

Making a Tough Choice: Teacher Target-Setting and Student Achievement in a Teacher Performance System Using Student Learning Objectives

Allison Atteberry 

University of Colorado Boulder

Sarah E. LaCour

University of Kentucky

The use of student learning objectives (SLOs) as part of teacher performance systems has gained traction quickly in the United States, yet little is known about how teachers select specific students' learning goals. When teachers are evaluated—and sometimes compensated—based on whether their students meet the very objectives the teachers set at the start of the year, there may be an incentive to set low targets. SLO systems rely on teachers' willingness and ability to set appropriately ambitious SLOs. We describe teachers' SLO target-setting behavior in one school-district. We document the accuracy/ambitiousness of targets and find that teachers regularly set targets that students did not meet. We also find that, within the same year, a student's spring test scores tend to be higher on the assessments for which they received higher targets. This raises the intriguing possibility that receiving higher targets might cause students to perform better than they otherwise would have.

Keywords: *descriptive analysis, educational policy, regression analyses, secondary data analysis, teacher assessment, teacher research*

OVER the past 15 years, U.S. federal initiatives like *Race to the Top* (RTTT) and the *Every Student Succeeds Act* (ESSA) have pushed states and districts to incorporate individual measures of student growth into teacher evaluation systems (Berg-Jacobson, 2016). That requirement is often met by linking teachers to their students' growth on end-of-year (EOY) statewide assessments (e.g., value-added measures [VAMs]), particularly in math and English language arts (ELA) in Grades 4 to 8. Yet this strategy raises at least two challenges. The first is logistical: VAMs can only be produced for teachers in tested subjects and grades—about 20% to 30% of a district's teachers. A second concern is that evaluation systems linked solely to EOY tests fail to incorporate teachers' professional expertise and knowledge of their students.

There have been a number of responses to these challenges but one that may address both is implementing student learning objectives (SLOs) systems. District SLO systems mandate a process by which teachers establish concrete, annual learning targets for their students, typically on some assessment chosen at the classroom, school, or district level. Since the SLO process does not necessarily rely on statewide tests, the approach can be used to cover a broader spectrum of teacher roles. Because teachers are more directly involved determining what reasonable goals should be, there

is more reliance on the teacher's professional judgment and their knowledge of the student's context. The act of setting clear goals, along with having ambitious expectations for learners, may help teachers inform and improve their daily practice. These aspects of SLOs make them a potential lever for addressing the shortcomings of teacher evaluation systems that rely solely on teacher VAMs described above.

Little is known about how SLO systems are implemented in practice. In particular, teachers are not necessarily trained on how to select appropriate learning objectives on the given assessment their school or district has adopted. Moreover, when teachers are evaluated based on whether their students meet the very objectives the teachers set at the beginning of the school year, there appears to be a perverse incentive to set low targets. The success of SLO systems relies on the premise that teachers are willing and able to set appropriately ambitious goals for their students each year.

We describe target-setting behaviors within an SLO system piloted for 4 years in a subset of schools in a southeastern school district. We describe this district's SLO system in greater detail below, but it is relevant to know at the outset that this district also used SLOs for a pay-for-performance policy; if fewer than 50% of a teacher's students did not meet the targets, they would not receive any bonus (up to \$5,000).



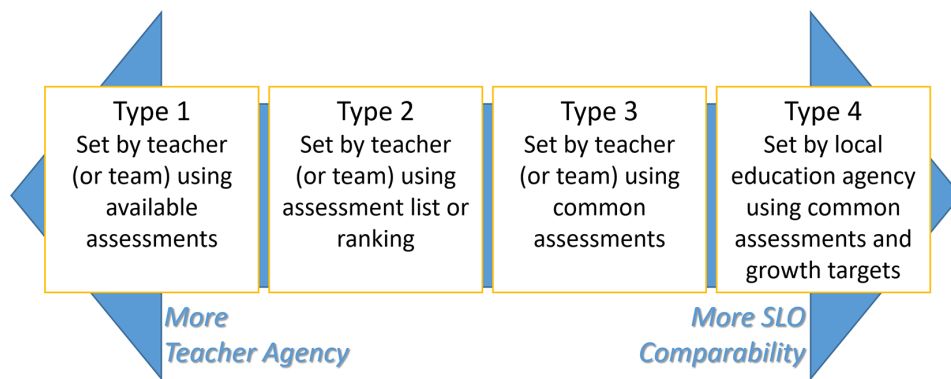


FIGURE 1. *Continuum of approaches to setting student learning objectives.*
 Note. Image adapted broadly, based on Lachlan-Haché et al. (2013).

How will teachers choose to enact a system that may run counter to their own interests (i.e., setting higher targets could lead to lower evaluations, losing a bonus)? The purpose of this article is therefore twofold: We provide new descriptive insights on how teachers set their SLOs for different students and on different assessments, and we examine associations between SLO targets and student outcomes.

SLO Policy Landscape

SLOs have gained traction quickly in part because several of the U.S. Department of Education’s grant initiatives (e.g., NCLB waivers, RTTT, the Teacher Incentive Fund [TIF] grant) have endorsed SLOs as a way to satisfy the requirement to measure student growth in nontested subjects and grades (U.S. Department of Education, 2012). There is considerable variability in what could constitute an SLO system, but the DoE has described the SLO process as

[a] participatory method of setting measurable goals, or objectives, based on the specific assignment or class, such as the students taught, the subject matter taught, the baseline performance of the students, and the measurable gain in student performance during the course of instruction. (RTTT Technical Assistance Network, 2010)

Inherent in this definition is room for flexibility about who chooses the assessments and the extent to which assessments must be comparable across classrooms and schools.

As of 2015, about 30 states already either recommend or require SLOs (Lachlan-Haché, 2015). We are aware of one report that gathers information on SLO implementation in educator evaluator systems across states (Lacireno-Paquet et al., 2014). The current district’s SLO system has many of the most common features reported therein. Of the 30 states studied, the report finds that 23 use SLOs to evaluate individual teachers, as opposed to teams of teachers or the entire school. In 26 states, the assessment used for SLOs was not chosen by teachers—rather, SLOs were based on a test used statewide ($N = 14$) or district/school-wide ($N = 12$). In 21 states, the SLO-setting process required that SLOs are approved by an external evaluator, usually a principal or

district leader. (A principal approves SLOs in the current district.) SLOs are sometimes linked to performance incentives, particularly among districts receiving TIF grants. It is not clear how common this practice is across the United States, but several other large districts have experimented with this, including Denver Public Schools, Charlotte-Mecklenburg schools, and Austin Independent School District (Lachlan-Haché et al., 2015).

In Figure 1, we reproduce a useful illustration initially developed by Lachlan-Haché et al. (2013) and widely adapted, which places SLO policies on a continuum: On the left side of Figure 1 are SLO approaches that emphasize greater teacher agency, while those on the right side of Figure 1 emphasize greater comparability. The current district’s SLO process falls on the right side of the Figure 1 continuum because teachers do not select an assessment of their choice.

Early research documents some of the challenges districts experience with SLO implementation (Delaware Department of Education, 2013; Donaldson et al., 2014; Lachlan-Haché, 2015; Lachlan-Haché et al., 2013; Lamb et al., 2013; Schmitt, Cornetto, Lamb, & Imes, 2009; Slotnik, Smith, & Liang, 2013).¹ However, because these systems are relatively new, little research has yet attempted to describe teacher goal-setting behavior.

Unique Contribution and Research Questions

One of our objectives is to fill this gap by documenting the accuracy and ambitiousness of the learning targets teachers set for their students, capturing the extent to which teachers are willing to set targets they may not reach (thereby reducing their bonus pay), and considering whether target-setting shifts across study years, varies by teacher effectiveness (proxied with VAMs), or is approached differently across the seven pilot schools.

The district we study provides a number of advantages for these questions. First, their particular version of an SLO process uses the *same* assessments for each grade level across all schools, enabling us to compare target setting on the same

assessments across classrooms. Second, this district piloted an SLO system in which teachers had to select and document very specific targets—that is, in the fall, teachers combined their knowledge of their students with a review of the student’s historical data to identify a score target for each child on each assessment (e.g., Grade 4 Teacher X sets as a goal that her Student Y will score a 221 on the math test at the end of this year).² This practice provides an opportunity to explore variability in objectives across students in the same class. Finally, in Grades 4 through 8, teachers set targets on at least two different EOY tests (both in math and ELA), providing an additional source of variability.³

The goal of the current article is to provide the field with a thorough description of target-setting in an SLO system where not meeting targets is linked to decreased incentive pay. Because these pilot schools were selected based on low prior performance and high percentages of free-/reduced-price lunch (FRPL) eligible students, we also examine SLO target setting in a context under pressure to raise student outcomes. We organize our analyses into three research objectives:

1. Describe accuracy/ambitiousness of targets, percent of targets met, and variation in target-setting behaviors across schools, teachers, and study years
2. Document whether receiving higher targets predicts higher EOY test scores, for similar/same students/schools
3. Assess possible evidence of differential target-setting by race/ethnicity, conditional on prior achievement profile

Next, we provide details on the district and its SLO policy. Our literature review focuses on the potential mechanisms through which goal-setting might influence student outcomes. We then describe the data and methods used to explore these research questions, present our findings, and conclude with a reflection on takeaways, limitations, and next steps.

The District and Its SLO Policy

The anonymous district in which this study takes place includes one of the largest cities and its surrounding area in a southeastern state. It has about 60 schools and serves around 50,000 students annually in a mix of urban, suburban, and rural communities. In 2012, the majority (55%) of the students in the district were non-White. Specifically, 36.9% of students were designated as Black, 6.5% as Asian, 4.9% as Hispanic, and 6.1% as mixed or unknown race/ethnicity. The district covers a wide geographic area and has both high-poverty and low-poverty communities.

In 2010, the district was awarded a TIF grant to implement and evaluate the impact of a new performance-based employee compensation system in high-needs schools. Eight of the district’s highest needs schools were selected to participate (more on sample below). Given TIF’s focus on incentive pay, the district needed a way to connect

compensation to teacher performance in nontested subjects and grades, and SLOs were selected to fill that role.

Beginning in 2011–2012, teachers created a “learning contract” with their principal for each student in their class. Teachers did so by first examining each student’s test scores from the prior spring.⁴ The teacher then combined that information with his/her own insight into the student’s context. By mid-November, the teacher committed to an annual growth target for each student.⁵ Teachers were not provided formal training on setting SLO targets. Because targets were reviewed by the principal, there was some oversight of teachers setting inappropriately low goals. We also will show that students did not meet the targets teachers set for them about 40% of the time, suggesting that teachers did not shy away from setting difficult-to-achieve targets.

The district opted to use uniform assessments across all students in the same grade. Each Grade 4 and 5 teacher selected score targets on four different assessments for each student each year—math and ELA scale score targets on the statewide standardized achievement test (pseudonym SSAT) and NWEA’s Measures of Academic Progress (MAP) spring math and ELA assessment.⁶ In Grades 6 to 8, a student’s math teacher set the two math targets (SSAT and MAP), while a different ELA teacher set the 2 ELA targets.⁷ SSAT results are high stakes since they play a central role in the state’s accountability policy, while the purpose of the MAP test is more formative. That said, since both the SSAT and MAP were used to determine the size of SLO incentive payments, both assessments had stakes attached to them in this context.

Teachers received incentive pay based on the percentage of their students who met their targets. If at least 50% of the teachers’ students met their target, they received a bonus between \$2500 and \$5000.⁸ If less than 50% of a teacher’s students met their target, they received no bonus. The average annual base salary for teachers was around \$49,500 (and an *SD* of about \$10,900), so that a \$5000 bonus is sizeable—about 10% of base pay for the average teacher.

Potential Mechanisms for SLO Impacts

A key driver of any link between SLO systems and improved student outcomes is the potential positive impacts of teachers setting high expectations for students. Researchers have long explored the positive association between teacher expectations and student outcomes (Jussim & Harber, 2005; Madon et al., 1997). Perhaps the best-known example is the Rosenthal and Jacobsen experiment from the mid-1960s in which the randomly selected students described to teachers as high-growth potential *did* exhibit higher EOY test scores (Rosenthal, 1987; Rosenthal & Jacobson, 1968).⁹

Of course, teachers’ expectations are usually not randomly generated. Just because the students for whom the teachers have higher expectations tend to perform better does not imply that high expectations *cause* high performance. Teacher expectations may correlate with achievement simply because

teachers can accurately anticipate subsequent performance. This ambiguity complicates any attempt to make a causal link between setting ambitious learning objectives and student outcomes (Jussim & Harber, 2005; Papageorge et al., 2020). To tackle this, Papageorge et al. (2020) leveraged data from a nationally-representative 2002 cohort of 10th graders—wherein two different teachers provided their educational expectations for the *same* student. This conditionally exogenous variation in expectations within student was linked to increased likelihood of college completion, suggesting that teacher expectations indeed become so-called self-fulfilling prophecies.¹⁰

Even if we accept the premise that teacher expectations exert an independent, positive impact on student outcomes, the underlying mechanisms are not entirely clear. High expectations might lead a teacher to alter the way they deliver instruction to a student, either consciously or unconsciously. For instance, Proctor (1984) found that

[t]eachers are less apt to direct instruction to low-expectation students, are less likely to be aware of, or more likely to tolerate, non-attending behavior on the part of such students, and tend to place fewer demands on them for classroom performance, homework assignments, and overall academic effort. (p. 472).

Teachers might also be directly communicating their high expectations to the given student, who may in turn alter their own behavior. The study by Papageorge et al. (2020) lends support to this potential pathway, since they found that high expectations led to students spending slightly more time on their homework. When teachers make their high expectations known, students might also find themselves reevaluating their *own* self-beliefs about what they can accomplish. In this sense, teachers' expressed beliefs about a student's capacity for growth also connects to the literature on the benefits of a growth mindset (see, e.g., Dweck, 2008).¹¹ Seaton (2018) provides a conceptual framework grounded in psychological theory for how a teacher's growth mindset can stimulate their students' growth mindsets, thus providing another potential pathway for positive effects of teacher expectations. Consistent with this idea, Papageorge et al. (2020) found that students exposed to higher teacher expectations also held higher expectations of their own educational prospects 2 years later.

Potential Bias in Teacher Expectations

Teacher expectations can also be harmful if they are shaped by implicit biases (Proctor, 1984). Indeed, it has been shown in many contexts that people have implicit association biases of which they themselves are not aware (Correll et al., 2007; Green et al., 2007; Tyler Eastman & Billings, 2001). Teachers may be expecting students to act or perform in accordance with their biases and may disregard contradictory evidence of change (Ferguson, 2003; Proctor, 1984; Rubie-Davies et al., 2006).

Burgess and Greaves (2013) detected troubling racialized patterns when comparing blinded versus nonblinded assessments of ethnic minority students' work. Likewise, Gershenson

et al. (2016) found that a Black student receives systematically lower expectations from non-Black teachers than Black teachers, particularly for male students in math. Their work is part of a growing body of literature that documents the importance of students being taught by teachers of their same race to academic and behavioral outcomes (e.g., Dee, 2004, 2005; Egalite et al., 2015; Gershenson et al., 2018; Holt & Gershenson, 2019). We cannot conduct a so-called race-match analysis because our data set does not include teacher demographics. However, given the potential for high expectations to improve student outcomes, we *do* explore differential target setting by student race/ethnicity.

Sample and Methods

Analytic Sample

The SLO process began in 2011–2012 as part of a TIF Initiative, which was targeted to high-needs schools. Of the district's 64 schools, 23 met the “high-needs” TIF definition—more than 50% of the school's students are FRPL-eligible—and eight schools were selected to participate (five elementary, two middle, one high). In this study, we examine all learning targets set in the seven SLO schools serving Grades 4 through 8 (i.e., all schools except the high school).

In Table 1, we present average pretreatment means (left) and trends (right) in school-level demographics and standardized test scores from 2007–2008 through 2010–2011—the pre-SLO period—for the group of seven SLO schools. To provide the reader with a sense of how these schools compared with other district schools serving Grades 4 through 8, we also present these descriptives for the other 16 high-needs schools, as well as the remaining 34 K–8 schools. Table 1 shows that the seven SLO schools served a higher percentage of FRPL-eligible students than did even their other high-need counterparts. The SLO schools were also, on average, lower performing on state tests than other schools in the district.

In Table 2, we present counts of unique students and teachers in the analytic sample, overall, by year, grade, school, and—for students—by race/ethnicity (the data set does not include teacher demographics). The teacher counts capture the unique number of teachers who are responsible for teaching and setting targets for students. In Grades 4 and 5, the number of unique ELA and math teachers is exactly the same because a single teacher provides all four targets (both tests, both subjects). In Grades 6 through 8, a different teacher provides a students' two math targets (both tests) than the teacher who provides that students' two ELA targets. For this reason, we separately analyze elementary and middle schools when looking across all four targets. In this sample, no students were assigned to the same teacher for more than 1 year.

Methods

Research Question 1: Target-Setting Descriptives. We begin by presenting descriptives for target setting in the current district. This includes the number of targets set and

TABLE 1

Compare Seven SLO Schools to District's 16 Other "High Needs" Schools, and All Other District Schools Serving Grades 4 to 8, Preonset of SLO System in 2012

| Descriptive | 2008–2011 Pretreatment means | | | 2008–2011 Pretreatment trends | | |
|-------------------------------------|------------------------------|----------------------------|------------------|-------------------------------|----------------------------|------------------|
| | Seven SLO schools | Other 16 high need schools | Other 34 schools | Seven SLO schools | Other 16 high need schools | Other 34 schools |
| Percent White | 6.8 | 28.2 | 61.4*** | -1.3 | -0.3 | -0.9 |
| Percent Black | 84.6 | 56.7 | 21.6*** | 4.2 | 1.7 | 0.7%** |
| Percent Latino/(a) | 3.1 | 8.5 | 5.4* | 0.5 | 0.9 | 0.5 |
| Percent Asian | 0.7 | 3.0 | 9.3*** | 0.2 | -0.1 | 0.9* |
| Percent FRPL-eligible | 77.4 | 67.5 | 27.0*** | 5.2 | 5.1 | 2.5** |
| Percent limited english proficiency | 2.7 | 10.4 | 8.6 | 1.1 | 0.2 | 0.9 |
| ELA statewide test score | -0.620 | -0.305 | 0.184*** | -0.020 | 0.001 | -0.002 |
| Math statewide test score | -0.658 | -0.263 | 0.184*** | 0.035 | -0.027 | -0.005 |
| History statewide test score | -0.635 | -0.314 | 0.180*** | -0.016 | 0.006 | 0.001 |
| Science statewide test score | -0.696 | -0.352 | 0.205*** | 0.024 | 0.049 | -0.015 |

Note. The seven SLO schools do not include high schools, and therefore we do not include the district's high schools in the comparison groups. Test scores on the statewide assessment, statewide standardized achievement test, have been standardized within this district, by subject, grade, and year, and reported school means are among students in Grades 3 to 8. The significance test corresponds to the null hypothesis that the three school groups do not have statistically distinguishable means of the given outcome, however, with only 56 schools, the tests are somewhat underpowered. SLO = student learning objectives; ELA = English language arts; FRPL = free/reduced-price lunch.

* $p < .05$. ** $p < .01$. *** $p < .001$.

TABLE 2

Counts of Unique Students and Teachers, Overall and by Year, Grade, School, and Student Race/Ethnicity

| Characteristic | Number of distinct . . . | | |
|----------------------------------|--------------------------|--------------------|---------------------|
| | . . . Students | . . . ELA teachers | . . . Math teachers |
| Overall | 5,123 | 158 | 155 |
| SY 2012 | 2,386 | 66 | 63 |
| SY 2013 | 2,404 | 63 | 63 |
| SY 2014 | 2,310 | 64 | 70 |
| SY 2015 | 2,432 | 75 | 76 |
| Grade 4 | 1,335 | 50 | 50 |
| Grade 5 | 1,368 | 39 | 39 |
| Grade 6 | 2,337 | 25 | 32 |
| Grade 7 | 2,252 | 34 | 44 |
| Grade 8 | 2,209 | 29 | 35 |
| Elementary School A | 289 | 14 | 14 |
| Elementary School B | 384 | 21 | 21 |
| Elementary School C | 427 | 24 | 24 |
| Elementary School D | 401 | 18 | 18 |
| Elementary School E | 428 | 20 | 20 |
| Middle School F | 2,002 | 40 | 43 |
| Middle School G | 1,960 | 50 | 49 |
| Student race/Ethnicity: Asian | 56 | N/A | N/A |
| Student race/Ethnicity: Black | 4,478 | N/A | N/A |
| Student race/Ethnicity: Hispanic | 186 | N/A | N/A |
| Student race/Ethnicity: White | 330 | N/A | N/A |

Note. Analytic sample of number of unique students who received targets during the 4 years of the study, along with the number of unique teachers who assign targets in the given subject. Note that the data set does not include teacher race/ethnicity or other teacher characteristics. ELA = English language arts; SY = school year.

met overall and by test/subject, year, grade, and school. We next compare the *target* scores teachers set for their students on each of the four assessments to students’ *actual* scores on those EOY assessments. This allows us to introduce two key terms to characterize the targets: the “accuracy” and “ambitiousness” of each target set for test t for student i . In terms of capturing accuracy, the closer a start-of-year target is to the EOY actual score, the more accurate the target was:

$$Accuracy_{it} = EOYscore_{it}^{target} - EOYscore_{it}^{actual} \quad (1)$$

In Equation (1), accuracy is equal to the EOY target score set for test t for student i , $EOYscore_{it}^{target}$, minus the actual EOY score, $EOYscore_{it}^{actual}$, on the same test t for student i .¹² When we use the term, “accurate,” we are capturing both the teacher’s ability to simply predict a student’s score at the end of the year *regardless* of the teacher’s expectations, along with their ability to anticipate how their expectations may influence their students’ performance. We cannot disentangle the two but both affect a teacher’s accuracy in predicting future student outcomes.

To capture target ambitiousness, we begin by estimating an expected EOY score based on student prior year achievement and characteristics.¹³ The higher a teacher sets the target for an EOY test score, relative to this statistical prediction, the more ambitious the target:

$$Ambitious_{it} = EOYscore_{it}^{target} - \widehat{EOYscore_{it}^{predicted}} \quad (2)$$

This comparison captures whether targets are above or below what students would be expected to score based on prior performance. We describe targets along these two dimensions overall, and whether targets are becoming more accurate or ambitious across over time. Finally, we examine variation in target-setting behaviors at the school and teacher level.

Research Question 2: Do Targets Predict EOY Scores? For two observationally similar students of the same prior skill, do we descriptively observe that having a higher target set at the start of the current year predicts higher test score performance at the end of the year? If a positive relationship exists, it suggests one of two mechanisms—or both—could be at work: One possibility is that teachers are able to bring to bear unobservable (to the district), local evidence to anticipate differences in how two students who are *observationally* similar will perform at the end of the year and set targets accordingly. Another possibility is that the experience of being given a higher target—that is, higher expectations—stimulates test score gains. If we find no conditional association between targets and scores, it suggests neither of these mechanisms is at play.

We adopt the following model to address this question, separately for each of the four tests on which targets are set for each student (MAP and SSAT, in math and ELA):

$$A_{igsy} = \alpha + \delta(TarScore_{igsy}) + \beta(A_{igs,y-1}) + (X_{i(y)})\gamma + (S_{sy})\Gamma + \Gamma_g + \Omega_y + \epsilon_{igsy} \quad (3a)$$

For instance, let A_{igsy} , be the observed math MAP achievement score for student i in grade g in school s in year y , standardized within subject-grade-year. The primary predictor of interest, $TarScore_{igsy}$, is the continuous MAP math scale score target set by student i ’s math teacher in grade g in school s at the start of year y . Importantly, we control for the student’s standardized achievement scores from the spring of the prior year, $A_{igs,y-1}$. We also include vectors of student ($X_{i(y)}$) and school-level (S_{sy}) time-invariant and time-varying covariates, and, as well as grade (Γ_g) and year (Ω_y) fixed effects.

In addition to the covariate-adjusted model (3a), we explore three other iterations of this model: The first includes teacher-by-year fixed effects (ϕ_{py}) to only compare students taught by the same teacher in the same year (Model 3b). The second adds student fixed effects (θ_i) to explore whether the same student appears to perform better in years in which a higher learning targets were set for them (Model 3c). For the third, we leverage the fact that, in every school year, four different targets are set for each student—SSAT and MAP tests in both math and ELA. Therefore, even within the same student year, variation exists in the targets set for the student. Up to this point, we have described estimating models separately for these four outcomes. For this final model, we stack all four EOY test scores as outcomes in the same model and predict them as a function of the target set for each particular test. We include dummies to track subject (j) and assessment (k) and replace student-year covariates with student-year fixed effects (ζ_{jy}) to ensure that $\hat{\delta}$ is estimated off of variation in the targets set across different tests and subjects for the same student in the same year (Model 3d). See Table 3 for a summary of these four models and the variation in targets they leverage.

Research Question 3: Differential Target-Setting by Race/Ethnicity. In light of research suggesting that teacher expectations may be susceptible to bias, and those expectations may in turn affect student outcomes, we explore whether teachers set lower targets for otherwise similar, non-White students. We begin with an overly-simplistic approach, predicting learning targets, $TarScore_{igsy}$, as a function of a set of race/ethnicity indicators (White students as the omitted category), along with grade and year fixed effects. We now return to estimating results separately by test and subject:

$$TarScore_{igsy} = \alpha + \delta_1(Black_i) + \delta_2(Asian_i) + \delta_3(Hispanic_i) + \delta_4(Other_i) + \Gamma_g + \Omega_y + \epsilon_{igsy} \quad (4a)$$

Equation (4a) provides a baseline for establishing observed differences in targets set for various race/ethnic groups relative to White students (the δ coefficients on indicators of

TABLE 3

Summary of Four Models for Research Question 3 Predicting Spring Test Scores as a Function of Variability in Targets Set for Spring

| RQ | Sample | Student-level covariates | School-level covariates | Fixed effects | Source of variation in target scores |
|------|--|---------------------------------|-------------------------|------------------------------------|---|
| (3a) | 7 SLO Schools During 4 Years of Target-Setting | Yes | Yes | Grade, Year | Among the 7 SLO Schools in 4 SLO years: Observably "similar" students in observably "similar" SLO schools who are given different Target scores (+ grade, year adjustments) |
| (3b) | " " | Yes | Yes | Grade, Year, Teacher-by-Year | Among the 7 SLO Schools in 4 SLO years: Observably "similar" students taught by the <u>same teacher in the same year</u> but their teacher gives them different Target scores (+ grade, year adjustments) |
| (3c) | " " | All but time-invariant controls | Yes | Grade, Year, School, Student | Among the 7 SLO Schools in 4 SLO years: <u>The same student in the same SLO school</u> who is given different Target scores in different years (+ grade, year adjustments) |
| (3d) | 7 SLO Schools in 4 SLO Years (4 Test-Subjects Stacked) | No, collinear | No, collinear | Subject, Test, Grade, Student-year | Among the 7 SLO Schools in 4 SLO years: The same student in the same year, who is given different target scores on different test-subject combinations (+ grade, adjustments) |

students' racial/ethnic categorization according to district records), but the observed, highly significant differences alone should not be interpreted as evidence of bias. The challenge here is that, for a variety of reasons related to long-standing structural inequities that powerfully shape how non-White students and their families experience life in the United States, non-White students *do* indeed perform lower on achievement tests. Teachers' targets that acknowledge these entrenched, historical achievement discrepancies should not—on their own—be interpreted as evidence of racial bias.

However, if non-White students are systematically assigned lower targets even when they have comparable prior achievement to their White classmates, then we should be more concerned about racial bias in target-setting. We therefore modify Equation (4a) by sequentially adding a vector of up to four prior-year test scores in both tests/subjects, $A_{igs,y-1}$ (4b), a vector of student demographics other than race,¹⁴ $X_{i(y)}$ (4c), and finally teacher-by-year fixed effects, ϕ_{py} (4c). If the estimated coefficients on student race/ethnicity dummies continue to be negative and significant across these models, then it would suggest that teachers tend to set lower targets, within their classes, for their non-White students than their White students with similar prior achievement and other characteristics.

Results

Research Question 1: Target-Setting Descriptives

Targets Set and Percentage Met. The sheer number of targets set in these seven schools over 4 years underscores the magnitude of the SLO system undertaking: In Table 4, we

show that between about 8,000 and 9,000 individual student learning targets were set per test and subject from 2011–2012 through 2014–2015, for a total of approximately 34,000 target scores. Table 4 also presents the percentage of those targets that were met, the average target score (in standard deviation [*SD*] units), and the average EOY score received. EOY scores are low—around -0.30 and -0.70 *SDs*—which reflects the fact that the SLO schools were particularly low performing (see Table 1).

By examining the percentage of targets that students ultimately met, we get a first glimpse at whether teachers were willing and/or able to set difficult-to-reach targets. For the SSAT test (upper panel of Table 4), we find that 43% of targets in ELA were met (61% in math). For the MAP test, 60% (ELA) and 65% (math) of targets were met. Overall, we see that students were more likely to meet math targets than ELA targets on both tests. We also disaggregate results in Table 4 by school year, grade, and school. There is some evidence that schools did not all approach target-setting uniformly. For SSAT ELA, only 37% of students met their targets in School B, while 49% of students in School E did so. To take another example, while 53% of School G students met their math MAP targets, 84% of students in School B did so. We do not see evidence that particular schools appear to have consistently high or low target attainment rates across the four assessments. For instance, while School B has the lowest rate of target attainment on the ELA SSAT test (and second to lowest on the SSAT math), School B has the *highest* rate of target attainment on the both the ELA and math MAP assessments.

The grade, year, and school means presented in Table 4 could obscure meaningful trends in target-setting over time in

TABLE 4

Number of Targets Set and Percentage Met Overall and by Year, Grade, and School

| Characteristic | SSAT test | | | | | | | |
|---------------------|-----------------------|---------------------------|--------------------------------|-------------------------------|-----------------------|---------------------------|--------------------------------|-------------------------------|
| | ELA | | | | Math | | | |
| | Number of targets set | Percentage of targets met | Mean target set, in <i>SDs</i> | Mean EOY score, in <i>SDs</i> | Number of targets set | Percentage of targets met | Mean target set, in <i>SDs</i> | Mean EOY score, in <i>SDs</i> |
| Overall | 8,737 | 43 | -0.33 | -0.46 | 8,113 | 61 | -0.73 | -0.50 |
| SY 2011–2012 | 2,249 | 50 | -0.41 | -0.40 | 1,944 | 38 | -0.24 | -0.53 |
| SY 2012–2013 | 2,319 | 32 | -0.04 | -0.46 | 2,162 | 63 | -0.82 | -0.52 |
| SY 2013–2014 | 1,811 | 45 | -0.45 | -0.51 | 1,678 | 76 | -1.04 | -0.51 |
| SY 2014–2015 | 2,358 | 45 | -0.45 | -0.49 | 2,329 | 67 | -0.83 | -0.44 |
| Grade 4 | 1,293 | 46 | -0.40 | -0.46 | 1,357 | 62 | -0.58 | -0.39 |
| Grade 5 | 1,383 | 42 | -0.27 | -0.44 | 1,386 | 40 | -0.16 | -0.37 |
| Grade 6 | 2,095 | 36 | -0.21 | -0.46 | 1,791 | 63 | -0.85 | -0.47 |
| Grade 7 | 2,122 | 50 | -0.48 | -0.47 | 1,825 | 57 | -0.74 | -0.55 |
| Grade 8 | 1,844 | 41 | -0.30 | -0.47 | 1,754 | 76 | -1.16 | -0.65 |
| Elementary School A | 420 | 49 | -0.37 | -0.39 | 420 | 60 | -0.48 | -0.34 |
| Elementary School B | 538 | 37 | -0.37 | -0.60 | 539 | 44 | -0.46 | -0.61 |
| Elementary School C | 510 | 43 | -0.45 | -0.52 | 568 | 54 | -0.41 | -0.35 |
| Elementary School D | 577 | 42 | -0.21 | -0.40 | 582 | 50 | -0.22 | -0.27 |
| Elementary School E | 631 | 50 | -0.29 | -0.34 | 634 | 49 | -0.30 | -0.33 |
| Middle School F | 3,182 | 44 | -0.17 | -0.27 | 3,172 | 83 | -1.21 | -0.55 |
| Middle School G | 2,879 | 41 | -0.51 | -0.68 | 2,198 | 40 | -0.48 | -0.57 |

| Characteristic | MAP test | | | | | | | |
|---------------------|-----------------------|---------------------------|------------------------------------|-----------------------------------|-----------------------|---------------------------|--------------------------------|-------------------------------|
| | ELA | | | | Math | | | |
| | Number of targets met | Percentage of targets met | Mean of targets set, in <i>SDs</i> | Mean of EOY scores, in <i>SDs</i> | Number of targets set | Percentage of targets met | Mean target set, in <i>SDs</i> | Mean EOY score, in <i>SDs</i> |
| Overall | 9,124 | 60 | -0.7 | -0.50 | 8,591 | 65 | -0.77 | -0.56 |
| SY 2011–2012 | 2,317 | 60 | -0.64 | -0.47 | 2,088 | 67 | -0.89 | -0.65 |
| SY 2012–2013 | 2,288 | 61 | -0.63 | -0.46 | 2,207 | 64 | -0.73 | -0.51 |
| SY 2013–2014 | 2,165 | 57 | -0.66 | -0.54 | 1,971 | 61 | -0.70 | -0.54 |
| SY 2014–2015 | 2,354 | 62 | -0.70 | -0.53 | 2,325 | 68 | -0.77 | -0.54 |
| Grade 4 | 1,335 | 69 | -0.86 | -0.56 | 1,293 | 79 | -0.95 | -0.51 |
| Grade 5 | 1,383 | 67 | -0.82 | -0.53 | 1,311 | 81 | -1.01 | -0.56 |
| Grade 6 | 2,281 | 56 | -0.57 | -0.51 | 2,182 | 58 | -0.65 | -0.56 |
| Grade 7 | 2,117 | 63 | -0.70 | -0.45 | 1,950 | 63 | -0.74 | -0.57 |
| Grade 8 | 2,008 | 52 | -0.47 | -0.47 | 1,855 | 55 | -0.67 | -0.58 |
| Elementary School A | 410 | 60 | -0.62 | -0.45 | 409 | 71 | -0.72 | -0.45 |
| Elementary School B | 537 | 73 | -0.95 | -0.56 | 418 | 84 | -1.13 | -0.57 |
| Elementary School C | 560 | 65 | -0.85 | -0.62 | 563 | 82 | -0.94 | -0.49 |
| Elementary School D | 580 | 72 | -0.81 | -0.45 | 583 | 83 | -0.95 | -0.44 |
| Elementary School E | 631 | 67 | -0.90 | -0.62 | 631 | 79 | -1.12 | -0.71 |
| Middle School F | 3,171 | 62 | -0.46 | -0.28 | 3,176 | 64 | -0.59 | -0.42 |
| Middle School G | 3,235 | 51 | -0.70 | -0.70 | 2,811 | 53 | -0.79 | -0.75 |

Note. Descriptive statistics are shown separately for SSAT tests (upper) and MAP tests (lower) and by subject (ELA left, math right). Results are reported overall and also disaggregated by year, grade, and SLO school. Columns report (A) the number of targets set for the given test and subject, (B) the percentage of those targets that were met, (C) the average target score set (after converting them to be standardized at the subject-grade-year), and (D) the average standardized EOY observed test score. A standardized score of -0.50, for example, would indicate that the group of students performed, on average, 50% of an *SD* lower on the given test than their peers district-wide, in the same subject, grade, and year. SLO = student learning objectives; ELA = English language arts; SY = school year; EOY = end-of-year; SSAT = statewide standardized achievement test.

school-grades. We are interested in examining within- and between-school variance in target-setting, however, a formal decomposition of variance is underpowered given that there are only seven schools. Figure 2 (SSAT) and Figure 3 (MAP) provide a visual look at the trends in percentage of targets met, as well as within-school variation across grades and subjects.

Taken together, the results in Table 4, Figure 2, and Figure 3 paint a picture of SLO target setting that generally did not guarantee that students would meet their targets. We can clearly infer from this that teachers either did not or could not act purely in strategic self-interest. We next examine the extent to which teachers appear able to select targets for their students that are near their EOY scores.

Target Accuracy. Recall that we characterize targets set at the beginning of the year for a given test as more accurate when they turn out to be closer to the actual EOY score. We are particularly interested in whether targets become more accurate as teachers gain experience with the SLO target-setting process. In Table 5, we therefore present descriptive statistics on the distribution of target accuracy, $Accuracy_{it}$, across the four tests and four study years. When $Accuracy_{it}$ is closer to zero, the target was more accurate, positive values of $Accuracy_{it}$ indicate the target was higher than the actual score, and negative values indicate that the target was lower than the actual score (refer back to Equation 1).

On average, teachers appear to be able to set target scores that are usually less than 20% of an SD away from scores the students ultimately received (Table 5). For instance, the average accuracy of 2011–2012 targets for the ELA SSAT was 0.01—very close to zero. On the other hand, the average accuracy of 2013–2014 targets for the math SSAT exam was -0.52 , suggesting that teachers generally underestimated those EOY scores by about half an SD that year. In most cases, we see that accuracy scores tend to be negative. The exception to this is for setting targets in ELA on the SSAT, wherein teachers overestimated actual scores, on average, in all 4 years. The 10 to 90 percentile range for accuracy is ± 0.93 SDs for ELA SSAT targets—that is, some teachers’ targets can be up about 1 SD above or below the student’s observed score. We see a similar range in accuracy across tests, subjects, and years.

We were particularly interested in whether accuracy improved over time, however, the results in Table 5 suggest this was not the case. For none of the four tests do we see a pattern of mean accuracy approaching closer to zero with additional years of the SLO policy. That said, we do observe relatively strong correlations (between .50 and .78) between the targets that teachers set and the scores students ultimately receive (last column of Table 5); teachers were able to set targets that generally aligned with how students would perform at the end of the year.

Target Ambitiousness. Recall that we define ambitiousness of a given target, Amb_{it} , as the standardized target score

minus a statistically-generated expected standardized score (Equation 2). In Figure 4, we present the distribution of target ambitiousness (the 1st–99th percentile range¹⁵) separately for each of the four tests. An ambitiousness score of zero indicates that the target perfectly corresponds to what a statistical model would predict based on a student’s prior achievement and demographics (for brevity, we refer to this as the statistically expected score). The solid line is the median of the distribution, and the dashed lines demarcate ± 1 SD around the median.

Given that part of the logic behind setting learning objectives is to formalize reasonable but high expectations, we were interested to find that the *typical* target tended to not be particularly ambitious (close to zero). With the exception of SSAT ELA, we show in Figure 4, that median ambitiousness is close to zero and even negative for MAP math (underestimating EOY scores). On the other hand, the distribution of target ambitiousness for the SSAT ELA test is centered right of zero (median = 0.27 SDs), indicating that the median teacher sets targets for this test that are nearly 30% of an SD higher than students’ statistically expected scores.

We also find quite large variation in ambitiousness, particularly for SSAT math targets (SD of ambitiousness = 0.83). Some targets are therefore quite ambitious (e.g., target scores between 1.2 and even 2 SDs higher than the statistically expected score), while other targets are equally unambitious. Some teachers certainly are setting “reach” targets: For the SSAT, 30% of the ELA targets were at least half an SD above statistically expected scores (24% for math). For the MAP test, we observe slightly fewer targets at least half an SD above statistically expected scores: 19% for ELA and only 11% for math.

We are particularly interested in whether teachers adjust the ambitiousness of the targets they set as they gain more experience with the SLO process. One could imagine that teachers feel compelled to set less ambitious targets once they realize that most of their students are not reaching them (in order to obtain larger financial bonuses). On the other hand, it is also possible teachers become more willing to push the envelope and set more ambitious goals as they become more comfortable with the process and recognize that they typically are able to set accurate goals.

We investigate this in Figure 5 by examining temporal patterns in average target ambitiousness, separately for three groups of teachers: (a) those who received no bonus in the first year they set targets (in red) because less than 50% of their students met their target, (b) teachers who received between \$2500 and \$4000 in their first year (orange) because between 50% and 80% of their students met their target, and (c) teachers who received \$4000 to \$5000 in their first year (green) because 80% or more of their students met their target. We display two interrelated factors in Figure 5: The Y -axis is the average target ambitiousness (in SDs) of teachers in the given group and year, with the $Y = 0$ dashed line

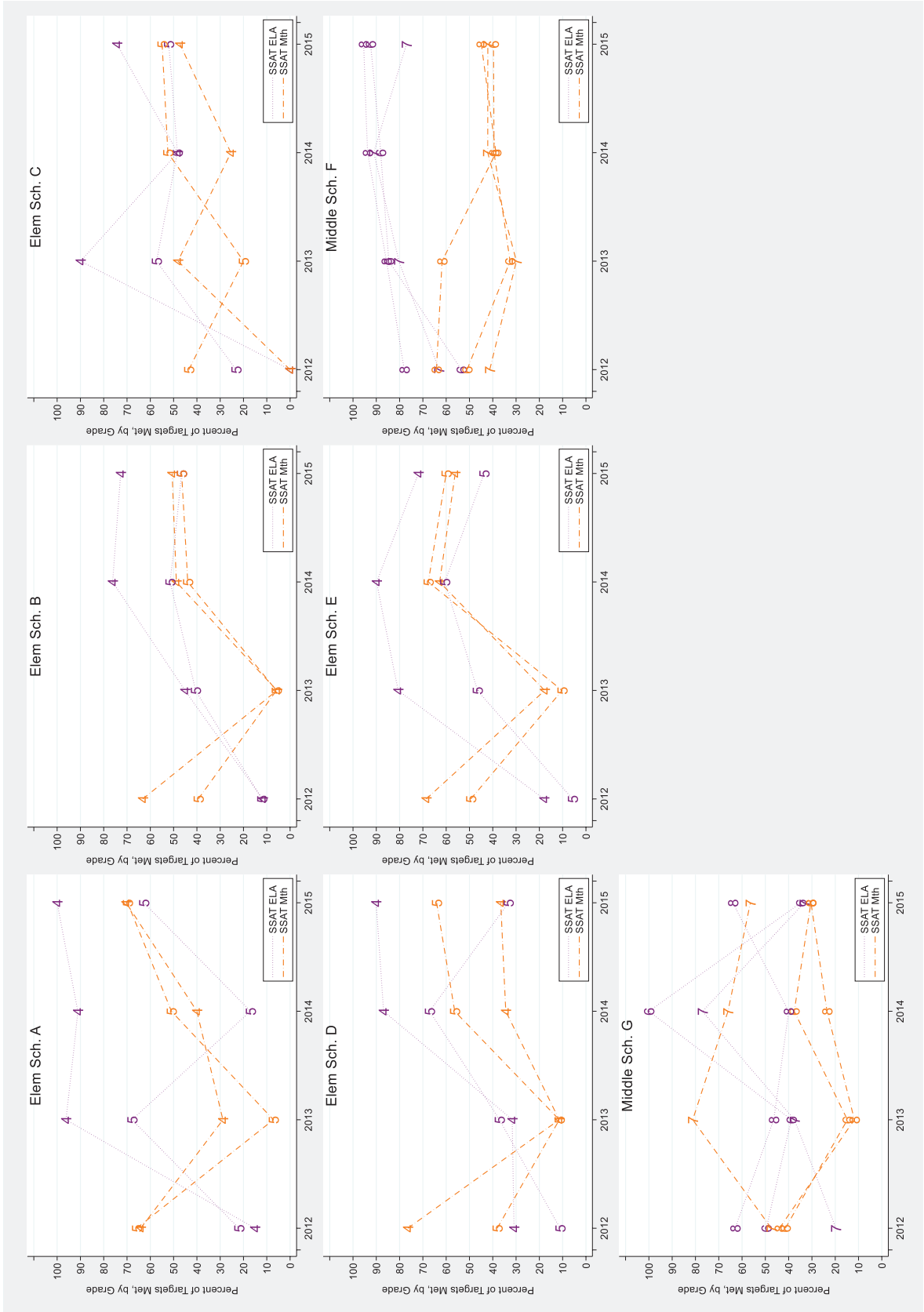


FIGURE 2. Percentage of statewide standardized achievement test (SSAT) targets met over time, by school and grade.

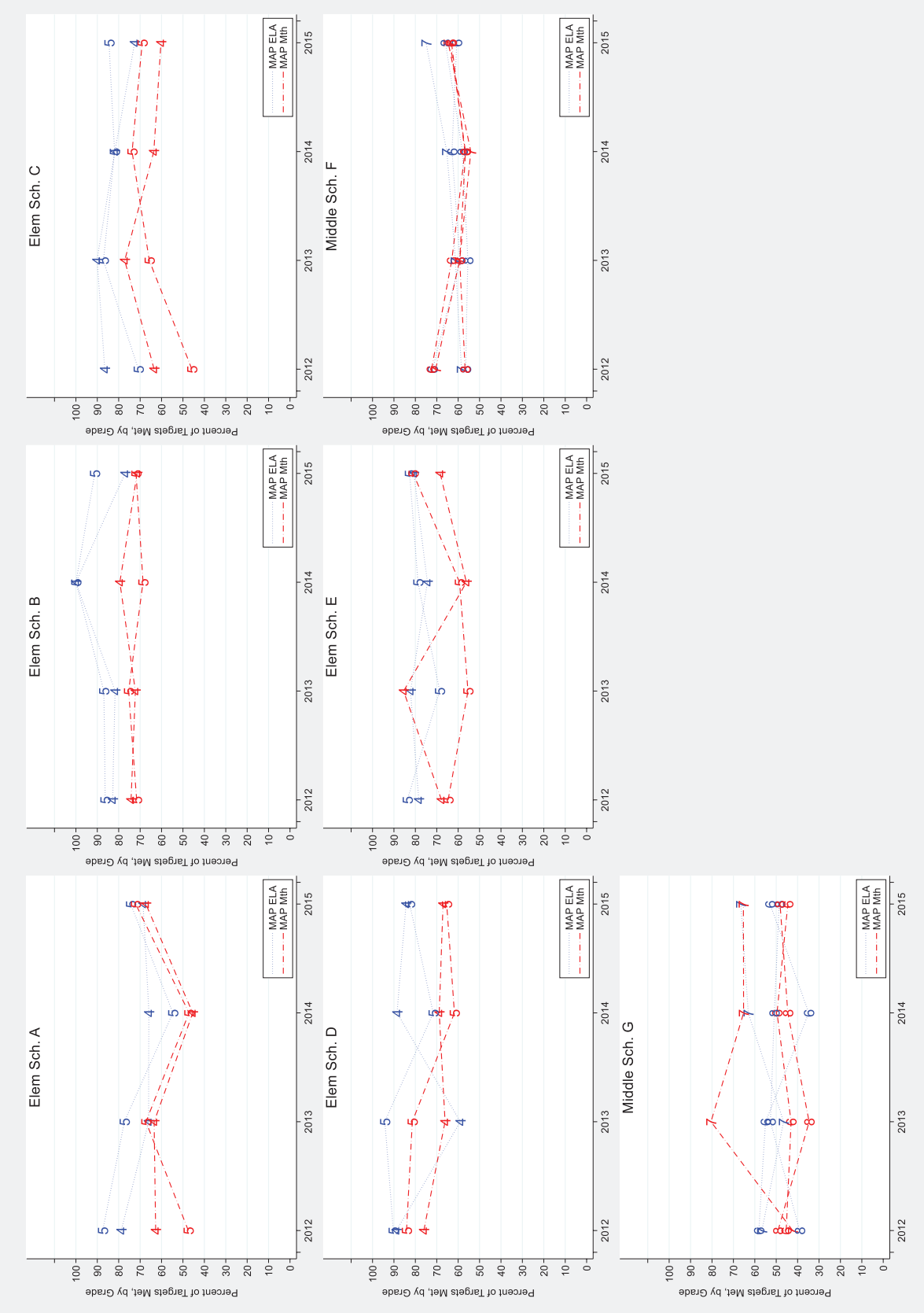


FIGURE 3. Percentage of Measures of Academic Progress (MAP) targets met over time, by school and grade.

TABLE 5

Descriptive Statistics for Accuracy of Targets (in SDs), by Test, Subject, and Year

| Test/subject/ year | Number of targets | Mean accuracy measure | SD of accuracy measure | Percentiles of accuracy measure distribution | | | | | Correlation of target and actual |
|-----------------------|----------------------|-----------------------------|------------------------------|--|-------|-------|------|------|--|
| | | | | 10th | 25th | 50th | 75th | 90th | |
| <i>SSAT test</i> | | | | | | | | | |
| ELA | | | | | | | | | |
| 2011–2012 | 2157.00 | 0.01 | 0.76 | –0.93 | –0.43 | 0.01 | 0.50 | 0.93 | 0.64 |
| 2012–2013 | 2267.00 | 0.42 | 0.91 | –0.76 | –0.17 | 0.44 | 1.02 | 1.54 | 0.57 |
| 2013–2014 | 1719.00 | 0.07 | 0.72 | –0.84 | –0.35 | 0.09 | 0.51 | 0.96 | 0.70 |
| 2014–2015 | 2249.00 | 0.05 | 0.64 | –0.73 | –0.33 | 0.08 | 0.47 | 0.84 | 0.72 |
| Math | | | | | | | | | |
| 2011–2012 | 1861.00 | 0.31 | 1.09 | –1.05 | –0.38 | 0.38 | 1.07 | 1.63 | 0.51 |
| 2012–2013 | 2108.00 | –0.30 | 0.86 | –1.37 | –0.88 | –0.27 | 0.29 | 0.78 | 0.50 |
| 2013–2014 | 1604.00 | –0.52 | 0.86 | –1.61 | –1.03 | –0.51 | 0.04 | 0.56 | 0.58 |
| 2014–2015 | 2224.00 | –0.38 | 0.89 | –1.56 | –0.97 | –0.34 | 0.25 | 0.72 | 0.50 |
| <i>MAP test</i> | | | | | | | | | |
| ELA | | | | | | | | | |
| 2011–2012 | 2125.00 | –0.14 | 0.78 | –1.07 | –0.58 | –0.14 | 0.30 | 0.76 | 0.67 |
| 2012–2013 | 2127.00 | –0.15 | 0.80 | –1.06 | –0.57 | –0.15 | 0.30 | 0.75 | 0.65 |
| 2013–2014 | 1930.00 | –0.09 | 0.87 | –1.02 | –0.48 | –0.07 | 0.34 | 0.75 | 0.63 |
| 2014–2015 | 2200.00 | –0.15 | 0.70 | –0.92 | –0.53 | –0.14 | 0.22 | 0.62 | 0.71 |
| Math | | | | | | | | | |
| 2011–2012 | 1947.00 | –0.22 | 0.62 | –0.92 | –0.53 | –0.18 | 0.15 | 0.46 | 0.73 |
| 2012–2013 | 2093.00 | –0.20 | 0.71 | –1.02 | –0.57 | –0.16 | 0.17 | 0.57 | 0.68 |
| 2013–2014 | 1858.00 | –0.13 | 0.61 | –0.78 | –0.45 | –0.12 | 0.23 | 0.51 | 0.76 |
| 2014–2015 | 2185.00 | –0.22 | 0.56 | –0.90 | –0.54 | –0.18 | 0.12 | 0.42 | 0.78 |

Note. Accuracy measures are calculated by subtracting the student’s observed EOY score from the target score set by the teacher for that student. Negative values of accuracy indicate that the target was lower than subsequently observed actual score. Positive values indicate that the target was higher than the actual score. Both the EOY observed and target scores have been standardized within the district at the subject-grade-year level. Results are shown separately for SSAT tests (upper panel) and MAP tests (lower panel) and by subject (ELA on left, math on right). ELA = English language arts; EOY = end-of-year; SSAT = statewide standardized achievement test; MAP = Measures of Academic Progress.

representing targets that were equal to the score a statistical model would predict. Then, each data point is labeled with the percentage of targets that were *met* in the given group. Here we can see how target performance tracked over time with the ambitiousness of targets.

If one could hope for any pattern, it would perhaps be to see teachers increasing the ambitiousness over time but simultaneously maintaining or even increasing the percentage of students who meet targets. We do not consistently observe this. Take as an example, SSAT ELA: Teachers who did not receive a bonus in their first year of target-setting (red) tended to set ambitious goals, but only 25% of their students’ targets were met. On the other hand, among the teachers who received large bonuses in their first year (green), those targets tended to be less ambitious (below $Y = 0$) but on average 97% of targets were met. We focus on how those three groups of teachers approached target-setting over the next 3 years. Unique to SSAT ELA, we see that all three

groups of teachers notably increased the ambitiousness of their targets in Year 2—even those who did not receive a bonus in Year 1. In all three groups, only an average of 22% to 33% of targets were met in Year 2. There seems to be a downward correction in Year 3 and 4, with all three groups then lowering the ambitiousness of their targets (though all still above $Y=0$) and therefore having more students meet their targets.

We see a clearer pattern of adjusting targets over time for SSAT math. Teachers who received large bonuses in Year 1 (green) set unambitious targets—about half an *SD* below statistical predictions, on average—and tended to have nearly all students reach their targets. This group adjusted their target-setting in Year 2, making their targets somewhat less *unambitious*, at the cost of having fewer students meet those targets (76%). Only by the fourth year does this group set targets with average positive ambitiousness. The opposite occurred for teachers who did *not*

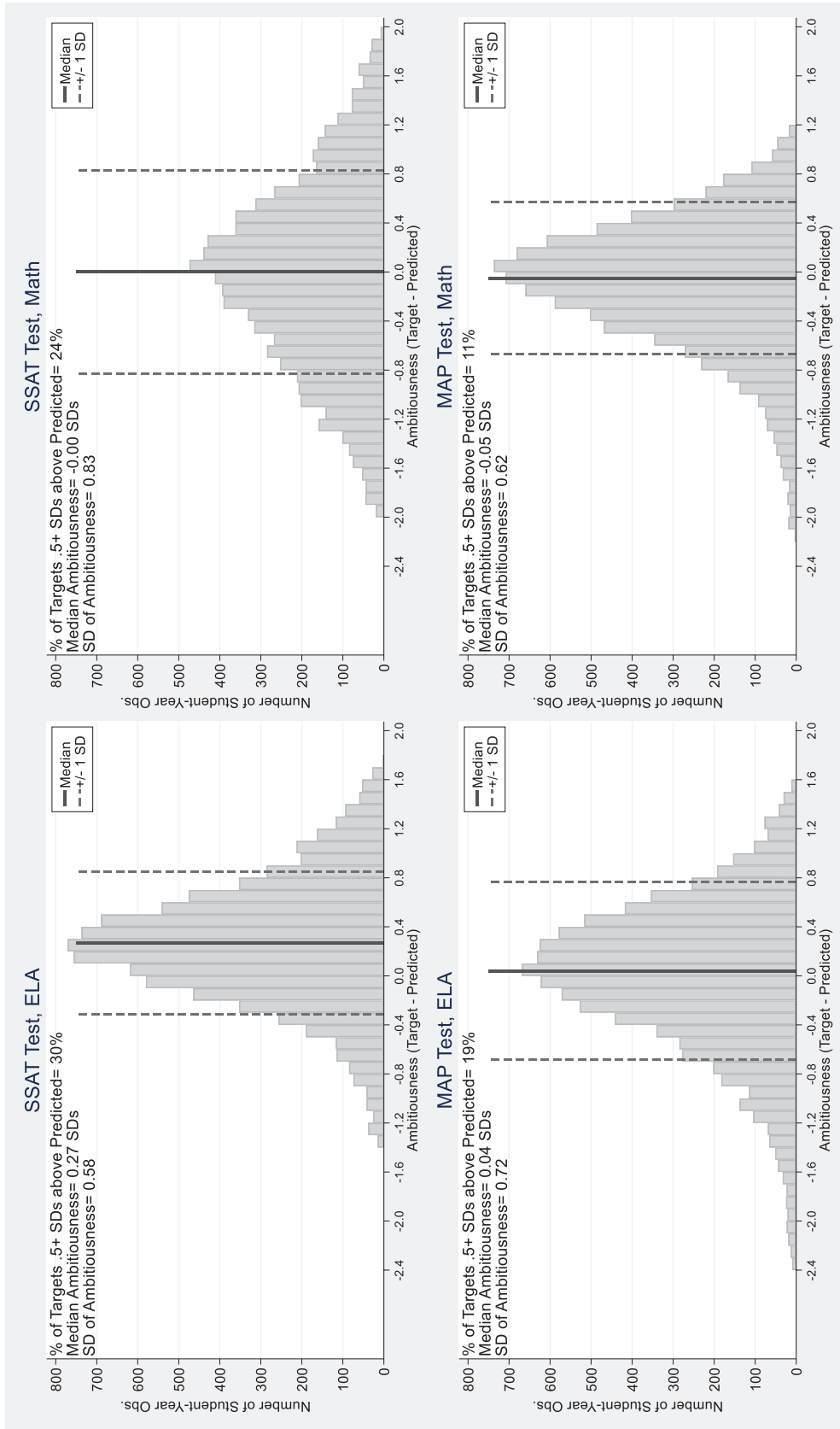


FIGURE 4. Distributions (1st-99th percentile range) of target ambitiousness (in SDs) separately for four tests.
Note. Target ambitiousness is calculated by subtracting a regression-adjusted prediction of a student's EOY test score from the target score set by the student's teacher for the EOY. Both scores have been standardized (i.e., we subtract the district-wide M and SD [within subject-grade-year] of the EOY scores in their original metric from all target and spring scores. The distributions of target ambitiousness are limited to the 1st to 99th percentile range to facilitate visualization of the main distribution of the ambitiousness measures, which were otherwise obscured by a small number of outliers. We find that, on closer examination of these outliers, they were potentially idiosyncratic data entry errors. EOY = end-of-year.

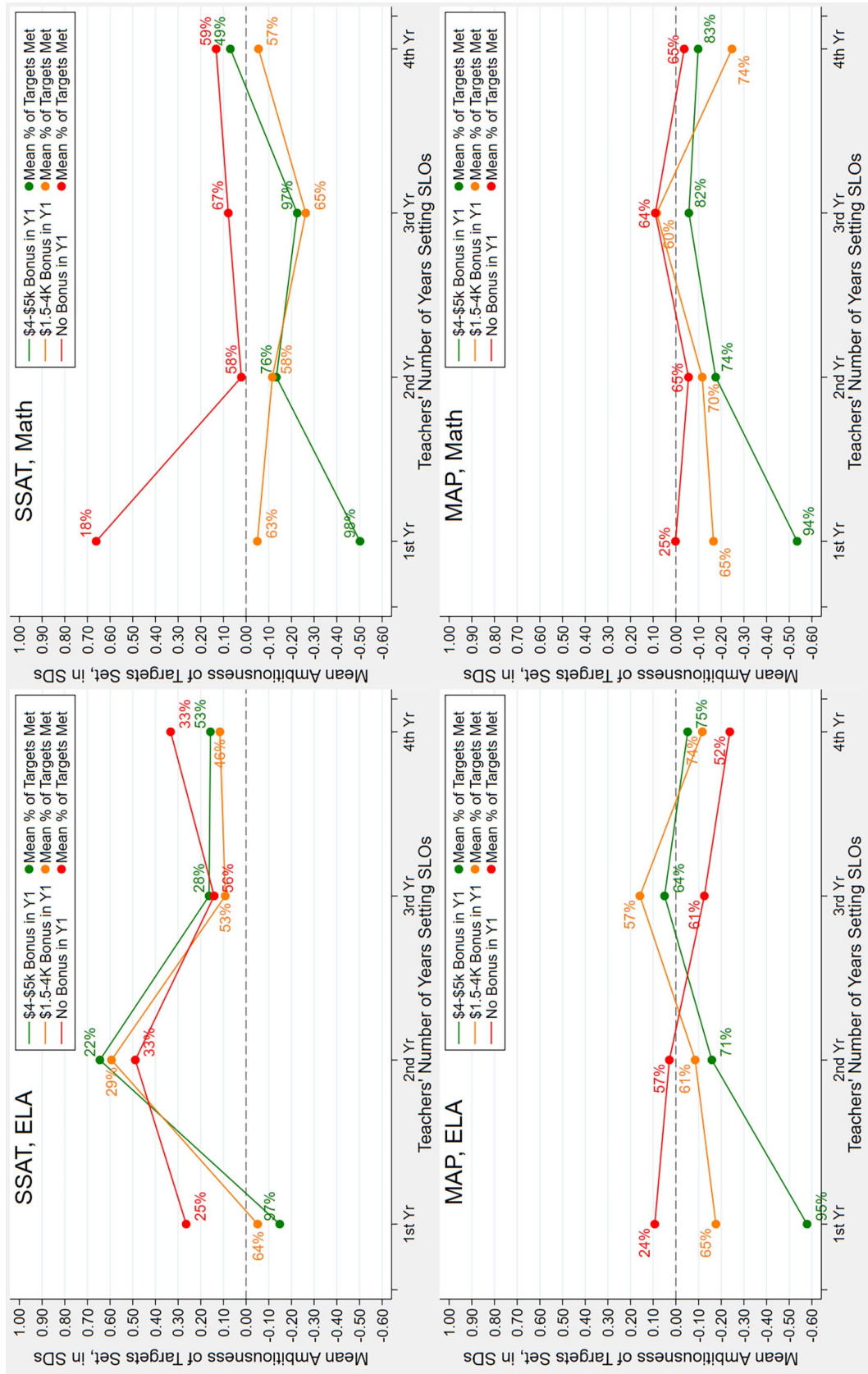


FIGURE 5. Mean ambitiousness of targets and mean percentage of targets met over time (data point labels), separately by teachers' bonus results in their first year of target-setting.

Note. A target's ambitiousness is calculated by subtracting a regression-adjusted prediction of a student's EOY test score (in SDs) from the target score set by the student's teacher (also in SDs). The Y-axis is the mean ambitiousness of targets set by the given group of teachers. The data point labels report the mean percentage of targets met for the given group of teachers. Teachers are grouped based on bonus outcome in their first year of setting targets. To receive a bonus of \$4,000+, at least 80% of a teacher's students must meet targets. If less than 50% of a teacher's students meet targets, teachers receive no bonus. EOY = end-of-year.

receive a bonus in their first year (red). They initially set *very* ambitious targets—their targets were, on average 0.60 *SDs* higher than statistical predictions—but on average only 18% of their students met those targets. These teachers appear to recalibrate in Year 2, by lowering target ambitiousness (though still positive) and therefore having 58% of those targets met. In MAP, we see similar patterns: Teachers who receive the largest bonuses (green) in their first year (due to relatively unambitious targets that were almost always met) tend to correct somewhat, by making targets less unambitious over time. In MAP ELA, teachers who initially received no bonus (red) set average targets in Year 1 above statistical predictions but subsequently set less and less ambitious targets over time.

Figure 5 paints a picture of SLO target-setting that involved a certain learning curve, in which teachers struggled to find the balance between setting ambitious targets while maintaining a high percentage of their students meeting those goals. We do see evidence that teachers whose targets seemed clearly too low in their first year adjust those targets upward in later years, despite a financial disincentive to do so. However, with the exception of SSAT ELA, we also observe that the average ambitiousness of targets often stayed near or below 0. That is, while it was relatively common to observe average ambitiousness in the -0.20 *SD* range, it was much less common to see averages in the $+0.20$ range.

Target-Setting and Teacher Value-Added. Because the practice of setting targets and adjusting instruction toward achieving those targets is part of the professional practice that constitutes good teaching, one might hypothesize that more effective or experienced teachers set targets that are more accurate and/or ambitious. The only imperfect proxy for teachers' effectiveness available to us comes from estimating teacher VAMs.¹⁶ Unfortunately, this data set does not contain teacher years of experience information (or teacher demographics).

We do not find evidence that teachers with higher VAMs (in the preceding school year) set more *accurate* targets. The correlations between teacher VAMs and the accuracy of targets are -0.01 (ELA, SSAT), -0.16 (ELA, MAP), $+0.14$ (math, SSAT), and -0.09 (math, MAP). However, there is some evidence of a modest, positive correlation between teacher effectiveness (VAMs) and the *ambitiousness* of targets. These correlations are $+0.21$ (ELA, SSAT), $+0.50$ (ELA, MAP), $+0.30$ (math, SSAT), and $+0.38$ (math, MAP). By definition, teachers with high VAMs (in the previous year) have a demonstrated ability to produce test scores that are above what the value-added model would predict. It therefore seems plausible that teachers with a record of producing higher than-expected test scores in prior years are also likely to set higher-than-expected (i.e., ambitious) learning targets.

Research Question 2: Do Targets Predict EOY Scores?

In unconditional models (not shown), we find that students with higher targets have higher standardized achievement scores on the order of 40% to 70% of an *SD* for every 1 *SD* increase in target scores. In Table 6, we go on to compare yearly observations of students who are similar on prior achievement, student and school covariates, and are in the same grade and year (Model 3a). For reasons that are not observed to the researcher, one of those students is assigned a higher learning target at the start of the year than their counterpart. We further condition on being taught by the same teacher in the same year in the given subject (Model 3b), and then leverage variation in targets set across years for the same student (Model 3c).

We generally find that, even for two otherwise similar students, receiving higher targets is associated with higher EOY scores. Controlling for prior scores, demographics, and school setting characteristics (Model 3a), two students who have a 1 *SD* difference in the SSAT ELA target scores that teachers set for them tend to exhibit differences in EOY scores of about 24% of an *SD* (17% of an *SD* for SSAT math). Those associations are a little stronger for the MAP test: Estimates from Model (3a) indicate that a 1 *SD* difference in target scores is associated with about 36% and 47% of an *SD* in EOY test scores in ELA and math, respectively. In Model 3b, we examine the correspondence between targets set and spring scores achieved among students taught by the same teacher in the same year. Again, the coefficients are large and statistically significant, ranging from 24% of an *SD* (SSAT math) to as high as 43% of an *SD* (MAP math).

In Model (3c) in Table 6, these associations remain significant but are smaller in magnitude, in the range of 4% to 13% of an *SD* difference in EOY scores per 1 *SD* difference in target scores. Even once we hold time-invariant student-level factors constant by comparing the same student's EOY performance over time as they receive different targets in different study years, we generally observe that higher targets correspond with higher EOY scores. However, it still may not be the case that receiving higher targets *caused* students to have improved EOY test scores; rather, their teachers could be considering information about the student's context in each year to adjust targets accordingly.

Because a primary confounder in the analysis above is that time-varying contextualizing factors may be observed by the teacher when they set annual goals but are not observed by the researcher, we also explore whether a student in any single year performs more strongly on the assessments for which he or she is given higher targets (Model 3d). As discussed above, this approach may still miss systemic confounders in target-setting across tests and subjects, but at least it attends to explanations like a student having a particularly hard year at home. Because we stack targets and EOY achievement outcomes across tests and subjects within

TABLE 6

Association Between Fall Targets Set for Spring Scores and Subsequent Spring Test Scores, Among Similar Students, Students With the Same Teacher, or for the Same Student Over Time (Prescores Imputed)

| | SSAT Test | | | | | |
|-----------------------|---------------------|---------------------|---------------------|---------------------|---------------------|-------------------|
| | ELA | | | Math | | |
| | (3a) | (3b) | (3c) | (3a) | (3b) | (3c) |
| Target set for spring | 0.239*** (0.014) | 0.366*** (0.017) | 0.095*** (0.017) | 0.169*** (0.013) | 0.239*** (0.014) | 0.036* (0.017) |
| Constant | 0.480* (0.231) | 1.962 (1.686) | -0.302 (0.863) | 0.066 (0.273) | 0.681 (1.613) | -1.774 (1.086) |
| R^2 | .538 | .598 | .892 | .405 | .564 | .841 |
| N | 8,379 | 8,127 | 8,379 | 7,786 | 7,737 | 7,786 |
| Covariates? | × | × | × | × | × | × |
| Teacher × Year FE's | | × | | | × | |
| Student FE's | | | × | | | × |

| | MAP test | | | | | |
|-----------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | ELA | | | Math | | |
| | (3a) | (3b) | (3c) | (3a) | (3b) | (3c) |
| Target set for spring | 0.360*** (0.010) | 0.358*** (0.011) | 0.065*** (0.015) | 0.468*** (0.010) | 0.426*** (0.010) | 0.129*** (0.015) |
| Constant | 1.053*** (0.236) | 3.289 (1.881) | -0.696 (0.920) | 0.502* (0.196) | 1.438 (1.569) | -1.133 (0.749) |
| R^2 | .567 | .602 | .879 | .651 | .691 | .910 |
| N | 8,358 | 8,097 | 8,358 | 8,060 | 8,010 | 8,060 |
| Covariates? | × | × | × | × | × | × |
| Teacher × Year FE's | | × | | | × | |
| Student FE's | | | × | | | × |

Note. Analytic sample comprises students in the seven district schools that implemented the SLO process in 2011–2012 through 2014–2015. All models have a baseline covariate vector (unless collinear with the fixed effects) that include grade and year fixed effects, a vector of four prior achievement scores (two subjects in each of two tests in the prior year), student covariates, and school covariates. Model 3a only includes that baseline covariate vector, Model 3b adds teacher-by-year fixed effects, and Model 3c replaces time-invariant student covariates with student fixed effects. Both the primary predictor of interest, the target score set in the fall for the EOY (spring) test and the outcome, the observed EOY test scores have been standardized (i.e., we subtract the district-wide mean and standard deviation (within subject-grade-year) of the EOY scores in their original metric from all target and spring scores. For the approximately 10% of observations missing at least 1 of the 4 prior-year achievement scores, we impute the mean and include a dummy variable indicating that the value was originally missing. As described in text, results are not substantively different if we instead exclude cases with missing data. ELA = English language arts; EOY = end-of-year; SSAT = statewide standardized achievement test; MAP = Measures of Academic Progress; SLO = student learning objectives; FE = fixed effects.

* $p < .05$. ** $p < .01$. *** $p < .001$.

student-year, results in Table 7 are no longer shown separately across these dimensions. We do, however, show the results overall (Column 1), and then separate the sample between elementary (Column 2) and middle schools (Column 3). The logic here is that variation within student-year target-setting may be a little different when the same teacher sets all four of the targets (i.e., in elementary school where the homeroom teacher is responsible for both math and ELA), versus in middle school grades where a separate teacher sets the math versus ELA targets.

In Table 7, we again see evidence that higher target scores correspond to higher EOY scores. In this case, we observe that a 1 *SD* positive difference in a student's targets set across tests/subjects in the same year is associated with a 22%

positive difference in EOY scores. It is worth noting that observing a 1 *SD* difference in targets set across the four tests for the same student in the same year is not uncommon; the average range across the tests is 1.4 *SDs*. Results are similar in both the elementary grades in which the same teacher is providing those four different target scores (coefficient is 0.19 *SDs*) and in middle grades in which different teachers provide the ELA and math targets (coefficient is 0.24 *SDs*).

Research Question 3. Differential Target-Setting by Race/Ethnicity

One might be concerned that not all students in the district would have equal access to these potential benefits, or

TABLE 7

Association Between Fall Targets Set for Spring Scores and Subsequent Spring Test Scores Observed, Across Different Tests for the Same Student in the Same Year (Within-Student, Pooled Sample)

| | Model (3d) | | |
|-----------------------|---------------------|-----------------------|---------------------|
| | All grades | Elementary grades 4–5 | Middle grades 6–8 |
| Target set for spring | 0.223*** (0.01) | 0.189*** (0.01) | 0.236*** (0.01) |
| Constant | -0.342*** (0.01) | -0.366*** (0.01) | -0.329*** (0.01) |
| R^2 | 0.723 | 0.715 | 0.728 |
| N | 28,058 | 9,585 | 18,473 |
| Test FE's | × | × | × |
| Subject FE's | × | × | × |
| Student-year FE's | × | × | × |

Note. Data set structured as long by repeated observations (up to four) across tests and subjects within year for each student. Analytic sample is composed of students in the seven district schools that implemented the SLO process in 2011–2012 through 2014–2015. Results are shown separately for all students (Column 1), students in elementary Grades 4 to 5 (Column 2), and middle school grade 5 to 8. Model (3d) includes test, subject, and student-year fixed effects. Both the primary predictor of interest, the target score set in the fall for the EOY (spring) test and the outcome, the observed EOY test scores have been standardized (i.e., we subtract the district-wide mean and standard deviation (within subject-grade-year) of the EOY scores in their original metric from all target and spring scores. FE = fixed effects; EOY = end-of-year; SLO = student learning objectives.

* $p < .05$. ** $p < .01$. *** $p < .001$

that teachers are not able to make these predictions equally well for all students. This could arise if teachers possess implicit or explicit biases about the capabilities of students of different races, e.g.,—a phenomenon that can arise when one does not attribute the observable differences in outcomes by race group to structural inequality but instead to underlying capacity. In Table 8, we present results across Models 4a through 4d, separately for SSAT targets (upper panel) and MAP targets (lower), and ELA (left) versus math (right). Again, we standardize the $TarScore_{ig\text{sy}}$ outcome variable.

According to the unconditional Model (4a) in Table 8, we find large, negative raw differences between the targets set for non-White students relative to White students across all racial/ethnic groups and all test-subject combinations (all coefficients are also statistically significant, with the exception of the difference in targets set for Latinx students, relative to White students on the SSAT math test). However, when we add the vector of previous achievement test scores to the equation (Model 4b), almost all the observed differences in targets set for White and non-White students are no longer statistically significant and are much closer to zero (i.e., no differences in targets set). This remains the case when we add the set of student characteristic controls in Model (4c) and teacher-by-year fixed effects in Model (4d). It appears that Black and Latinx students in this district do not receive systematically different learning targets than their White counterparts when they possess similar test score profiles.

The possible exception to the lack of significant differences in racial groups' targets received is for Asian students on both ELA tests (but neither math test). For the ELA

targets that Asian students receive in this district, we observe that in every model, their targets are significantly lower than their otherwise similar White counterparts. For the ELA targets on the SSAT, Asian students received target scores that were 23% of an SD lower than White students with the same prior test scores, other student characteristics, and in the same school (Model 4d). The Asian White differences from Model (4d) are even larger on the MAP test for ELA, where we observe targets 47% of an SD lower for Asian students than their White counterparts. We find the lower targets for Asian students on ELA tests to be concerning and warranting of further investigation, though the population of Asian students in this district is small (<10%).

Conclusions

We explore a SLO system as implemented in one southeastern school district over a 4-year period. SLO systems serve many purposes. For instance, SLOs can address the limitation of using VAMs to evaluate teachers, which can only be calculated in tested subjects and grades. Our results come from SLOs in tested subjects/grades, and it is important to acknowledge that target-setting behaviors might be different in untested grades. The act of setting targets itself could change how teachers use data to inform their practice, monitor their students' success, and reflect on what seems to work in their classroom. The SLO theory of action posits that, through thoughtful reflection on data and/or by setting high expectations for all students, learning outcomes will improve over time.

While a burgeoning literature describes the implementation process for SLO systems and documents the many

TABLE 8
Conditional Differences in Targets Set by Race/Ethnicity, Separately for Four Tests

| | SSAT Test | | | | | | | |
|---------------|--------------------|--------------------|-------------------|--------------------|--------------------|-------------------|-------------------|--------------------|
| | ELA | | | | Math | | | |
| | (4a) | (4b) | (4c) | (4d) | (4a) | (4b) | (4c) | (4d) |
| Black | -0.49*** (0.04) | -0.03 (0.02) | -0.03 (0.02) | -0.03 (0.02) | -0.24*** (0.04) | 0.04 (0.03) | 0.03 (0.03) | -0.01 (0.02) |
| Asian | -0.84*** (0.11) | -0.30*** (0.06) | -0.20** (0.06) | -0.23*** (0.06) | -0.21* (0.11) | 0.02 (0.08) | 0.03 (0.09) | -0.01 (0.07) |
| Latinx | -0.38*** (0.06) | -0.05 (0.03) | 0.01 (0.04) | 0.01 (0.03) | -0.09 (0.07) | -0.03 (0.05) | -0.03 (0.06) | -0.03 (0.04) |
| Other/unknown | -0.79*** (0.12) | -0.01 (0.06) | 0.11 (0.10) | 0.11 (0.08) | -0.49*** (0.13) | 0.12 (0.09) | 0.31* (0.16) | 0.18 (0.11) |
| Constant | -0.02 (0.05) | -0.03 (0.03) | 0.03 (0.03) | 0.22*** (0.03) | 0.12* (0.05) | 0.37*** (0.04) | 0.33*** (0.04) | -0.29*** (0.03) |
| R^2 | .065 | .743 | .745 | .83 | .174 | .559 | .56 | .786 |
| N | 8,727 | 8,727 | 8,685 | 8,335 | 8,105 | 8,105 | 8,061 | 7,916 |

| | MAP Test | | | | | | | |
|---------------------|--------------------|----------------------|----------------------------|------------------------|--------------------|----------------------|-------------------------|------------------------|
| | ELA | | | | Math | | | |
| | (4a) | (4b) | (4c) | (4d) | (4a) | (4b) | (4c) | (4d) |
| Black | -0.46*** (0.04) | -0.06 (0.03) | -0.06* (0.03) | -0.04 (0.03) | -0.48*** (0.04) | -0.11*** (0.03) | -0.11*** (0.03) | -0.07** (0.03) |
| Asian | -1.02*** (0.11) | -0.50*** (0.08) | -0.47*** (0.09) | -0.47*** (0.09) | -0.59*** (0.09) | -0.18** (0.07) | -0.14 (0.08) | -0.09 (0.07) |
| Latinx | -0.30*** (0.07) | -0.02 (0.05) | 0.00 (0.05) | 0.03 (0.05) | -0.30*** (0.06) | -0.09* (0.04) | -0.05 (0.05) | -0.01 (0.04) |
| Other/unknown | -0.68*** (0.11) | -0.13 (0.09) | 0.12 (0.14) | 0.17 (0.13) | -0.81*** (0.11) | -0.24** (0.08) | 0.12 (0.13) | 0.13 (0.11) |
| Constant | -0.41*** (0.05) | -0.33*** (0.04) | -0.19*** (0.04) | 0.05 (0.06) | -0.61*** (0.05) | -0.48*** (0.03) | -0.38*** (0.04) | -0.27*** (0.04) |
| R^2 | .039 | .473 | .479 | .56 | .052 | .516 | .52 | .627 |
| N | 9,114 | 9,114 | 9,060 | 8,702 | 8,583 | 8,583 | 8,536 | 8,393 |
| Model specification | Grade, year FE's | + Prior achievements | + Other student covariates | + Teacher-by-year FE's | Grade, year FE's | + Prior Achievements | + Other stud covariates | + Teacher-by-year FE's |

Note. The outcome, target score set in the fall for the EOY (spring) test, has been standardized (i.e., we subtract the district-wide mean and standard deviation [within subject-grade-year] of the EOY scores in their original metric from all target and spring scores). Model (4a) only includes grade and year fixed effects. We add a vector of prior achievement in both tests in both subjects in 2 prior years (Model 4b), a vector of student demographics other than race including gender, free or reduced-price lunch status, and limited English proficiency (Model 4c), and finally teacher-by-year fixed effects (Model 4d). ELA = English language arts; EOY = end-of-year; SSAT = statewide standardized achievement test; MAP = Measures of Academic Progress; SLO = student learning objectives; FE = fixed effects. * $p < .05$. ** $p < .01$. *** $p < .001$.

challenges that often arise, there is little evidence about actual teacher practices in target setting. Teachers face a choice: They can set for ambitious targets in hopes that doing so improves their students' outcomes, or they can set less ambitious targets that are easier to meet, which in turn increases bonus payments. A priori, we might anticipate a tendency to act in one's own self-interest, and given that teaching is generally not a well-paid occupation, an up to \$5,000 annual bonus may feel particularly high-stakes for teachers. Moreover, we also might expect teachers to hone their ability to set accurate and/or attainable goals as they become familiar with the system. Another possibility is that, with the limited information at their disposal in the fall, teachers are unable to predict EOY test scores with the degree of accuracy that this SLO system presumes is possible. In any of these scenarios, the hypothesized mechanisms for SLO systems would be undermined.

We find that many teachers set targets for their students that they did not always attain. Depending on the test, teachers only met, on average, between 43% and 65% of the learning targets they selected for their students. Teachers exhibited the ability to establish targets that broadly corresponded to later performance, generally within 0.30 *SDs* of EOY scores. However, the targets teachers set for students varied dramatically in terms of their ambitiousness. For instance, while 29% of targets set for the SSAT ELA were at least 0.50 *SDs* above a regression-adjusted predicted score, we see equal numbers of targets that were at least 0.50 *SDs* below the predicted scores. We also must grapple with the finding that, even when teachers set targets that were lower than would be expected based on prior performance, they often could not get students to meet those so-labeled unambitious targets (see Figure 5). We nonetheless find evidence supporting the idea that students have higher EOY test scores when teachers set higher targets for them.

Finally, if racial/ethnic biases are embedded within target setting, which in turn may affect student outcomes, then we worry that SLO processes could inadvertently exacerbate achievement disparities. We do not find, among students of the same prior skill, that White, Black, or Latinx students receive systematically different targets from one another. There is some concerning evidence that suggests the Asian students receive lower targets on ELA tests than their White counterparts (though not on math tests). These findings warrant continued attention, but they should be interpreted with caution since only 6% of the students in this district are Asian.

Taken together, the current study takes a step forward in our understanding of how teachers might react to and implement a SLO process. Because teachers' success in reaching these objectives often plays a role in how they are evaluated—and sometimes how much they are paid—it is not at all clear how teachers will approach this task. Given the increasing use of SLOs, both as a complement to and

substitute for other measures of teaching effectiveness, it is important for the field to have empirical evidence about whether the theory of action behind SLO systems is consistent with SLO implementation in practice.

Acknowledgments

We are grateful Martin West and Mark Long for their feedback on drafts of this article presented at the APPAM and AEFPP conferences, respectively. Support has also been provided by IES Grant R305B100009 to the University of Virginia. The views expressed in the article are solely those of the authors. Any errors are attributable to the authors.

ORCID iD

Allison Atteberry  <https://orcid.org/0000-0002-9409-4372>

Notes

1. Despite the potential promise of SLOs according to the theoretical framing presented above, effective implementation of SLO systems has proven challenging. For example, finding the data necessary to create the SLOs can prove time consuming, especially where teachers must design the assessments (Donaldson et al., 2014; Lachlan-Haché, 2015; Lamb & Schmitt, 2012; Schmitt, 2013; The New Teacher Project, 2012). In fact, assessment design is one of the areas in which teachers report they do not feel qualified to do what is asked of them with SLO implementation (Lachlan-Haché, 2015; Lamb & Schmitt, 2012; Schmitt, 2013). It can also be difficult to communicate best practices in assessment selection or design (Delaware Department of Education, 2013; Donaldson et al., 2014; Lachlan-Haché, 2015; Lachlan-Haché et al., 2013; Lamb et al., 2013; Schmitt et al., 2009; Slotnik et al., 2013). Issues related to assessment design and selection are of less relevance in the current context, wherein the district centrally selected which assessments were used for each grade level in every SLO school. In our interactions with district personnel, it is clear that the providing student-level historical test score data, recording the thousands of learning objectives set, and later evaluating whether those learning objectives were met was a large undertaking.

2. In some versions of SLOs, teachers set learning targets for the classroom overall (e.g., the percentage of students who will meet a given threshold) rather than setting targets for individual students.

3. In this district, SLOs were set in all grades; however, data are only available in Grades 4 through 8.

4. Prior-year test scores were provided to teachers by the district as an Excel spreadsheet populated with data only for the students in the teacher's class. The spreadsheets included four test scores per student from the preceding spring—the student's EOY state standardized tests in ELA and math, and the student's test scores on the Measure of Academic Progress (MAP, a computer-adaptive benchmark assessment) in ELA and math. Though MAP is often also administered in the fall, no fall MAP scores were not included on the spreadsheets. The scores were presented only in a table format (i.e., not graphed). Teachers were provided with guidance on how test scores could be translated into performance categories or benchmarks, but scores were only provided in their continuous form.

5. In about 6% of cases, teachers revised their targets later in the year. They were allowed to do so either in the case of

excessive student absences (20% of the year); or changes to student's environment (outside teacher control) arose that affects his/her ability to meet the goal. Examples include homelessness; loss of a parent, grandparent or guardian; change of guardian; guardian arrest; involvement with the juvenile justice system; illness/hospitalization.

6. The MAP assessment is used as a supplementary tool to aid schools' in improving their instruction. NWEA's MAP test is computer adaptive and is designed so that its scores can be expressed on a vertical scale. However, in practice we standardize all scores within subject-grade-year in this article. Though the NWEA is often administered by schools to their students in the fall, winter, and spring of the school year, teachers were not provided with their students' fall NWEA test scores for the purposes of setting targets. They received only the MAP and SSAT test scores from the prior spring. NWEA reports marginal reliabilities in the low to mid .90s (NWEA Technical Manual, 2011), and the SSAT technical manual reports reliabilities between .87 and .93 (State Technical Manual, 2015).

7. Teachers also set targets for K–3 students on different tests, however, data are only available for Grades 4 through 8.

8. Exact bonus amounts were based on the percentage of students meeting the target. For instance, if 80% of students met their target, then the teacher would receive 80% of the maximum bonus of \$5000.

9. Teachers were told by the researchers at the start of the school-year which of their students (who were, in reality, randomly selected) exhibited “unusual potential for intellectual growth” according to a pretest and would likely exhibit significant growth academically within the year. Indeed, when the researchers assessed the students at the end of the school year, the randomly selected students *did* exhibit higher test scores.

10. Hill and Jones (2017) conducted a similar analysis on grade 3 through 8 student achievement outcomes and also found large, positive causal effects of teacher expectations, particularly in lower grades.

11. Students with a growth mindset believe that intelligence is something one can develop with effort, rather than a fixed trait, and research has shown both that interventions can improve students' growth mindset, which in turn improved their educational outcomes (Yeager et al., 2019).

12. Though teachers set targets in the original scale of each test, we have standardized all scores (including the targets) across the district, within subject, grade, and year.

13. We predict EOY scores as a function of student demographic variables, as well as all available prior-year test scores in both subjects on both MAP and SSAT tests, with a third-order polynomial for each. We include students with idiosyncratically missing prior test scores by imputing the mean test-subject-grade-year test score and including a dummy variable indicating missingness. About 10% of observations have at least 1 missing prescore. To ensure our findings are robust to the decision to impute, we also replicate relevant analyses instead by simply excluding cases with missing prescores and find quite similar results, available on request. R^2 values for these models were between .75 (in earliest school year) and .87 (in latest school years). We then use the model coefficients to estimate “expected” scores for each student in each school year. In the approximately 6% of cases in which teachers changed the targets midyear, we used the higher target, typically the target set earlier in the fall.

14. This vector includes gender, free-/reduced-price lunch status, and any observed designation of limited English proficiency.

15. For the purpose of visualization, focusing on the 1st to 99th percentile range ensures that the distribution is not obscured by a few outliers. We find that, on closer examination of these outliers, they were possibly driven by idiosyncratic data entry errors. For instance, 34 of the 9,124 MAP ELA targets set (.37%) are between 3 and 6 *SDs* in absolute value from the average ambitiousness score, and it is possible that teachers accidentally assigned targets on a different scale than the EOY test was reported in. Since we cannot definitively account for these unusual targets, we do not remove them from the analytic sample. However, since there are so few of them (less than half a percentage), they do not affect our results.

16. We use a relatively common specification of a teacher value-added model: Student outcomes in a given year are modeled as a function of lagged student achievement in the preceding 2 years across up to four content areas (ELA, math, history, and science), student demographics, school demographics, grade fixed effects, and teacher fixed effects. The coefficients on the teacher indicator variables become the VAMs themselves, and we calculate these separately for each year and subject area.

References

- Berg-Jacobson, A. (2016). *Teacher effectiveness in the Every Student Succeeds Act: A discussion guide*. American Institute for Research. https://www.air.org/sites/default/files/downloads/report/TeacherEffectiveness_ESSA.pdf
- Burgess, S., & Greaves, E. (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities. *Journal of Labor Economics*, 31(3), 535–576. <https://doi.org/10.1086/669340>
- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2007). The influence of stereotypes on decisions to shoot. *European Journal of Social Psychology*, 37(6), 1102–1117. <https://doi.org/10.1002/ejsp.450>
- Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, 86(1), 195–210. <https://doi.org/10.1162/003465304323023750>
- Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95(2), 158–165. <https://doi.org/10.1257/000282805774670446>
- Delaware Department of Education. (2013). *Continuous improvement: A report on “year one” of the revised DPAS-II educator evaluation system*. https://www.doe.k12.de.us/cms/lib/DE01922744/Centricity/domain/271/present%20and%20reports/DPAS_II_Year_One_Report_2013.pdf
- Donaldson, M., Cobb, C., LeChasseur, K., Gabriel, R., Gonzales, R., Woulfin, S., & Makuch, A. (2014). *An evaluation of the pilot implementation of Connecticut's system for educator evaluation and development*. Center for Education Policy Analysis.
- Dweck, C. S. (2008). *Mindset: The new psychology of success*. Random House Digital.
- Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review*, 45(April), 44–52. <https://doi.org/10.1016/j.econedurev.2015.01.007>
- Ferguson, R. F. (2003). Teachers' perceptions and expectations and the Black-White test score gap. *Urban Education*, 38(4), 460–507. <https://doi.org/10.1177/0042085903038004006>

- Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, 52, 209–224. <https://doi.org/10.1016/j.econedurev.2016.03.002>
- Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M. R. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, 22(9), 1231–1238. <https://doi.org/10.1007/s11606-007-0258-5>
- Hill, A., & Jones, D. (2017). *Self-fulfilling prophecies in the classroom*. SSRN. <http://doi.org/10.2139/ssrn.3016029>
- Holt, S. B., & Gershenson, S. (2019). The impact of demographic representation on absences and suspensions. *Policy Studies Journal*, 47(4), 1069–1099. <https://doi.org/10.1111/psj.12229>
- Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, 9(2), 131–155. https://doi.org/10.1207/s15327957pspr0902_3
- Lachlan-Haché, L. (2015). *The art and science of student learning objectives: A research synthesis*. American Institutes for Research.
- Lachlan-Haché, L., Matlach, L., Guiden, A., & Castro, M. (2015). *What we know about SLOs: An annotated bibliography of research on and evaluations of student learning objectives*. <https://www.air.org/sites/default/files/downloads/report/Student-Learning-Objectives-Annotated-Bibliography-April-2015.pdf>
- Lachlan-Haché, L., Matlach, L., Reese, K., Cushing, E., & Mean, M. (2013). *Student learning objectives: Early lessons from the Teacher Incentive Fund*. Teacher Incentive Fund Technical Assistance Network.
- Lacireno-Paquet, N., Morgan, C., & Mello, D. (2014). *How states use student learning objectives in teacher evaluation systems: A review of state websites (REL 2014-013)*. Institute of Education Sciences & Regional Educational Laboratory Northeast & Islands. https://ies.ed.gov/ncee/edlabs/regions/northeast/pdf/REL_2014013.pdf
- Lamb, L., & Schmitt, L. (2012). *AISD REACH program update, 2010–2011: Participant feedback* (DRE Report No. 10.86 RB). Austin Independent School District.
- Lamb, L., Schmitt, L., Gross, R., & Cornetto, K. (2013). *Austin Independent School District (AISD) pilot teacher appraisal system update: 2012–2013 focus group and survey summary* (DRE Publication No. 12.70). Austin Independent School District.
- Madon, S., Jussim, L., & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology*, 72(4), 791–809. <https://doi.org/10.1037/0022-3514.72.4.791>
- The New Teacher Project. (2012). *Summer report: Creating a culture of excellence in Indiana schools*. Indiana Department of Education.
- Papageorge, N. W., Gershenson, S., & Kang, K. M. (2020). Teacher expectations matter. *Review of Economics and Statistics*, 102(2), 234–251. https://doi.org/10.1162/rest_a_00838
- Proctor, C. P. (1984). Teacher expectations: A model for school improvement. *Elementary School Journal*, 84(4), 469–481. <https://doi.org/10.2307/1001371>
- Rosenthal, R. (1987). “Pygmalion” effects: Existence, magnitude, and social importance. *Educational Researcher*, 16(9), 37–41. <https://doi.org/10.2307/1175728>
- Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *Urban Review*, 3(1), 16–20. <https://doi.org/10.1007/BF02322211>
- RTTT Technical Assistance Network. (2010). *Measuring student growth for teachers in non-tested grades and subjects: A primer*. ICF International.
- Rubie-Davies, C., Hattie, J., & Hamilton, R. (2006). Expecting the best for students: Teacher expectations and academic outcomes. *British Journal of Educational Psychology*, 76(3), 429–444. <https://doi.org/10.1348/000709905X53589>
- Schmitt, L. (2013). *AISD REACH program update, 2012–2013: Peer observation* (DRE Report No. 12.89). Austin Independent School District.
- Schmitt, L., Cornetto, K., Lamb, L., & Imes, A. (2009). *AISD Reach Year 2 Evaluation Report I, 2008–2009* (DRE Publication No. 08.53). Austin Independent School District.
- Seaton, F. S. (2018). Empowering teachers to implement a growth mindset. *Educational Psychology in Practice*, 34(1), 41–57. <https://doi.org/10.1080/02667363.2017.1382333>
- Slotnik, W. J., Smith, M. D., & Liang, G. (2013). *Focus on Rhode Island: Student learning objectives and evaluation*. <http://www.ctacusa.com/wp-content/uploads/2013/11/FocusOnRhodeIsland.pdf>
- Tyler Eastman, S., & Billings, A. C. (2001). Biased voices of sports: Racial and gender stereotyping in college basketball announcing. *Howard Journal of Communication*, 12(4), 183–201. <https://doi.org/10.1080/106461701753287714>
- U.S. Department of Education. (2012). *Teacher Incentive Fund guidance*. U.S. Department of Education. <http://www2.ed.gov/programs/teacherincentive/faqs2012.pdf>
- Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C. P., Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Yang, S. M., Carvalho, C. M., . . . Hinojosa, C. P. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, 573(7774), 364–369. <https://doi.org/10.1038/s41586-019-1466-y>

Authors

ALLISON ATTEBERRY is an assistant professor of research and evaluation methodology at the University of Colorado Boulder School of Education. Her research focuses on policies and interventions that are intended to help provide effective teachers to the students who need them most.

SARAH E. LACOUR is an assistant professor at the University of Kentucky College of Education. Her research primarily uses quasi-experimental designs to explore the impacts of law and larger educational policy on students and communities.