High School Value-Added and College Outcomes

Evan Totty*

Abstract

This paper contributes to the sparse literature on the lasting impact of teacher and school value-added on adult outcomes by estimating value-added scores for high schools and linking these scores to a student-level dataset on college performance. After controlling for detailed student and high school characteristics, one standard deviation increase in high school value-added increases the probability of graduating from college by six percentage points and final GPA by 0.05-0.08 points on a 4.0 scale. Most of the GPA impact occurs in early semesters. There is some evidence that the impact may be largest for male students and Black students.

Keywords: value-added, high school, college, graduation, GPA

*Purdue University, Department of Economics, 403 W State Street, West Lafayette, IN 47907. Email: evan.scott.totty@gmail.com. Phone: 731-441-0230. This work was completed while in graduate school at Purdue University; I have since taken a position as an Economist with the United States Census Bureau.

Electronic copy available at: https://ssrn.com/abstract=3203759

1 Introduction

Many states and school districts have recently begun using value-added models to evaluate teachers and schools; at least 30 states had introduced or were developing some form of a value-added model as of 2010 (Blank, 2010). These models are intended to estimate the impact that teachers and schools have on student test scores. This typically involves estimating the gains or losses in terms of normalized test scores that students experience while under the instruction of a particular teacher or while enrolled in a particular school. Some states and school districts have tied raises, tenure, and school funding to value-added scores. Some also release these value-added scores to the public.

The use of value-added models to evaluate teachers and schools has caused two debates. The first is whether value-added models provide unbiased estimates of the impact that teachers and schools have on student test scores. The second is a public policy debate of whether these value-added models should actually be a component of the evaluation process for teachers and schools. Much work has been done on the first debate. Proponents of using value-added to select, retain, and sort teachers have argued that doing so can produce significant gains in students achievement (Gordon et al., 2006; Hanushek, 2009; Goldhaber and Hansen, 2010; Kraft, 2015), while others have raised concerns that value-added models may be biased by an inability to capture the effects of students sorting into classrooms and schools based on unobservable student, family, and socio-economic characteristics that are correlated with test scores (Baker et al., 2010; Rothstein, 2010; Paufler and Amrein-Beardsley, 2014). The same benefits and concerns have been raised with respect to measuring school effectiveness using value-added models (Raudenbush and Willms, 1995; Meyer, 1997; McCaffrey et al., 2004; Rubin et al., 2004; Reardon and Raudenbush, 2009; Deming, 2014; Angrist et al., 2017). However, there is evidence and a growing consensus that modern value-added models do a good job of accounting for these unobservable characteristics, primarily through the use of lagged test scores, and providing unbiased estimates of teacher and school value-added (Kane and Staiger, 2008; Kane et al., 2013; Deming, 2014; Chetty et al., 014a; Altonji and Mansfield, 2014; Koedel et al., 2015; Angrist et al., 2017).

Electronic copy available at: https://ssrn.com/abstract=3203759

The second debate has not approached a consensus and has received comparatively little attention. The debate is whether value-added models are an appropriate evaluation method for teachers and schools, even if they do provide unbiased estimates of the impact that teachers and schools have on student test scores. One critique against the use of value-added models to evaluate schools and teachers is that value-added may be too unstable from year-to-year to be used for high-stakes personnel decisions (Baker et al., 2010; Newton et al., 2010). Another concern is whether there is any lasting impact of high value-added teachers and schools on the adult outcomes of their students; while test scores are the most direct quantifiable outcome measure for the impact of teachers and schools, higher test scores are desirable mostly because of the assumption that they are predictive of success later in life. In the case of value-added, this assumption may not be true. Getting a higher test score because a student is in a higher value-added classroom or school could translate to better adult outcomes if the test score improvement represents an impact that the teacher or school had on the student's cognitive skills[1]. It could also occur through a "transcript effect," in which higher test scores ultimately track students into more and better universities. Alternatively, higher value-added teachers and schools could just be better at teaching to the test, without having any impact on students' cognitive skills or adult outcomes.

The goal of this paper is to determine whether value-added translates to better adult outcomes. Specifically, the research question is whether students from higher value-added high schools are more likely to do well in college and graduate from college. This research question implies a more fundamental research question related to the issue raised above about the mechanism by which schools impact test scores. That is, are schools impacting students' cognitive skills or are they simply teaching to the test? If test score gains represent increases in cognitive skills, then one would expect students who go to higher value-added high schools to end up doing better in college. If value-added gains represent teaching to the test, then it is not clear that there should be any relationship between value-added and college performance once other selection effects are

[1]It is common in the value-added literature to think of test scores as measuring "cognitive skills" and to think of "non-cognitive skills" as all of the skills that are not captured by test scores, such as adaptability, self-restraint, motivation, and persistence (Jackson, 2012). These non-cognitive skills could also be thought of as "non-test-score skills," and these two terms are used interchangeably below.

controlled for.

The lasting impact of value-added on adult outcomes was first addressed in Chetty et al. (014b), which found that high value-added teachers have positive lasting impacts on college attendance, college quality, and early adulthood earnings. The current paper adds to this line of work and extends the analysis from Chetty et al. (014b) in three ways: (1) by analyzing value-added that occurs at a different level of the education production system (schools versus teachers), (2) by analyzing value-added that occurs at a different time in the student's academic career (high school versus grades 4-8), and (3) by analyzing different adult outcome variables (college performance versus college attendance and quality). The results extend the existing knowledge about the population of teachers and schools for whom value-added models can be a useful method. This is important, given that it can be problematic to generalize results about some groups of teachers and schools to all teachers and schools (Jackson, 2014). The results also speak to the mechanism through which value-added may affect adult outcomes by analyzing performance within college, rather than just selection into colleges. Understanding the mechanism provides important insights for the debate about whether value-added scores should be included in the teacher and school evaluation process.

This paper addresses the lasting impact of value-added on adult outcomes by using K-12 test score data for students in Florida to estimate value-added scores for high schools and then linking these scores to a dataset on college performance for students at three Florida universities. The high school value-added scores are estimated using value-added models that have been shown to provide unbiased estimates of value-added scores in prior work. The models appear to do a good job of accounting for sorting of students into classrooms; tests show that the high school value-added scores are uncorrelated with 7th grade test scores of students who eventually attended the high school, conditional on observable student and high school characteristics included in the value-added model. These value-added scores, and other observable characteristics about the high school, are then linked to a dataset on college performance that includes college transcript data, detailed information on the student, and a high school identifier. Because the K-12 test score dataset and the college performance dataset can only be linked at the high school level, rather than student

Electronic copy available at: https://ssrn.com/abstract=3203759

level, the matched dataset of analysis lacks the ideal student-level pre-high school test scores for use as controls. However, college students at the sample of universities in the dataset who are from higher value-added high schools do not appear to be selected based on non-test-score skills relative to students from lower value-added high schools in any way that can be observed.

Analysis of the matched dataset shows that higher value-added high schools have positive and statistically significant impacts on college graduation rates and GPA; after controlling for detailed student characteristics and high school characteristics, one standard deviation increase in high school value-added increases the probability of graduating college by six percentage points and final cumulative GPA by 0.05-0.08 points on a 4.0 scale. This represents a 9.71% increase form the sample mean graduation rate and a 1.95%-3.11% increase from the sample mean final cumulative GPA. Both results are statistically significant and robust to different sets of controls for students, high schools, and sorting effects. Analysis by semester shows that the impact on GPA tends to be largest in earlier semesters and decrease over time. Results also suggest that the impact may be larger for Black students than White students and male students than female students, although the estimates are not measured precisely enough to rule out that the impact is the same across gender and race.

The remainder of the paper is organized as follows: Section 2 provides a review of related literature. Section 3 describes the datasets used in the study. Section 4 discusses the conceptual framework on which modern value-added models are based. Section 5 describes the empirical approach to estimating value-added scores for high schools and the impact of value-added on college performance. Section 6 discusses summary statistics for the data and tests of the identification assumptions. Section 7 presents and discusses the results. Section 8 concludes.

2 Literature review

Many papers have studied the lasting impact of high value-added K-12 teachers on student performance in later K-12 grades (Kane and Staiger, 2008; Brian A. Jacob and Sims, 2010; Rothstein,

Electronic copy available at: https://ssrn.com/abstract=3203759

2010; Cascio and Staiger, 2012; Chetty et al., 014a). However, only Chetty et al. (014b) have studied the lasting impact that high value-added teachers or schools have on post-secondary outcomes.

Chetty et al. (014b) study the lasting impact of value-added measured at the teacher level in grades 4-8 and find statistically significant positive impacts of having a higher value-added teacher on college attendance, college quality, and early adulthood earnings. There are three important differences between Chetty et al. (014b) and this study. The first two are related to the level and timing of the value-added. First, whereas Chetty et al. (014b) estimate value-added at the teacher level, this study estimates school-level value-added scores. This is an important contribution because, while the recent value-added literature has become increasingly focused on teacher-level estimates due to the increasing availability of data linking students to teachers, school-level value-added has a long history in the literature itself. Furthermore, school-level value-added is used as a policy tool by some states and school districts for determining school funding, principal evaluation, and as a supplement into the value-added component of teacher evaluations when teachers have missing value-added scores (Raudenbush and Willms, 1995; Meyer, 1997; Ladd and Walsh, 2002; Reardon and Raudenbush, 2009; Chiang et al., 2016).

The second difference is with respect to the timing of the value-added: whereas Chetty et al. (014b) estimate value-added in grades 4-8, this study estimates value-added in high school. Studies have shown that earlier years of a student's schooling are the most influential (Heckman, 2006). Therefore, while the results from Chetty et al. (014b) suggest that value-added is an effective way to evaluate teachers and schools in grades 4-8, it is unknown whether value-added is still an appropriate method for evaluating teachers and schools at the high school level. Jackson (2014) discusses issues related to estimating value-added for high school teachers, rather than K-8 teachers, and concludes that either (a) teachers are less influential in high school than elementary schools or (b) value-added is a poor method for evaluating high school teachers. Therefore, the results of this study address the unresolved issue of whether value-added methods should be used for evaluation of teachers and schools at the high school level.

The third difference between this study and Chetty et al. (014b) is the post-secondary outcome

Electronic copy available at: https://ssrn.com/abstract=3203759

variables: while Chetty et al. (014b) show that high value-added teachers have lasting impacts on college attendance and college quality, this paper studies the lasting impact of value-added on performance within college. This is an important distinction for determining the mechanism through which value-added impacts post-secondary outcomes. There are two potential mechanisms for the impact of value-added on college attendance and quality found in Chetty et al. (014b): (1) a lasting impact of high value-added teachers on students' cognitive skills and (2) a "transcript effect," in which students ultimately track into more and better universities because of higher test scores on their transcript, regardless of whether there were any lasting impacts on the students' cognitive skills[2]. Performance in college is less likely to be affected by a "transcript effect," and is more likely to identify whether there is any lasting impact of high value-added teachers and schools on student's cognitive skills, specifically. Understanding the mechanism more thoroughly can help inform the relative importance of policies that emphasizes test scores specifically versus potentially more well-rounded policies that emphasize other ways teachers and schools can influence cognitive skills in addition to test scores.

3 Data

The analysis below is made possible by linking K-12 student-level data with a dataset on college performance at several universities in the United States. The datasets cannot be linked at the student level. Rather, they can only be linked at the high school level. Therefore, the K-12 student-level dataset is used to estimate value-added scores for all high schools in the dataset. These value-added scores are then matched to the dataset on college performance by matching the high school that each college student attended to the estimated value-added score for that high school.

[2]This could apply to the positive impact of teacher value-added on earnings as well; Chetty et al. (014b) do not control for college attendance or quality in the earnings regressions, so it is unknown whether this impact occurs independently of the impact on college attendance and quality or through the impact on college attendance and quality. Additionally, while colleges may not review transcripts from grades 4-8, the "transcript effect" could accumulate through high school if higher test scores in grades 4-8 place the student into a track with higher value-added teachers and schools in high school.

Electronic copy available at: https://ssrn.com/abstract=3203759

3.1 K-12 test score data

The K-12 student-level dataset comes from the PK20 Education Data Warehouse at the Florida Department of Education. Beginning in 1998, all students in Florida in grades 4, 5, 8, and 10 took end-of-grade exams in math and reading, known as the Florida Comprehensive Assessment Test (FCAT). Other subjects and grades were added over time. In order to have a consistent model, the high school exam scores used in the value-added models below are always the 10th grade math and reading exam scores and the pre-high school exam scores are always the 8th grade math and reading exam scores.

As discussed below, the college outcome dataset stopped tracking students in 2005. Therefore, the K-12 dataset is used to measure school value-added for the years 2000-2005[3]. During this time, approximately 99% of students in the data have a record of taking the reading exam in high school, while 91% have a record of taking the math exam. Thus, Florida and the FCAT data provides a nearly ideal case for studying value-added at the high school level, because nearly every high school student has a record of taking the FCAT exams and all students take them in the same grade. Alternatively, using end-of-course high school exams in Algebra 1 or English 1, for example, that are available in datasets from other states such as North Carolina, would be problematic because not all students take these exams in high school and the students who do take them do so in different grades. This adds additional selection and omitted variable concerns, also known as "track treatment bias," as discussed in Jackson (2014).

The dataset also contains information on race, ethnicity, gender, and receipt of free-or-reduced lunch. In total, the dataset contains full information on 677,468 high school students from 2000-2005. End-of-grade exams in math and reading during 10th grade are used as the dependent variables in the value-added model discussed in Section 5. The independent variables are the 8th grade math and reading scores and the demographic characteristics listed above. Mean demographic characteristics for the student's high school, the mean fraction of students on free-or-reduced lunch,

[3]As discussed below, the value-added model makes use of 10th grade test scores and 8th grade test scores. Because the 8th grade and 10th grade FCAT exams began in 1998, students in the dataset do not have scores from both grades until the year 2000.

Electronic copy available at: https://ssrn.com/abstract=3203759

and the mean incoming 8th grade test scores for each high school are also included in order to help account for sorting of students into schools[4].

All test scores are normalized by subject, grade, and year before estimation. After estimating the value-added scores, these value-added scores and the other high school characteristics are matched to the dataset on college performance by matching on high school. As described above, not every high school student has a record for each high school exam. Additionally, there is variation in the fraction of students with high school exam records across high schools. The fraction of students taking one of these exams at each high school could be a proxy for the ability and socio-economic status of students at the schools. Therefore, the fraction of students taking each exam at each high school is also used as an independent variable in the value-added models described below. Summary statistics for the K-12 students used in the value-added models are shown in Table 1.

3.2 College performance data

The dataset on college performance comes from the Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD). MIDFIELD grew out of what was an agreement among universities in the Southeastern University and College Coalition for Engineering Education to construct a longitudinal database containing student records from each university. Additional data and universities were added to the database over time. The purpose of the database is to provide data that can be used to study engineering student persistence. However, MIDFIELD collects data on students from many majors in addition to engineering. The major information is fairly detailed and diverse, with 51 unique majors included in the dataset.

MIDFIELD contains complete academic records for all undergraduate students at 11 large universities in the United States. These records go back to 1987 for some schools, but fewer years for others. The dataset tracks students at these universities until 2005. There is information on 1,248,363 total college students in the dataset across all universities and years. The dataset includes

[4]A small number of students have 10th grade math and reading scores for multiple schools. For these students, only test scores from the high school in which the student originally enrolled are included in the analysis, because movement across schools during high school is likely endogenous.

detailed student-level information including start date, final date, graduation, GPA by semester, final GPA, majors, transfers, ACT scores, SAT scores, high school GPA, high school ID, high school peer economic status, ethnicity, gender, and age[5]. Of the 11 universities in the dataset, three are from Florida: the University of Florida, Florida State University, and Florida A&M. The high school ID in MIDFIELD is used to match value-added scores estimated from the K-12 dataset to the high schools in the MIDFIELD dataset.

3.3 Dataset of analysis

The matching of value-added scores to MIDFIELD is based on whether the students in MIDFIELD attended a high school with a value-added score during the years for which the value-added score was estimated. As discussed in more detailed in Section 5, the analysis is based a high school's average value-added score over the time frame for which the value-added models can be estimated, which begins in 2000. Because MIDFIELD stopped tracking students in 2005, the value-added models are only estimated through 2005. Therefore, a student is matched to a value-added score if they attended a high school for which a value-added score has been estimated and they attended that school during the years 2000-2005.

MIDFIELD provides the college enrollment date, but not the high school graduation date. Therefore, for the matching process it is assumed that a student in the MIDFIELD dataset graduated high school during the year they enrolled in college if they are 17-19 years old at the time of enrollment[6]. In summary, a match between a MIDFIELD student and high school value-added score occurs if: (1) a value-added score for the student's high school exists, (2) the student's first semester occurred in the years 2000-2005, and (3) the student's first semester occurs at age 17-19. The matches are further restricted to students at one of the three Florida universities in the data,

[5]High school peer economic status is a variable created by MIDFIELD that attempts to measure average socio-economic status of other students who attended the same high school at the same time as the student in the MIDFIELD dataset. The variable is based on demographic information, test score information, and free-or-reduced lunch information.

[6]Students who enrolled in college before the age of 17 or after the age of 19 are excluded from the analysis because they represent non-traditional students for whom there is likey additional endogeneity.

Electronic copy available at: https://ssrn.com/abstract=3203759

students who enrolled at these universities as freshmen, and full-time students. Of the 1,248,363 students in the MIDFIELD dataset, 42,063 were able to be matched to a value-added score using the steps described above.

If a student meets the match criteria listed above, then the student's high school's value-added score, mean demographic characteristics, mean fraction of students on free-or-reduced lunch, mean incoming 8th grade test scores, and mean fraction of students taking high school exams are matched from the K-12 dataset to the MIDFIELD dataset. The fraction of students receiving free or reduced lunch, fraction of students with a record of taking the high school exams, and mean demographic characteristics serve as a measure of the school's socio-economic setting, while the mean incoming 8th grade test scores help account for sorting of students into schools based on other unobserved dimensions of ability that are unrelated to those socio-economic characteristics. Table 2 shows summary statistics for the matched dataset of analysis[7].

4 Model background

4.1 High school value-added model

Value-added modeling typically involves linear regression analysis that predicts student test score performance based on a combination of prior achievement and demographics. The details for the specification and estimation of value-added scores and their impact on college outcomes are discussed in Section 5. First, a statistical framework for the impact of schools on student achievement adopted from Todd and Wolpin (2003) is discussed. Under some conditions discussed below, the linear regression value-added models can be derived from this framework. A more detailed discussion of the framework from Todd and Wolpin (2003) and it's relation to the empirical value-added literature can be found in Koedel et al. (2015).

[7]The sample size for the graduation variable is smaller than the other variables because MIDFIELD stops tracking students in 2005. Thus, only students who were in the data for at least five years were included in the sample for the graduation variable, in order to allow at least five years for the student to graduate. Students in this sample were counted as having graduated if they graduated in 5 or fewer years after enrolling.

The framework assumes that student achievement in a given year is a function of school inputs and other factors:

$$A_{it} = A[S_i(t), F_i(t), \alpha_{i0}, \varepsilon_{it}], \tag{1}$$

where A_{it} is a measure of achievement for student i at time t, which is a function of the history of school inputs ($S_i(t)$), family inputs ($F_i(t)$), ability (α_{i0}), and an idiosyncratic error term (ε_{it}), where $S_i(t)$ and $F_i(t)$ represent a detailed history for each year until time t.

The intuition behind the common value-added models is that lagged achievement measures can be a sufficient statistic for detailed input histories and ability[8]. In this setting, the achievement production function can be re-written as a function of contemporaneous inputs and a baseline test score:

$$A_{ijt} = A_t[s_{j(i,t)}, f_{it}, A_{i,t-1}[S_i(t-1), F_i(t-1), \alpha_{i0}], \varepsilon_{ijt}], \tag{2}$$

which models current academic achievement as a function of current school inputs ($s_{j(i,t)}$, which is a measure of school quality for school j that student i attends in year t), current family inputs (f_{it}), lagged achievement ($A_{i,t-1}$), and an idiosyncratic error term. The key assumption is that lagged inputs and unobservable individual ability can be accounted for by the presence of lagged achievement scores.

Applied value-added work traditionally assumes that the arguments in (2) are additively separable, which leads to the equation:

$$A_{ijt} = s_{j(i,t)} + \alpha f_{it} + \gamma A_{i,t-1} + \varepsilon_{ijt}. \tag{3}$$

Difficult-to-measure prior input histories and unobservable individual ability are not in (3) because they are assumed to be captured in the lagged achievement score. Due to the inability of researchers

[8]For consistency with the discussion in the rest of the paper, ability could be referred to as skills, and skills could include both cognitive and non-cognitive skills.

Electronic copy available at: https://ssrn.com/abstract=3203759

to observe family inputs such as parental effort and involvement in school work, applied value-added models typically make one additional assumption: current family inputs do not respond to contemporaneous school inputs. Then, the model of student achievement is further simplified to:

$$A_{ijt} = s_{j(i,t)} + \gamma A_{i,t-1} + \varepsilon_{ijt}, \tag{4}$$

which is the basis for applied value-added work. There are a variety of estimation methods that can be used to recover the school inputs from this equation. These methods are discussed in the next section.

Several additional conditions are required in order to link the value-added model in equation (4) to the underlying structural cumulative achievement model in (1) beyond the assumptions about lagged achievement scores and family inputs. Among them are grade invariance in the education production function, geometric decay in the impact of prior inputs, and geometric decay in the impact of unobservable individual ability that is equal to the rate of geometric decay for prior inputs. A detailed discussion of these conditions and their relevance in empirical work can be found in Todd and Wolpin (2003) and Sass et al. (2014). Sass et al. (2014) show that these conditions are usually not met. However, as they point out, the failure of typical value-added models to link to a structural interpretation says little about the value of the estimated value-added scores. Ultimately, the extent to which value-added scores capture useful information about school inputs is an empirical question. Another assumption of this model that is standard in applied value-added work is that school quality is time-invariant, conditional on the roster and experience of the teachers and administrators in the school. This rules out the possibility that school value-added fluctuates from year-to-year for reasons unrelated to school faculty composition effects or that it depends on characteristics of the students[9].

[9]Alternatively, school quality could be thought of as consisting of three components: (1) persistent real quality, (2) non-persistent real quality, and (3) non-persistent non-real quality (i.e., estimation error) (Koedel et al., 2015). The first component is represented by $s_{j(i,t)}$ in equation (4), and is the part that can be estimated and is the focus of most value-added studies. Non-persistent quality can still exist in the error term of equation (4).

4.2 School versus teacher value-added

The literature on value-added modeling has become increasingly focused on teacher-level value-added as data linking students and teachers has become more readily available. In the case of teacher value-added, 'school inputs' in equation (4) above becomes a measure of teacher impact on test scores. As discussed above, one contribution of this paper to the work started by Chetty et al. (014b) is the use of value-added at the school level, given that school-level value-added scores also have a long history in the literature and are used by many states and school districts.

The use of school-level value-added in this paper is a necessity, given that the college outcome data only allows students to be linked to high schools, rather than teachers. However, the use of school-level value-added is likely the most appropriate approach at the high school level. Within any school, there is variation in teacher quality. For earlier grades, when students often only have instruction from a single teacher, the entire school's value-added score would be a very noisy measure of the value-added that each individual student receives. In high school, however, students take courses with many different instructors. Over the course of their high school career, students will likely receive instruction from a large fraction of the school's teachers. This makes the use of teacher value-added less appropriate, as a student may take only one course from a specific teacher in their entire high school career.

Furthermore, while several papers have estimated high school value-added at the teacher level (Aaronson et al., 2007; Koedel, 2009; Slater et al., 2012; Goldhaber and Tseng, 2013), Jackson (2014) has shown that estimating value-added at the teacher level is more difficult for high school teachers than elementary and middle school teachers. He concludes that value-added may ultimately be a poor evaluation tool for high school teachers, due to the fact that high school teachers teach different tracks (groups of courses), which introduces "track treatment bias" into the value-added model. This extra bias may be less of a concern at the school level, because students from all tracks are combined into the same model to produce a comprehensive value-added score for the school. Therefore, in addition to filling in a gap in the value-added literature, measuring value-added at the

Electronic copy available at: https://ssrn.com/abstract=3203759

school level may be a more appropriate approach for high schools[10].

When only school value-added is included in the model without teacher value-added, the school value-added is likely made up of an accumulation of teacher value-added, but also likely captures the effects of the administration, leadership, and polices associated with the school itself. In this sense, the interpretation for a school value-added score can be thought of as the accumulated value-added that the student receives from all of the various teachers, administrators, and policies associated with the school.

In order to isolate the value-added associated with a school's teachers, administrators, and policies, it is necessary to control for other school characteristics that may be correlated with value-added and student outcomes. These could include socio-demographic characteristics of the students and neighborhood or access to funding and resources that may promote student learning. As described above, the dataset of analysis has many high school-level control variables that should help to estimate value-added that is independent of the students and resources available to the school. Additional information on school-level or district-level funding associated with schools would have been a valuable additional control, but no such information was available in either the K-12 administrative dataset or the MIDFIELD dataset.

4.3 College achievement model

An analogous model to the model for academic achievement shown in equation (1) can be written for college achievement. College achievement is a function of school inputs, family inputs, ability, and an idiosyncratic error term:

$$A_{it}^{C} = A^{C}[S_i(t), F_i(t), \alpha_{i0}^{C}, \nu_{it}^{C}]. \tag{5}$$

The significant difference between this equation and that in (1) is that there is a new ability term,

[10]School-level value-added will still be a noisy measure of the value-added that an individual student receives while in high school, because the student will likely neither take courses from every teacher nor take the same number of courses from each teacher. Similarly, "track treatment bias" may also impact school-level value-added and may do so in heterogeneous ways across schools, depending on school-level curriculum. Thus, the hope is not that school-level value-added removes these issues associated with high school value-added, but rather that it reduces them.

α_{i0}^{C}, that is specific to college outcomes. The two models are allowed to have different abilities because of the difference in outcome measures between the high school value-added achievement models and the college achievement outcomes used below. Some of the unobserved ability that impacts standardized test scores will likely also impact college GPA and graduation; both involve cognitive skills. Therefore, α_{i0} and α_{i0}^{C} may be correlated. However, college GPA and graduation may be more dependent non-cognitive skills in addition to cognitive skills.

Relying on the same assumptions discussed above for the high school model, equation (5) can be rewritten as a linear function of high school quality and a proxy that is a sufficient statistic for detailed school and family histories and ability:

$$A_{ijt}^{C} = \beta s_{j(i)} + \lambda A_{i,t-1}^{C} + \nu_{ijt}^{C}. \tag{6}$$

Details on the proxy and the interpretation of β are be discussed in Section 5.2.

5 Methodology

5.1 Estimating the impact of high schools on test scores

The methodology for estimating school-level value-added scores follows directly from the other studies in the literature. The approach is based on changes in student test scores before and after exposure to a particular teacher or school as students matriculate through their school system. While many studies have shown that this approach combined with controls for prior test scores and other characteristics does a good job of producing unbiased value-added estimates, much of this literature has focused on teacher-level value-added scores (Kane and Staiger, 2008; Kane et al., 2013; Chetty et al., 014a)[11]. It is possible that there are uniques selection threats that make the teacher value-added methodology less appropriate for estimating school-level value-added. For example, sorting

[11]A more detailed discussion of the somewhat surprising result that controlling for observable characteristics in value-added models appears to also remove bias due to sorting on unobservable characteristics is provided in Altonji and Mansfield (2014).

of students based on unobserved skills may be greater across high schools than across teachers within K-8 schools. There could also be different omitted variable bias associated with school-level value-added. As mentioned in Section 4, school value-added is likely an accumulation of teacher value-added, administration effects, and policies associated with the school itself. Access to resources and funding could be more correlated with school than just teacher value-added if they impact not only teacher value-added, but also policies implemented by the school.

Two recent papers have tested the validity of school-level value-added models, specifically: Deming (2014) and Angrist et al. (2017). Both papers exploit school choice lottery outcomes in order to test out-of-sample predictions of value-added scores for students who are randomly assigned to schools. Deming (2014) draws positive conclusions about school value-added models: he fails to reject the hypothesis that school value-added estimates are unbiased when one year of prior test scores are included in the model and multiple years of outcome data are used to produce the school's value-added score. On the other hand, Angrist et al. (2017) detects moderate levels of bias in traditional value-added models. However, the authors determine this bias to be modest enough that value-added estimates are still highly correlated with school effectiveness and policy decisions based on value-added models that control for lagged achievement can still generate substantial achievement gains.

Thus, there is growing evidence that models which control for prior test scores and other characteristics produce useful estimates of effectiveness for schools, in addition to teachers. Therefore, this paper follows the literature and uses lagged-achievement value-added models along with other covariates. The model is as follows:

$$A_{ijt} = \gamma f(A_{i,t-1}) + \phi X_{it} + \phi_2 \bar{X}_{j(it)} + \epsilon_{ijt} \tag{7}$$

$$\epsilon_{ijt} = s_{j(i,t)} + u_{it}, \tag{8}$$

where A_{ijt} is the 10th grade high school test score (either math or reading) for student i at high

Electronic copy available at: https://ssrn.com/abstract=3203759

school j in year t, $f(A_{i,t-1})$ is a control function of lagged test scores that includes a cubic in both 8th grade reading and math scores, and X_{it} is a vector of controls for the student's gender, race, ethnicity, and free-or-reduced lunch status. This is the same model specification used throughout the modern value-added literature (Kane and Staiger, 2008; Deming, 2014; Chetty et al., 014a), which uses free-or-reduced lunch status to proxy for the student's family income, lagged test scores to proxy for skills and history of school and family inputs, and gender, race, and ethnicity to control for additional socio-economic factors correlated with test scores and student sorting. The model above also includes $\bar{X}_{j(it)}$, which is a vector of characteristics for the student's high school, including the fraction of the school population that is male, Black, Asian, any other race, and Hispanic, as well as the fraction that is on free-or-reduced lunch, the mean incoming 8th grade test scores, the fraction that take the high school math and reading exams. and peer economic status. These high school-level controls are intended to help further account for potential sorting of students into high schools. The error term is then decomposed into a measure of school quality ($s_{j(i,t)}$) and a student-year idiosyncratic error term (u_{it}). The school quality term is the "value-added" for school j that needs to be recovered.

There are multiple methods for recovering the school quality term in (8) (Koedel et al., 2015). One methodology decision is the use of gain-score versus lagged-score models. The model presented in (7)-(8) is a lagged-score model, whereas a gain-score model would be one in which the the lagged-score was moved to the left-hand side of the equation such that the dependent variable becomes $(A_{ijt} - A_{i,t-1})$. Thus, the gains-score model imposes the restriction that the parameter on the lagged-score in equation (7) is equal to one. A benefit of the gains-score model is that it avoids the complication of having an independent variable that is measured with error. However, the current consensus is that lagged-score models outperform the gain-score specification (Koedel et al., 2015).

Another methodology decision is the use of one-step versus two-step models. The model presented above, which includes the school value-added in the residual, is a two-step model, whereas a model with a fixed effect for each school in equation (7) would be a one-step model.

Electronic copy available at: https://ssrn.com/abstract=3203759

There is not a preferred approach between these two options in the literature, although the two-step model may be the most common (Kane et al., 2013; Deming, 2014; Chetty et al., 014a; Koedel et al., 2015). Correlations between the two have been shown to be as high as 0.98 (Chetty et al., 014a).

Two other methodological decisions are the use of shrinkage estimators and the combination of value-added scores across years. It is common in the literature to apply a Bayesian shrinkage factor that shrinks the value-added scores toward zero based on how noisy the value-added score is across years for a given teacher/school (Kane and Staiger, 2008; Deming, 2014; Chetty et al., 014a). The benefit of shrinkage is that it reduces the estimation-error variance, which helps account for the fact that the value-added scores are measured with error when they are used as right-hand-side variables (Jacob and Lefgren, 2008; Kane and Staiger, 2008; Harris and Sass, 2011; Kane et al., 2013; Koedel et al., 2015). Lefgren and Sims (2012) show that variation in value-added across both years and subjects can be useful when performing shrinkage; to the extent that teacher effectiveness is correlated across subjects, the best guess of a teacher's ability to increase reading scores should also incorporate information about the teacher's ability to increase math scores. Therefore, Lefgren and Sims (2012) present a shrinkage estimation procedure that incorporates information across both subjects and years using optimal subject weights. The results in this paper use the Lefgren and Sims (2012) approach.

The model illustrated above is used to estimate school value-added scores for each year from 2000-2005. These estimates are then combined across years to produce a single value-added estimate for each school. It has been shown that much of the variation within teacher or school value-added over time is noise (Kane and Staiger, 2002; McCaffrey et al., 2009). Consequently, many studies have found that value-added scores based on multiple years of data are more reliable than single-year valued-added scores (McCaffrey et al., 2009; Goldhaber and Hansen, 2013; Koedel and Betts, 2011; Deming, 2014). Thus, the results in this paper use combined value-added scores across multiple years in order to leverage a larger degree of true variation in school quality.

The complete steps for estimating the value-added scores are as follows:

Electronic copy available at: https://ssrn.com/abstract=3203759

Step 1: Estimate equation (7), recover the residuals, $\hat{\epsilon}_{ijt}$, and calculate the average residual for each school, $\bar{\epsilon}_{jt}$, by averaging the residuals for all students who took a 10th grade exam at high school j in year t. This step is performed separately for each subject and year of high school test scores, such that each school ends up with an average residual for each subject and each year in the data, $\bar{\epsilon}_{jkt}$, where k is the subject.

Step 2: Shrink the school-subject-year residuals. The shrinkage approach from Lefgren and Sims (2012) can be performed by ordinary least squares (OLS) regressions of a school's value-added estimate for one subject-year on it's value-added estimates for all previous subject-years:

$$\bar{\epsilon}_{jkt} = \pi_0 + \pi_1 \bar{\epsilon}_{j,1,-t} + \pi_2 \bar{\epsilon}_{j,2,-t} + \ldots + \pi_K \bar{\epsilon}_{j,K,-t} + \mu_{jkt},$$

where $\bar{\epsilon}_{j,k,-t}$ is a vector of average subject-year residuals for all years prior to year t for school j and subject k, and π_k is a vector of coefficients that represents the prediction weights from the theoretical shrinkage function from equation (7) in Lefgren and Sims (2012). The intuition for this procedure is as follows: if the value-added scores for a particular school are imprecisely measured and noisy, then there is less correlation over time and subjects. This causes estimates of π_k that are closer to zero and therefore shrinks the estimated value-added score for that school toward zero. Including separate regressors for each subject and for each distance from the current year's value-added score allows for a flexible shrinkage function in which some subjects and some years can have greater weight than others. The predicted value for each school-subject-year based on these regressions, $\bar{\epsilon}^*_{jkt}$, serves as the post-shrinkage estimate of the school-subject-year residuals.

Step 3: Average the post-shrinkage estimates across subjects and years. In addition to averaging value-added scores across years, as discussed above, they are also averaged across math and reading to produce a single value-added score for each high school. Lefgren and Sims (2012) also use this composite-style value-added measure in their comparison of value-added models, citing the fact that it may be more representative of the type of statistics that school districts wish to use. Letting $\bar{\epsilon}^*_{jkt}$ be the post-shrinkage predicted school-subject-year residuals for school j in subject k and year t, the average school value-added estimate is:

$$\hat{s}_j = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{K}\sum_{k=1}^{K}\bar{\epsilon}^*_{jkt}.$$

The test scores for each subject are normalized by year to have a mean of zero and standard deviation of one before being used in the value-added models. Therefore, the estimates of value-added, $\hat{s}_j$, are scaled in units of student test score standard deviations. For example, if a high school had an average estimated value-added of 0.20, then that high school increases student test scores by 0.20 standard deviations, on average, across subjects and years.

The identification assumption of the value-added model is that students are not sorted into schools based on unobserved determinants of test scores that have not been accounted for. If this assumption fails, then the estimate of school quality captures both school quality and the correlation between the portion of school quality that is orthogonal to the other covariates and the unobserved determinants of test scores. This assumption allows for sorting of students into schools; sorting into high schools based on characteristics such as skills and family characteristics does occur. The identification assumption just requires that the lagged test scores and observable characteristics of the student and high school are sufficient to account for this sorting. Section 6.1 discusses a test of this assumption.

5.2 Estimating the impact of high schools on college performance

Before moving forward, it is instructive to consider the mechanism by which high school value-added may impact college performance in more detail. As mentioned in the introduction, it is common to think of test scores as measuring "cognitive skills" and "non-cognitive skills" as all of the skills that are not captured by test scores, such as adaptability, self-restraint, motivation, and persistence (Jackson, 2012). College GPA and college graduation are likely impacted by both a student's cognitive and non-cognitive skills; this is where the mechanism occurs. If high value-added schools increase cognitive skills, then this increase in cognitive skills should translate to improved performance in college. Alternatively, if high value-added schools are teaching to the test rather than increasing cognitive skills, then students from higher value-added schools should perform no better in college than students from lower value-added high schools, conditional on other characteristics. There is also a threat of negative selection bias due to "transcript effects" if students with less skills who otherwise would not have been accepted to the sample of universities in MIDFIELD are now admitted because they went to a higher value-added high school and got higher test scores, and students who would have attended these universities are admitted to better universities.

The ideal specification for estimating the impact of attending a higher value-added high school on college performance would include student-level controls, such as demographics; high school-level characteristics, in order to isolate the effect of high school value-added from other high school characteristics that may be correlated with value-added and college performance; and student-level proxies for cognitive and non-cognitive skills before beginning high school. The first two sets of ideal controls exist: the dataset of analysis includes student-level demographics and many high school-level controls, either matched from the K-12 administrative dataset or from MIDFIELD. But there are no available proxies for pre-high school cognitive and non-cognitive skills in the MIDFIELD dataset. A pre-high school standardized test score and GPA could be good proxies for cognitive and non-cognitive skills, respectively, but neither exists in the MIDFIELD dataset.

Thus, the specification for estimating the impact of attending a higher value-added high school on college performance takes on the following form:

$$A_{ij}^{C} = \beta \hat{s}_{j(i)} + \phi_1 X_i + \phi_2 \bar{X}_{j(i)} + v_{ij}, \tag{9}$$

where A_{ij}^{C} is the college outcome variable (either GPA or graduation) for student i who attended high school j, $\hat{s}_{j(i)}$ is the value-added score for the high school j that student i attended, X_i is a vector of controls for the student's race, ethnicity, and gender, and $\bar{X}_{j(it)}$ is a vector of characteristics for the student's high school, intended to address issues related to sorting of students into schools based on skills. These school-level characteristics include the fraction of the school population that is male, Black, Asian, any other race, and Hispanic, as well as the fraction on free-or-reduced lunch, the fraction that takes the high school math and reading exams, the average incoming 8th grade test scores, and peer economic status. Each of the high school-level characteristics are the same ones used in equation (7). The analysis in Section 7 also includes fixed effects for each year, university, and major. The coefficient β in this regression captures the correlation between high school value-added and college performance, conditional on student characteristics and high school characteristics. Equation (9) is directly comparable to equation (6), except with controls for student characteristics and mean characteristics for the student's high school added and no proxy for pre-high school skills.

However, there is one potentially useful variable as a proxy for pre-high school non-cognitive skills: high school GPA. If high school GPA is determined more by non-cognitive skills than cognitive skills, and if there is no correlation between the impact that high schools have on cognitive skills and the impact that high schools have on non-cognitive skills, then high school GPA could still be a useful proxy for a student's pre-high school non-cognitive skills. The best evidence on this comes from Jackson (2012), who finds that most of the variability in high school grades is uncorrelated with high school tests scores and that 9th grade teachers who are good at raising test scores are not necessarily good at raising non-test-score factors such as grades. The two assumptions needed in order for high school GPA to remain a useful proxy for pre-high school non-cognitive skills that is uncorrelated with high school value-added are not trivial. And the caveat that Jackson is focused on teacher-level, rather than school-level, value-added precludes strong conclusions with respect to this application. Nonetheless, the evidence from Jackson (2012) is encouraging given that a large part of school-level value-added is likely determined by the value-added of its teachers.

Furthermore, correlation between value-added and GPA seems even less likely in this application, because value-added is measured at the school level. Whereas standardized test scores measure a student's performance relative to students from other schools, high school GPA is to some degree more a measure of performance relative to peers within one's school and relative to a school's own grading standards. Because value-added in this paper is measured at the high school

Electronic copy available at: https://ssrn.com/abstract=3203759

level and every student in the high school is theoretically exposed to the same "value-added," the link between high school value-added and GPA is weak. For example, even if high schools that have positive effects on standardized test scores did also have positive effects on non-cognitive skills and grades, GPA would still remain largely unchanged if grades are based on a curve or if the curriculum is made more rigorous to match students' higher skills. A test of this assumption in Section 6.2 provides evidence to support the use of high school GPA as a proxy for non-cognitive skills.

If high school GPA is not a valid proxy for pre-high school non-cognitive skills, then it is most likely because high schools that are good at raising test scores are also good at raising non-test-score skills. In this case of positive correlation between a school's impact on cognitive and non-cognitive skills, if two students from different schools in the MIDFIELD dataset have the same high school GPA, then the student from the higher value-added high school would have had a lower pre-high school GPA and thus lower pre-high school non-cognitive skills, on average. Therefore, if high school GPA is not a valid proxy for pre-high school GPA, it likely generates negative bias.

Bias generated by the lack of a valid pre-high school proxy for cognitive skills is also likely to be negative. Tests in Section 6.1 provide evidence that, across high schools, value-added scores are not correlated with 7th grade test scores. This suggests no correlation across high schools between pre-high school cognitive skills and high school value-added for the entire population of students. The concern then is that students who happen to show up in the MIDFIELD dataset could be selected based on unobserved skills. The most likely story is that students from higher value-added schools are negatively selected with respect to pre-high school cognitive skills, due to "transcript effects." In this case, the lack of a pre-high school proxy for cognitive skills suggests negative bias, if any, due to students who show up in the dataset from higher value-added high schools having lower pre-high school cognitive skills.

Ultimately, it is informative to show results with and without the use of high school GPA as a proxy for pre-high school non-cognitive skills. If the estimated effect of value-added on college performance decreases with the addition of high school GPA, it could suggest the removal of

Electronic copy available at: https://ssrn.com/abstract=3203759

positive bias due to students from higher value-added high schools having higher non-cognitive skills or it could suggest negative bias caused by high school GPA being an invalid control due to positive correlation between schools' impacts on test scores and non-cognitive skills. If the estimated effect increases, it could suggest the removal of negative selection bias due to "transcript effects" or it could suggest positive bias caused by high school GPA being an invalid control due to negative correlation between schools' impacts on test scores and non-cognitive skills. The first explanation for each case seems more plausible given the intuition and existing knowledge about the correlation between the ability of teachers to raise cognitive and non-cognitive skills, but it is not possible to rule out one story or the other.

6 Testing the assumptions for analysis

6.1 Sorting of students into high schools

As discussed in Section 5.1, the key identification assumption of the value-added model is that the portion of high school value-added scores that is orthogonal to the other covariates is uncorrelated with unobserved and unaccounted for determinants of student test scores, such as unobserved skills. This section provides an indirect test of this assumption by testing the correlation between high school value-added scores and a variable that is typically left out of value-added models: twice-lagged test scores.

The summary statistics in Table 1 and Table 2 show that student- and high school-level characteristics appear to be fairly balanced across below- and above-median value-added schools, although there are some noticeable differences. For example, students from below-median value-added high schools are more likely to be Black. Some correlation between value-added and observable characteristics is not surprising, because sorting of students into high schools is inevitable. The key assumption is that differences in observable characteristics of the students and high schools can fully explain this sorting of students into high schools and can therefore be used to account for the potential selection bias associated with this sorting. This assumption can be tested indirectly using the intuition originally outlined by Rothstein (2009, 2010) and also by Chetty et al. (014a).

The intuition is to test the correlation between high school value-added scores and prior test scores that are not included in the value-added model. The value-added models control for once-lagged test scores, $A_{i,t-1}$, so they cannot be used to test for selection. But twice-lagged scores, $A_{i,t-2}$, can be used. If once-lagged test scores and observable student and school characteristics can account for the selection, then high school value-added scores should be uncorrelated with 7th grade test scores, conditional on the observable high school and student characteristics included in the value-added model. If 7th grade test scores are correlated with the value-added of the student's future high school, then this would provide strong evidence that the value-added model is not sufficiently accounting for the unobservable characteristics on which students sort into high schools. Similarly, adding 7th grade test scores to the value-added model should not significantly change the estimates of the value-added scores. The tendency for value-added scores to be correlated with and appear to "impact" prior test scores has been one critique of value-added models (Rothstein, 2009, 2010).

Florida began administering the FCAT to 7th grade students in 1999. The test is based on a regression with 7th grade test scores in reading or math as the dependent variable and the average value-added score of the high school that the 7th grader would eventually attend as an independent variable. All of the other independent variables from the value-added model are also included, which are student-level controls for gender, race, ethnicity, free-or-reduced lunch status, cubic in 8th grade reading and math scores, and high school-level controls for mean incoming 8th grade test scores, fraction male, fraction Black, fraction Asian, fraction Hispanic, fraction on free-or-reduced lunch, fraction taking each high school exam, and peer economic status. Year fixed effects are also included. The sample for the test is students who are included in the value-added models and also have 7th grade FCAT scores. Because 7th grade FCAT exams were administered beginning in 1999, students who took 10th grade FCAT exams prior to 2002 do not have 7th grade FCAT scores. Thus, the test is based on students who have 10th grade FCAT scores during the years 2002-2005, whereas the value-added models are estimated for students with 10th grade FCAT scores for the years 2000-2005.

Electronic copy available at: https://ssrn.com/abstract=3203759

Table 3 shows the results. Columns (1)-(2) show the coefficient for the high school value-added variable in regressions with high school math and reading scores as the dependent variable, respectively. As expected, attending a higher value-added high school has a positive and statistically significant impact on high school test scores[12].

Columns (3)-(4) replace the dependent variables with 7th grade math and reading scores, which produces the indirect test for bias described above. For both math and reading, the coefficient for the high school value-added variable is close to zero and not statistically significant. Thus, there is no appearance of an impact of high school value-added on 7th grade test scores. Columns (5)-(6) are based on the same specification from columns (1)-(2), except with 7th grade math and reading scores added as independent variables. Although 7th grade math and reading scores are significant predictors of high school math and reading scores, the coefficient for the impact of high school value-added on high school math and reading scores remains fairly similar when these controls are added.

Testing for evidence of endogeneity bias due to sorting of students into schools based on unobserved student skills by testing for an effect of value-added on prior scores has been one of the most relevant critiques of value-added models in past work (Rothstein, 2009, 2010). The value-added scores for Florida high schools appear to past this test. Combined with similar results from Chetty et al. (014a), it appears that value-added models which make use of rich student-level and school-level data, combine data across many years, and use a shrinkage step in the estimation process may produce more reliable estimates than some other models that have been used in past work by researchers and policymakers.

[12]The fact that the coefficient for value-added is larger than one, which implies that attending a high school that raises test scores by one standard deviation causes the student's high school exam score to increase by *more than one standard deviation*, is likely the result of auto-correlation due to the fact that each student's high school exam score is included in both the dependent variable and the calculation of value-added for their high school in the right-hand-side variable. This issue is only related to the "first-stage" test in columns (1)-(2); it does not impact the value-added scores themselves or the main tests of interest in columns (3)-(4).

Electronic copy available at: https://ssrn.com/abstract=3203759

6.2 Selection of students in the MIDFIELD data

Before analyzing the impact of high school value-added on college outcomes, this section discusses results for the effect of value-added on high school outcomes for SAT scores, ACT scores, and high school GPA. Whereas the purpose of the previous section was to test for bias in value-added scores due to student selection and sorting at the high school level, these results essentially serve as a first stage test and a selection test for the the students who happen to show up in the MIDFIELD data; on the one hand, high schools that are good at raising state-wide standardized test scores likely also raise nation-wide SAT and ACT test scores because of overlap in test material, while on the other hand, because high school GPA is measured based on within-school grading standards and relative performance, the theoretical link between high school-level value-added and high school-level GPA is less strong.

Columns (1) and (2) of Table 4 show the impact of attending a higher value-added high school on SAT and ACT test scores, respectively. Each column also controls for student-level covariates, high school-level covariates, and year fixed effects. Both results are positive and statistically significant; multiplying the coefficients by 0.109 in order to get the effect of one standard deviation increase in value-added, the results suggest that one standard deviation increase in value-added increases SAT scores by 13.57 points and ACT scores by 0.61 points. This positive effect is not surprising, given the likely overlap between state-wide standardized math and reading tests and nation-wide SAT and ACT exams.

However, a threat to the analysis below is that students who show up in the MIDFIELD data may be selected in a way that is correlated with both their school's value-added and their unobserved skills. Because the theoretical link between high school value-added and high school GPA is less strong, a positive effect of high school value-added on high school GPA may actually be evidence of a selected sample of students from high value-added high schools; whereas attending a higher value-added high school should give you an advantage on both state- and nation-wide standardized exams over students who attend lower value-added high schools, the same advantage may not occur for GPA because GPA is more of a within-school metric of performance relative to classmates and grading standards and all students in the high school are theoretically exposed to the same "value-added."

Therefore, column (3) of Table 4 shows the effect of high school value-added on high school GPA. The coefficient is small and negative and not statistically significant; there appears to be no effect of high school value-added on high school GPA, conditional on covariates for student and high school characteristics and year fixed effects. In addition to being consistent with the logic presented above that the link between high school value-added and high school GPA may be weak, this is also consistent with the results from Jackson (2012), which finds that most of the variability in high school grades is uncorrelated with high school tests scores and that 9th grade teachers who are good at raising test scores are not necessarily good raising non-test-score factors such as grades. This is an important finding for two reasons: (1) this validates the use of high school GPA as a potentially valuable proxy for non-test-score skills that is independent of high school value-added and (2) assuming high school GPA is a proxy for non-test-score skills that is independent of high school value-added, there is no direct evidence that students in the MIDFIELD data from higher value-added high schools are selected along any dimension related to non-test-score skills. If anything, the negative coefficient for value-added in the high school GPA regression suggests that students in the MIDFIELD dataset from higher value-added high schools may have slightly lower non-test-score skills. This is potentially consistent with the story that attending a higher value-added high school may generate a "transcript effect," in which high value-added high schools are able to get students of a given skill level into better universities.

An important caveat is that this analysis on ACT/SAT scores and high school GPA is only performed for students from Florida high schools who show up at one of the three Florida universities in the MIDFIELD sample. It could be the case that a combination of value-added impacting on high school GPA and selection into the MIDFIELD sample are both occurring in a way that has offsetting effects for the high school GPA results and thus obscures the relationship between high school value-added and high school GPA. For example, if higher value-added high schools do increase school-wide GPA and other non-test-score skills, but students from higher value-added

Electronic copy available at: https://ssrn.com/abstract=3203759

high schools who show up in the sample are negatively selected with respect to non-test-score skills due to "transcript effects," then there could appear to be no relationship between high school value-added and high school GPA. Alternatively, higher value-added high schools could increase non-test-score skills, but also have tougher grading standards than lower value-added high schools. Both of these examples could result in no observed relationship between high school value-added and high school GPA in the sample while masking important biases and selection effects.

These stories have conflicting implications about the direction of potential bias due to controlling for high school GPA and would invalidate its use. Ultimately, controlling for high school GPA may not be very problematic at all, since the theoretical link between school-level value-added and within-school, or within-course, grades is not obvious. Therefore, it is important to show results without high school GPA as a proxy for non-test-score skills and to show how the results change when high school GPA is added as an additional control variable.

7 Results

As described in Section 3.3, the analysis is limited to students at one of the three universities from MIDFIELD that are located in Florida, students who enrolled in these universities as freshmen, and full-time students. All of the analysis below controls for the student-level covariates shown in Table 2 at a minimum. Standard errors are clustered by high school.

7.1 Graduation

Table 5 shows the impact of attending a higher value-added high school on the probability of graduating from college. Results are based on a logit model and the coefficients are the marginal effects. Each column uses a different combination of covariates and fixed effects to control for potential endogeneity bias associated with the high school value-added variable.

Column (1) is the least saturated, with no controls for other characteristics of the high school, university or major selection, or skills; it only includes the student-level covariates shown in the

Electronic copy available at: https://ssrn.com/abstract=3203759

table and year fixed effects. Student-level covariates control for student-level characteristics that might be correlated with graduation and value-added scores. Year fixed effects control for changes in the graduation rate over time, which is important if students from higher value-added high schools were more likely to enroll during high or low graduation rate periods. For this specification, the coefficient for the high school value-added variables is positive and statistically significant at the 5% level. The coefficient magnitude of 0.62 means that one unit increase in high school value-added increases the probability of graduating by 62 percentage points. This coefficient is very large because the standard deviation of the value-added scores is very small. The bottom of Table 5 shows a more appropriate marginal effect, which is the effect of one standard deviation increase in high school value-added. This is computed my multiplying the coefficient for the value-added variable by the standard deviation of high school value-added scores, shown in Table 1 to be 0.109. Thus, the effect of one standard deviation increase in high school value-added is a seven percentage point increase in the probability of graduating from college. This represents a 11.32% increase from the sample mean graduation rate.

Column (2) adds controls for the other high school-level characteristics shown in Table 2. Including these other school-level characteristics is important in order to control for other characteristics of the high school that may be correlated with graduation in order to isolate the effect of value-added. The coefficient for value-added decreases by a small amount from 0.62 to 0.55, but remains statistically significant at the 5% level. The new estimate corresponds to a six percentage point increase in the probability of graduating from college based one standard deviation increase in high school value-added, which represents a 9.71% increase from the sample mean graduation rate.

Columns (3)-(4) add controls for university and major selection. Column (3) includes separate fixed effects for each and column (4) includes university-by-major fixed effects. Including these fixed effects is important if students from higher value-added schools tend to select into different universities or majors within the linked dataset. The results are nearly unchanged from column (2): the estimated coefficient for value-added is 0.53 in column (3) and 0.57 in column (4). These

Electronic copy available at: https://ssrn.com/abstract=3203759

represent an increase in the probability of graduating by approximately six percentage points, or 9.71% from the sample mean, similar to column (2).

Column (5) is the most saturated specification. In addition to controls for student-level demographics, school-level characteristics, major selection, and university selection, it also controls for high school GPA. High school GPA is used with the intention to proxy for non-test-score skills. This is an important control if students from higher value-added high schools are selected in any way with respect to non-test-score skills, although previous tests showed no direct evidence of this potential bias. As discussed previously, there are reasons why including high school GPA may not be appropriate. Nonetheless, in column (5) the estimated effect of attending a higher value-added high school is practically unchanged once again: the estimated value-added coefficient is 0.57, the same as column (4), and remains statistically significant at the 5% level.

7.2 GPA

Table 6 shows the impact of attending a higher value-added high school on final cumulative GPA. This includes both students who did and did not graduate. Results are based on OLS regressions. As in Table 5, each column uses a different combination of covariates and fixed effects to control for potential endogeneity bias associated with the high school value-added variable. The columns have the same progression as Table 5: column (1) only controls for student-level covariates and year fixed effects, column (2) adds high school-level covariates, columns (3) and (4) add university and major fixed effects, and column (5) adds high school GPA.

When controlling for only student-level covariates and year fixed effects, the estimated coefficient for value-added is 0.86, statistically significant at the 5% level. The coefficient magnitude of 0.86 means that one unit increase in high school value-added increases final cumulative GPA by 0.86 points on a 4.0 scale. Alternatively, one standard deviation increase in high school value-added increase GPA by 0.09 points. This represents a 3.5% increase from the sample mean final cumulative GPA.

Similar to the results for graduation, the estimated coefficient decreases only a small amount when high school-level covariates are added in column (2): the coefficient becomes 0.77, statistically significant at the 1% level. This corresponds to a 0.08 increase in final cumulative GPA from one standard deviation increase in high school value-added, which is a 3.11% increase from the sample mean.

The estimated effect decreases a bit more when university and major fixed effects are added. For the most flexible specification, with university-by-major fixed effects, the estimated coefficient is 0.48, statistically significant at the 5% level. This represents a 0.05 increase in final cumulative GPA from one standard deviation increase in high school value-added or a 1.95% increase from the sample mean.

Finally, when high school GPA is added as a proxy for non-test-score skills in column (5), the estimated coefficient increases back to 0.77 and remains statistically significant at the 5% level. This corresponds to a 0.08 increase in final cumulative GPA from one standard deviation increase in high school value-added, which is a 3.11% increase from the sample mean.

In this case, the inclusion of high school GPA as a proxy for non-test-score skills does affect the interpretation of the results somewhat: adding high school GPA to the specification than already includes university-by-major fixed effects increases the estimated effect by approximately 60%. Related to the earlier discussion on the merits of controlling for GPA, a larger estimated effect when controlling for high school GPA could be consistent with removal of negative bias due to students from higher value-added schools in the MIDFIELD sample having lower unobserved non-test-score skills due to "transcript effects" pushing lower-skill students into better or more universities. It could also be consistent with higher value-added high schools having tougher grading standards, which would generate positive bias in the estimated effect. It is not possible to distinguish between the two stories, but it is reassuring to note that the increase relative to the mean final cumulative GPA is similar in magnitude regardless of whether high school GPA is included: approximately 2% without high school GPA and 3% without it. Both estimates are statistically significant at the 5% level.

Electronic copy available at: https://ssrn.com/abstract=3203759

7.3 Results by semester

Several studies have documented fade out or decay in the effect of value-added over time (Kane and Staiger, 2008; Brian A. Jacob and Sims, 2010; Rothstein, 2010; Cascio and Staiger, 2012; Chetty et al., 014a). To test similar patterns in the college outcomes and to get a better idea of how the impact of value-added occurs, Table 7 and Table 8 show the impact of attending a higher value-added high school on college GPA by semester for the first 6 semesters. Results are based on the specification in column (5) of Table 6, which includes student-level covariates, high school-level covariates, university-by-major fixed effects for the student's major in the given semester, high school GPA, and enrollment year fixed effects.

Table 7 shows that the impact of high school value-added on GPA tends to be larger in the earlier semesters. While the relationship between semester and impact size is not linear, the largest impacts occur in semesters 1-4, and the impacts are no longer statistically significant after semester 4.

However, in addition to the outcome semester changing across columns, the sample also changes as some students drop out and others have missing data for some semesters. Therefore, Table 8 uses a consistent sample of students across semesters by including only students who had GPA observations for each of semesters 1-6, thus dropping those who drop out of college before semester 6 and those with missing data for one or more semesters. This reduces the sample size in the early semesters by almost half, which produces more noisy estimates; the impact in semester 1 becomes much smaller and statistically insignificant, while the other estimates remain similar to those from Table 7. Excluding the small impact in semester 1, these results show the same general pattern as before: the largest impacts occur in the earlier semesters, while the impacts become smaller and not statistically significant in later semesters.

7.4 Gender and racial heterogeneity

Chetty et al. (014b) find some evidence that the impact of teacher value-added in grades 4-8 may be larger for female and non-minority students, although the differences are often not large enough to

Electronic copy available at: https://ssrn.com/abstract=3203759

reject equality of results. Therefore, Table 9 shows how the impact of value-added on graduation and GPA varies by gender and race. Each column-by-panel represents a separate regression and reports five stacked numbers: the coefficient for the value-added variable, the standard error, the coefficient translated into the impact of one standard deviation increase in value-added, the percentage increase over the sample mean that the one-standard-deviation-increase impact represents, and the number of observations. The percentage increase over the same mean for each regression is based on the sample mean for the group specified in the panel. All of the results are based on specifications that also control for student-level covariates, high school-level covariates, university-by-major fixed effects, high school GPA, and year fixed effects.

Column (1) shows the impact on graduation separately by gender and race. Sample size becomes more of a concern when regressions are run separately by gender and race, particularly for the graduation results that had significantly fewer observations even when all genders and races were pooled together. Nonetheless, the results are suggestive of differential effects by gender and race.

Panel I and Panel II show the impact separately for males and females. The coefficient in column (1) for is 0.83 for males and 0.36 for females, compared to 0.57 for the combined results from Table 5. This corresponds to an increase in the probability of graduating by approximately nine percentage points for males and three percentage points for females based on one standard deviation increase in value-added. The standard error is much larger than in Table 5 for males, and the impact is only statistically significant for males at the 10% level.

Panel III and Panel IV show the effect separately for students who are White and students who are Black. The coefficient for students who are White is 0.47 compared to 0.56 for students who are Black. This corresponds to an increase in the probability of graduating by approximately 5 percentage points for White students and 6 percentage points for Black students based on one standard deviation increase in value-added, although neither result is statistically significant. This difference is even larger in percentage gains because of the difference in the sample mean graduation rate between students who are White and students who are Black: it represents a 7.26% increase

Electronic copy available at: https://ssrn.com/abstract=3203759

from the sample mean gradation rate for students who are White and a 14.57% increase for students who are Black. These results are suggestive of larger graduation effects for male students and students who are Black, although none of the impacts are measured very precisely and it is therefore difficult to make conclusions about relative effect size.

Columns (2)-(8) show the impact on GPA. There is again evidence to suggest that the impact on GPA may be larger for both male students compared to female students and students who are Black compared to students who are White. Male students have a coefficient of 1.09 in semester 1, which translates to a 0.12 increase in first-semester GPA from one standard deviation increase in high school value-added, statistically significant at the 5% level. This impact slowly fades out over time, and is no longer statistically significant after semester 3. For females, the coefficient is only 0.54 in semester 1, which is not statistically significant and does not change much across columns, except for column (6) in which the coefficient jumps to 1.14 and becomes statistically significant. The coefficient for final cumulative GPA is also larger for male students than female students: 0.98 for males compared to 0.65 for females, which translates to a 0.11 increase and 0.07 increase in final GPA from one standard deviation increase in value-added for males and females, respectively. The estimate is statistically significant at the 5% level for males and females. However, because of the size of the standard errors, it is again not possible to rule out that the magnitude of the impact of high school value-added on GPA is the same for males and females.

The difference in the impact on GPA is even more pronounced for students who are Black in comparison to students who are White. Students who are Black have a positive and statistically significant coefficient of 1.45 in semester 1, which translates to a 0.16 increase in first-semester GPA from one standard deviation increase in high school value-added. This impact slowly fades out over time, but remains statistically significant until semester 4. For students who are White, the coefficient is only 0.42 in semester 1, which translates to a 0.05 increase in first-semester GPA from one standard deviation increase in value-added. The estimate is not statistically significant and does not change much across columns. Finally, the coefficient for the impact of high school value-added on final cumulative GPA is 1.34 for students who are Black, statistically significant

Electronic copy available at: https://ssrn.com/abstract=3203759

at the 1% level, which translates to a 0.15 increase in final cumulative GPA from one standard deviation increase in value-added. The coefficient for the impact on final cumulative GPA is 0.44 for students who are White, which translates to a 0.05 increase and is not statistically significant. Once again, the impact is even larger in percentage gains, due to the difference in sample mean GPA between students who are White and students who are Black: the impact on final cumulative GPA translates to a 1.83% increase over the sample mean for students who are White compared to a 7.16% increase over the sample mean for students who are Black.

These results are inconsistent with Chetty et al. (014b) in that they are suggestive of larger impacts for male students and students who are Black, rather than female and non-minority students. However, they are similar in that the differences are not measured precisely enough to rule out equality of results. Either set of results could be supported by theory. If school inputs and family inputs are complimentary in education production, then non-minority students, who are of higher socio-economic status on average, could stand to benefit more from higher value-added teachers and schools. Alternatively, substitutability of school and family inputs in education production could suggest that the lasting impact of higher value-added high schools would be larger for minority students. Heterogeneous impacts of value-added remain an unresolved question.

8 Conclusion

This paper addresses the policy-relevant question of whether attending a higher value-added high school has a lasting impact on adult outcomes. The use of value-added models has sparked debate about whether they provide unbiased estimates of the impact that teachers and schools have on test scores and debate about whether they are an appropriate method for evaluation purposes even if they are unbiased. While the growing consensus is that modern value-added models can provide unbiased, or at least useful, estimates of teacher and even school impacts on test scores (Deming, 2014; Angrist et al., 2017), the use of these methods in determining teacher pay, sorting, and school resources is still heavily debated. One relevant question in addressing this debate is whether higher

Electronic copy available at: https://ssrn.com/abstract=3203759

value-added teachers and schools are inducing lasting positive impacts on the adult outcomes of their students; do higher test scores represent the impact of teachers and schools on students' cognitive skills or represent teachers and schools who are good at teaching to the test?

This study addresses the question by analyzing whether attending a higher value-added high school raises college GPA or the probability of graduating from college. This is possible by measuring value-added scores for high schools in Florida and then linking these scores to data on college performance for students at three Florida universities. Tests suggest that the value-added scores are not significantly biased due to sorting of students into schools based on unobservable and unaccounted for determinants of test scores. The linked dataset lacks the ideal student-level pre-high school controls. However, while students who show up in the college performance dataset from higher value-added high schools do have higher ACT and SAT test scores, they do not appear to be selected with respect to non-test-score skills based on available information.

Results suggest that attending a higher value-added high school has a positive and statistically significant impact on both graduation and GPA; after controlling for detailed student and high school characteristics, one standard deviation increase in high school value-added increases the probability of graduating from college by six percentage points and increases final cumulative GPA by 0.05-0.08 points on a 4.0 scale, depending on whether high school GPA is included as a control variable. These results represent an increase over the sample mean by approximately 10% and 2%-3%, respectively. The results remain statistically significant and broadly consistent across different sets of controls. The results for GPA translate to a one-letter-grade increase in 3.12 courses during a student's college career. Much of this gain occurs in early semesters. There is also some evidence that the impact may be larger for male students and students who are Black. The estimates for male students and students who are Black suggest an increase in final cumulative GPA of 0.11 points and 0.15 points, respectively, from one standard deviation increase in high school value-added. This translates to a one-letter-grade increase in four classes for male students and six classes for students who are Black.

The magnitudes in this study are generally consistent with those in Chetty et al. (014b). Across

Electronic copy available at: https://ssrn.com/abstract=3203759

the range of outcome variables in their study, they find that one standard deviation increase in teacher value-added in grades 4-8 increases adult outcomes by 1-5% over the sample mean outcome, depending on the outcome variable. The results in this study fall in a similar range. The relatively large 10% increase in the probability of graduating from college could be due to the fact that high school value-added occurs much closer to the timing of the outcome variables in this study than the timing of the value-added and outcomes in Chetty et al. (014b), given the abundance of evidence in the literature that value-added fades out over time.

This paper extends the results from Chetty et al. (014b) in three important ways: (1) it shows that school-level value-added scores represent lasting impacts of teachers and schools; much of the recent literature has focused on teacher-level value-added, even though school-level value-added is also used for school funding, principal evaluation, and teacher evaluation, (2) it shows that value-added at the high school level represents lasting impacts of teachers and schools. Previous work has suggested that either value-added is not a useful method at the high school level or high school teachers are less impactful than elementary school teachers (Jackson, 2014). These results support the latter conclusion over the former, although more work should be done to compare the relative effectiveness of elementary versus high school teachers in terms of value-added and to determine what factors may contribute to differences in effectiveness, and (3) it studies the impact on different outcome variables. This is significant because, while the impact on college attendance, quality, and earnings found in Chetty et al. (014b) could be attributable to either an impact on cognitive skills or a "transcript effect," college performance is more likely to be attributable to cognitive skills specifically.

As discussed in Chetty et al. (014b), these results do not imply that value-added models should be the only or the premier method for evaluating teachers and schools. Other important issues should be considered, such as how teachers may dynamically respond to policies that emphasize test scores and how the lasting impact of value-added compares to the lasting impact of other teacher and school attributes that can be identified via other evaluation methods such as classroom observation or principal evaluation. Nonetheless, the results in this study confirm that value-added

Electronic copy available at: https://ssrn.com/abstract=3203759

models can be a useful method for identifying effective teachers and schools at many levels.

Electronic copy available at: https://ssrn.com/abstract=3203759

References

Aaronson, D., L. Barrow, and W. Sander (2007). Teachers and student achievement in the chicago public high schools. *Journal of Labor Economics 25*(1), 95–135.

Altonji, J. G. and R. K. Mansfield (2014). Group-average observables as controls for sorting on unobservables when estimating group treatment effects: the case for school and neighborhood effects. Technical Report NBER Working Paper No. 20781.

Angrist, J. D., P. D. Hull, P. A. Pathak, and C. R. Walters (2017). Leveraging lotteries for school value-added: Testing and estimation. *The Quarterly Journal of Economics 132*(2), 871–919.

Baker, E. L., P. E. Barton, L. Darling-Hammond, E. Haertel, H. F. Ladd, R. L. Linn, D. Ravitch, R. Rothstein, R. J. Shavelson, and L. A. Shepard (2010). Problems with the use of student test scores to evaluate teachers. Technical Report Briefing Paper No. 278, Economic Policy Institute.

Blank, R. K. (2010). State growth models for school accountability. progress on developing and reporting measures of student growth. Technical Report Council of Chief State School Supervisors: Washington, DC.

Brian A. Jacob, L. L. and D. P. Sims (2010). The persistence of teacher-induced learning gains. *Journal of Human Resources 45*(4), 915–943.

Cascio, E. U. and D. O. Staiger (2012). Knowledge, tests, and fadeout in educational interventions. Technical Report NBER Working Paper 18038.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review 104*(9), 2593–2632.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014b). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review 104*(9), 2633–2679.

Electronic copy available at: https://ssrn.com/abstract=3203759

Chiang, H., S. Lipscomb, and B. Gill (2016). Is school value-added indicative of principal quality? *Education Finance and Policy 11*(3), 283–209.

Deming, D. J. (2014). Using school choice lotteries to test measures of school effectiveness. *American Economics Review: Papers and Proceedings 104*(5), 406–411.

Goldhaber, D. and M. Hansen (2010). Using performance on the job to inform teacher tenure decisions. *The American Economic Review 100*(2), 250–255.

Goldhaber, D. and M. Hansen (2013). Is it just a bad class? assessing the stability of measured teacher performance. *Economica 80*(319), 589–612.

Goldhaber, D. and P. G. F. Tseng (2013). Teacher value-added at the high school level. *Educational Evaluation and Policy Analysis 35*(2), 220–236.

Gordon, R., T. J. Kane, and D. O. Staiger (2006). Identifying effective teachers using performance on the job. Technical Report The Hamilton Project White Paper 2006-01.

Hanushek, E. A. (2009). Teacher deselection. In D. Goldhaber and J. Hannaway (Eds.), *Creating a New Teaching Profession*, pp. 165–180. Washington, DC: Urban Institute Press.

Harris, D. N. and T. R. Sass (2011). Skills, productivity and the evaluation of teacher performance. *Economics of Education Review 40*, 183–204.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science 312*, 1900–1902.

Jackson, C. K. (2012). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in north carolina. Technical Report NBER Working Paper 18624.

Jackson, C. K. (2014). Teacher quality at the high school level: The importance of accounting for tracks. *Journal of Labor Economics 32*(4), 645–684.

Electronic copy available at: https://ssrn.com/abstract=3203759

Jacob, B. A. and L. Lefgren (2008). Can principals identify effective teachers? evidence on subjective performance evaluations in education. *Journal of Labor Economics 26*(1), 101–136.

Kane, T., D. McCaffrey, T. Miller, and D.O.Staiger (2013). Have we identified effective teachers? validating measures of effective teaching using random assignment. Technical Report Bill and Melinda Gates Foundation: Seattle, WA.

Kane, T. J. and D. O. Staiger (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives 16*(4), 91–114.

Kane, T. J. and D. O. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. Technical Report NBER Working Paper No. 14607.

Koedel, C. (2009). An empirical analysis of teacher spillover effects in secondary school. *Economics of Education Review 28*(6), 682–692.

Koedel, C. and J. Betts (2011). Does student sorting invalidate value-added models of teacher effectiveness? an extended analysis of the rothstein critique. *Education Finance and Policy 6*(1), 18–42.

Koedel, C., K. Mihaly, and J. E. Rockoff (2015). Value-added modeling: A review. *Economics of Education Review 47*, 180–195.

Kraft, M. A. (2015). Teacher layoffs, teacher quality, and student achievement: Evidence form a discretionary layoff policy. *Education Finance and Policy 10*, 4467–4507.

Ladd, H. F. and R. P. Walsh (2002). Implementing value-added measures of school effectiveness: Getting the incentives right. *Economics of Education Review 21*(1), 1–17.

Lefgren, L. and D. Sims (2012). Using subject test scores efficiently to predict teacher value-added. *Education Evaluation and Policy Analysis 34*(1), 109–121.

Electronic copy available at: https://ssrn.com/abstract=3203759

McCaffrey, D. F., J. Lockwood, D. Koretz, T. A. Lewis, and L. Hamilton (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics 29*(1), 67–101.

McCaffrey, D. F., T. R. Sass, J. Lockwood, and K. Mihaly (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy 4*(4), 572–606.

Meyer, R. H. (1997). Value-added indicators of school performance: A primer. *Economics of Education Review 16*(3), 283–301.

Newton, X. A., L. Darlilng-Hammond, E. Haertel, and E. Thomas (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis 18*(3), 1–27.

Paufler, N. A. and A. Amrein-Beardsley (2014). The random assignment of students into elementary classrooms: Implications for value-added analysis and interpretations. *American Educational Research Journal 51*(2), 328–362.

Raudenbush, S. W. and J. D. Willms (1995). The estimation of school effects. *Journal of Education and Behavioral Statistics 20*(4).

Reardon, S. F. and S. W. Raudenbush (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy 4*(4).

Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy 4*(4), 537–571.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics 125*(1), 175–214.

Rubin, D. B., E. A. Stuart, and E. L. Zanutto (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics 29*(1), 103–116.

Electronic copy available at: https://ssrn.com/abstract=3203759

Sass, T. R., A. Semykina, and D. N. Harris (2014). Value-added models and the measurement of teacher productivity. *Economics of Education Review 38*(1), 9–23.

Slater, H., N. M. Davies, and S. Burgess (2012). Do teachers matter? measuring the variation in teacher effectiveness in england. *Oxford Bulletin of Economics and Statistics 74*(5), 629–645.

Todd, P. E. and K. I. Wolpin (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal 113*, F3–F33.

Electronic copy available at: https://ssrn.com/abstract=3203759

Table 1: Summary statistics for Florida education data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Below-median VA		*Above-median VA*		*All Students*		
	Mean	SD	Mean	SD	Mean	SD	N
Value-added					*0.001*	*0.109*	*508*
Female	0.500	0.500	0.505	0.500	0.502	0.500	677468
Black	0.230	0.421	0.192	0.394	0.211	0.408	677468
Hispanic	0.196	0.397	0.212	0.409	0.204	0.403	677468
Asian	0.023	0.151	0.023	0.151	0.023	0.151	677468
Other race	0.021	0.143	0.022	0.146	0.021	0.144	677468
Free/reduced lunch	0.309	0.462	0.306	0.461	0.307	0.461	677468
8th grade math score	0.133	0.939	0.129	0.915	0.131	0.927	677468
8th grade reading score	0.132	0.941	0.131	0.934	0.132	0.937	677468
HS - fraction male	0.449	0.035	0.450	0.036	0.450	0.035	677468
HS - fraction Black	0.255	0.232	0.216	0.193	0.236	0.214	677468
HS - fraction Asian	0.023	0.017	0.023	0.020	0.023	0.018	677468
HS - fraction Hispanic	0.221	0.197	0.232	0.249	0.227	0.224	677468
HS - fraction on free/reduced lunch	0.335	0.140	0.332	0.139	0.334	0.139	677468
HS - mean incoming 8th grade math scores	-0.004	0.366	0.002	0.328	-0.001	0.348	677468
HS - mean incoming 8th grade reading scores	-0.014	0.342	-0.009	0.324	-0.012	0.333	677468
HS - fraction of students taking math exam	0.911	0.018	0.913	0.020	0.912	0.019	677468
HS - fraction of student taking reading exam	0.989	0.009	0.989	0.011	0.989	0.010	677468

Summary statistics are shown for all students in the administrative data from Florida who were used to estimate the value-added models. Summary statistics are shown separately according to whether the student is from a below- (columns (1)-(2)) or above-median (columns (3)-(4)) value-added high school, as well as for all students combined (columns (5)-(7)). The top row shows the mean and standard deviation of the value-added scores, as well as the number of high schools for which a value-added score could be estimated. Value-added is scaled in units of student test score standard deviations. *Female, Black, Hispanic, Asian, Other race, Free/reduced lunch, 8th grade math score* and *8th grade reading score* are student-level independent variables in the value-added model. The other variables listed, beginning with "HS -", are high school-level independent variables. Additional information on the Florida administrative data can be found in Section 3.1 and additional information on the value-added model can be found in Section 5.1.

Table 2: Summary statistics for the dataset of analysis

	(1) *Below-median VA*	(2)	(3) *Above-median VA*	(4)	(5)	(6) *All Students*	(7)
	Mean	SD	Mean	SD	Mean	SD	N
Value-added					*0.001*	*0.097*	*365*
Female	0.584	0.493	0.582	0.493	0.583	0.493	42063
Black	0.261	0.439	0.207	0.405	0.236	0.425	42063
Hispanic	0.090	0.286	0.096	0.295	0.093	0.290	42063
Asian	0.047	0.212	0.057	0.231	0.052	0.221	42063
Other race	0.009	0.092	0.007	0.086	0.008	0.089	42063
High school GPA	3.650	0.538	3.689	0.522	3.668	0.531	41926
Graduated	0.603	0.489	0.635	0.481	0.618	0.486	8991
Final cumulative GPA	2.534	0.835	2.611	0.808	2.570	0.824	42063
HS - peer economic status	83.426	11.789	83.543	10.102	83.480	11.043	42037
HS - fraction male	0.450	0.032	0.452	0.029	0.451	0.031	42063
HS - fraction Black	0.279	0.245	0.220	0.189	0.252	0.223	42063
HS - fraction Asian	0.028	0.019	0.032	0.028	0.030	0.024	42063
HS - fraction Hispanic	0.190	0.166	0.194	0.206	0.192	0.185	42063
HS - fraction on free/reduced lunch	0.295	0.142	0.283	0.134	0.289	0.139	42063
HS - mean incoming 8th grade math scores	0.094	0.349	0.131	0.369	0.111	0.359	42063
HS - mean incoming 8th grade reading scores	0.106	0.375	0.144	0.374	0.124	0.375	42063
HS - fraction of students taking math exam	0.914	0.011	0.916	0.012	0.915	0.012	42063
HS - fraction of students taking reading exam	0.990	0.005	0.991	0.005	0.990	0.005	42063

Summary statistics are shown for all students in the MIDFIELD data who were matched to a value-added score. Summary statistics are shown separately according to whether the student is from a below- (columns (1)-(2)) or above-median (columns (3)-(4)) value-added high school, as well as for all students combined (columns (5)-(7)). The top row shows the mean and standard deviation of the value-added scores for the schools that were matched to students in the MIDFIELD data, as well as the number of schools that were matched to a student. Value-added is scaled in units of student test score standard deviations. *Female, Black, Hispanic, Asian, Other race, High school GPA, Graduated* and *Final GPA* are student-level variables from the MIDFIELD data. The other variables listed, beginning with "HS -", are high school-level independent variables. The variable *HS - peer economic status* comes from the MIDFIELD dataset, while the other high school-level variables were matched to the MIDFIELD students along with the value-added score for the high school. Additional information on the MIDFIELD data and this matched dataset of analysis can be found in Section 3.2 and Section 3.3.

Table 3: Test for bias in value-added scores due to sorting of students into high schools

	(1) High school math	(2) High school reading	(3) 7th grade math	(4) 7th grade reading	(5) High school math	(6) High school reading
Value-added	1.43***	1.36***	0.01	-0.07	1.24***	1.12***
	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)	(0.11)
7th grade math score					0.23***	0.06***
					(0.01)	(0.01)
7th grade reading score					0.02***	0.27***
					(0.01)	(0.01)
Observations	450324	451878	464704	464356	421347	422564
R-squared	0.705	0.638	0.679	0.736	0.732	0.671

Results are based on students who were included in the value-added models and also have 7th grade exam scores. This corresponds to students with 10th grade exams scores during the years 2002-2005. Each column reports the coefficient from an OLS regression for the impact of attending a higher value-added high school on K-12 test scores. The dependent K-12 test score is shown in the column heading. Each column also controls for the same independent variables included in the value-added model: gender, race, ethnicity, free-or-reduced lunch, a cubic in 8th grade math and reading scores, and high school-level controls for mean incoming 8th grade test scores, fraction male, fraction Black, fraction Asian, fraction Hispanic, fraction on free-or-reduced lunch, fraction taking each of the high school exams, and peer economic status. Year fixed effects are also included. Columns (1)-(2) show the impact of value-added on the different high school test scores used to estimate school value-added scores. Columns (3)-(4) show the impact of high school value-added on 7th grade test scores. Columns (5)-(6) show the impact of high school value-added on the high school test scores with additional controls for 7th grade test scores. Standard errors, shown in parentheses, are clustered at the high school level.

Electronic copy available at: https://ssrn.com/abstract=3203759

Table 4: Impact of high school value-added on high school outcomes

	(1) SAT	(2) ACT	(3) High school GPA
Value-added	124.52*	5.62***	-0.20
	(68.71)	(1.62)	(0.30)
Female	-35.65***	-0.81***	0.15***
	(1.85)	(0.05)	(0.01)
Black	-200.50***	-5.88***	-0.56***
	(4.69)	(0.11)	(0.01)
Hispanic	-44.94***	-1.11***	-0.07***
	(3.08)	(0.08)	(0.01)
Asian	-5.32	-0.14	0.08***
	(4.10)	(0.10)	(0.01)
Other race	-16.04*	-0.54***	-0.11***
	(8.65)	(0.17)	(0.03)
Year FE	Yes	Yes	Yes
High school controls	Yes	Yes	Yes
Observations	25,883	38,721	41,900
R-squared	0.332	0.409	0.271

Each column reports the coefficients from an OLS regression for the effect of attending a higher value-added high school on high school outcomes for students in the MIDFIELD database. Standard errors, shown in parentheses, are clustered at the high school level. Additional information on sample restriction can be found in Section 6.

Electronic copy available at: https://ssrn.com/abstract=3203759

Table 5: Impact of high school value-added on college graduation

	(1)	(2)	(3)	(4)	(5)
Value-added	0.62**	0.54**	0.53**	0.57**	0.57**
	(0.26)	(0.26)	(0.25)	(0.25)	(0.29)
Female	0.09***	0.09***	0.10***	0.11***	0.07***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Black	-0.26***	-0.25***	-0.11***	-0.10***	-0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Hispanic	-0.06***	-0.07***	-0.05***	-0.06***	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Asian	-0.01	-0.02	-0.02	-0.03	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Other race	-0.13***	-0.13***	-0.13***	-0.14***	-0.10**
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)
High school GPA					0.25***
					(0.01)
Impact of 1 SD increase in VA	*0.07*	*0.06*	*0.06*	*0.06*	*0.06*
Percent change over graduation mean	*11.32%*	*9.71%*	*9.71%*	*9.71%*	*9.71%*
Year FE	Yes	Yes	Yes	Yes	Yes
High school controls	No	Yes	Yes	Yes	Yes
University FE	No	No	Yes	No	No
Major FE	No	No	Yes	No	No
University-by-major FE	No	No	No	Yes	Yes
Observations	8991	8988	8980	8750	8638
Pseudo r-squared	0.051	0.054	0.134	0.111	0.158

Each column reports the marginal effects from a logit model for the impact of attending a higher value-added high school on college graduation. Column (1) shows the correlation between high school value-added and college graduation with student-level covariates and year fixed effects. Column (2) adds the high school-level covariates shown in Table 2. Columns (3)-(4) add major and university fixed effects. Column (5) adds high school GPA as a control for non-test-score skills. The table also shows the results in terms of one standard deviation increase in value-added and the percent change over the mean of the graduation variable that this increase represents. These are constructed by multiplying the coefficient on the value-added variable by the standard deviation of the value-added scores in Table 1 and dividing the resulting number by the mean of the graduation variable in Table 2 then multiplying by 100, respectively. Standard errors, shown in parentheses, are clustered at the high school level. Additional information on sample restriction can be found in Section 7.

Electronic copy available at: https://ssrn.com/abstract=3203759

Table 6: Impact of high school value-added on final cumulative college GPA

	(1)	(2)	(3)	(4)	(5)
Value-added	0.86**	0.77***	0.52**	0.48**	0.77**
	(0.39)	(0.27)	(0.23)	(0.23)	(0.31)
Female	0.25***	0.25***	0.22***	0.22***	0.11***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Black	-0.65***	-0.60***	-0.30***	-0.29***	-0.10***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Hispanic	-0.11***	-0.12***	-0.11***	-0.11***	-0.06***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Asian	-0.00	-0.01	-0.08***	-0.09***	-0.10***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Other race	-0.22***	-0.21***	-0.23***	-0.23***	-0.15***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
High school GPA					0.68***
					(0.01)
Impact of 1 SD VA increase	*0.09*	*0.08*	*0.06*	*0.05*	*0.08*
Percent change over mean GPA	*3.50%*	*3.11%*	*2.33%*	*1.95%*	*3.11%*
Year FE	Yes	Yes	Yes	Yes	Yes
High school controls	No	Yes	Yes	Yes	Yes
University FE	No	No	Yes	No	No
Major FE	No	No	Yes	No	No
University-by-major FE	No	No	No	Yes	Yes
Observations	42063	42037	42037	42037	41900
R-squared	0.128	0.137	0.248	0.260	0.376

Each column reports the coefficients from an OLS regression for the impact of attending a higher value-added high school on final cumulative college GPA. Column (1) shows the correlation between high school value-added and final cumulative GPA with student-level covariates and year fixed effects. Column (2) adds the high school-level covariates shown in Table 2. Columns (3)-(4) add major and university fixed effects. Column (5) adds high school GPA as a control for non-test-score ability. The table also shows the results in terms of one standard deviation increase in value-added and the percent change over the mean of the final cumulative GPA variable that this increase represents. These are constructed by multiplying the coefficient on the value-added variable by the standard deviation of the value-added scores in Table 1 and dividing the resulting number by the mean of the final cumulative GPA variable in Table 2 then multiplying by 100, respectively. Standard errors, shown in parentheses, are clustered at the high school level. Additional information on sample restriction can be found in Section 7.

Electronic copy available at: https://ssrn.com/abstract=3203759

Table 7: Impact of high school value-added on college GPA by semester

	(1) Semester 1	(2) Semester 2	(3) Semester 3	(4) Semester 4	(5) Semester 5	(6) Semester 6	(7) Final GPA
Value-added	0.75**	0.82**	0.75**	0.63*	0.34	0.47	0.77**
	(0.37)	(0.39)	(0.36)	(0.37)	(0.36)	(0.39)	(0.31)
Female	0.07***	0.07***	0.09***	0.12***	0.11***	0.09***	0.11***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Black	-0.01	-0.05**	-0.12***	-0.14***	-0.21***	-0.20***	-0.10***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Hispanic	-0.05***	-0.04**	-0.05***	-0.11***	-0.08***	-0.05**	-0.06***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Asian	-0.10***	-0.10***	-0.09***	-0.10***	-0.13***	-0.13***	-0.10***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)
Other race	-0.14**	-0.11**	-0.09	-0.18***	-0.16**	-0.09	-0.15***
	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)	(0.04)
High school GPA	0.57***	0.86***	0.73***	0.64***	0.60***	0.55***	0.68***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
University-by-major FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	41322	40150	33719	29715	26256	22197	41900
R-squared	0.284	0.271	0.243	0.217	0.190	0.164	0.376

Each column is a separate regression showing the impact of attending a higher value-added high school on college GPA by semester, as indicated in the column headings. In addition to the independent variables shown in the table, each column also controls for the high school-level covariates shown in Table 2, university-by-major fixed effects based on the student's major in the given semester, and year fixed effects. Standard errors are clustered by high school.

Table 8: Impact of high school value-added on college GPA by semester, consistent sample

	(1) Semester 1	(2) Semester 2	(3) Semester 3	(4) Semester 4	(5) Semester 5	(6) Semester 6	(7) Final GPA
Value-added	0.27	0.69*	0.90**	0.74**	0.19	0.47	0.59**
	(0.39)	(0.38)	(0.36)	(0.33)	(0.33)	(0.39)	(0.26)
Female	0.02*	0.02**	0.07***	0.10***	0.09***	0.09***	0.09***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Black	-0.05**	-0.09***	-0.13***	-0.17***	-0.23***	-0.20***	-0.17***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Hispanic	-0.06***	-0.04*	-0.04**	-0.11***	-0.09***	-0.06**	-0.07***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Asian	-0.15***	-0.09***	-0.10***	-0.11***	-0.15***	-0.13***	-0.12***
	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Other race	-0.18***	-0.15***	-0.04	-0.20***	-0.15**	-0.09	-0.14***
	(0.07)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.04)
High school GPA	0.48***	0.75***	0.64***	0.58***	0.59***	0.55***	0.54***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
University-by-major FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21380	21380	21380	21380	21380	21380	21380
R-squared	0.276	0.253	0.224	0.204	0.191	0.165	0.403

The table above reproduces the results from Table 7, except with a consistent sample of students across columns. This limits the sample to students who have GPA observations for each of semesters 1-6 and final cumulative GPA.

Table 9: Impact of high school value-added on graduation and college GPA by semester, by race and gender

	(1) Graduated	(2) Semester 1	(3) Semester 2	(4) Semester 3	(5) Semester 4	(6) Semester 5	(7) Semester 6	(8) Final GPA
Panel I: Male	0.83*	1.09**	0.96**	0.98*	0.70	0.31	-0.49	0.98**
	(0.50)	(0.48)	(0.44)	(0.50)	(0.49)	(0.47)	(0.49)	(0.39)
	0.09	*0.12*	*0.10*	*0.11*	*0.08*	*0.03*	*-0.05*	*0.11*
	15.79%	*4.61%*	*3.93%*	*4.31%*	*3.15%*	*1.17%*	*-1.92%*	*4.50%*
	3563	17257	16753	13969	12264	10782	9117	17471
Panel II: Female	0.36	0.54	0.75*	0.65*	0.56	0.33	1.14**	0.65**
	(0.32)	(0.37)	(0.43)	(0.38)	(0.40)	(0.41)	(0.53)	(0.32)
	0.03	*0.06*	*0.08*	*0.07*	*0.06*	*0.04*	*0.12*	*0.07*
	4.62%	*2.20%*	*2.94%*	*2.54%*	*2.16%*	*1.43%*	*4.27%*	*2.63%*
	5044	24065	23397	19750	17451	15474	13080	24429
Panel III: White	0.47	0.42	0.44	0.58	0.56	0.36	0.73	0.44
	(0.32)	(0.43)	(0.44)	(0.46)	(0.46)	(0.44)	(0.53)	(0.39)
	0.05	*0.05*	*0.05*	*0.06*	*0.06*	*0.04*	*0.08*	*0.05*
	7.26%	*1.76%*	*1.78%*	*2.11%*	*2.11%*	*1.40%*	*2.79%*	*1.83%*
	5535	25231	24667	20636	18394	16401	13921	25619
Panel IV: Black	0.56	1.45***	1.33***	1.10**	0.98	0.60	0.61	1.34***
	(0.49)	(0.52)	(0.51)	(0.52)	(0.61)	(0.77)	(0.68)	(0.38)
	0.06	*0.16*	*0.14*	*0.12*	*0.11*	*0.07*	*0.07*	*0.15*
	14.57%	*7.34%*	*5.54%*	*5.32%*	*4.98%*	*3.15%*	*3.06%*	*7.16%*
	1856	9783	9344	7857	6759	5809	4831	9906

Each panel-by-column is a separate regression for the impact attending a higher value-added high school on the outcome variable shown in the column heading, separately by gender and race as shown in each panel. Each panel-by-column reports five stacked numbers, which represent the coefficient for the value-added variable, the standard error, the coefficient translated into the impact of one standard deviation increase in value-added, the percentage increase over the sample mean that the one-standard-deviation-increase impact represents, and the number of observations. The percentage increase over the sample mean for each regression is based on the sample mean for the group specified in the panel. The regressions controls for the same independent variables as previous tables: *Female, Black, Hispanic, Asian, Other race, High school GPA*, as well as high school-level controls, university-by-major fixed effects based on the student's major in the given semester, and year fixed effects..