that it is important to ensure that special education is appropriately targeted. More research is needed to understand the mechanisms the underlie heterogeneity in the effect of being placed in special education.

The rest of the paper proceeds as follows. I present background information on special education, accountability, and other relevant policies in Section 2. In Section 3 I summarize prior research. Section 4 describes the data and sample. Next, in Section 5, I discuss the method used to estimate school responses to AYP incentives and the results of this analysis presented. In Section 6, I present the method used to estimate the effect of special education on achievement, the results of the analysis, and an investigation into potential mechanisms. Section 7 concludes.

2. Policy Background

In this section I discuss several policies and institutions. These include special education, No Child Left Behind, and North Carolina's state accountability policy. I then describe the resulting incentives to alter the special education population.

2.1 Special Education

Under the Individuals with Disabilities Education Act (IDEA) public schools must provide a free and appropriate public education (FAPE), delivered in the least restrictive setting possible, to students who are diagnosed with one of 13 categories of disability (e.g. specific learning disability, autism, visual impairment) that impedes their ability to learn or participate in other age-appropriate activities.[7] The nature and extent of services vary widely depending on the student's needs – one student might receive weekly speech therapy, while another attends regular classes accompanied by a 1-on-1 aide, and a third

[7] Some students who do not have one of the 13 impairments listed under the IDEA receive accommodations under Section 504 of the Rehabilitation Act of 1973, which covers a broader set of conditions with a looser legal framework. I focus on students who are covered by the IDEA.

student spends his time in a separate program. Special education students are lower-performing on average than their regular education peers, but there is substantial overlap between the two groups in terms of achievement, as shown in Figure 1. In the first panel, the distribution of math scores for students in special education is drawn with a bold solid line, and that for students not in special education with a thin dashed line. The x-axis is in standard deviation units relative to the average scores across all students.[8] The second panel of Figure 1 displays a similar pattern using reading scores.

Before a student is placed into special education someone – often a parent or teacher – notices that the student is struggling and requests a disability assessment. The school then conducts a disability evaluation, which assesses the students' abilities and needs across multiple dimensions. A group of stakeholders meets to establish an individualized education program (IEP), which details the services and supports the student will receive, as well as the setting in which they will be provided. These meetings include parents, teachers, administrators, specialists, and sometimes lawyers or the students themselves. Diagnoses are reviewed at least every 3 years, and IEPs every year.

Most states provide funding for special education to local education agencies (LEAs, essentially school districts) either based on the number of students enrolled or the number of special education students. North Carolina is alone in providing special education funding calculated as a set dollar amount multiplied by either the number of students with IEPs or 12.5 percent of LEA membership, whichever is smaller (Morrill, 2016). This unusual funding structure makes North Carolina a uniquely useful setting, as it can offer insights into school behavior under both common funding mechanisms. Schools in states that provide funding based on the total number of students face funding incentives similar to those faced by schools in North Carolina LEAs with more than 12.5 percent of their students in special education. Schools in states that provide funding based on the number of special education

[8] These scores reflect the normalization described in detail in section 4. Dropping scores from alternate tests, rather than normalizing them to be comparable with regular tests, results in very similar patterns.

students are in situations more similar to those of schools in North Carolina LEAs below the 12.5 percent funding cap.

2.2 No Child Left Behind

No Child Left Behind, a federal law enacted in 2002, required states to establish testing programs and evaluate schools based on students' math and reading performance. States were given some leeway in determining implementation details. This paper focuses on NCLB as implemented in North Carolina, primarily from 2006-2007 – 2010-2011, so I concentrate here on characterizing that version of the policy. This time restriction allows me to work with a relatively consistent set of policies and offers advantages in data availability, described in Section 4.

Under NCLB, each school receiving Title I funding was accountable for the performance and participation of the overall population of students as well as several subgroups defined by race/ethnicity, income, and disability status. About half of US public schools receive Title I funding, which is available to schools and districts with relatively high poverty as measured by participation in the Federal School Lunch Program. For a school to achieve Adequate Yearly Progress (AYP), at least 95% of the students in each group were required to contribute scores, the percentage of students demonstrating proficiency needed to meet target levels, and the school had to show progress on the other academic indicator (OAI): attendance and/or graduation rate. If a subgroup had fewer than 40 students it was not considered, with the exception of the full student sample.

Schools faced no consequences in their first year of AYP failure, but those that failed to make AYP in subsequent years could face sanctions. These included being forced to allow their students to choose a different school, to provide extra services, or to undergo major restructuring, depending on the number of consecutive years AYP had not been achieved. Adequate Yearly Progress was determined separately for both reading and math, such that a school could be in year 1 of AYP failure for one subject and year 3 for the other. Consequences were based on the higher of these two numbers.

There were several conditions under which a school that had not achieved all the AYP requirements would be treated as though it had done so. Schools that achieved the participation requirement and made progress in the other two measures could receive “safe harbor” and avoid sanctions. Consequences also were withheld if the proficiency threshold was within a 95% confidence interval of the actual level of proficiency or if the proficiency of the population eligible for or receiving Title I targeted assistance met the threshold. Students who had exited LEP status or special education in the previous two years could also be included in proficiency counts. All of these details resulted in variation across schools, years, and subjects in whether an AYP failure would result in sanctions, in addition to variation across subgroups in whether the subgroup could be expected to make AYP. I exploit this variation, in addition to differences across students in subgroup membership and expected performance, in order to identify school reactions to incentives and the effect of special education on student achievement.

2.3 State Accountability

In North Carolina, NCLB operated in tandem with the state’s own “ABC” program, which was first implemented in 1996 and continued with slight alterations over the period studied. Under “the ABCs,” schools were labelled with various positive and negative terms based on individual student growth and the percentage passing. For example, in 2006-2007, schools that met AYP, met expected growth, and had at least 90% proficiency across grades and subjects were labelled as Honor Schools of Excellence (NCDPI 2007).[9] Among schools with at least expected growth, those with at least 90% proficiency were labelled as Schools of Excellence, those with 80-89% proficiency as Schools of Distinction, those with 60-79% proficiency as Schools of Progress, and those with less than 60% proficiency as Priority Schools. For those schools not making expected growth, those with at least 60%

[9] A school met expected growth if, on average, students at least maintained their achievement level from the prior year.

proficiency were labelled No Recognition Schools and those with 50-59% proficiency were Priority Schools. Those with less than 50% proficiency that did not make expected growth were labelled as Low-Performing Schools, and were provided with state assistance such as additional professional development. Schools that achieved high or expected growth could receive teacher bonuses.

While many of the principles underlying NCLB and the ABCs are similar, the precise inputs to the accountability formulae, the relative emphasis on growth vs. proficiency, and the cut-points are different across the two regimes. The ABCs also does not hold schools accountable for the performance of subgroups defined based on demographics or special education status. Thus, incentives created by the ABCs should be uniform across students with similar performance, regardless of their subgroup membership or the AYP performance of their school. However, it is possible that school responses to AYP incentives would have been different if they were not also trying to react to the state program, as the ABCs incentivizes increases in the percent proficient regardless of subgroup membership or the school's past performance.

Two additional features of the North Carolina context deserve mention. First, schools that did not receive Title I funding were evaluated under the NCLB standards. Their performance, overall and for subgroups, was announced, but there were no consequences directly tied to whether or not these schools made AYP. These schools were evaluated according to NCLB standards in all states, but states varied in how they used this information. I focus on the experiences of Title I schools, which faced stronger incentives that were consistent across states. Second, North Carolina qualified for an early waiver to the standard NCLB framework beginning in 2005-2006. This allowed students to count towards the school's proficiency rate if they either performed above the cut score or were exhibiting growth that suggested they would reach proficiency within four years of their initial test. This was uncommon when introduced by became more common elsewhere by the end of NCLB. I incorporate this detail into one measure of school expectations of student performance and find that is does not alter my results substantially. While these two policy details are important for understanding my analysis, they do not seriously limit its generalizability.

2.4 Resulting Incentives

Prior to NCLB, accountability programs allowed schools to exclude special education students from their accountability populations and often from testing. Thus, a school could appear to have improved its performance by steering low-performing students into special education.

In contrast, there is no such option under NCLB and subsequent policies. Instead, schools could encourage special education for “bubble” students who are expected to be close to the passing threshold as a way of targeting services and supports to them. This could improve the school’s AYP performance if it expected to otherwise fail to make AYP for at least one subgroup of which the student was a member. Schools that expected to fail to achieve AYP for the SWD subgroup had an incentive to try to improve its performance. One way of doing this would be to change the group’s composition by encouraging special education for students who were expected to pass and/or discouraging those who were expected to fail. This strategy would be most useful to schools that were close to the AYP threshold, so they could change their rating by moving a relatively small number of students.

3. Prior Literature

My analysis is related to two strands of prior literature. The first has analyzed how schools respond to incentives. The part of this literature that is most closely related to my has addressed accountability policy incentives to alter special education placement. All previous work has focused on those in force before NCLB. These policies, examined by Jacob (2005), Hanushek and Raymond (2005), Cullen and Reback (2006), Figlio and Getzler (2006), Cohen (2007) and Bokhari and Schneider (2011), held schools accountable for student performance but, importantly, allowed special education students to be excluded from the accountability population and often from testing. Thus, a school could increase its chance of passing by placing low-performing students into special education, and schools facing accountability pressure did just that. When programs monitored the performance of subgroups, members of those groups at risk of not meeting benchmarks were more likely than other students to enter special education, especially for students whose exclusion from accountability improved the school's performance (Cullen and Reback 2006). Using state-level variation in the roll-out of accountability policies pre-NCLB, Hanushek and Raymond (2005) found no evidence that these policies increased special education rolls, but Bokhari and Schneider (2011) found that accountability systems that provided rewards for good performance increased the number of ADHD diagnosis in the public school population, as well as the use of medication. None of this research has considered responses to more recent accountability policies, which were crafted in part to eliminate these incentives.

More broadly, a rich literature has explored how schools responded to NCLB. These responses include focusing on tested grades and subjects, focusing on the needs of "bubble" students whose passing status might change as a result, altering the testing pool through discipline, and even altering the content of school lunches on testing days (e.g. Figlio, 2006; Figlio and Winicki, 2008; Griffith and Scharmann, 2008; Krieg, 2008; Reback, 2008; Byrd-Blake et al., 2010; Dee and Jacob, 2010; Ladd and Lauren, 2010; Neal and Schanzenbach, 2010). I add to this literature by considering responses along a different margin, that of special education placements.

Prior work also has shown that schools respond to financial incentives to alter the size and composition of their special education population. These incentives can come in the form of state funding formulas (Cullen, 2003; Kwak, 2010; Mahitivanichcha and Parrish, 2005; Morrill 2016) or voucher programs open only to students with disabilities (Winters and Greene, 2011; Chakrabarti, 2013). I contribute to this research by presenting evidence on how financial and accountability incentives interact.

The second strand of literature addresses how students are affected by special education assignment. Only two previous papers have applied rigorous research designs to the question directly. [10] Hanushek, Kain, and Rivkin (2002) used student fixed effects in a panel dataset from Texas in the 1990s and found small but significant gains in test scores in years in which students were in special education. Their identification strategy requires two assumptions that I am able to relax. First, they assume that any omitted variables that are correlated with both special education status and achievement are static, which would not be true if students' impairments change over time. Second, they must assume that changes in achievement do not cause changes in special education placement, at least after controlling for observable factors. This would be of particular concern if students are more likely to be placed in special education when struggling and to leave special education when performing well, so that regression to the mean would appear as a positive effect of special education.

In an unpublished working paper, Cohen (2007) used the accountability policy implemented by Chicago Public Schools in 1996 to construct instruments for special education placement. She found evidence that schools responded to incentives to place low-achieving students into special education but was not able to detect effects on attendance, graduation, or GPA. Her analysis also rests on an assumption that schools trying to improve the percentage of students scoring at grade level would not do anything, aside from altering special education placement, that would affect the outcomes of low-achieving students.[11] While this may be true for the very lowest achieving, who saw the greatest

[10] While not a direct analysis of the effect of special education on achievement, Setren (2016) demonstrates that special education students who win charter lotteries experience gains similar to those of their classmates who were not previously in special education, despite charters' practice of removing special education classifications at a high rate. Multivariate regressions suggest the removal of special education classifications is either not harmful or improves scores.

[11] I would need to make a similar assumption in order to use the incentives to target services to bubble as instruments for special education. For this reason I do not do so, and instead only use the incentives to select the special education population as instruments.

increase in special education placement, it is less likely for students only slightly below the average, whose probability of being in special education also increased.

While previous work has addressed many ways that schools responded to a variety of NCLB incentives, and school responses to the special education incentives in accountability policies that existed prior to NCLB, this paper is the first to consider school responses to the incentives in NCLB to alter the special education population. In doing so I am able to evaluate to what extent recent accountability policies have solved the problems identified in the earlier literature on incentives to alter the special education population. I also provide an estimate of the effect of special education on student achievement, contributing to a small literature based on strong assumptions that I relax. My estimates are local to students whose placements can be altered by schools. While this limits their generalizability it also means that they are relevant to the very group of students for whom it is most important to know the effect of special education.

4. Data

I use restricted-access student-level information from the North Carolina Education Research Data Center (NCERDC) and public use school-level information from the North Carolina Department of Education and the Common Core of Data. Student-level files provide year-by-year information on tests taken, standardized test scores, testing accommodations, disability classifications, and demographics. These files are linkable across years to create a panel that includes the universe of North Carolina public school students who were in tested grades during the years I examine. I also draw information on which schools a student attended in which grades and year from the student-level files. To the student data I add information on school characteristics and the number of years of AYP failure each school had in math and reading for all students and subgroups.

Based on their scores, students are assigned to one of four achievement levels numbered 1-4, defined by whether students have mastered grade-level content sufficiently to be prepared for the next grade.[12] Students in achievement levels 3 and 4 are proficient, while those in levels 1 and 2 are not. Under North Carolina's growth model, students who are in levels 1 or 2 but are on track to be proficient within 4 years of initial testing can be considered proficient for purposes of determining AYP. In my main specification, I assume schools expect students to perform about as well in the current year as in the past year, so this growth component is not relevant. However, I take the growth model into account when considering whether schools treated students who would need small gains to achieve proficiency differently from those who would need larger gains and find similar results. Thus, this modeling assumption does not drive my results.

North Carolina offered a series of alternate tests to students for whom the standard test was inappropriate, including special education students whose IEPs specified that they would take these tests. The alternate tests mean I have access to information on almost all students in the tested grades, but scores must be standardized to allow for comparisons across tests. To do so, I first assume that students who scored at a given achievement level cutoff have the same achievement – that is, students who just received passing scores for a given grade, year, and subject had the same achievement, regardless of test taken.[13] Then I assume the distance between achievement levels has the same

[12] North Carolina has since switched to a 5-category classification, but used this 4-category system during the period I consider.

[13] While taking an alternate test might improve the score of a student who would struggle to demonstrate their knowledge on a standard test, the most common of these tests, the NCEXTEND2, evaluated students relative to grade-level standards. For example, the reading form "uses shorter reading selections, simplified language, and fewer test items and item responses (foils/answer choices) to assess students on grade-level content" (North Carolina Public Schools, 2009, p 5). The cut scores between achievement levels were selected through a similar procedure for both the NCEXCEND2 and the regular end of grade tests. First, a group of students who met the eligibility criteria piloted the tests. Then the teachers of these students were asked to use their knowledge of the students' classroom performance to categorize them into achievement levels. Test makers noted the percentage of students expected to score in each achievement level, and set cut scores accordingly. That is, if the teachers reported that 15 percent of the 4th grade students tested were in achievement level 1 the test makers set the cut off between levels 1 and 2 such that the lowest 15 percent of scores were in level 1 (North Carolina Public Schools, 2009). Cut scores were then reviewed and approved by a panel of policy makers and stakeholders. To the extent that this assumption is incorrect my results would be biased towards finding positive effects of special education, suggesting that its true negative effects are even larger than estimated.

meaning for all tests for a given grade, year, and subject – that is, students who scored halfway between the level one and level two cut points have the same achievement, regardless of test taken. Finally, I form z-scores by subtracting the mean and dividing by the standard deviation for each grade, year, subject combination. This produces a set of scores with mean zero and standard deviation one for each grade, year, and subject.

Special education serves students with a variety of disabilities. The distribution of diagnoses for special education students in North Carolina in grades 4-8 during my sample period is reported in Table 1. Schools are unlikely to be able to influence the special education placement of students with many impairments, such as a visual impairment or traumatic brain injury. I consider the likelihood of being in one of two broad categories of diagnoses – those that are likely to be relatively malleable and those that seem especially difficult for schools to alter. Malleable impairments are speech and language impairments, learning disabilities, emotional and behavioral disorders, and other impairments (which includes ADHD). Non-malleable impairments are autism, intellectual disability, developmental disabilities, sensory disabilities, traumatic brain injury, orthopedic impairments, and multiple disabilities. I focus on malleable diagnoses defined this way in my main analysis. Estimates that include students with autism in the malleable group appear in Appendix Table A.6 and are similar to my main results. I also use the non-malleable diagnoses to conduct a falsification test.

Data from the alternative tests are available beginning in 2006, so I begin my analysis in 2007 to have at least one previous year of data for students taking the alternative tests. This restriction also allows me to analyze a consistent policy environment, as North Carolina began using student gains in its AYP calculation in 2005-2006. I also exclude third-grade students, as most do not have a prior year test score.

I remove from the analysis sample students who are missing information on current special education status, current or previous standardized test scores, or the performance of their school and subgroups in the past.[14] I loosen the requirement to have current-year scores when considering whether schools altered the testing population in response to incentives. Students with incomplete data are only included in the sample for the years and subjects for which complete information is available, resulting in an unbalanced panel. I investigate the relationship between incentives and attrition in section 5.4. In baseline specifications, I drop all those who ever appear in the data with a non-malleable diagnosis, so as not to confuse changes of diagnosis with movement in and out of special education. This restriction is altered when considering the effects of incentives on having a non-malleable diagnosis, and as a robustness check in Appendix Table A.2. Results are not sensitive to the exclusion of students who had a non-malleable diagnosis at some point in time.

My main sample, described in Table 2, comprises about 1.3 million student-year observations, representing about 700,000 students in seventeen hundred schools.[15] About 10.5 percent of the students in my sample were in special education. While about 13 percent of students in North Carolina were in special education, my main sample excludes those who had a non-malleable diagnosis at any point in time, decreasing the percent in special education. About half were female, and 57 percent were eligible for free or reduced-price lunch. Slightly less than half the sample identified as White/Caucasian, 31 percent as Black, and 12 percent as Hispanic. A majority had passed their standardized test in the previous year, about 69 percent in reading and 73 percent in math. Twenty two percent were also in schools that had failed to achieve AYP thresholds in math for at least one subgroup in the previous year, and 18 percent were in schools that had failed to do so in reading for at least one group. Students in special education were significantly less likely to have passed their test in the previous year. They were also more likely to be male and more likely to be low-income. This demonstrates the disadvantaged

[14] Of students in the grades and years considered, 8.9 percent are missing prior-year test sores. This includes students who are in their first year in North Carolina Public Schools. Among those with prior-year test scores, 1.4 percent are missing information on prior school performance. Of those with data on prior test scores and prior school performance 0.4 percent are missing information on school Title I status.

[15] Descriptive statistics for those who had a non-malleable diagnosis at some point in time appear in Table A.3.

nature of the special education population, which is part of the empirical challenge of identifying the causal effect of placement.

5. School Responses to Accountability Incentives

5.1 Method

I first examine school responses to accountability incentives, then consider the effect of special education placement on student outcomes. I assume that schools make decisions about special education placements at least once a school year. This is consistent with US Department of Education regulations that require IEPs to be reviewed every 12 months or more often if necessary (US Department of Education, 2000). When making these choices, they may consider a wide array of information about their students but only take AYP incentives for the **current** year into account. This assumption would be violated if a school tried to slow its improvement this year in order to make improving next year easier. While it is likely that schools would want to plan ahead, it seems less likely that they would be able to do so effectively. This assumption allows me to consider a static model in which schools respond to current incentives. If it is incorrect my estimates will not reflect all responses to accountability incentives but would still reflect current year responses to current year incentives - a relevant parameter.

I address the question of how NCLB incentives altered disability classifications by testing two main hypotheses about school responses to incentives as outlined in Section 2.4. First, schools that are at risk of failing to make AYP may use special education as a way to target extra services to “bubble students” whose passing status could reasonably be changed.

$$(1)\ SE_{igjt} = \beta_1 X_{igjt} + \sum_{s=r,m}[\ \beta_{2s} Incentive_{igjst} + \sum_{a=2,3}[\ \beta_{3sa} Bubble_{igjsat} + \beta_{4sa} Incentive_{igjst} * Bubble_{igjsat}]] + \ \gamma_{gt} + \ \sigma_j + \varepsilon_{igjt}\ .$$

In Equation (1), an indicator for whether student *i* in grade *g* in school *j* in year *t* is in special education (SE_{igjt}) is a function of observed characteristics of student *i* (X_{igjt}) as well as whether student *i* is a bubble student ($Bubble_{igjsat}$) just above the passing threshold (a = 3) or just below (a = 2) in reading (*s = r)* or math *(s = m),* whether the school has an accountability incentive to ensure that student *i* passes ($Incentive_{igjst}$) as defined in the next paragraph, and an interaction between those final terms for each subject. All models include year by grade fixed effects (γ_{gt}). Main estimates include school fixed effects (σ_j); estimates without school fixed effects appear in Appendix B. Student characteristics include prior scores in math and reading as well as indicators for lagged special education status, LEP status, free or reduced-price lunch eligibility, and the racial and ethnic categories used by NCLB.

I consider a school to have an incentive to improve the likelihood of student *i* passing if the school failed to make AYP in the previous year for any subgroup – not including the SWD group– into which student *i* falls and would face sanctions for future AYP failures. For example, if student *i* is economically disadvantaged and Hispanic the school would have an incentive to ensure that the student passes if the school was at risk of failing AYP for all students, Hispanic students, or economically disadvantaged students, and would face consequences for doing so. As discussed in more detail in section 2.2, some schools that did not achieve AYP requirements avoided AYP failure status. Schools with these types of failures in the past year would not receive immediate sanctions if they failed in the current year.

I define bubble students in two ways. First, I consider that schools may target their actions bluntly based on the achievement levels in which students scored in the previous year. To do this I define bubble status as having scored in achievement level 2 or achievement level 3 in the previous year. Because schools may treat students who previously passed differently from those who previously failed, I include separate terms for being in level 2 or level 3.

Second, I consider that schools may target their actions more precisely to students particularly close to the passing threshold. In doing so, I incorporate North Carolina's gain score model, under which any student who was on a trajectory to reach proficiency within 4 years of their first test could be counted as passing for AYP targets. I define an inverse measure of the amount by which a student's score would need to rise or fall in order for them to just count as passing for AYP determinations. I begin by defining the student's distance from counting as passing. For students who passed and students who failed but are in at least 7th grade, this is the absolute difference between the student's score and the test's cut score. Students in sixth grade or lower who failed the previous year can count as passing in the current year if they improve enough to be on a trajectory to be proficient by seventh grade. I approximate this needed improvement as their distance from the cut score divided by the number of years left before seventh grade. I then use the distance from counting as passing to construct an inverse distance measure standardized across grades, years, and subjects. To do this I calculate the largest distance to counting as passing for each grade-year-subject combination. The standardized measure of the inverse distance to counting as passing is the largest distance for the grade-year-subject minus the student's distance, divided by the largest distance for the grade-year-subject. In my models, I test whether this measure of distance matters for either students who previously scored in level 2 or in level 3.

If schools use special education to target services to almost-passing (just-passing) students, I would expect to find positive coefficients on β_{4r2} and β_{4m2} (β_{4r3} and β_{4m3}) when using the simple definitions of bubble group membership. To the extent that schools focus on those very close to passing even among level 2 (level 3) students, I would also find positive values when defining bubble status based on the inverse distance from counting as passing for level 2 (level 3) students.

Second, schools that are at risk of failing to make AYP for the SWD subgroup could attempt to select their special education population to be relatively high performing. I test for this possibility using the following model:

$$(2)\quad SE_{igjt} = \beta_1 X_{igjt} + \sum_{s=r,m}[\beta_{2s}\widehat{Pass}_{igjst} + \beta_{3s}MarginalSWD_{jst} + \beta_{4s}\widehat{Pass}_{igjst} * MarginalSWD_{jst}] + \gamma_{gt} + \sigma_j + \varepsilon_{igjt}\ .$$

The special education status of student *i* still depends on student characteristics and year-by-grade fixed effects. Now, the school's incentive is an interaction between the student's predicted performance ($\widehat{Pass_{igjst}}$) in each subject and whether the school is at the margin of failing to make AYP in that subject for the SWD subgroup ($MarginalSWD_{jst}$). I assume that a school's best prediction about a student's performance this year is their score last year (or, alternately, whether they were proficient last year). To define to what extent schools are at the margin of failing to make AYP in a subject for the SWD subgroup, I create an inverse measure of the amount a school would have to improve their performance to meet AYP.

$$(3)\ \ MarginalSWD_{jst} = \left(1 - \left(\frac{Threshold_{jst-1} - PercentProfSWD_{jst-1}}{Threshold_{jst-1}}\right)\right) * \ FailedSWD_{jst-1}$$

In Equation (3), the degree to which school *j*'s SWD subgroup is marginal to passing in subject *s* and year *t* ($MarginalSWD_{jst}$) is defined based on the schools' percent proficient in the previous year ($PercentProfSWD_{jst-1}$), the AYP threshold for that subject and year ($Threshold_{jst-1}$), and an indicator for whether the school had failed to make AYP for the SWD group in the previous year ($FailedSWD_{jst-1}$). Figure 2 illustrates the MarginalSWD measure for a sample of schools that had failed to make AYP in the previous year. The measure takes on a value of 0 for schools in which no SWD students passed in the previous year and climbs linearly with the percent of SWD students passing until reaching a value of nearly 1 for schools just below the AYP threshold. The measure is 0 for schools in which the SWD group achieved AYP, either by having a passing rate at or over the threshold, or through one of the alternate calculations discussed earlier.

In reality, many schools that fail in one group fail in more than one – roughly half of schools that failed in the SWD subgroup also failed in another group and most schools that fail in another group fail in the SWD subgroup. As a result, many schools are faced with both incentives simultaneously. Some of the students in these schools will also be the targets of both incentives; consider a student who just passed and whose school failed to make AYP both for the SWD group and a demographic subgroup of which the student is a member. For this reason, I model them together as in equation (4) below.

$$(4)\ SE_{igjt} = \beta_1 X_{igjt} + \sum_{s=r,m}[\ \beta_{2s} Incentive_{igjst} + \sum_{a=2,3}[\ \beta_{3sa} Bubble_{igjsat} + \beta_{4sa} Incentive_{igjst} * Bubble_{igjsat}] + \ \beta_{5s}\widehat{Score}_{igjst} + \beta_{6s} MarginalSWD_{jst} + \ \beta_{7s}\widehat{Score}_{igjst} MarginalSWD_{jst}] + \ \ \gamma_{gt} + \ \sigma_j + \varepsilon_{igjt}$$

This model identifies causal effects of accountability-related incentives on placement under two assumptions about parallel trends. For the first hypothesis, the assumption is that placement differences between bubble and non-bubble students when schools do not have an AYP incentive to improve the students' likelihood of passing reflect the differences by bubble status that would exist for students whose schools have an AYP incentive to improve their likelihood of passing, in the absence of that incentive. Suppose for a given year and subject NCLB required 84 percent proficiency for all groups. The parallel trends assumption would be violated if, in the absence of AYP incentives, the change in placement that a bubble student experiences when one of the subgroups to which they belong goes from having at least 84 percent proficiency to less than 84 percent proficiency was different from the change experienced by a non-bubble student.

For the second hypothesis, the assumption is that differences in placement by prior test score would not vary with the SWD group's passing rate relative to the AYP threshold in the absence of AYP incentives. This would be violated if, in the absence of AYP incentives, the change in placement a previously passing student would experience if their school's SWD group went from above 84 percent proficiency to below was different from that of a previously failing student.

Students and parents have their own incentives surrounding special education, and schools face other sources of pressure. However, most student and parent incentives do not change around student passing thresholds, and those that do should not change around schools' AYP thresholds. That is, a student's parents may want them to be in special education to improve their performance, and this will appear in coefficients on prior score or demographics. Some of those families will probably push harder for placement if their student is struggling, or perhaps if they appear to be almost doing "well enough" but need a slight boost. Schools may similarly use scores to identify struggling students. Both of these responses will appear in the coefficient on expected score or being a "bubble" student. In the absence of accountability incentives, these reasons for special education classification do not change when the school is at risk of failing to make AYP or when the student is important to the school's effort to do so. Similarly, in a world without NCLB, schools' identification of struggling students by their score should not depend on the student's subgroup membership or the SWD group's performance. Thus, the interaction terms that identify school responses to incentives and form my instruments should reflect only school responses as a result of NCLB incentives.

5.2 Placement Responses to AYP Incentives

Figure 3 illustrates the residual percentage of students in the main analysis sample who had a malleable diagnosis after controlling for demographics, prior score, and year-by-grade fixed effects, by the student's distance to a passing score in math the previous year. The first series, marked with a solid line, includes all those whose school expected to fail AYP for a group of which that student was a member, and the school would face consequences for such a failure ($Incentive_{ijmt} = 1$). The second series, marked with a dashed line, includes those whose school either did not expect to fail AYP for any group of which that student was a member or did not face sanctions for such a failure ($Incentive_{ijmt} = 0$).[16] Most students (about 69 percent) whose school had previously failed to make AYP for a group of which they were a member also failed to make AYP for the SWD group, so had positive values of $MarginalSWD_{jst}$. Nearly all students whose schools did not have an incentive to improve their

[16] Schools that failed to meet AYP thresholds in the previous year could have avoided AYP failure because the subgroups that failed contained fewer than 40 students, because the threshold was within a 95 percent confidence interval around the passing rate, or because the subgroup met other AYP requirements and had improved its passing rate by at last 10 percent from the previous year.

likelihood of passing also did not have an incentive to improve the performance of the SWD group (97 percent). As a result the figure illustrates responses to both incentives simultaneously.

Students who previously received a failing score or just passed were more likely to be in special education with a malleable diagnosis if their school did not have an incentive to improve their performance. Those who had previously passed by more than about a standard deviation were about as likely to be in special education regardless of whether their school had an incentive to improve their performance. There is a similar pattern in reading, illustrated in Figure 4. If hypothesis 2 is correct we would expect the relatively likelihood of being in special education for previously-passing versus previously failing students to be higher in schools with incentives to improve the performance of the SWD subgroup. Figures 3 and 4 suggest that this is the case, and that the lion's share of the selection takes place through discouraging special education for previously-failing students, rather than encouraging special education for previously-passing students.

If schools use special education to target supports and services to bubble students when the school would benefit from their achieving a passing score, as suggested by hypothesis 1, we would expect the presence of an AYP incentive to increase residual malleable diagnoses close to the passing threshold. There is a noticeable bump in malleable diagnoses in Figure 3, peaking around a quarter of a standard deviation below the passing threshold. The pattern appears more dramatic for those whose schools had an AYP incentive to improve their likelihood of passing, but is also present for those without such an incentive. A broadly similar pattern holds in reading, as illustrated in Figure 4.

I now turn to a more systematic examination of how schools respond to accountability incentives using Equation (4). The next several tables are structured similarly. Each column displays estimates from a single regression. In the odd columns, "bubble" status is defined using binary indicators of achievement level, while the even columns test whether the distance from the cut sore matters for students in levels 2 or 3, as detailed in the previous section. In columns 1 and 2, I test whether schools select students based on prior passing status when trying to improve the performance of the SWD subgroup; columns 3 and 4 use prior score. Student demographics, year-by-grade fixed effects, and school fixed effects are included in all regressions. Estimates without school fixed effects appear in Appendix B.

Table 3 presents the effects of NCLB incentives on the likelihood of being in special education with a malleable diagnosis. I will start by discussing the second hypothesis, that schools select the special education population to be relatively high-performing when the school would benefit from improving the performance of the SWD group. Students who had previously passed in reading were 2.2 percentage points more likely to be in special education when their school had an incentive to improve the performance of the SWD group in reading, as shown in columns 1 and 2. I find no significant evidence of selection based on math performance with this specification. Considering score rather than passing status, as in columns 3 and 4, suggests that a one standard deviation higher prior reading score increased the likelihood of being in special education by about 1 percentage point when the student's school had just failed to achieve AYP for the SWD group. A math score that was one standard deviation higher increased the likelihood that a student would be in special education by 0.7 percentage points when the student's school had just failed to achieve AYP in math for the SWD group.

Next, I consider evidence of the first hypothesis. I find that schools encouraged special education placement for students who were close to the passing threshold in reading, whether above or below. In column 1, when bubble status is measured by having previously scored in level 2 or 3, a student who had previously scored in level 2 (level 3) in reading would be 1.4 (2.5) percentage points more likely to be in special education if their school would benefit from their passing. These estimates are robust to the way selection of high-performing students is parameterized. I also find evidence that schools targeted those closer to the passing threshold more strongly than those farther away, as shown in columns 2 and 4.

Schools appear to have discouraged students who had scored in level 2 in math from special education when the school would benefit from that student passing, and not changed their treatment of level 3 students in math. This may reflect schools believing that there are other options available to raise the math scores of almost-passing students. It could also be a manifestation of school attempts to control the size of the special education population. To investigate this, it is useful to compare schools that had at least 12.5 percent of their student body in special education, and would not receive additional state aid to support further special education placements, with those schools with a smaller proportion in special education. As shown in Table A.1, schools with at least 12.5 percent of their students in special education appear to discourage placement more strongly for students who had scored in level 2 in math than do schools with a smaller proportion of students in special education. Schools with more than 12.5 percent of their students in special education also encourage special education less strongly for students who scored in level 2 in reading, in comparison to schools with smaller special education populations.

5.3 Heterogeneity

One way of verifying that the estimates from Table 3 reflect school responses to AYP incentives is to compare these reactions across groups that faced stronger and weaker incentives. If estimates reflect a causal relationship rather than omitted variables those who faced stronger incentives should have exhibited larger reactions, or at least not smaller.

While a school's incentives surrounding a student do not necessarily depend on that student's underlying impairment, a school's ability to influence whether the student is in special education does. Schools should have much less influence on diagnoses for which it would be difficult to not place a student in special education – say a student who is blind or uses a wheelchair – than on the relatively malleable diagnoses I consider in my main analyses. Table 4 displays the results of relaxing the sample restriction that excluded those who had ever had a non-malleable diagnosis and estimating the effect of incentives on non-malleable diagnosis. I find some evidence that students who had just passed in the previous year were less likely to have a non-malleable diagnosis when their school would benefit from their passing. However, these effects are quite small in comparison to the results in Table 3, and no other coefficients are significant. This does however highlight the fact that some of the diagnoses I am classifying as non-malleable can be influenced by schools, just to a lesser extent than the malleable diagnoses.

I also examine reactions by schools that did not receive Title I funding. These schools were still required to test their students in accordance with NCLB mandates, and performance was reported publicly. As such they did have incentives to perform well according to the NCLB metrics, but these incentives were much lower than for Title I schools, which faced the possibility of sanctions. Unfortunately for this analysis, very few non-Title I schools failed to achieve AYP, so there is little variation in the data, resulting in imprecise estimates. These results appear in Table A4, and suggest that schools that did not receive Title I funding did not use special education to target services to level 3 students whose passing would benefit the school. There is some evidence of targeting away from level 2 students, particularly in reading, and selection against prior-passing students in reading, both of which are somewhat puzzling. They may reflect school attempts to keep the special education population below the 12.5 percent funding threshold, while using placement to respond to other priorities. Non-Title-I schools do appear to have selected students based on their math performance when the school expected its SWD group to fail. These point estimates are larger than those for Title I schools, but are quite imprecise.

5.4 Test taking and selection

One of the primary innovations of NCLB was its testing requirement – schools were required to have at least 95 percent of students contributing scores, both overall and in each accountable subgroup. This drastically decreased the scope for schools to select the tested population but did not eliminate it entirely. Schools could still potentially do more to encourage test-day attendance for some students than others, and those with especially good attendance might be able to actively discourage some particularly low-scoring students. Selection of the test taking population would be another way for schools to respond to accountability incentives that would not appear in my main analysis. It also would limit my ability to use these incentives as instruments for special education placement in the next section. Suppose special education had no effect on student achievement, but the same accountability pressures that influenced selection into special education drove schools to change the tested population. In this case it would be possible to find effects of special education – in either direction depending on how selection into testing took place.

I investigate whether accountability incentives predict the likelihood a student appears in the data with a valid test score. Results appear in Table 5. I find no evidence that schools respond to AYP incentives by altering the tested population. This suggests that my analysis of the effect of special education on achievement outcomes is not subject to bias due to sample selection, something that previous research on the subject likely suffered from and was not able to analyze directly. It also suggests that the combination of NCLB's testing requirements and the introduction of alternate tests succeeded in discouraging schools from excluding their special education students from testing.

6 Effects of Special Education on Student Achievement

6.1 Method

Special education is not randomly assigned, and students who receive special education are systematically different from those who do not. As a result, simple comparisons of the outcomes of students in and out of special education would not reflect the causal effect of placement. To overcome this problem I use incentives from the previous section as instruments for special education. In order for the incentives to be valid instruments they must obey the exclusion restriction. That is, they must not affect outcomes through some mechanism other than special education placement. I do not expect this the exclusion restriction to hold in the case of the incentives from the first hypothesis, as prior work suggests that schools are able to use other efforts to target resources to students who they wish to pass (Reback 2008). For this reason, I do not use the incentives from the first hypothesis as instruments.

The incentives from the second hypothesis make more suitable instruments. For these instruments the exclusion restriction holds as long as schools do not change their allocation of resources to students who were selected into special education to change the composition of the SWD group. This could either take the form of focusing extra energy on those students who were placed in special education in order to be especially certain that they passed or withdrawing resources the students would otherwise have been provided. The restriction would not be violated if the school is making other efforts to improve the test performance of students in special education in general. To the extent that special education under accountability pressure is different from that without, it could limit generalizability. However, my results would still apply to any situation where schools make a particular attempt to increase the percentage of students in special education achieving proficient scores. Similarly, if some resources or opportunities are not provided to students in special education, perhaps due to time or scheduling constraints, this would not bias my results. Rather, it would mean that a move into special education entailed not only the addition of services detailed in the student's IEP but a loss of other services provided to students who were not in special education. If for some reason this withdrawal of resources only took place when the school was under accountability pressure the result might not fully generalize to environments where special education was an additive service rather than the exchange of one set of services for another. It is not clear why this would be true, and even then the exclusion restriction would not be violated. Finally, it is possible that students who are selected into special education based on scoring well in one year experience a decrease in score in the next. This reversion to the mean would only confound my estimates if it somehow occurred for students in schools with an incentive to improve the performance of the SWD group and not for those in schools without that incentive or vice versa. There is no reason to believe that this is the case.

It is also necessary for the instruments to satisfy a monotonicity assumption. This would be violated if some students I have labelled as incentive targets were not seen as such by their schools and instead were discouraged from being in special education in order to "make space" for others. This would require a school to believe that students who failed in the previous year were more likely to pass in the current year than those who passed in the previous year, which seems extremely unlikely.

That is (4) serves as the first stage of a model with the second stage:

$$
\begin{aligned}
(5)\ Score_{igjt} &= \beta_1 X_{igjt} \\
&+ \sum_{s=r,m} [\, \beta_{2s} Incentive_{igjst} + \sum\nolimits_{a=2,3} [\, \beta_{3sa} Bubble_{igjsat} \\
&+ \beta_{4sa} Incentive_{igjst} * Bubble_{igjsat}] + \beta_{5s} \widehat{Score}_{igjst} + \beta_{6s} MarginalSWD_{jst}] \\
&+ \beta_7 \widehat{SE}_{igjt} + \gamma_{gt} + \sigma_j + \varepsilon_{igjt}
\end{aligned}
$$

In Equation (5) $Score_{ijt}$ is the relevant current-year test score or other achievement outcome. Special education status (SE_{igjt}) is a simple indicator for whether the student is in special education with a malleable diagnosis in the current year. The model also includes student characteristics as well as year-by-grade fixed effects and school fixed effects. Estimates without school fixed effects appear in Appendix B.

6.2 Effects of special education on achievement

I use Equation (5) to estimate the effect of being placed in special education on a marginal student's same-year achievement. Relevant first-stage estimates appear in Table 3. Incentives to select the SWD group to be relatively high performing in reading form strong instruments, as shown in Table 6. I do not find evidence that special education has an effect on achievement for students whose placement is altered by incentives to select the SWD population to have a higher passing rate in reading.

These estimates are very noisy, so I cannot rule out large effects in either direction. Math incentives form weaker instruments; only those based on prior score are strong. However, these estimates suggest that special education hurts math achievement for those students whose special education placement is altered to improve the math achievement of their school's SWD group. Being placed in special education lowers the math scores by about 1.2 standard deviations for this group of students. The point estimates for reading scores for this group are also negative, though smaller and not significant at conventional levels. A 1.2 standard deviation effect on test score is extremely large, roughly equivalent to falling from the 75th percentile to the 25th. About 4 percent of students in my sample experience a year-to-year change at least this large. However a 1.2 standard deviation effect could reflect a 0.6 standard deviation fall from a student who would otherwise have experienced a 0.6 standard deviation gain. About 28 percent of my sample experiences a year-to-year score change of at least 0.6 standard deviations. It is also important to note that my estimates are noisy, so I am unable to rule out smaller effects.

The fact that special education can hurt the achievement of some students is surprising given prior findings of the effect of special education on achievement. However, my results are consistent with the findings by Setren (2016) that students who had a special education placement before entering a charter school saw gains similar to those of their non-special education classmates despite losing special education designations at a fairly high rate. The differences between my estimates and those in Hanushek, Kain, and Rivkin (2002) could be driven by differences in settings – location, time period, grades, etc. They could also be the result of or differences between the local average treatment effect (LATE) I estimate – that of of being placed in (or not being placed in) special education as a result of AYP incentives – and the parameters estimated in prior work.

To investigate the first possibility, I replicate the main analysis from Hanushek, Kain, and Rivkin (2002) on my sample of students in North Carolina in the NCLB era. This specification uses student fixed effects to control for unobserved differences between those who receive special education and those who do not, and measures the effect of special education placement on gain scores.[17] Here the change in score for student *i* in grade *g* in school *j* in year t (ΔA_{igjt}) is a function of the student's special education status in that year (SE_{igjt}), student characteristics (X_{igjt}), school characteristics (D_{gjt}), a student fixed effect (γ_i), a school fixed effect (δ_j), cohort by grade dummies (ω_{gt}), and an error term (ϵ_{igjt}):

$$(6)\ \Delta A_{igjt} = SE_{igjt}\lambda + X_{igjt}\beta + D_{gjt}\theta + \gamma_i + \delta_j + \omega_{gt} + \epsilon_{igjt}.$$

Student characteristics include free or reduced-price lunch eligibility and an indicator for whether the student changed schools that year; school characteristics include the percentage of students who were Black, the percentage Hispanic, and the percentage eligible for free or reduced-price lunch. I estimate the model as written, and removing the student fixed effects but adding controls for student race and gender. The resulting estimates appear in Table A5. Using this student fixed effect specification, I find small but positive effects of special education on student achievement in math, about twice the size of those found by Hanushek, Kain, and Rivkin. However, there is substantial variation across diagnosis groups and student ability as defined by third-grade test scores. Students diagnosed with learning disabilities or other health impairments experienced especially large gains, while there is no significant effect for those with autism. My sample includes a smaller proportion of students with learning disabilities and larger proportions with other health impairments (the classification used for ADD and ADHD) and autism. Students who scored in lower achievement levels in third grade saw greater gains in special education than did those who started with higher test scores. Taken together, these suggest that the differences between my estimates and those in previous work are not primarily driven by differences in sample.

[17] I do not have the power necessary to include student fixed effects in my IV analysis.

6.3 Mechanisms

Why would my estimates be so different from those found in previous research? My estimates reflect the local average treatment effect (LATE) of being placed in (or not being placed in) special education as a result of AYP incentives. Recalling Figures 4 and 5, this mostly takes the form of students who had previously received a failing score either leaving special education or never entering it in the first place. Previous research has focused on the average treatment effect (ATE) for students who move in and out of special education or the LATE for students who were placed in special education because they were low-performing and attended schools that faced accountability pressure under a pre-NCLB system.

It is possible that the marginal special education students in my setting are harmed by the stigma, low-expectations, or lower-achieving peer group in a way that the average special education student is not. It could also be that schools were intentional in pushing low-achieving students out of special education, particularly targeting those for whom it was the worst fit. These students might be easier to discourage from special education, and any gains they experienced by not being in special education could have the added benefit of helping the school achieve AYP in other subgroups. It could also be that when schools were forced to serve these students' needs outside of special education they turned to alternatives that were even better, or otherwise changed how they supported student learning. I explore several possibilities below. I use Equation (4) when considering mechanisms operating through school reactions, as I want to capture any potential channels through which incentives alter test scores. When considering student reactions to being placed in special education, I estimate Equation (5).

First, I consider whether schools substituted other supports for special education. Schools could be particularly likely to do so when trying to discourage special education for certain students, so might need to make a case that there is scope to meet the student's needs outside of special education. Supports might include extra time or attention, which is not observable in the data, or grade retention, which is. Students who are held back have another chance to master that grade's content. Recent research suggests that being held back in third grade improves the performance of students who struggle in reading, by 23% of a standard deviation in reading and 30% of a standard deviation in math (Schwerdt, West, & Winters, 2017). Students who are held back are also able to take an easier test – say the 4th grade test rather than the 5th grade test they would have taken if not held back – on which they are compared with younger students. A student taking the 4th grade test rather than the 5th grade test would be expected to perform roughly a half standard deviation better (North Carolina Public Schools, 2009). Taking these two effects together, a retained student would be expected to receive test scores that were roughly 75 percent of a standard deviation better than if they had not been retained.

Table 7 displays the effect of accountability incentives on the likelihood that a student is promoted. I find no evidence that schools respond to selection incentives by changing promotion behavior, although schools are more likely to hold back students who scored in level 3 in reading when the school would benefit from their passing.

Schools could also attempt to change their accountability populations by encouraging or discouraging student movement in and out of schools. This could affect achievement either if students systematically move into better (or worse) schools or because changing schools is generally disruptive. If schools pushed students out (or held on to them) anytime a student was placed in special education this would be part of the policy effect of being in special education; if it only occurred in the presence of NCLB incentives it would be part of the policy effect of being in special education under NCLB. I examine whether previously passing students are more likely to be in a new school when their school faces an incentive to alter the SWD population, with results appearing in Table 8. I find no evidence that schools respond on this margin when trying to improve the performance of the SWD subgroup, although students who previously scored in level 2 in math are more likely to be in a new school when their school would benefit from their passing.

It is also possible that students respond to their placement by changing their level of effort. This could be a reaction to stigma or low expectations, which might be particularly marked for students whose placements are altered by incentives. While many aspects of effort are difficult to observe, I use information on absences to determine whether special education affects one fundamental aspect of effort – attendance. The reading incentives form strong instruments in this sample, but the math incentives do not, as shown in Table 9. However, their pattern of signs and significance is similar to that using the reading incentives as instruments.

I find no evidence that special education increases overall absences or excused absences, but I do find evidence that it increases unexcused absences and instances of being tardy. Attending school less often would be expected to lower achievement; that this effect appears only for unexcused absences and tardiness suggests that it may also reflect a loss of engagement or effort, which would independently lower achievement.

In sum, I do not find evidence that observable changes in schools' other investments in students drive the negative effects of special education on achievement for marginal students. However, it appears that students react to being placed in special education in ways that have negative implications for achievement. Prior work has shown that students in special education have worse attendance and report lower engagement with school and peers (Bear, Clever, and Proctor, 1991, Lackaye and Margalit 2006, Stiefel et al. 2017). My results suggest that these differences are at least in part causal rather than purely correlational, and highlight the need for a better understanding of how to mitigate the negative consequences of special education placement.

7. Conclusion

I examine school responses to AYP incentives to classify particular students as disabled in order to either target resources to students close to the proficiency threshold or to change the composition of the students with disabilities (SWD) subgroup. I use variation across schools in their past performance in the subjects and subgroups relevant to AYP, and across students in their prior scores and subgroup membership in order to isolate school responses.

I find evidence that schools discourage special education classification for students who have previously failed their reading or math test when the school benefits from improving the passing rate for the students with disabilities subgroup. I also find evidence that schools use special education to target resources to students near the passing threshold in reading when the school would benefit from their passing. However, students who just passed in math are unaffected, and those who just failed are less likely to be in special education when their school would benefit from their passing. This likely reflects two factors. First, state accountability rewards schools for the percentage of students passing, without a focus on subgroups or a single high-stakes threshold, so that schools already attempt to improve the scores of almost- and just-passing students regardless of AYP incentives. Second, North Carolina's formula for funding special education incentivizes schools to limit the size of their special education population, so schools may discourage placement for some in order to "make space" for others.

While it is important to understand how schools have responded to policy incentives, it is not clear what those responses mean for students. Either over classification or under classification is at best an inefficiency and at worst an impediment to student learning and development. Without knowing the underlying need for special education services, it is unclear which prevails. It also is possible for the wrong students to be targeted even if neither over classification nor under classification occurs. I find that, for students whose placement is driven by their schools' incentives to alter the SWD population to be higher performing in math, special education is harmful to math achievement. Effects on reading achievement are consistently negative for this group but not significant. For students whose placement is influenced by the school's incentive to improve the performance of the SWD population in reading, there is no significant effect on reading or math scores.

This raises the question of why special education has different effects on these two groups of students. Differences across groups could be driven by differences in either the beneficial or detrimental effects of special education placement, or both. One possibility is that services and supports unique to special education have more scope to improve the performance of students who are low-performing in reading than in math. This seems plausible, as reading performance is generally harder for schools to alter, and the alterations in the NCEXTEND2 might be especially valuable to struggling readers. Another possibility is that schools are better at discouraging placement for students who are low-performing in math but would not be well-served by special education than they are at discouraging placement for similar students who are low-performing in reading. This could either be because such students are more difficult to identify or because schools have less discretion in their placement.

School reactions to selection incentives mostly take the form of discouraging previously low-performing students from entering or remaining in special education. Thus, my results suggest that schools faced with accountability pressure are rationally using special education placement to serve their own goals, with benefits to some of the students affected. While schools that do not face accountability pressure might also benefit from discouraging special education placement for some students, it may be costly to do so. This could be because providing alternative supports is expensive and not defrayed by additional state funding, because identifying who would do better without special education is difficult, or because there are strong pressures to place low-performing students into special education.

My main estimates do not reflect the average benefit of special education for all students who receive it, but rather marginal special education students. These are the students whose placement can reasonably be altered by the action of stakeholders or plausible changes to identification and classification procedures. As such, their experiences are the ones relevant for setting accountability policy. Importantly, while the previous literature supports a policy of providing as much special education as budgets allowed, my finding suggests that placing a student into special education can in fact be harmful to achievement. Thus, it is crucial to target special education services to students who will benefit from them.

These are to my knowledge the first estimates of how schools responded to AYP incentives to alter which students received special education. In comparison with earlier accountability regimes, NCLB appears to have eliminated one method of gaming the system - removing low-achieving students from the accountability population - and replaced it with others – targeting special education to students near the passing threshold in reading and manipulating subgroup composition. Although NCLB is no longer in force, current accountability policies impose similar incentives on schools with a continued emphasis on the percentage of students meeting targets and the use of a students with disabilities subgroup. These facts – that schools can and will manipulate special education placement in the face of NCLB-style incentives and that some students can be hurt by special education placement- are important for policymakers and stakeholders to consider as accountability and special education policies continue to evolve in the future.

References

Angrist, J. & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton, NJ: Princeton University Press.

Bear, G.G., Clever, A., & Proctor, A. (1991). Self-perceptions of non-handicapped children and children with learning disabilities in integrated classrooms. *The Journal of Special Education, 24*(4), 409-426.

Bokhari, F. & Schneider, H. (2011). School accountability laws and the consumption of psychostimulants. *Journal of Health Economics, 30*(2), 355-72.

Byrd-Blake, M., Afolayan, M. O., Hunt, J. W., Fabunmi, M., Pryor, B. W., & Leander, R. (2010). Morale of teachers in high poverty schools: A post-NCLB mixed methods analysis. *Education and Urban Society, 42*, 450–472.

Chakrabarti, R. (2013). Accountability with voucher threats, responses, and the test-taking population: Regression discontinuity evidence from Florida. Federal Reserve Bank of New York Working Paper.

Cohen, J. (2007). Causes and consequences of special education placement: Evidence from Chicago Public Schools. Brookings Institution Working Paper.

Cullen, J. (2003). The impact of fiscal incentives on student disability rates. *Journal of Public Economics, 87*(7-8), 1557-1589.

Cullen, J. & Reback, R. (2006). Tinkering toward accolades: school gaming under a performance accountability system. In T. J. Gronberg and D. W. Jansen (Eds.), *Improving School Accountability (Advances in Applied Microeconomics, Volume 14)* (pp. 1 - 34). Emerald Group Publishing Limited.

Dee, T. S., & Jacob, B. A. (2010). The impact of No Child Left Behind on students, teachers, and schools. Brookings Papers on Economic Activity, 149–194.

Federal Education Budget Project (2014). Background and analysis: Individuals with Disabilities Education Act – Cost impact on local school districts. Retrieved from http://febp.newamerica.net/background-analysis/individuals-disabilities-education-act-cost-impact-local-school-districts

Figlio, D. (2006). Testing, crime, and punishment. *Journal of Public Economics, 90* (2006), 837-851.

Figlio, D. & Getzler, L. (2002). Accountability, ability, and disability: Gaming the system. In T. J. Gronberg and D. W. Jansen (Eds.) *Improving School Accountability (Advances in Applied Microeconomics, Volume 14)* (pp. 35 – 49). Emerald Group Publishing Limited.

Figlio, D. & Winicki, J. (2008). Food for thought: the effect of school accountability plans on school nutrition. *Journal of Public Economics, 89*(2–3), 381-394.

Hanushek, E., Kain, J., & Rivkin, S. (2002). Inferring program effects for special populations: Does special education raise achievement for students with disabilities? *The Review of Economics and Statistics, 84*(4), 584-599.

Hanushek, E. & Raymond, M. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management, 24*(2), 297-327.

Jacob, B. (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.

Kwak, S. (2010). The impact of intergovernmental incentives on student disability rates. *Public Finance Review, 38*(1), 41-73.

Griffith, G., & Scharmann, L. (2008). Initial impacts of No Child Left Behind on elementary science education. *Journal of Elementary Science Education, 20*, 35–48.

Lackaye, T. D. & Margalit, M. (2006). Comparisons of achievement, effort, and self-perceptions among studens with learning disabilities and their peers from different achievement levels. *Journal of Learning Disabilities, 39*(5). 432-446.

Ladd, H. F. & Lauen, D. L. (2010). Status versus growth: The distributional effects of school accountability policies. *Journal of Policy Analysis and Management, 29*, 426–450.

Mahitivanichcha, K. & Parrish, T. (2005). The implications of fiscal incentives on identification rates and placement in special education: Formulas for influencing best practice. *Journal of Education Finance, 31*(1), 1-22.

Morrill, M.S. (2016). Special education financing and ADHD medications: A bitter pill to swallow. Working paper. Retrieved from https://sites.google.com/a/ncsu.edu/msmorrill/files/Morrill_BitterPill.pdf?attredirects=0

Neal, D. & Schanzenbach, D. (2010). Left behind by design: proficiency counts and test-based accountability. *The Review of Economics and Statistics, 92*(2), 263-283.

North Carolina Public Schools (2009). North Carolina Extend2 end-of-grade reading comprehension tests technical report. Edition 2. Retrieved from http://www.ncpublicschools.org/docs/accountability/testing/reports/extend2readingtechmanual.pdf

U.S. Department of Education, Office of Special Education and Rehabilitative Services (2010). A guide to the Individualized Education Program. Jessup, MD: ED Pubs.

U.S. Department of Health and Human Services, Office for Civil Rights (2006). Your rights under Section 504 of the Rehabilitation Act. Retrieved from https://archive.hhs.gov/ocr/504.pdf

Reback, R. (2008). Teaching to the rating: school accountability and the distribution of student achievement. *Journal of Public Economics, 92*(5-6), 352-353.

Schwerdt, G., West, M. R., Winters, M.A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics, 152,* 154-169.

Setren, E. (2016). Special education and English language learner students in Boston charter schools: Impact and classification. Working Paper. Retrieved from http://economics.mit.edu/grad/esetren/research

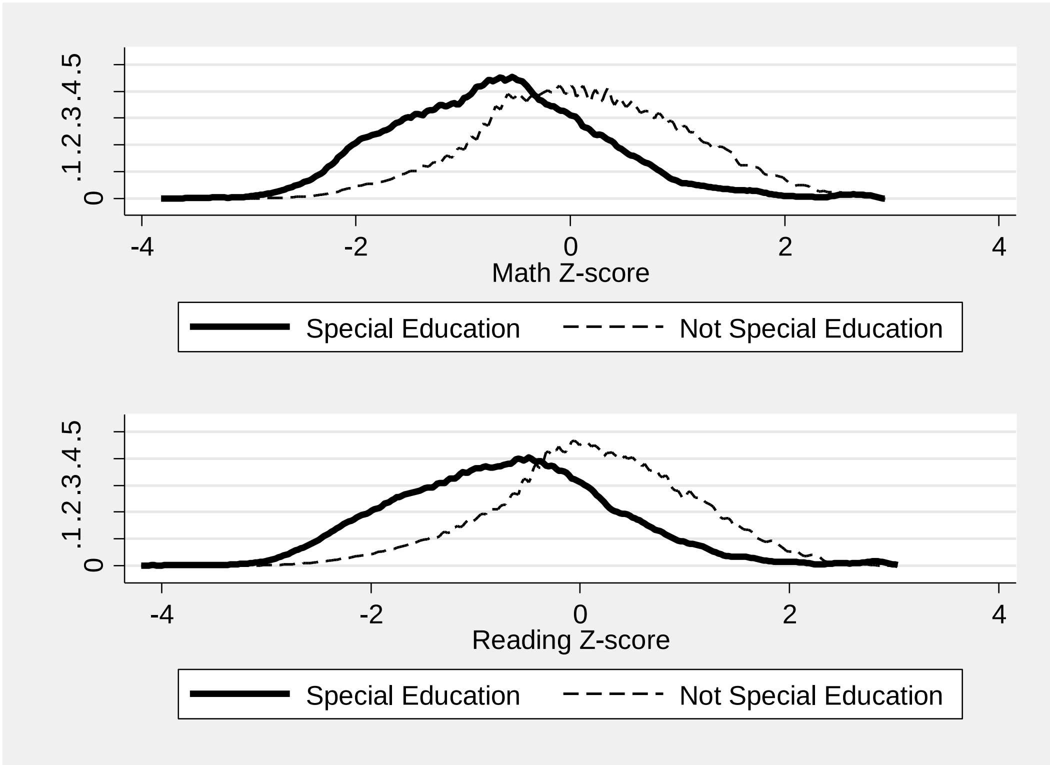
Stiefel, L., Shiferaw, M., Schwartz, A. E., and Gottfried, M. (2017). Is special education improving? Evidence on segregation, outcomes, and spending from New York City. IESP Working Paper No. 02-01.

U.S. Department of Education, National Center for Education Statistics. (2013). Digest of Education Statistics, 2012 (NCES 2014-015), Table 48.

U.S. Department of Education, National Center for Education Statistics. (2016). Digest of Education Statistics, 2014 (NCES 2016-006), Table 236.60.

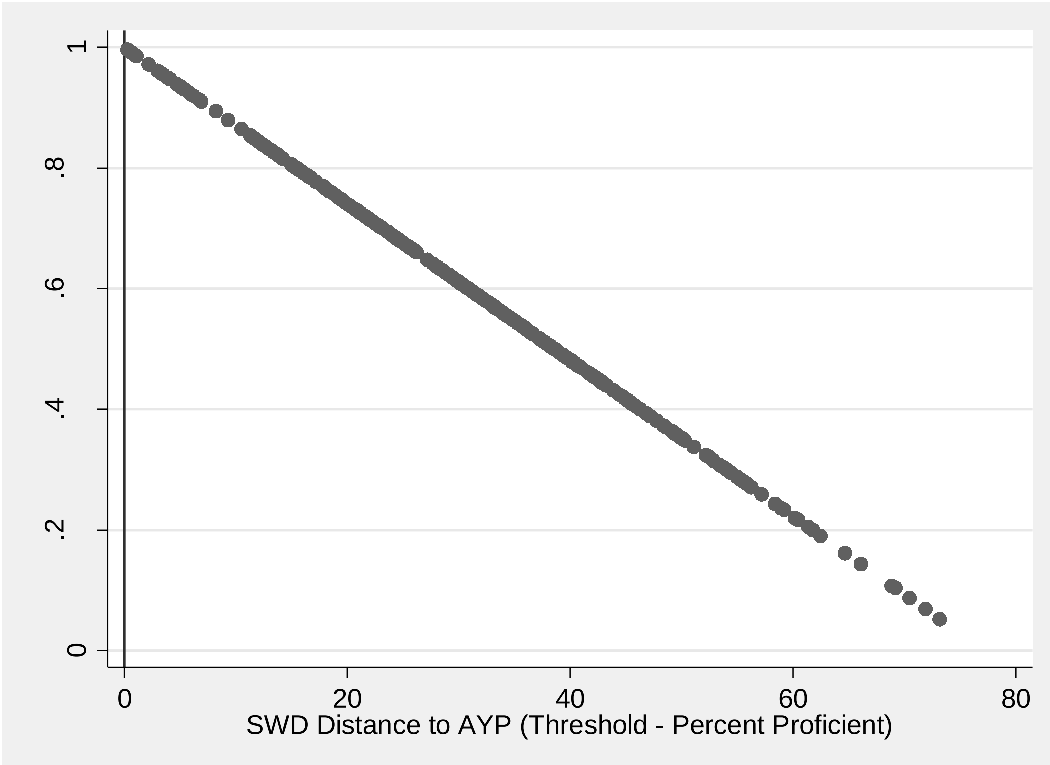
Winters, M. & Greene, J. (2011). Public school response to special education vouchers: The impact of Florida's McKay scholarship program on disability diagnosis and student achievement in public schools. *Educational Evaluation and Policy Analysis, 33*(2), 138-158.

Figure 1. Math and Reading Scores by Special Education Status



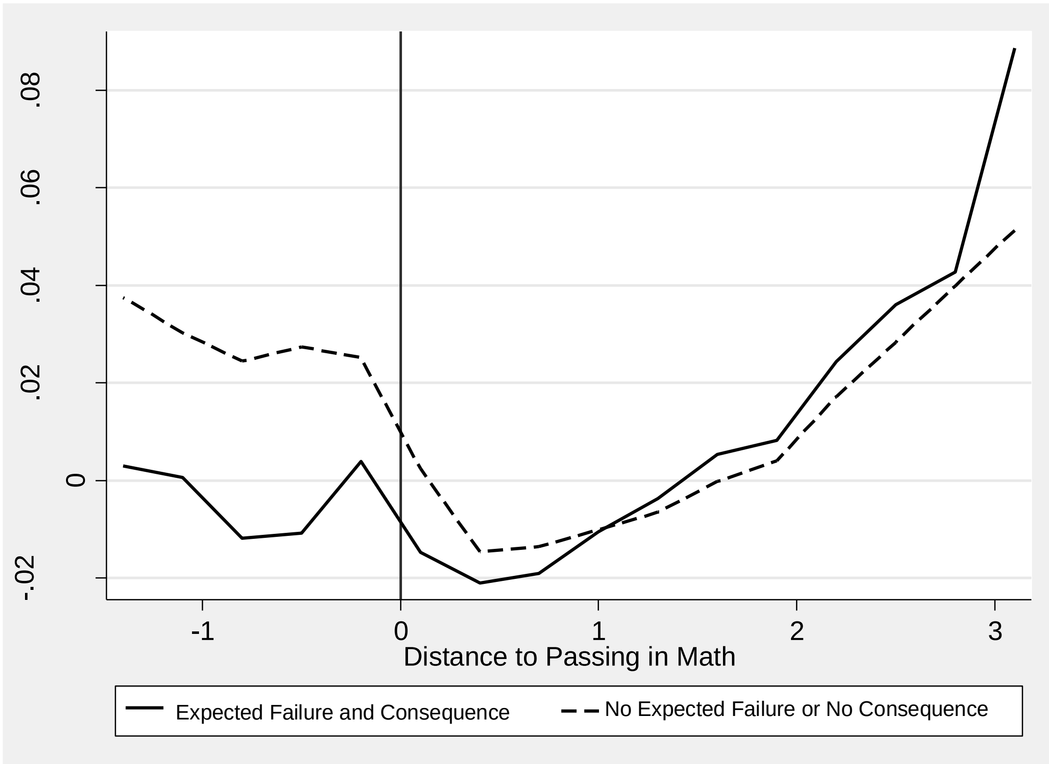
Notes: Figure shows the distribution of math and reading scores, standardized to have mean zero and standard deviation 1 for each grade-year-subject grouping, graphed separately by special education status.

Figure 2. SWD Incentive Instrument



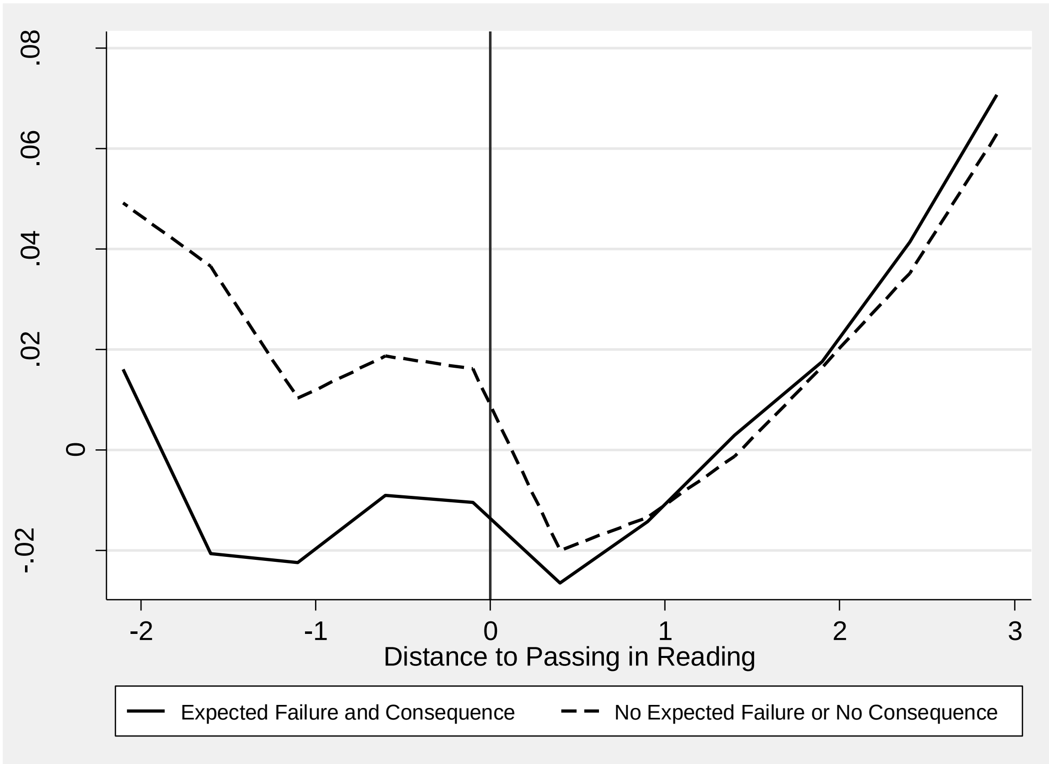
Notes: This figure depicts the SWD incentive instrument as described in Equation (3).

Figure 3. Residual Percent with Malleable Diagnosis by Distance to Passing Score in Math and Accountability Incentives



Notes: This figure displays the residual percent of students with a malleable diagnosis, after controlling for prior score and demographics, by the student's distance from the passing threshold measured in standard deviations. The first series, marked with the solid line, includes students whose school expected to fail AYP for at least one group of which the student was a member, and would potentially face consequences for doing so. The second series, marked with the dashed line, includes all other students.

Figure 4. Residual Percent with Malleable Diagnosis by Distance to Passing Score in Reading and Accountability Incentives



Notes: This figure displays the residual percent of students with a malleable diagnosis, after controlling for prior score and demographics, by the student's distance from the passing threshold measured in standard deviations. The first series, marked with the solid line, includes students whose school expected to fail AYP for at least one group of which the student was a member, and would potentially face consequences for doing so. The second series, marked with the dashed line, includes all other students.

Table 1. Distribution of Diagnoses in Special Education

	Percent of Students	Percent of Special Education Students
Autism	0.7	4.9
Deaf-Blindness	0.0	0.0
Developmental Delay	0.0	0.0
Emotional Disturbance	0.6	4.7
Hearing Impairment	0.1	1.1
Intellectual Disability	1.9	14.1
Multiple Disabilities	0.1	1.1
Orthopedic Impairment	0.1	0.5
Other Health Impairment	2.6	19.4
Specific Learning Disability	5.9	44.8
Speech or Language Impairment	1.2	8.8
Traumatic Brain Injury	0.0	0.2
Visual Impairment	0.0	0.3
Total	13.2	100

Notes: Table reports the percent in special education by diagnosis for the sample of students in North Carolina Title I schools in grades 4-8 in years 2006-7 – 20010-11. Diagnoses shaded in grey (emotional disturbance, other health impairment, specific learning disability, speech or language impairment) are included in the "malleable impairment" category.

Table 2. Descriptive Statistics for the Main Analysis Sample

	All		Not Special Education		Special Education	
	Mean	SD	Mean	SD	Mean	SD
Special Education	0.105	0.306				
Prior Pass Reading	0.692	0.462	0.730	0.444	0.364	0.481
Prior Pass Math	0.732	0.443	0.764	0.425	0.463	0.499
Native American	0.023	0.151	0.023	0.150	0.024	0.154
Asian	0.017	0.130	0.018	0.135	0.007	0.085
Hispanic	0.124	0.329	0.126	0.332	0.104	0.306
Black	0.311	0.463	0.307	0.461	0.351	0.477
White	0.488	0.500	0.489	0.500	0.476	0.499
Other	0.037	0.188	0.037	0.188	0.037	0.190
Female	0.497	0.500	0.516	0.500	0.331	0.470
Free or Reduced-Price Lunch	0.570	0.495	0.556	0.497	0.691	0.462
School failed in math	0.221	0.415	0.220	0.414	0.222	0.416
School failed in reading	0.184	0.387	0.184	0.388	0.184	0.387
Math Score	-0.100	0.946	-0.024	0.923	-0.743	0.889
Prior math score	-0.109	0.948	-0.030	0.923	-0.779	0.894
Reading Score	-0.105	0.962	-0.022	0.924	-0.814	0.979
Prior reading score	-0.107	0.965	-0.015	0.922	-0.896	0.959
N	1,298,002		1,161,922		136,080	

Notes: This table presents descriptive information on the main analysis sample as described in the text. Current and prior scores are in standard deviation units.

Table 3. Effect of Accountability Incentives on Special Education Placement

	(1)	(2)	(3)	(4)
Reading Level 2*Incentive	0.0142**		0.00930**	
	(0.00302)		(0.00283)	
Math Level 2*Incentive	-0.0115**		-0.0133**	
	(0.00353)		(0.00300)	
Reading Level 3*Incentive	0.0252**		0.0239**	
	(0.00285)		(0.00272)	
Math Level 3*Incentive	-0.000835		-0.000850	
	(0.00193)		(0.00222)	
Reading Level 2*Distance*Incentive		0.0143**		0.0101**
		(0.00319)		(0.00299)
Math Level 2*Distance*Incentive		-0.0139**		-0.0140**
		(0.00388)		(0.00335)
Reading Level 3*Distance*Incentive		0.0277**		0.0284**
		(0.00323)		(0.00312)
Math Level 3*Distance*Incentive		-0.00352		-0.000355
		(0.00233)		(0.00277)
Reading Prior Pass*SWD Incentive	0.0221**	0.0223**		
	(0.00463)	(0.00469)		
Math Prior Pass*SWD Incentive	0.0125	0.0139		
	(0.00726)	(0.00756)		
Reading Prior Score*SWD Incentive			0.0110**	0.0115**
			(0.00254)	(0.00258)
Math Prior Score*SWD Incentive			0.00682**	0.00706**
			(0.00262)	(0.00273)
N	1199737	1199737	1199737	1199737

Notes: This table displays results from 4 linear probability models following Equation (4), one in each column, in which being in special education with a malleable diagnosis is the dependent variable. Each model includes demographic controls, year-by-grade fixed effects, school-level fixed effects, and the main effects of school incentives and prior student performance. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table 4. Effect of Accountability Incentives on Having a Non-Malleable Diagnosis

	(1)	(2)	(3)	(4)
Reading Level 2*Incentive	0.00113		0.000556	
	(0.00184)		(0.00178)	
Math Level 2*Incentive	0.00035		-0.00121	
	(0.00222)		(0.00202)	
Reading Level 3*Incentive	-0.00205		-0.00390*	
	(0.00176)		(0.00170)	
Math Level 3*Incentive	-0.00053		-0.00365*	
	(0.00153)		(0.00169)	
Reading Level 2*Distance*Incentive		0.000717		-0.0000361
		(0.00194)		(0.00189)
Math Level 2*Distance*Incentive		-0.00099		-0.000853
		(0.00238)		(0.00221)
Reading Level 3*Distance*Incentive		-0.00282		-0.00476*
		(0.00200)		(0.00193)
Math Level 3*Distance*Incentive		-0.00064		-0.00282
		(0.00175)		(0.00201)
Reading Prior Pass*SWD Incentive	-0.00039	-0.00041		
	(0.00257)	(0.00259)		
Math Prior Pass*SWD Incentive	-0.00357	-0.00475		
	(0.00392)	(0.00399)		
Reading Prior Score*SWD Incentive			0.00253	0.00261
			(0.00147)	(0.00197)
Math Prior Score*SWD Incentive			-0.00184	-0.00189
			(0.00156)	(0.00162)
N	1237846	1237846	1237846	1237846

Notes: This table displays results from 4 linear probability models following Equation (4), one in each column, in which being in special education with a non-malleable diagnosis is the dependent variable. Each model includes demographic controls, year-by-grade fixed effects, school-level fixed effects, and the main effects of school incentives and prior student performance. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table 5 Effect of Accountability Incentives on Not Having a Valid Score

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Reading	Reading	Reading	Reading	Math	Math	Math	Math
Reading Level 2*Incentive	0.00157		0.00217		0.00150		0.00212	
	(0.00418)		(0.00416)		(0.00419)		(0.00417)	
Math Level 2*Incentive	0.00535		0.00368		0.00542		0.00373	
	(0.00398)		(0.00353)		(0.00398)		(0.00352)	
Reading Level 3*Incentive	-0.00221		-0.00368		-0.00230		-0.00379	
	(0.00391)		(0.00382)		(0.00391)		(0.00383)	
Math Level 3*Incentive	-0.00194		-0.000234		-0.00190		-0.000180	
	(0.00276)		(0.00276)		(0.00276)		(0.00275)	
Reading Level 2*Distance*Incentive		-0.00234		-0.00398		-0.00239		-0.00406
		(0.00447)		(0.00436)		(0.00448)		(0.00436)
Math Level 2*Distance*Incentive		-0.00173		0.000448		-0.00170		0.000519
		(0.00321)		(0.00324)		(0.00320)		(0.00323)
Reading Level 3*Distance*Incentive		0.00215		0.00281		0.00212		0.00280
		(0.00466)		(0.00464)		(0.00466)		(0.00465)
Math Level 3*Distance*Incentive		0.00662		0.00483		0.00673		0.00491
		(0.00438)		(0.00393)		(0.00437)		(0.00392)
Reading Prior Pass*SWD Incentive	-0.00548	-0.00550			-0.00561	-0.00564		
	(0.00448)	(0.00451)			(0.00448)	(0.00451)		
Math Prior Pass*SWD Incentive	0.00947	0.00950			0.00960	0.00967		
	(0.00653)	(0.00646)			(0.00653)	(0.00647)		
Reading Prior Score*SWD Incentive			0.000142	0.0000902			0.0000824	0.0000268
			(0.00226)	(0.00226)			(0.00226)	(0.00226)
Math Prior Score*SWD Incentive			0.00156	0.00161			0.00160	0.00166
			(0.00227)	(0.00228)			(0.00227)	(0.00229)
N	1199737	1199737	1199737	1199737	1199737	1199737	1199737	1199737

Notes: This table displays results from 8 linear probability models following Equation (4), one in each column, in which not having a valid reading or math score is the dependent variable. Each model includes demographic controls, year-by-grade fixed effects, school-level fixed effects, and the main effects of school incentives and prior student performance. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table 6. Effect of Special Education on Student Achievement

Panel 1: Reading score, reading instrument				
	(1)	(2)	(3)	(4)
Special Education	0.275	0.577	0.291	0.581
	(0.389)	(0.344)	(0.386)	(0.329)
F-Statistic	25.35	24.78	25.39	26.03
N	1199505	1199505	1199505	1199505
Panel 2: Math score, reading instrument				
Special Education	0.0451	0.171	0.0107	0.249
	(0.343)	(0.320)	(0.334)	(0.309)
F-Statistic	25.33	24.77	25.37	26.02
N	1199496	1199496	1199496	1199496
Panel 3: Reading score, math instrument				
Special Education	-1.827	-0.770	-1.714	-0.685
	(1.034)	(0.484)	(0.947)	(0.470)
F-Statistic	5.778	14.28	6.388	14.09
Panel 4: Math score, math instrument				
Special Education	-1.501	-1.289*	-1.623	-1.129*
	(0.941)	(0.580)	(0.917)	(0.555)
F-Statistic	5.777	14.28	6.388	14.09
Instruments:				
Prior Pass*SWD Incentive	Y		Y	
Prior Score*SWD Incentive		Y		Y
Controls:				
Levels*Incentive	Y	Y		
Levels*Distance*Incentive			Y	Y

Notes: This table displays results from 16 linear IV models following Equation (5), one in each column and panel, in which being in special education with a malleable diagnosis is instrumented by selection incentives as noted and math or reading z-score is the dependent variable. All specifications include demographic controls, year-by-grade fixed effects, and school-level fixed effects, and I control for hypothesis 1 incentives as noted. Standard errors in parentheses are clustered at the school level. Kleibergen-Papp F-statistics from the first stage are reported. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table 7. Effect of Accountability Incentives on Being Promoted to the Next Grade

	(1)	(2)	(3)	(4)
Reading Level 2*Incentive	-0.000228		0.000139	
	(0.00155)		(0.00149)	
Math Level 2*Incentive	0.000468		0.00276	
	(0.00204)		(0.00166)	
Reading Level 3*Incentive	-0.00374*		-0.00345*	
	(0.00173)		(0.00170)	
Math Level 3*Incentive	-0.000745		0.00276	
	(0.00128)		(0.00150)	
Reading Level 2*Distance*Incentive		-0.000462		-0.000183
		(0.00165)		(0.00159)
Math Level 2*Distance*Incentive		0.00125		0.00257
		(0.00227)		(0.00188)
Reading Level 3*Distance*Incentive		-0.00437*		-0.00446*
		(0.00204)		(0.00200)
Math Level 3*Distance*Incentive		-0.000961		0.00227
		(0.00154)		(0.00192)
Reading Prior Pass*SWD Incentive	-0.00231	-0.00255		
	(0.00282)	(0.00286)		
Math Prior Pass*SWD Incentive	0.00315	0.00383		
	(0.00488)	(0.00505)		
Reading Prior Score*SWD Incentive			0.00173	0.00179
			(0.00196)	(0.00195)
Math Prior Score*SWD Incentive			-0.000277	-0.000177
			(0.00307)	(0.00303)
N	1199720	1199720	1199720	1199720

Notes: This table displays results from 4 linear probability models following Equation (4), one in each column, in which being promoted to the next grade is the dependent variable. Each model includes demographic controls, year-by-grade fixed effects, school-level fixed effects, and the main effects of school incentives and prior student performance. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table 8. Effect of Accountability Incentives on Changing Schools

	(1)	(2)	(3)	(4)
Reading Level 2*Incentive	-0.000313		-0.000997	
	(0.00280)		(0.00280)	
Math Level 2*Incentive	-0.00629*		-0.00465	
	(0.00291)		(0.00284)	
Reading Level 3*Incentive	0.00302		0.00115	
	(0.00360)		(0.00351)	
Math Level 3*Incentive	0.00424		0.00386	
	(0.00285)		(0.00205)	
Reading Level 2*Distance*Incentive		-0.000902		-0.00159
		(0.00305)		(0.00308)
Math Level 2*Distance*Incentive		-0.00793*		-0.00556
		(0.00319)		(0.00327)
Reading Level 3*Distance*Incentive		0.00235		0.000530
		(0.00402)		(0.00389)
Math Level 3*Distance*Incentive		0.00476		0.00451
		(0.00368)		(0.00262)
Reading Prior Pass*SWD Incentive	-0.00146	-0.00107		
	(0.00414)	(0.00417)		
Math Prior Pass*SWD Incentive	-0.00372	-0.00453		
	(0.00915)	(0.00929)		
Reading Prior Score*SWD Incentive			0.00173	0.00179
			(0.00196)	(0.00195)
Math Prior Score*SWD Incentive			-0.000277	-0.000177
			(0.00307)	(0.00303)
N	1199737	1199737	1199737	1199737

Notes: This table displays results from 4 linear probability models following Equation (4), one in each column, in which being in a new school in the current year is the dependent variable. Each model includes demographic controls, year-by-grade fixed effects, school-level fixed effects, and the main effects of school incentives and prior student performance. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table 9. Effect of Special Education on Attendance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel 1. Percent of Days Absent								
	-0.0138	3.023	-0.0115	-0.0198	0.140	0.0832	0.146	0.0882
	(0.0352)	(0.0355)	(0.0344)	(0.0338)	(0.103)	(0.0651)	(0.106)	(0.0659)
Panel 2. Percent of Days Excused Absence								
	-0.00404	-0.0109	-0.00448	-0.0109	0.0462	0.0224	0.0417	0.0209
	(.0107)	(0.0115)	(0.0106)	(0.0111)	(0.0300)	(0.0173)	(0.0281)	(0.0174)
Panel 3. Percent of Days Unexcused Absence								
	0.0203	0.0286*	0.0197	0.0267*	0.0903	0.0619*	0.0859	0.0591*
	(0.0118)	(0.0135)	(0.0117)	(0.0128)	(0.0472)	(0.0271)	(0.0454)	(0.0271)
Panel 4. Times Tardy per Day Enrolled								
	0.0190*	0.0238*	0.0191*	0.0229*	0.0157	0.0479	0.0493	0.0475
	(0.00926)	(0.0115)	(0.00940)	(0.0111)	(0.0311)	(0.0266)	(0.0328)	(0.0269)
F-statistic	19.34	18.04	19.56	19.22	4.89	9.881	4.927	9.606
N	928511	928511	928511	928511	928511	928511	928511	928511
Instruments:								
	Reading	Reading	Reading	Reading	Math	Math	Math	Math
Prior Pass * SWD Incentive	Y		Y		Y		Y	
Prior Score * SWD Incentive		Y		Y		Y		Y
Controls:								
Levels * Incentive	Y	Y			Y	Y		
Levels * Distance*Incentive			Y	Y			Y	Y

Notes: This table displays results from 24 linear IV models following Equation (5), one in each column and panel, in which being in special education with a malleable diagnosis is instrumented by selection incentives as noted. All specifications include demographic controls, year-by-grade fixed effects, and school-level fixed effects, and I control for hypothesis 1 incentives as noted. Standard errors in parentheses are clustered at the school level. Kleibergen-Papp F-statistics from the first stage are reported. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table A.1 Effect of Accountability Incentives on Special Education Placement by Percentage in Special Education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reading Level 2*Incentive	0.0113**		0.00530		0.0179**		0.0142**	
	(0.00420)		(0.00387)		(0.00412)		(0.00389)	
Math Level 2*Incentive	-0.0156**		-0.0163**		-0.00755		-0.0104**	
	(0.00486)		(0.00415)		(0.00471)		(0.00393)	
Reading Level 3*Incentive	0.0228**		0.0223**		0.0312**		0.0290**	
	(0.00378)		(0.00364)		(0.00423)		(0.00392)	
Math Level 3*Incentive	-0.00232		-0.00213		-0.00117		0.000735	
	(0.00272)		(0.00314)		(0.00242)		(0.00283)	
Reading Level 2*Distance*Incentive		0.0111*		0.00567		0.0187**		0.0159**
		(0.00443)		(0.00409)		(0.00437)		(0.00414)
Math Level 2*Distance*Incentive		-0.0186**		-0.0179**		-0.00914		-0.0103*
		(0.00533)		(0.00459)		(0.00522)		(0.00443)
Reading Level 3*Distance*Incentive		0.0244**		0.0261**		0.0354**		0.0349**
		(0.00427)		(0.00416)		(0.00475)		(0.00450)
Math Level 3*Distance*Incentive		-0.00579		-0.00237		-0.00392		0.00182
		(0.00327)		(0.00387)		(0.00290)		(0.00361)
Reading Prior Pass*SWD Incentive	0.00882	0.00860			0.0132*	0.0128*		
	(0.00573)	(0.00575)			(0.00573)	(0.00574)		
Math Prior Pass*SWD Incentive	0.0157	0.0174			0.0230*	0.0249*		
	(0.00955)	(0.00982)			(0.00970)	(0.01000)		
Reading Prior Score*SWD Incentive			0.00329	0.00361			0.00551	0.00585
			(0.00308)	(0.00312)			(0.00312)	(0.00316)
Math Prior Score*SWD Incentive			0.00947**	0.0100**			0.0113**	0.0119**
			(0.00315)	(0.00329)			(0.00323)	(0.00338)
N	625785	625785	625785	625785	573952	573952	573952	573952
Sample	>= 12.5	>= 12.5	>= 12.5	>= 12.5	<12.5	<12.5	<12.5	<12.5

Notes: This table displays results from 8 linear probability models following Equation (4), one in each column, in which being in special education with a malleable diagnosis is the dependent variable. Each model includes demographic controls, year-by-grade fixed effects, school-level fixed effects, and the main effects of school incentives and prior student performance. The sample is divided by whether the school had more or less than 12.5 percent of its student body in special education. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table A.2 Effect of Accountability Incentives on Malleable Diagnoses, Including Those with Non-Malleable Diagnoses in Sample

	(1)	(2)	(3)	(4)
Reading Level 2*Incentive	0.0115**		0.00685**	
	(0.00274)		(0.00259)	
Math Level 2*Incentive	-0.0127**		-0.0139**	
	(0.00320)		(0.00279)	
Reading Level 3*Incentive	0.0220**		0.0204**	
	(0.00266)		(0.00253)	
Math Level 3*Incentive	-0.00289		-0.00210	
	(0.00183)		(0.00203)	
Reading Level 2*Distance*Incentive		0.0115**		0.00744**
		(0.00290)		(0.00274)
Math Level 2*Distance*Incentive		-0.0146**		-0.0150**
		(0.00351)		(0.00308)
Reading Level 3*Distance*Incentive		0.0241**		0.0241**
		(0.00300)		(0.00290)
Math Level 3*Distance*Incentive		-0.00570**		-0.00247
		(0.00220)		(0.00250)
Reading Prior Pass*SWD Incentive	0.0193**	0.0194**		
	(0.00438)	(0.00442)		
Math Prior Pass*SWD Incentive	0.0123	0.0140*		
	(0.00638)	(0.00661)		
Reading Prior Score*SWD Incentive			0.00964**	0.0101**
			(0.00235)	(0.00238)
Math Prior Score*SWD Incentive			0.00652**	0.00661**
			(0.00233)	(0.00240)
N	1237846	1237846	1237846	1237846

Notes: This table displays results from 4 linear probability models following Equation (4), one in each column, in which being in special education with a malleable diagnosis is the dependent variable. Each model includes demographic controls, year-by-grade fixed effects, school-level fixed effects, and the main effects of school incentives and prior student performance. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table A.3 Descriptive Statistics for Students with Non-Malleable Diagnoses

	Mean	SD
Special Education	1	0
Prior Pass Reading	0.351838	0.47755
Prior Pass Math	0.399542	0.489811
Native American	0.03758	0.19018
Asian	0.008678	0.092753
Hispanic	0.075343	0.263948
Black	0.445011	0.496974
White	0.406301	0.491149
Other	0.027087	0.162338
Female	0.345264	0.475461
Free or Reduced-Price Lunch	0.714748	0.451541
School failed in math	0.281229	0.449605
School failed in reading	0.231421	0.421746
Math Score	-0.83819	1.111153
Prior math score	-0.87208	1.185548
Reading Score	-0.70621	1.198559
Prior reading score	-0.79773	1.21959
N	38,026	

Notes: Table presents descriptive statistics for the sample of students with non-malleable diagnoses. These students are excluded from the main analysis sample. Current and prior test scores are presented in standard deviation units.

Table A.4 Effect of Accountability Incentives on Special Education Placement – Non-Title I Schools

	(1)	(2)	(3)	(4)
Reading Level 2*Incentive	-0.0492**		-0.0418**	
	(0.0156)		(0.0130)	
Math Level 2*Incentive	-0.0149		-0.0207**	
	(0.00765)		(0.00655)	
Reading Level 3*Incentive	-0.00528		-0.0156	
	(0.00643)		(0.00866)	
Math Level 3*Incentive	0.00594		0.00508	
	(0.00531)		(0.00463)	
Reading Level 2*Distance*Incentive		-0.0566**		-0.0490**
		(0.0166)		(0.0138)
Math Level 2*Distance*Incentive		-0.0182*		-0.0232**
		(0.00893)		(0.00765)
Reading Level 3*Distance*Incentive		-0.00867		-0.0205*
		(0.00730)		(0.0100)
Math Level 3*Distance*Incentive		0.00338		0.00585
		(0.00598)		(0.00552)
Reading Prior Pass*SWD Incentive	-0.0556*	-0.0547*		
	(0.0245)	(0.0246)		
Math Prior Pass*SWD Incentive	0.0289	0.0323		
	(0.0198)	(0.0202)		
Reading Prior Score*SWD Incentive			-0.0172	-0.0177*
			(0.00884)	(0.00893)
Math Prior Score*SWD Incentive			0.0156*	0.0164*
			(0.00767)	(0.00790)
N	1047084	1047084	1047084	1047084

Notes: This table displays results from 4 linear probability models following Equation (4), one in each column, in which being in special education with a malleable diagnosis is the dependent variable. Each model includes demographic controls, year-by-grade fixed effects, school-level fixed effects, and the main effects of school incentives and prior student performance. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table A.5 Effects of Special Education on Gains in Math Scores, Hanushek, Kain, & Rivkin model

	No Student FE	Student FE	N
	(1)	(2)	
All	0.0261**	0.0663**	2,311,728
	(0.0028)	(0.0068)	
Diagnosis			
Learning Disabled	0.0635**	0.0930**	157,442
	(0.0052)	(0.0107)	
Speech/Language	0.0272**	0.0280*	70,901
	(0.0051)	(0.0126)	
Emotional/Behavioral	0.0923**	0.1030	14,560
	(0.0213)	(0.0527)	
Other Health (ADHD)	0.0818**	0.1065**	72,986
	(0.0087)	(0.0192)	
Autism	0.0429	0.0865	17,071
	(0.0323)	(0.0801)	
Previous Achievement Level			
1	0.0563**	0.1991**	68,653
	(0.0061)	(0.0223)	
2	-0.0027	0.1461**	248,366
	(0.0035)	(0.0129)	
3	-0.0717**	0.0723**	673,307
	(0.0028)	(0.0094)	
4	-0.1672**	0.0297	320,893
	(0.0081)	(0.0163)	

Notes: This table presents the results of 20 models following Equation (6), in which the gain in math z-score is the dependent variable. Standard errors in parentheses are clustered at the school level. A * denotes significance at the 0.05 level and **denotes significance at the 0.01 level.

Table A.6 Effect of Accountability Incentives on Special Education Placement – Alternate Malleable Definition

	(1)	(2)	(3)	(4)
Reading Level 2*Incentive	0.0167**		0.0108**	
	(0.00337)		(0.00312)	
Math Level 2*Incentive	-0.0118**		-0.0144**	
	(0.00397)		(0.00332)	
Reading Level 3*Incentive	0.0281**		0.0264**	
	(0.00328)		(0.00313)	
Math Level 3*Incentive	0.000585		0.000593	
	(0.00220)		(0.00247)	
Reading Level 2*Distance*Incentive		0.0165**		0.0115**
		(0.00357)		(0.00331)
Math Level 2*Distance*Incentive		-0.0147**		-0.0151**
		(0.00438)		(0.00372)
Reading Level 3*Distance*Incentive		0.0306**		0.0312**
		(0.00372)		(0.00360)
Math Level 3*Distance*Incentive		-0.00250		0.00151
		(0.00264)		(0.00310)
Reading Prior Pass*SWD Incentive	0.0261**	0.0262**		
	(0.00543)	(0.00555)		
Math Prior Pass*SWD Incentive	0.0163	0.0177*		
	(0.00854)	(0.00888)		
Reading Prior Score*SWD Incentive			0.0127**	0.0133**
			(0.00307)	(0.00311)
Math Prior Score*SWD Incentive			0.00841**	0.00873**
			(0.00308)	(0.00321)
N	1199737	1199737	1199737	1199737

Notes: This table displays results from 4 linear probability models following Equation (4), one in each column, in which being in special education with a malleable or autism diagnosis is the dependent variable. Each model includes demographic controls, year-by-grade fixed effects, school-level fixed effects, and the main effects of school incentives and prior student performance. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table B.1 Effect of Accountability Incentives on Special Education Placement, No School Fixed Effects

	(1)	(2)	(3)	(4)
Reading Level 2*Incentive	0.0142**		0.00973**	
	(0.00300)		(0.00281)	
Math Level 2*Incentive	-0.0114**		-0.0125**	
	(0.00349)		(0.00296)	
Reading Level 3*Incentive	0.0263**		0.0245**	
	(0.00291)		(0.00279)	
Math Level 3*Incentive	0.00173		-0.000747	
	(0.00199)		(0.00223)	
Reading Level 2*Distance*Incentive		0.0141**		0.0104**
		(0.00317)		(0.00296)
Math Level 2*Distance*Incentive		-0.0139**		-0.0132**
		(0.00382)		(0.00328)
Reading Level 3*Distance*Incentive		0.0289**		0.0289**
		(0.00327)		(0.00317)
Math Level 3*Distance*Incentive		-0.000291		-0.000314
		(0.00237)		(0.00277)
Reading Prior Pass*SWD Incentive	0.0156**	0.0158**		
	(0.00462)	(0.00469)		
Math Prior Pass*SWD Incentive	0.00192	0.00306		
	(0.00741)	(0.00771)		
Reading Prior Score*SWD Incentive			0.00773**	0.00829**
			(0.00249)	(0.00253)
Math Prior Score*SWD Incentive			0.00308	0.00329
			(0.00262)	(0.00273)
N	1199737	1199737	1199737	1199737

Notes: This table displays results from 4 linear probability models following Equation (4), one in each column, in which being in special education with a malleable diagnosis is the dependent variable. Each model includes demographic controls, year-by-grade fixed effects, and the main effects of school incentives and prior student performance. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table B.2 Effect of Accountability Incentives on Having a Non-Malleable Diagnosis, No School Fixed Effects

	(1)	(2)	(3)	(4)
Reading Level 2*Incentive	0.00180		0.0000917	
	(0.00187)		(0.00180)	
Math Level 2*Incentive	0.000895		-0.00193	
	(0.00232)		(0.00206)	
Reading Level 3*Incentive	0.00343*		0.000614	
	(0.00171)		(0.00165)	
Math Level 3*Incentive	0.0000975		-0.00236	
	(0.00161)		(0.00176)	
Reading Level 2*Distance*Incentive		0.00103		-0.000804
		(0.00198)		(0.00191)
Math Level 2*Distance*Incentive		-0.00125		-0.00189
		(0.00249)		(0.00229)
Reading Level 3*Distance*Incentive		0.00278		0.000169
		(0.00194)		(0.00189)
Math Level 3*Distance*Incentive		-0.000725		-0.00129
		(0.00182)		(0.00211)
Reading Prior Pass*SWD Incentive	0.00294	0.00301		
	(0.00252)	(0.00254)		
Math Prior Pass*SWD Incentive	0.00406	0.00285		
	(0.00434)	(0.00443)		
Reading Prior Score*SWD Incentive			0.00459**	0.00477**
			(0.00143)	(0.00144)
Math Prior Score*SWD Incentive			0.001000	0.000966
			(0.00168)	(0.00174)
N	1237846	1237846	1237846	1237846

Notes: This table displays results from 4 linear probability models following Equation (4), one in each column, in which being in special education with a non-malleable diagnosis is the dependent variable. Each model includes demographic controls, year-by-grade fixed effects, and the main effects of school incentives and prior student performance. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table B.3 Effect of Accountability Incentives on Special Education Placement – Non-Title I Schools, No School Fixed Effects

	(1)	(2)	(3)	(4)
Reading Level 2*Incentive	-0.0490**		-0.0375**	
	(0.0162)		(0.0122)	
Math Level 2*Incentive	-0.0155		-0.0203**	
	(0.00793)		(0.00666)	
Reading Level 3*Incentive	-0.00139		-0.0151	
	(0.00605)		(0.00889)	
Math Level 3*Incentive	0.00906		0.00627	
	(0.00524)		(0.00446)	
Reading Level 2*Distance*Incentive		-0.0568**		-0.0443**
		(0.0172)		(0.0129)
Math Level 2*Distance*Incentive		-0.0199*		-0.0235**
		(0.00924)		(0.00777)
Reading Level 3*Distance*Incentive		-0.00349		-0.0194
		(0.00684)		(0.0101)
Math Level 3*Distance*Incentive		0.00724		0.00689
		(0.00595)		(0.00546)
Reading Prior Pass*SWD Incentive	-0.0729**	-0.0730**		
	(0.0271)	(0.0273)		
Math Prior Pass*SWD Incentive	0.0149	0.0169		
	(0.0208)	(0.0209)		
Reading Prior Score*SWD Incentive			-0.0179*	-0.0184*
			(0.00798)	(0.00810)
Math Prior Score*SWD Incentive			0.00886	0.00937
			(0.00768)	(0.00790)
N	1047084	1047084	1047084	1047084

Notes: This table displays results from 4 linear probability models following Equation (4), one in each column, in which being in special education with a malleable diagnosis is the dependent variable. Each model includes demographic controls, year-by-grade fixed effects, and the main effects of school incentives and prior student performance. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table B.4 Effect of Special Education on Student Achievement, No School Fixed Effects

Panel 1: Reading score, reading instrument				
	(1)	(2)	(3)	(4)
Special Education	-1.050	-0.653	-1.004	-0.573
	(0.685)	(0.535)	(0.673)	(0.494)
F-Statistic	11.69	11.86	11.85	13.00
N	1199506	1199506	1199506	1199506
Panel 2: Math score, reading instrument				
Special Education	-1.313	-1.421*	-1.362*	-1.215*
	(0.670)	(0.631)	(0.665)	(0.570)
F-Statistic	11.68	11.86	11.84	13.00
N	1199497	1199497	1199497	1199497
Panel 3: Reading score, math instrument				
Special Education	-9.904	-3.441	-8.343	-3.138
	(13.51)	(2.019)	(9.585)	(1.835)
F-Statistic	0.569	4.056	0.817	4.263
Panel 4: Math score, math instrument				
Special Education	-13.06	-6.556	-11.71	-5.982
	(17.48)	(3.400)	(13.14)	(3.060)
F-Statistic	0.569	4.054	0.818	4.261
Instruments:				
Prior Pass*SWD Incentive	Y		Y	
Prior Score*SWD Incentive		Y		Y
Controls:				
Levels*Incentive	Y	Y		
Levels*Distance*Incentive			Y	Y
School FE	N	N	N	N

Notes: This table displays results from 16 linear IV models following Equation (5), one in each column and panel, in which being in special education with a malleable diagnosis is instrumented by selection incentives as noted and math or reading z-score is the dependent variable. All specifications include demographic controls, and I control for hypothesis 1 incentives as noted. Standard errors in parentheses are clustered at the school level. Kleibergen-Papp F-statistics from the first stage are reported. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table B.5 Effect of Accountability Incentives on Being Promoted to the Next Grade, No School Fixed Effects

	(1)	(2)	(3)	(4)
Reading Level 2*Incentive	0.000116		0.000449	
	(0.00156)		(0.00150)	
Math Level 2*Incentive	-0.000391		0.00220	
	(0.00220)		(0.00176)	
Reading Level 3*Incentive	-0.00351*		-0.00317	
	(0.00175)		(0.00171)	
Math Level 3*Incentive	-0.000889		0.00240	
	(0.00130)		(0.00148)	
Reading Level 2*Distance*Incentive		-0.0000532		0.000216
		(0.00166)		(0.00159)
Math Level 2*Distance*Incentive		0.000473		0.00204
		(0.00241)		(0.00197)
Reading Level 3*Distance*Incentive		-0.00403*		-0.00406*
		(0.00205)		(0.00202)
Math Level 3*Distance*Incentive		-0.00115		0.00182
		(0.00155)		(0.00189)
Reading Prior Pass*SWD Incentive	-0.00200	-0.00225		
	(0.00285)	(0.00289)		
Math Prior Pass*SWD Incentive	0.00175	0.00261		
	(0.00505)	(0.00518)		
Reading Prior Score*SWD Incentive			-0.000951	-0.00116
			(0.00138)	(0.00139)
Math Prior Score*SWD Incentive			0.00101	0.000951
			(0.00151)	(0.00160)
N	1199720	1199720	1199720	1199720

Notes: This table displays results from 4 linear probability models following Equation (4), one in each column, in which being promoted to the next grade is the dependent variable. All models include demographic controls and year-by-grade fixed effects. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table B.6 Effect of Accountability Incentives on Changing Schools, No School Fixed Effects

	(1)	(2)	(3)	(4)
Reading Level 2*Incentive	0.000966		-0.00164	
	(0.00348)		(0.00336)	
Math Level 2*Incentive	-0.00535		-0.00190	
	(0.00455)		(0.00421)	
Reading Level 3*Incentive	0.00369		0.00377	
	(0.00443)		(0.00455)	
Math Level 3*Incentive	-0.00522		-0.00436	
	(0.00374)		(0.00249)	
Reading Level 2*Distance*Incentive		0.000508		-0.00232
		(0.00369)		(0.00362)
Math Level 2*Distance*Incentive		-0.00666		-0.00277
		(0.00480)		(0.00479)
Reading Level 3*Distance*Incentive		0.00243		0.00290
		(0.00491)		(0.00494)
Math Level 3*Distance*Incentive		-0.00798		-0.00595
		(0.00513)		(0.00336)
Reading Prior Pass*SWD Incentive	0.0142	0.0144		
	(0.00829)	(0.00832)		
Math Prior Pass*SWD Incentive	0.00115	0.00188		
	(0.0126)	(0.0132)		
Reading Prior Score*SWD Incentive			0.00658*	0.00666*
			(0.00326)	(0.00326)
Math Prior Score*SWD Incentive			0.00609	0.00579
			(0.00463)	(0.00466)
N	1199737	1199737	1199737	1199737

Notes: This table displays results from 4 linear probability models following Equation (4), one in each column, in which being in a new school in the current year is the dependent variable. All models include demographic controls and school-level fixed effects are included as noted. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table B.7 Effect of Special Education on Attendance, No School FE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel 1. Percent of Days Absent								
	0.0924	0.0586	0.0919	0.0579	0.650	0.326	0.624	0.325
	(0.0705)	(0.0679)	(0.0692)	(0.0629)	(0.759)	(0.243)	(0.698)	(0.237)
Panel 2. Percent of Days Excused Absence								
	-0.0201	-0.0470	-0.0238	-0.0456	0.209	0.0671	0.167	0.0562
	(0.0309)	(0.0370)	(0.0297)	(0.0343)	(0.240)	(0.0607)	(0.188)	(0.0557)
Panel 3. Percent of Days Unexcused Absence								
	0.0319	0.0197	0.0263	0.0159	0.422	0.183	0.359	0.165
	(0.0383)	(0.0387)	(0.0362)	(0.0357)	(0.461)	(0.118)	(0.378)	(0.107)
Panel 4. Times Tardy per Day Enrolled								
	0.0176	0.0303	0.0192	0.0295	0.139	0.116	0.152	0.115
	(0.0157)	(0.0233)	(0.0163)	(0.0222)	(0.166)	(0.0893)	(0.173)	(0.0883)
F-statistic	9.498	8.896	9.703	9.964	0.875	2.939	0.962	3.023
N	928512	928512	928512	928512	928512	928512	928512	928512
Instruments:								
	Reading	Reading	Reading	Reading	Math	Math	Math	Math
Prior Pass * SWD Incentive	Y		Y		Y		Y	
Prior Score * SWD Incentive		Y		Y		Y		Y
Controls:								
Levels * Incentive	Y	Y			Y	Y		
Levels * Distance*Incentive			Y	Y			Y	Y

Notes: This table displays results from 32 linear IV models following Equation (5), one in each column and panel, in which being in special education with a malleable diagnosis is instrumented by selection incentives as noted. All specifications include demographic controls and year-by-grade fixed effects, and I control for hypothesis 1 incentives as noted. Standard errors in parentheses are clustered at the school level. Kleibergen-Papp F-statistics from the first stage are reported. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table B.8 Effect of Accountability Incentives on Special Education Placement by Percentage in Special Education, No School Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reading Level 2*Incentive	0.0109**		0.00574		0.0182**		0.0150**	
	(0.00420)		(0.00386)		(0.00410)		(0.00386)	
Math Level 2*Incentive	-0.0157**		-0.0155**		-0.00716		-0.00893*	
	(0.00482)		(0.00413)		(0.00470)		(0.00395)	
Reading Level 3*Incentive	0.0239**		0.0229**		0.0325**		0.0299**	
	(0.00387)		(0.00374)		(0.00423)		(0.00393)	
Math Level 3*Incentive	0.000882		-0.00184		0.000298		0.000890	
	(0.00275)		(0.00314)		(0.00250)		(0.00286)	
Reading Level 2*Distance*Incentive		0.0105*		0.00605		0.0188**		0.0165**
		(0.00443)		(0.00408)		(0.00433)		(0.00410)
Math Level 2*Distance*Incentive		-0.0187**		-0.0170**		-0.00879		-0.00872*
		(0.00529)		(0.00457)		(0.00516)		(0.00439)
Reading Level 3*Distance*Incentive		0.0256**		0.0266**		0.0366**		0.0357**
		(0.00435)		(0.00424)		(0.00476)		(0.00451)
Math Level 3*Distance*Incentive		-0.00170		-0.00213		-0.00201		0.00203
		(0.00330)		(0.00388)		(0.00296)		(0.00359)
Reading Prior Pass*SWD Incentive	0.0235**	0.0243**			0.00882	0.00860		
	(0.00677)	(0.00687)			(0.00573)	(0.00575)		
Math Prior Pass*SWD Incentive	-0.00201	-0.000742			0.0157	0.0174		
	(0.00995)	(0.0104)			(0.00955)	(0.00982)		
Reading Prior Score*SWD Incentive			0.0122**	0.0130**			0.00329	0.00361
			(0.00352)	(0.00358)			(0.00308)	(0.00312)
Math Prior Score*SWD Incentive			0.00186	0.00177			0.00947**	0.0100**
			(0.00348)	(0.00362)			(0.00315)	(0.00329)
N	625785	625785	625785	625785	573952	573952	573952	573952
Sample	>=12.5	>=12.5	>=12.5	>=12.5	<12.5	<12.5	<12.5	<12.5

Notes: This table displays results from 8 linear probability models following Equation (4), one in each column, in which being in special education with a malleable diagnosis is the dependent variable. All models include demographic controls, year-by-grade fixed effects, and the main effects of school incentives and prior student performance. The sample is divided by whether the school had more or less than 12.5 percent of its student body in special education. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table B.9 Effect of Accountability Incentives on Malleable Diagnoses, Including Those with Non-Malleable Diagnoses in Sample, No School Fixed Effects

	(1)	(2)	(3)	(4)
Reading Level 2*Incentive	0.0115**		0.00753**	
	(0.00274)		(0.00258)	
Math Level 2*Incentive	-0.0128**		-0.0131**	
	(0.00316)		(0.00275)	
Reading Level 3*Incentive	0.0231**		0.0210**	
	(0.00275)		(0.00262)	
Math Level 3*Incentive	-0.000488		-0.00182	
	(0.00191)		(0.00205)	
Reading Level 2*Distance*Incentive		0.0113**		0.00804**
		(0.00289)		(0.00273)
Math Level 2*Distance*Incentive		-0.0147**		-0.0141**
		(0.00345)		(0.00302)
Reading Level 3*Distance*Incentive		0.0253**		0.0247**
		(0.00308)		(0.00297)
Math Level 3*Distance*Incentive		-0.00266		-0.00221
		(0.00227)		(0.00250)
Reading Prior Pass*SWD Incentive	0.0129**	0.0131**		
	(0.00439)	(0.00444)		
Math Prior Pass*SWD Incentive	0.00207	0.00358		
	(0.00659)	(0.00681)		
Reading Prior Score*SWD Incentive			0.00652**	0.00695**
			(0.00231)	(0.00234)
Math Prior Score*SWD Incentive			0.00290	0.00299
			(0.00234)	(0.00241)
N	1237846	1237846	1237846	1237846

Notes: This table displays results from 4 linear probability models following Equation (4), one in each column, in which being in special education with a malleable diagnosis is the dependent variable. All models include demographic controls, year-by-grade fixed effects, and the main effects of school incentives and prior student performance. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table B.10 Effect of Accountability Incentives on Not Having a Valid Score, No School Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Reading	Reading	Reading	Reading	Math	Math	Math	Math
Reading Level 2*Incentive	0.00159		0.00216		0.00153		0.00211	
	(0.00417)		(0.00415)		(0.00418)		(0.00416)	
Math Level 2*Incentive	0.00498		0.00364		0.00505		0.00369	
	(0.00396)		(0.00351)		(0.00396)		(0.00350)	
Reading Level 3*Incentive	-0.00246		-0.00392		-0.00255		-0.00402	
	(0.00390)		(0.00382)		(0.00390)		(0.00383)	
Math Level 3*Incentive	-0.00140		-0.000136		-0.00138		-0.0000954	
	(0.00277)		(0.00276)		(0.00277)		(0.00275)	
Reading Level 2*Distance*Incentive		-0.00268		-0.00430		-0.00272		-0.00437
		(0.00446)		(0.00435)		(0.00446)		(0.00436)
Math Level 2*Distance*Incentive		-0.00109		0.000508		-0.00106		0.000561
		(0.00322)		(0.00324)		(0.00321)		(0.00323)
Reading Level 3*Distance*Incentive		0.00204		0.00267		0.00201		0.00267
		(0.00464)		(0.00463)		(0.00465)		(0.00463)
Math Level 3*Distance*Incentive		0.00610		0.00470		0.00620		0.00477
		(0.00435)		(0.00391)		(0.00435)		(0.00390)
Reading Prior Pass*SWD Incentive	-0.00523	-0.00526			-0.00537	-0.00541		
	(0.00446)	(0.00449)			(0.00446)	(0.00449)		
Math Prior Pass*SWD Incentive	0.00717	0.00707			0.00725	0.00718		
	(0.00648)	(0.00641)			(0.00649)	(0.00642)		
Reading Prior Score*SWD Incentive			0.000431	0.000372			0.000370	0.000308
			(0.00226)	(0.00225)			(0.00226)	(0.00226)
Math Prior Score*SWD Incentive			0.000775	0.000810			0.000798	0.000840
			(0.00226)	(0.00227)			(0.00226)	(0.00227)
N	1199473	1199473	1199473	1199473	1199473	1199473	1199473	1199473

Notes: This table displays results from 8 linear probability models following Equation (4), one in each column, in which not having a valid reading or math score is the dependent variable. All models include demographic controls, year-by-grade fixed effects, and the main effects of school incentives and prior student performance. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.

Table B.11 Effect of Accountability Incentives on Special Education Placement – Alternate Malleable Definition, No School FE

	(1)	(2)	(3)	(4)
Reading Level 2*Incentive	0.0137**		0.00793**	
	(0.00323)		(0.00301)	
Math Level 2*Incentive	-0.0186**		-0.0181**	
	(0.00371)		(0.00307)	
Reading Level 3*Incentive	0.0268**		0.0240**	
	(0.00332)		(0.00319)	
Math Level 3*Incentive	0.000799		-0.00206	
	(0.00228)		(0.00240)	
Reading Level 2*Distance*Incentive		0.0130**		0.00798*
		(0.00343)		(0.00320)
Math Level 2*Distance*Incentive		-0.0212**		-0.0192**
		(0.00403)		(0.00339)
Reading Level 3*Distance*Incentive		0.0288**		0.0279**
		(0.00375)		(0.00364)
Math Level 3*Distance*Incentive		-0.00157		-0.00205
		(0.00269)		(0.00294)
Reading Prior Pass*SWD Incentive	0.0202**	0.0205**		
	(0.00535)	(0.00544)		
Math Prior Pass*SWD Incentive	-0.00237	-0.000684		
	(0.00851)	(0.00873)		
Reading Prior Score*SWD Incentive			0.0108**	0.0113**
			(0.00291)	(0.00294)
Math Prior Score*SWD Incentive			0.00257	0.00275
			(0.00301)	(0.00310)
N	1237846	1237846	1237846	1237846

Notes: This table displays results from 4 linear probability models following Equation (4), one in each column, in which being in special education with a malleable or autism diagnosis is the dependent variable. Each model includes demographic controls, year-by-grade fixed effects, and the main effects of school incentives and prior student performance. Standard errors in parentheses are clustered at the school level. * denotes significance at the 0.05 level, ** at the 0.01 level.