Differences in Teachers' Assessments of Students by English Learner Status

By Maria Zhu*

Prior research has shown that teachers' perceptions of students matter for academic outcomes.[1] Teachers' perceptions may operate directly through recommendations for grade retention or different academic tracks or indirectly through the so-called "Pygmalion effect." One potential concern arising from these findings is that teachers may have inaccurate perceptions of students due to negative stereotypes or biases, which could exacerbate existing inequalities.

This paper provides new evidence on differences in teachers' perceptions of English Learner (EL) students compared to non-EL students using data on elementary and middle students in North Carolina. In this setting, students take standardized tests to assess their achievement at the end of the year. Additionally, teachers also provide assessments of achievement during this time. Teachers are asked to evaluate their students on their mastery of math and reading for the same skills that standardized tests in these subjects evaluate. I focus on teacher assessments in math, a subject that does whose mastery does not directly pertain to English mastery, to assess whether language proficiency colors teachers' perceptions. To address the possibility that disparities in teachers' assessments are driven by differences in achievement between EL and non-EL students, I focus the analysis on comparisons of EL and non-EL students with the same underlying achievement, as measured by standardized test scores. Additionally, I restrict the analysis to comparisons students to peers in the same class to control for unobserved factors varying at the classroom level that may affect teacher assessments.

Results indicate teachers are more significantly more negative in their assessments of EL students compared to non-EL peers in the same classes with the same level of standardized test-based achievement. Teachers are more likely to assess an EL student as below grade level proficiency in math achievement by 1.1 to 2.3 percentage points compared to non-EL peers. This effects represent an increase in the propensity to judge a student as below proficiency by 4.5 to 9.4 percent of the baseline rate of assessing students as below proficiency.

This paper relates to a growing body of work on group-level disparities in teachers' assessments of achievement, which thus far has focused race and gender differences (Lavy, 2008; Burgess and Greaves, 2013; Botelho, Madeira and Rangel, 2015; Lindahl, 2016; Shi and Zhu, 2023). This paper provides novel evidence on disparities in teachers' assessments by English proficiency. These disparities are important to understand, as the population as EL students in the US is growing rapidly (Flynn and Hill, 2005), and many of these students come from disadvantaged backgrounds. Results suggest that challenges EL students face in the classroom due to language barriers are compounded by negative teacher perceptions of EL students.

I. Data and descriptive statistics

This paper uses administrative data from the North Carolina Education Research Data Center (NCERDC), covering all public school students in North Carolina. Key to this study, in addition to observing demographic characteristics of students, I observe two measures of student achievement. First, students take standardized tests in math and reading at the end of the school year, which are multiple-choice and

* Syracuse University, Department of Economics, mzhu33@syr.edu.

[1] See Rosenthal and Jacobson (1968); Papageorge, Gershenson and Kang (2020); Hill and Jones (2021).

machine-scored. Second, students receive assessments in math and reading from their teachers as well. Teacher assessments are collected the end of the year, before standardized test results come out. Teachers assess students on a scale of 1 to 4:

1) **Insufficient mastery**: Students performing at this level do not have sufficient mastery of knowledge and skills in this subject area to be successful at the next grade level.
2) **Inconsistent mastery**: Students performing at this level demonstrate inconsistent mastery of knowledge and skills in this subject area and are minimally prepared to be successful at the next grade level.
3) **Consistent mastery**: Students performing at this level consistently demonstrate mastery of grade level subject matter and skills and are well prepared for the next grade level.
4) **Superior performance**: Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient at grade level work.

An assessment of 3 or above is considered to be proficient mastery for the grade level. Teachers' assessments are low-stakes evaluations that do not directly affect students. North Carolina uses teacher assessments as one of multiple inputs to help calibrate standardized test score interpretation. Teacher assessments are intended to assess mastery of the same material that standardized tests cover. Furthermore, teachers are explicitly instructed to assess students on mastery of the subject material, rather than behavioral characteristics such as absences and failure to turn in homework.

This study focuses on students in grades 3-8 from 2007-2013, which is the sample for which both standardized test score and teacher assessment is available. Table 1 presents descriptive statistics for the sample by EL status. EL students are those who are classified by the state as having "limited English proficiency," and approximately 7 percent of students in the sample are EL students. Column (1) displays characteristics of non-EL students in the sample, and column (2) displays characteristics of EL students. EL students are much more likely to be Hispanic or Asian and less likely to be White or Black, compared to non-EL students. While 5 percent of non-EL students are Hispanic and 2 percent are Asian, these numbers for EL students are 83 percent and 10 percent, respectively. EL students are also much more likely to be classified as economically disadvantaged, with 48 percent of non-EL students being economically disadvantaged compared to 86 percent of EL students.

Next, Figure 1 displays math standardized test score distributions. Test scores are normalized within year and grade to have a mean of zero and standard deviation of one. The figure shows that the distribution of test scores for non-EL students is first order stochastically dominant to the distribution of test scores for EL students for virtually the entire support, indicating non-EL students are higher performing overall compared to EL students.

II. Empirical strategy

The following equation estimates differences in teacher assessments of EL students compared to non-EL students:

$$(1) \qquad Y_{ic} = \beta EL_{ic} + f(Test_{ic}) + \mathbf{X}'_{ic}\Gamma + \zeta_c + \epsilon_{ic}$$

where Y_{ic} is an indicator variable taking a value of one if a teacher rates student i in class c as being below proficiency level (i.e., assesses the student's achievement at a level 1 or 2) and zero otherwise. The variable EL_{ic} is an indicator variable taking a value of one if the student is an EL student and zero otherwise. To address the possibility disparities in teacher assessments reflect achievement differences by EL status, I control for underlying standardized test scores, $Test_{ic}$, using a third order polynomial. The vector $\mathbf{X}'_{ic}$ controls for student race/ethnicity, gender, and economic disadvantage status. Finally, ζ_c is a class fixed effect. By construction, each observation in a given class is in the same school, teacher, course, and year. Thus, the inclusion of class fixed effects controls possibility that student sorting along these dimensions drives differences in teacher assessments by EL status. For example, class fixed effects address the concern that schools with high concentrations of EL students have different assessment standards than those with lower concentrations or that both the share of EL students and and assessment standards have changed over time.

The coefficient of interest, β, captures differences teachers' assessments of EL students compared to non-EL peers in the same class with the same achievement, as measured by standardized test scores. The key identifying assumption is that conditional on students' test scores, there are no unobserved factors correlated with EL status that affect differences in teachers' assessments of students by EL status.

One potential concern with the empirical strategy is that results will be biased by measurement error in standardized test scores. Measurement error can arise from a number of sources, such as test instrument noise (e.g., randomness in test question selection), testing conditions (e.g., illness on day of test), and idiosyncratic factors (e.g., luck in guessing question answers). This is concerning because classical measurement error in one regressor of interest will also bias coefficient estimates of other, non-mis-measured variables (DeGroot and Schervish, 2011). Furthermore, studies have found that standardized tests measure achievement with a sizable amount of error (Boyd et al., 2013).

Previous studies have used instrumental variables to address measurement error issues in test scores. A valid instrument needs to be strongly correlated with test scores and only influence teacher assessments through test scores. Commonly used instruments for standardized test scores include test scores in a different subject or lagged test scores in the same subject (Zabel, 2008; Hanushek and Rivkin, 2009; Botelho, Madeira and Rangel, 2015; Bond and Lang, 2017). I use these instruments to correct for measurement error in this study as well, and I discuss advantages and limitations of each candidate instrument in the next section.

III. Results

Table 2 reports estimation results of Equation 1. Column (1) displays ordinary least squares estimation results, and columns (2)-(6) display instrumental variables estimation results, with each column using a different instrument. OLS estimates indicate teachers are 5.2 percentage points more likely to evaluate an EL student as not meeting math proficiency standards compared to non-EL peers in the same course with the same standardized test scores. This is a sizable magnitude, considering the overall propensity for teachers to assess students as below proficiency standards in math is 0.246.

Next, I correct for measurement error in test scores using various instruments. First, I instrument for standardized test scores in math using math test scores from the prior year in column (2) and math scores from two years go in column (3). This instrument is plausibly valid since teachers are told explicitly to assess students on achievement at the time of assessment. However, a drawback of this instrument is that teachers' assessments may still be influenced by their impressions of students at the beginning of the year, which is likely reflected in prior achievement scores. To address this, I next instrument for standardized test scores in math using standardized tests in reading in column (4). A limitation of using reading test scores in the same year is that there may be correlated errors between math and reading scores, if for example, a student is sick during the week of testing. Finally, I instrument for math test scores using lagged and twice lagged reading test scores in columns (5) and (6), respectively. These instruments are arguably the least likely to suffer from the above concerns.

IV estimates consistently show across instruments that teachers are more likely to judge EL students as being below proficiency levels in math compared to non-EL peers in the same classes with the same standardized test-based achievement. However, the magnitude of these gaps is less than half of that of OLS estimates, indicating correcting for measurement error is important in this setting. IV estimates indicate teachers are more likely to assess an EL student as below proficiency in math achievement by 1.1 to 2.3 percentage points compared to non-EL peers. These effects represent an increase in the propensity to judge a student as below proficiency by 4.5 to 9.4 percent of the baseline rate of teachers assessing students as below proficiency.

IV. Conclusion

This paper provides novel evidence that teachers assess EL students as being lower performing in math, compared to non-EL students with the same underlying achievement, as measured by standardized tests. These findings suggest EL students face disadvantages in perceived achievement even in subjects that are not directly related to language skills. These findings are especially troubling since teachers' perceptions of students matter for academic outcomes, and results indicate disparities in teachers' assessments may exacerbate existing achievement gaps.

REFERENCES

Bond, Timothy, and Kevin Lang. 2017. "The Black-White Education Scaled Test-Score Gap in Grades K-7." *Journal of Human Resources*, 53: 0916–8242R.

Botelho, Fernando, Ricardo A. Madeira, and Marcos A. Rangel. 2015. "Racial Discrimination in Grading: Evidence from Brazil." *American Economic Journal: Applied Economics*, 7(4): 37–52.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2013. "Measuring Test Measurement Error: A General Approach." *Journal of Educational and Behavioral Statistics*, 38(6): 629–663. Publisher: American Educational Research Association.

Burgess, Simon, and Ellen Greaves. 2013. "Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities." *Journal of Labor Economics*, 31(3): 535–576.

DeGroot, Morris H., and Mark J. Schervish. 2011. *Probability and Statistics.* . 4th edition ed., Boston:Pearson.

Flynn, Kathleen, and Jane Hill. 2005. "English Language Learners: A Growing Population."

Hanushek, Eric, and Steven Rivkin. 2009. "Harming the best: How schools affect the black-white achievement gap -." *Journal of Policy Analysis and Management*, 28(3): 366–393.

Hill, Andrew J., and Daniel B. Jones. 2021. "Self-Fulfilling Prophecies in the Classroom." *Journal of Human Capital.* Publisher: The University of Chicago PressChicago, IL.

Lavy, Victor. 2008. "Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment." *Journal of Public Economics*, 92(10): 2083–2105.

Lindahl, Erica. 2016. "Are teacher assessments biased? – evidence from Sweden." *Education Economics*, 24(2): 224–238.

Papageorge, Nicholas W., Seth Gershenson, and Kyung Min Kang. 2020. "Teacher Expectations Matter." *The Review of Economics and Statistics*, 102(2): 234–251.

Rosenthal, Robert, and Lenore Jacobson. 1968. "Pygmalion in the classroom." *The Urban Review*, 3(1): 16–20.

Shi, Ying, and Maria Zhu. 2023. ""Model minorities" in the classroom? Positive evaluation bias towards Asian students and its consequences." *Journal of Public Economics*, 220: 104838.

Zabel, Jeffrey E. 2008. "The Impact of Peer Effects on Student Outcomes in New York City Public Schools." *Education Finance and Policy*, 3: 197–249. Publisher: MIT Press.

Table 1—Student Characteristics by English Learner Status

	Non-EL (1)	EL (2)
White	0.58	0.03
Black	0.29	0.02
Hispanic	0.05	0.83
Asian	0.02	0.10
Other race	0.06	0.01
Female	0.49	0.46
Econ. disadv.	0.48	0.86
N	4,154,180	298,801

Note: Observations represent student-year level observations for students in grades 3-8 between 2007-2013.

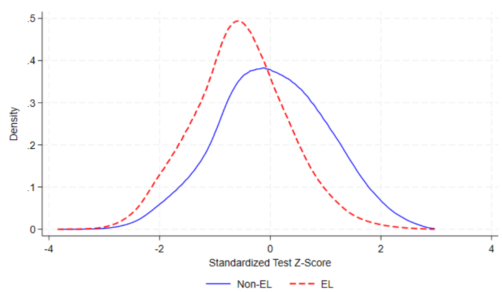

Figure 1. Math Test Scores Distributions by English Learner Status

Note: Standardized tests scores have been normalized within year and grade to have a mean of zero and standard deviation of one.

Table 2—Differences in Teachers' Propensities to Rate Students as Below Proficient in Math by English Learner Status

		Instrumental variables				
	OLS (1)	Lagged Subject (2)	Twice Lagged Subject (3)	Other Subject (4)	Lagged Other Subject (5)	Twice Lagged Other Subject (6)
English Learner	0.052	0.017	0.023	0.019	0.011	0.013
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
N	2,360,880	2,360,880	2,360,880	2,360,880	2,360,880	2,360,880
Baseline	0.246	0.246	0.246	0.246	0.246	0.246
First stage F-stat		15,009	10,649	7,254	5,229	3,900

Note: Standard errors are in parentheses clustered at the teacher level. The outcome is an indicator variable taking a value of one if a teacher rates a students as below proficiency level (i.e, the teacher assesses the student's achievement as 1 or 2) and zero otherwise. All specifications include third-order polynomial controls for standardized test scores and classroom fixed effects, as well as controls for student race/ethnicity, gender, and economic disadvantage status. Sample is restricted to individuals who have information available for other subject test scores, lagged same- and other- subject test scores, and twice lagged same- and other-subject test scores in order to have a consistent sample across specifications. Kleibergen-Paap F-statistics reported.