Teacher Quality Gaps by Disability and Socioeconomic Status: Evidence From Los Angeles

Ijun Lai[1], W. Jesse Wood[2], Scott A. Imberman[2,3], Nathan D. Jones[4], and Katharine O. Strunk[2]

Although most students with disabilities (SWDs) receive instruction from general education teachers, little empirical work has investigated whether these students have suitable access to high-quality teachers. We explore the differences in teacher quality experienced by SWDs and students without disabilities (non-SWDs) in the Los Angeles Unified School District, examining how access varies within schools as well as across school-level disadvantage rates. We leverage several different indicators of teacher effectiveness for general education teachers who instruct both SWDs and non-SWDs. We find that SWDs are significantly more likely to have teachers with lower math value-added (–0.024 standard deviations) than their non-SWD peers, and we find emerging gaps in teacher evaluation scores and exposure to novice teachers. In general, these gaps do not vary by school-level disadvantage.

Keywords: correlational analysis; equity; regression analysis; secondary data analysis; special education; statistics; teacher characteristics

Since the 1975 passage of the Individuals With Disabilities Education Act (IDEA), federal law has dictated that students with disabilities (SWDs) be provided with a "free and appropriate public education in the least restrictive environment." In 2017, the U.S. Supreme Court established an even higher standard for special education with its decision in *Endrew F. v. Douglas County School District RE-1*. The case stressed the need to ensure equitable outcomes for SWDs, who continue to lag behind their nondisabled peers in math and reading achievement (e.g., Chudowsky et al., 2009; Schulte et al., 2016; Schulte & Stevens, 2015). Today, roughly 6.4 million public school students in the United States receive special education services, and the majority receive their primary instruction in general education (U.S. Department of Education, 2019). Yet, to date, little empirical work has examined teacher quality for SWDs in the general education classroom context, raising critical questions about teacher quality gaps (TQGs) between students with disabilities in general education classrooms and their peers.

The teacher quality literature has repeatedly documented the unequal distribution of teachers—both within and across schools. Research shows that within schools, students sorting to teachers of varying quality often depends on their academic and behavioral histories (e.g., Kalogrides & Loeb, 2013; Lankford et al., 2002). Furthermore, Goldhaber et al. (2015, 2018) establish that substantial gaps exist in teacher quality across the socioeconomic distribution; low-income students consistently have less access to highly qualified teachers, likely because disadvantaged schools have more difficulty attracting and retaining high-quality teachers (e.g., Boyd et al., 2005). Unfortunately, we do not have a parallel literature on TQGs experienced by students with disabilities.

It is unclear, based on existing research, whether to expect between- or within-school TQGs between SWDs in general education classrooms and their non-SWD (i.e., students without disabilities) peers. Nonetheless, there is some evidence to suggest that schools distribute SWDs to teachers in nonrandom ways. One recent study using North Carolina data (Gilmour & Henry,

[1]Mathematica, Oakland, CA, USA
[2]Michigan State University, East Lansing, MI, USA
[3]NBER, Cambridge, MA, USA
[4]Boston University, Boston, MA, USA

Educational Researcher, Vol. XX No. X, pp. 1–12
DOI: 10.3102/0013189X20955170
Article reuse guidelines: sagepub.com/journals-permissions

2020) found that SWDs were more likely to have classmates with lower prior academic performance than their non-SWD peers. However, it is not clear whether this sorting results in differential access to high-quality teachers. On the one hand, it is possible that SWDs are seen as more "difficult" students to teach and, consequently, more likely to be assigned to lower-quality teachers (e.g. Clotfelter et al., 2006). At the same time, because in many states and districts there is a higher level of accountability or attention paid to the opportunities given to SWDs (see, for example, Swaak, 2020), districts may proactively assign SWDs to particularly effective teachers. Indeed, using a variety of teacher quality measures, Gilmour and Henry (2018) find little evidence of TQGs for SWDs overall in North Carolina. The gaps they did observe were due to within-school differences in teachers' prior academic achievement. They also noted that a focus on SWDs overall masked some heterogeneity across disability subcategories.

To ensure equitable outcomes for SWDs, it is also important to identify between-school variation in teacher quality. The question of how TQGs vary by school-level disadvantage was not addressed in Gilmour and Henry (2018). We hypothesize that TQGs in high-poverty schools may be felt even more acutely by SWDs. Gilmour and Wehby (2019) demonstrate that the likelihood of teacher turnover increases with the number of SWDs in the classroom. Given the higher concentration of SWDs in higher-poverty schools, this finding suggests that SWDs in disadvantaged schools may be even less likely to access high-quality teachers than both non-SWDs in high-poverty schools and SWDs in low-poverty schools. However, to date, no studies have directly examined whether, within and across schools of varying income levels, TQGs exist across students with and without disabilities.

To promote more equitable outcomes for SWDs, we join these two strands of research and ask the following research questions in the Los Angeles Unified School District (LAUSD) context:

1) Does teacher quality vary across schools with differing degrees of disadvantage?
2) Are there SWD vs. non-SWD TQGs?
3) Do SWD vs. non-SWD gaps vary by school-level disadvantage?
4) Do TQGs across school-level disadvantages vary by specific disability type?

This article makes two primary contributions to the literature on TQGs among SWDs. First, we include multiple teacher quality indicators including value-added measures (VAMs), teachers' observation-based performance ratings, hiring scores, and teacher experience (novice status). This allows us to examine whether TQGs exist across a large, urban context (LAUSD) using an expanded range of quality measures. Second, we examine whether the small overall quality gaps found by Gilmour and Henry (2018) might mask school-level variation related to students' socioeconomic status. If higher-quality teachers are sorting into schools with fewer low-socioeconomic-status students, we might find that TQGs are exacerbated across schools within the same district. Therefore, it is important to understand how TQGs might differ across schools with different degrees of student disadvantage rather than just overall differences within a district or state.

Data

Context

LAUSD is the second largest district in the country with approximately 570,000 K–12 students and 24,000 K–12 teachers in 2018. LAUSD administrative data provide detailed information on student and teacher characteristics and allow for student-teacher matches. The district is also economically and racially diverse, providing variation in teacher quality, school level wealth and achievement, disability status, and student characteristics.

The LAUSD classroom assignment process allows researchers a unique opportunity to study whether SWD TQGs exist when assignments are mostly teacher-driven, focused on balancing classroom sizes, and based on credentials. At the elementary level, teachers may submit requests for track and grade level positions. Teachers may be assigned to their preferred classes based on district seniority, though principals may dispute a specific assignment if they believe it is not in the best interest of the school. At the secondary level, teachers submit requests for department selection, and principals consult with department heads to assign teachers to classes. See Online Appendix A for more details about teacher assignments (online appendices and online appendix tables are available on the journal website).

Elementary classroom rosters are created at the end of the school year by grade-level teams. After the start of school, the teams can call meetings to ensure that students are equitably distributed across classrooms. If they find the distribution unequal, they can recommend changes to the principal. Student-teacher pairings for elementary school classes and core classes in middle school are "fairly randomized," and placement adjustments are mostly around balancing class sizes (LAUSD, personal communication, December 12, 2019).

In LAUSD, approximately 60% of students with mild or moderate disabilities spend most of their days in general education classrooms (Swaak, 2020). The decision of whether a student should be included in general education is made on a case-by-case basis by a team of school personnel and outside professionals, in collaboration with parents and driven by the child's individualized needs. Then, among SWDs who are educated in the general education classroom, placement procedures into specific teachers' classrooms are decided by school personnel.

Sample

This study uses student- and teacher-level matched administrative data from School Year (SY) 2014–2015 through SY2017–2018 provided by LAUSD's Office of Data and Accountability and the Division of Human Resources. Our sample includes all kindergarten through 8th-grade students attending mainstream public schools.[1] The data are at the student-year level and include demographic information such as disability status (detailed below), race/ethnicity, gender, free or reduced-price lunch (FRL), and English Language Learner (ELL) status, as well as state standardized math and English Language Arts (ELA) test scores for students in Grades 3 through 8. We normalize each subject's test scores to have a mean of zero and standard deviation of one for each grade-year combination. We include both math and ELA scores because of evidence that SWDs may have different challenges in each subject (e.g., Child et al., 2019; Fuchs et al., 2016). The data also contain teachers' demographic information, educational background, and contract status (i.e., pretenure and permanent). Additionally, the teacher files include teachers' final evaluation scores as well as observation subcomponent scores and, for teachers hired since 2013 to 2014, their hiring scores from the district's teacher screening system.

Students are linked to teachers through a transcript file, which provides details on students' classroom placements and teacher for each class period.[2] The final dataset is restricted to students who are linked to at least one teacher. Our study focuses on SWDs who are taught by general education teachers since the majority of SWDs are in general education classrooms for most of their school days. Our overall sample consists of 1,175,666 student-year observations, or 13,107 unique teachers in 619 schools.

Variables of Interest

Disability status. We created indicator variables for four disability subcategories—autism, specific learning disability, speech/language impairment, and other. The categories reflect the disability groups that have high incidence rates in LAUSD and represent students with a range of needs. These categories are not mutually exclusive since students may have multiple disabilities.[3]

School characteristics. Since previous literature has shown that teacher quality can vary across schools with different characteristics, we generate school-level characteristics at the year level and then average across the 4 years in our panel (SY2014–2015 through SY2017–2018). Our main analysis focuses on FRL status.[4] We split schools into three categories based on their 4-year FRL average: less than 70% FRL, 70% to <95%, and 95% to 100%. We chose these FRL categories based on a combination of how previous literature has examined the FRL distribution and the distribution of FRL students within LAUSD, which skews towards high rates of poverty.[5] Grouping schools in this manner allows us to compare students with and without disabilities at schools with similar demographics, while also observing how these differences may vary across schools with different student characteristics.

Teacher characteristics. The literature suggests that teacher input variables, such as teachers' educational histories and credentials, are poorly correlated with teacher effectiveness (e.g., Angrist & Guryan, 2008; Chingos & Peterson, 2011; Kane et al., 2008). Much of the more recent literature has advocated for using teacher output measures as indicators of teacher quality, such as VAMs and teacher evaluation scores (e.g., Aaronson et al., 2007; Rivkin et al., 2005). Additional research has shown that exposure to early career teachers has negative impacts on student performance (Clotfelter et al., 2007; Rice, 2010; Ladd & Sorensen, 2017; Staiger & Rockoff, 2010). Consequently, our main analysis focuses on four aspects of teacher quality: VAMs of teachers' contributions to student achievement gains, teachers' ratings on their observation-based performance evaluations, teachers' initial hiring scores, and new teacher status (in first 2 years).

VAMs. We calculate VAMs for teachers teaching fourth through eighth grade. Following Chetty et al. (2014), we use a multistep calculation to create our value-added estimator.[6] First, to create residualized test scores, we regress scores on student, classroom, and grade demographics, including student-, classroom- and grade-level averages for race, gender, FRL status, ELL status, disability status, testing accommodations,[7] and previous test scores (cubed) in both ELA and math. Next, we average residualized test scores across all students for each teacher j in year t. We then calculate forecasting coefficients, which minimizes the mean squared error of the test-score predictions. Finally, we use data for teacher j in years outside of t to predict the value-added for teacher j in year t. See Chetty et al. (2014) and Online Appendix B for a more detailed description.

Teacher evaluation scores. Beginning in SY2014–2015, LAUSD teachers have been evaluated via classroom observations. All teachers new to a school are evaluated during their first 2 years. After the first 2 years, teachers are evaluated at least every other year, but some veteran teachers may extend the time between evaluations to up to 5 years.[8] In evaluation years, teachers are observed one or two times and receive scores on between 7 and 15 subcomponents from the Teaching and Learning Framework (depending on the year), as well as an overall evaluation score. Three subcomponents are required for all teachers across all years, whereas others are selected by teachers before the observation period. Since observation components varied by academic year and across teachers, we take the average score across all subcomponents and standardize by year.[9] In our sample, about 25% to 30% of our teachers are evaluated every year. Teachers who do not pass their evaluation are reevaluated in the following year. We use teacher's evaluation scores from the prior year (or, for those who were not evaluated in the prior year, the most recent evaluation score) to create our teacher evaluation score measure.

Hiring scores. In SY2014–2015, LAUSD introduced a new teacher screening system. For teachers hired/rehired SY2014–2015, we have composite hiring scores based on application information (e.g., licensure, grade point averages), professional references, writing samples, interviews, and sample lesson demonstrations. Details can be found in Online Appendix C and Bruno and Strunk (2019). All hiring scores are standardized by year.

Experience. The current literature suggests that, on average, early career teachers rapidly improve their effectiveness over their first few years (e.g., Kane et al., 2008; Papay & Kraft, 2015; Rivkin et al., 2005), but that new teachers are generally lower quality than more experienced teachers. Consequently, we examine students' exposure to novice teachers, which we define as having 2 or fewer years of experience.[10]

Teacher characteristics. Table 1 provides average general education teacher characteristics across FRL school bins and teacher quality measures.[11] Panels A through D highlight that each teacher quality measure is coming from a different teacher subsample, with "Novice" teachers as the most inclusive sample. Specifically, only teachers who taught Grades 4 through 8 will have VAM scores, only teachers who have been evaluated will have evaluation scores, and only teachers who have been hired (or rehired) since SY2014–2015 will have hiring scores. To give a sense of how the sample varies across teacher quality measures, we also include descriptive statistics about the share of teachers with valid measures and the average score for these measures. For example, Panel A documents that about 32% of math teachers in our VAM sample have a valid teacher evaluation score and that the average z-scored evaluation score for this sample is 0.16.

There are many patterns that are consistent across all subsamples. Most notably, students at lower-FRL (i.e., higher-income) schools tend to have more White and female teachers than students at higher-FRL schools. Additionally, on average, students in lower-FRL schools tend to be exposed to teachers with higher teacher evaluation scores and hiring scores than students attending higher-FRL schools.

Methods

We examine average teacher characteristics and whether these differ by student disability status. Our analytic approach is similar to that used in previous TQG literature (e.g., Clotfelter et al., 2005; Goldhaber et al., 2015). We estimate bivariate regressions of the following form:

$$Y_{ijs} = \beta_0 + \beta_1 Disability_{ijs} + \varepsilon_{ijs}, \quad (1)$$

where Y_{ijs} represents the teacher quality measure of interest (i.e., ≤2 years of experience) for student *i* matched with teacher *j* at school *s*. We run the model separately by each disadvantage bin (FRL < 70%, 70%–<95%, and 95+%). For our main results, $Disability_{ijs}$ is our SWD indicator variable. For our subgroup analysis, *Disability* represents one of three disability subgroups (Specific Learning Disability, Autism, or Speech/Language Impairment),[12] and zeros are given for non-SWDs. Standard errors are clustered at the school level. The model allows us to calculate exposure rates to students with disabilities for teachers across different quality measures, as well as the exposure rates for their nondisabled peers, and to test if this difference (captured in β_1) is statistically significant. We also run a school fixed effects model to test the stability of our coefficients when focusing on within-school variation. Since results across the models are similar, we conclude that much of the TQGs are driven by within-school sorting, and only report coefficients from the unadjusted model (Equation 1). We include variance decomposition information in Table 3 and Online Appendix Table 2 to examine how much variation in TQGs that occurs within and across schools.

In addition to TQGs within FRL bins, we are interested in whether TQGs differ across bins (referred to as "disadvantage gaps" from this point forward). Specifically, we evaluate whether any of the TQGs are significantly different from the TQG in schools with less than 70% FRL students. To do this, we pool observations across two bins (with the most advantaged school as the reference bin) and estimate the following equation:

$$\begin{aligned} Y_{ijs} = \beta_0 + \beta_1 Disability_{ijs} + \beta_2 Adv\ Sch_{ijs} \\ + \beta_3 (Disability_{ijs} * Adv\ Sch_{ijs}) + \varepsilon_{ijs}, \end{aligned} \quad (2)$$

where, again, Y_{ijs} represents the teacher quality measure of interest (i.e., ≤2 years of experience) for student *i* in teacher *j* at school *s* and $Disability_{ijs}$ is an indicator variable for students with disabilities. $Adv\ Sch_{ijs}$ is an indicator variable for the most advantaged school bin (FRL <70%). The $Disability_{ijs}*Adv\ Sch_{ijs}$ interaction measures the TQG differences between the two school disadvantage bins and tests whether this difference is statistically significant. In the interest of space, we only display the *p*-value associated with $Disability_{ijs}*Adv\ Sch_{ijs}$ in our tables.

Results

Research Question 1: Does Teacher Quality in LAUSD Vary Across Schools With Differing Degrees of Disadvantage?

Table 2 presents the mean and standard deviation for each general education teacher characteristic, split by subject, within our three FRL bins. Consistent with previous studies (e.g., Clotfelter et al., 2007; Goldhaber et al., 2018), we find some evidence of increasing exposure to lower-quality teachers as we move down the column from most to least advantaged schools. For example, the average math teacher evaluation score for students attending the most advantaged schools in our sample (<70% FRL) was 0.361 (measured in standard deviation units), whereas the average score at the least advantaged school (≥95% FRL) was 0.161. For students in the middle FRL bin (70%–<95% FRL), we find significant disadvantage gaps for both math and ELA VAMs. These students tend to be exposed to lower-VAM teachers than students attending the lowest-FRL schools, although we find no significant differences between students in the lowest and highest FRL bins. Additionally, students attending the middle and highest FRL bins tends to be exposed to teachers with lower teacher evaluation scores than their peers attending the lowest-FRL schools. Aside from ELA teachers in the highest FRL bin, which have significantly lower hiring scores than their peers in the lowest FRL bin, hiring scores seem to be evenly distributed across subject and FRL bins. In contrast to studies in other contexts (e.g., Boyd et al., 2008), novice teachers are relatively equitably distributed across LAUSD schools regardless of school-level disadvantage.

Research Question 2: Are There SWD vs. Non-SWD TQGs? Do These Gaps Vary by School-Level Disadvantage?

Table 3 presents overall SWD versus non-SWD TQGs. We begin by examining the average teacher quality for SWD in general education classrooms and then the average teacher quality for their non-SWD peers across each teacher quality measure.

Table 1
Teacher Characteristics in Math Classes by Teacher Sample and Free or Reduced-Price Lunch (FRL) School Bins

	Panel A. Value-Added Measure (VAM) Sample				Panel B. Teacher Eval Sample				Panel C. Hiring Score Sample				Panel D. Novice Sample			
FRL Group	≤70%	70%–95%	≥95%	Overall	≤70%	70%–95%	≥95%	Overall	≤70%	70%–95%	≥95%	Overall	≤70%	70%–95%	≥95%	Overall
No. of schools	121	133	341	595	121	128	351	600	106	98	284	488	123	134	362	619
General education teacher characteristics																
No. of teachers	836	993	3,121	4,950	983	1,214	4,094	6,291	257	217	785	1,259	2,423	2,477	8,207	13,107
No. of teacher-years	2,764	3,398	10,253	16,415	1,783	2,338	7,749	11,870	593	512	1,783	2,888	7,519	8,087	26,226	41,832
Mean experience	9.41	9.27	9.31	9.32	8.82	8.92	9.11	9.02	4.65	4.81	5.14	4.98	9.36	9.36	9.37	9.36
%Novice	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.34	0.31	0.29	0.31	0.03	0.03	0.03	0.03
%Masters or higher	0.34	0.40	0.39	0.38	0.34	0.37	0.36	0.36	0.39	0.39	0.36	0.37	0.34	0.36	0.36	0.36
%Female	0.72	0.68	0.66	0.68	0.78	0.76	0.73	0.74	0.90	0.80	0.85	0.85	0.80	0.75	0.75	0.76
%White	0.53	0.36	0.25	0.33	0.52	0.31	0.22	0.29	0.62	0.28	0.25	0.33	0.53	0.35	0.23	0.31
%Black	0.06	0.10	0.10	0.09	0.06	0.09	0.09	0.08	0.02	0.05	0.09	0.07	0.06	0.10	0.09	0.08
%Hispanic	0.19	0.32	0.53	0.42	0.22	0.40	0.58	0.48	0.19	0.41	0.51	0.42	0.20	0.36	0.56	0.45
%Have VAM	1.00	1.00	1.00	1.00	0.61	0.61	0.56	0.58	0.46	0.63	0.57	0.56	0.53	0.58	0.51	0.52
Mean VAM	−0.02	−0.14	−0.05	−0.07	−0.11	−0.19	−0.13	−0.14	−0.04	−0.20	−0.08	−0.10	−0.02	−0.14	−0.05	−0.07
%Have teacher eval	0.29	0.30	0.33	0.32	1.00	1.00	1.00	1.00	0.57	0.61	0.57	0.58	0.25	0.29	0.30	0.29
Mean teacher eval	0.35	0.14	0.11	0.16	0.36	0.19	0.16	0.20	0.28	0.13	0.16	0.18	0.36	0.19	0.16	0.20
%Have hiring score	0.06	0.07	0.08	0.08	0.17	0.15	0.13	0.14	1.00	1.00	1.00	1.00	0.07	0.07	0.07	0.07
Mean hiring score	0.10	0.08	−0.08	−0.01	0.18	0.06	−0.03	0.03	0.13	0.06	−0.04	0.01	0.13	0.06	−0.04	0.01

Note. Observations are at student-teacher cell level pooled across school years 2014–2015 to 2017–2018. Each column represents teacher characteristics in schools binned by the percent of FRL-eligible students. A school's FRL bin is defined by taking a 4-year average of the percent of FRL-eligible students. Experience represents years of experience and is top-coded at 10 years. Novice is defined by any teacher with fewer than 2 years of experience. VAM, Teacher Eval, and Hiring Score are z-scored measures for value-added, evaluation scores, and hiring scores, respectively. %Have indicates what percent of the given sample has a VAM, Teacher Eval, or Hiring Score measure. Panels A, B, C, and D represent the VAM, Teacher Eval, Hiring Score, and Novice samples, respectively. Panel D also represents the population sample.

Table 2
Average Teacher Quality, by Subject and Free or Reduced-Price Lunch (FRL) Bin

	Math				English Language Arts (ELA)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	VAM	Teach Eval	Hiring Score	Novice	VAM	Teach Eval	Hiring Score	Novice
A. <70% FRL teacher quality	–0.016	0.361	0.123	0.034	–0.12	0.323	0.232	0.032
SD	(0.529)	(0.729)	(1.084)	(0.180)	(0.908)	(0.773)	(0.875)	(0.176)
n	118,370	55,682	16,546	223,871	131,167	56,593	17,853	235,623
B. 70%–<95% FRL teacher quality	–0.135	0.19	0.061	0.03	–0.367	0.202	0.159	0.031
SD	(0.620)	(0.826)	(0.710)	(0.172)	(1.095)	(0.852)	(0.888)	(0.175)
n	141,686	69,980	16,777	245,351	148,669	72,322	19,787	250,140
Disadv. gap [*p*-value]	[.023]	[.01]	[.48]	[.85]	[.048]	[.045]	[.499]	[.822]
C. ≥ 95% FRL teacher quality	–0.055	0.161	–0.039	0.031	–0.23	0.172	–0.01	0.036
SD	(0.617)	(0.854)	(0.928)	(0.173)	(1.126)	(0.865)	(0.916)	(0.186)
n	357,985	213,663	49,676	708,089	371,134	216,808	60,273	715,865
Disadv. gap [*p*-value]	[.289]	[.00]	[.114]	[.754]	[.257]	[.001]	[.004]	[.14]

Note. Disadv. Gap represents difference in teacher quality of FRL bin from the least disadvantaged bin (<70% FRL). Observations are at student-teacher cell level pooled across school years 2014–2015 to 2017–2018. A school's FRL bin is defined by taking a 4-year average of the percent of FRL-eligible students. Novice is defined by any teacher with fewer than 2 years of experience. VAM (value-added measure), Teacher Eval, and Hiring Score are z-scored measures.

Table 3
SWD vs. Non-SWD Teacher Quality Gaps

	(1)	(2)	(3)	(4)
	VAM	Teach Eval	Hiring Score	Novice
Panel A. Math				
SWD teacher quality	–0.088	0.174	–0.002	0.033
Non-SWD teacher quality	–0.064	0.202	0.015	0.031
Teacher quality gap	–0.024***	–0.028*	–0.017	0.002
SE	(0.007)	(0.011)	(0.021)	(0.001)
SWD *n*	50,521	27,276	6,706	91,062
Non-SWD *n*	567,520	312,049	76,293	1,086,249
Between variance	*0.253*	*0.341*	*0.475*	*0.332*
Within variance	*0.747*	*0.659*	*0.525*	*0.668*
Panel B. English Language Arts (ELA)				
SWD teacher quality	–0.260	0.164	0.084	0.040
Non-SWD teacher quality	–0.237	0.207	0.067	0.034
Teacher quality gap	–0.023	–0.042***	0.017	0.006***
SE	(0.015)	(0.013)	(0.024)	(0.001)
SWD *n*	53,604	28,401	8,940	93,437
Non-SWD *n*	597,366	317,322	88,973	1,108,191
Between variance	*0.463*	*0.34*	*0.45*	*0.406*
Within variance	*0.537*	*0.66*	*0.55*	*0.594*

Note. Between/within variance calculated from Equation (1) with the addition of school fixed effects. These represent the variance of teacher quality gaps between and within schools. Standard errors clustered at school level. SWD stands for students with disabilities. Teacher quality gap represents the difference between the SWD and non-SWD exposure rates. Observations are at student-teacher cell level pooled across school years 2014–2015 to 2017–2018. Novice is defined by any teacher with fewer than 2 years of experience. VAM (value-added measure), Teacher Eval, and Hiring Score are z-scored measures.
*$p < .05$. ***$p < .001$.

The third and fourth rows present the quality gap and the corresponding standard error. Panel A presents results for math teachers, while panel B displays ELA results. We find that, relative to students without disabilities, SWDs are assigned to lower-quality math teachers in terms of VAMs and teacher evaluation scores. On average, SWDs in general education

Table 4
SWD With GET vs. Non-SWD Teacher Quality Gaps by Subject and Free or Reduced-Price Lunch (FRL) Bins

	Math				English Language Arts (ELA)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	VAM	Teach Eval	Hiring Score	Novice	VAM	Teach Eval	Hiring Score	Novice
A. <70% FRL								
TQG	–0.047***	–0.028	–0.019	0.000	0.001	–0.085**	0.032	0.001
SE	(0.014)	(0.026)	(0.050)	(0.003)	(0.024)	(0.028)	(0.042)	(0.003)
SWD *n*	9,258	4,397	1,266	16,865	9,839	4,318	1,467	17,366
Non-SWD *n*	109,112	51,285	15,280	207,006	121,328	52,275	16,386	218,257
B. 70%–<95% FRL								
TQG	–0.014	–0.041	–0.001	0.007	–0.012	–0.072*	0.084	0.008*
SE	(0.018)	(0.024)	(0.035)	(0.004)	(0.037)	(0.030)	(0.068)	(0.004)
SWD *n*	12,152	5,913	1,429	20,005	13,112	6,294	1,915	20,751
Non-SWD *n*	129,534	64,067	15,348	225,346	135,557	66,028	17,872	229,389
Disadv. gap [*p*-value]	[.143]	[.71]	[.757]	[.19]	[.76]	[.752]	[.521]	[.163]
C. ≥95% FRL								
TQG	–0.019*	–0.023	–0.021	0.001	–0.027	–0.019	–0.002	0.007***
SE	(0.008)	(0.014)	(0.029)	(0.001)	(0.020)	(0.015)	(0.028)	(0.002)
SWD *n*	29,111	16,966	4,011	54,192	30,653	17,789	5,558	55,320
Non-SWD *n*	328,874	196,697	45,665	653,897	340,481	199,019	54,715	660,545
Disadv. gap [*p*-value]	[.083]	[.879]	[.974]	[.865]	[.368]	[.037]	[.487]	[.111]

Note. TQG stands for teacher quality gap. GET stands for general education teacher. SWD stands for students with disabilities. Disadv. Gap represents how similar the FRL bins are from the least disadvantaged bin (<70% FRL). Observations are at student-teacher cell level pooled across school years 2014–2015 to 2017–2018. A school's FRL bin is defined by taking a 4-year average of the percent of FRL-eligible students. Exp represents years of experience and is top-coded at 10 years. Novice is defined by any teacher with fewer than 2 years of experience. VAM (value-added measure), Teacher Eval, and Hiring Score are z-scored measures.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

classrooms have math teachers with 0.024 standard deviation (*SD*) lower VAMs and 0.028 *SD* lower teacher evaluation scores than their non-SWD peers. There are no significant gaps in experience (novice status) or hiring scores. For ELA teachers, SWDs tend to be assigned to teachers with 0.042 *SD* lower evaluation scores and are more likely to be assigned a novice teacher (0.6 percentage point).

The last two rows of each panel present estimates from a model adding a school fixed effect, which allows us to examine how much of the TQG variance is due to within- or between-school factors. For math, approximately two-thirds of the gaps in the VAM, teacher evaluation, and novice measures are driven by within-school differences, suggesting that the gaps are mostly a function of the within-school distribution of teachers to SWDs and non-SWDs. For hiring scores, the across-school differences are larger, but 53% of the variance remains within. One possible explanation for this difference may be that higher turnover rates at certain schools drive the increase in across-school variation for this measure. Similar patterns are found for ELA, although the distribution for VAMs is more evenly distributed than in math (53.7% and 74.4% due to within-school variation, respectively).

Tables 2 and 3 show that overall, there are TQGs by school-level disadvantage and, for SWD in general education classrooms, significant differences by disability status in average teacher VAMs, teacher evaluation scores, and hiring scores. However, these findings are unable to shed light on how these factors interact. The rest of the article explores how teacher quality varies when we examine student disability status and school poverty levels simultaneously.

Research Question 3: Do SWD vs. Non-SWD Gaps Vary by School-Level Disadvantage?

Table 4 presents the mean teacher quality scores and quality gaps across disability status and school disadvantage, with each column representing a different teacher characteristic of interest. Columns (1) through (4) present the results for math teachers, while columns (5) through (8) present our findings for ELA teachers. Panel A presents the results for the most advantaged (<70% FRL) bin. We find significant SWD versus non-SWD TQGs for math VAM (–0.047 *SD*) and ELA teacher evaluation score (–0.085 *SD*).

Panels B and C present our findings for the middle (70%–<95% FRL) and most disadvantaged (≥95% FRL) schools, respectively. As in panel A, we see that SWDs have teachers with lower math VAMs (Panel B: –0.014, Panel C: –0.019), although this difference is only significant for the highest-poverty schools. We find no statistically significant math TQGs based on the other teacher characteristics. In ELA, we find SWDs in the middle FRL bin are exposed to teachers with lower evaluation scores (–0.072 *SD*) and are more likely to have teachers with 2 or fewer years of experience (0.8 percentage point difference) than their non-SWD peers. Aside from an increased likelihood of having a novice teacher (0.7 percentage points), SWDs in the highest-FRL bin are exposed to ELA teachers with similar qualities as their non-SWD peers. In subgroup analyses (Online Appendix Tables 2A–2D), we find that much of the TQGs are being driven by middle school grades (sixth through eighth grades) rather than earlier grades.

The last row of panels B and C displays the *p*-values for the disadvantage gap when we compare TQGs from each bin to the lowest-FRL bin. Our estimates suggest that, overall, TQGs are similar across FRL bins. The one exception is a significantly larger ELA teacher evaluation TQG in the lowest-FRL bin (–0.085) compared to the highest-FRL bin (–0.019).

Online Appendix Table 3 presents the between- and within-school variance decomposition within disadvantage bins. In general, the ratio of within- to between-school variance is similar across bins, although there are differing patterns for novice teachers. Within the lowest-poverty schools, TQGs for novice math teachers are mostly driven by within-school sorting (96%), whereas TQGs in the highest-poverty schools are more evenly split. For ELA, novice teacher disparities are mostly driven by within-school differences for the middle-FRL bin (73.6%), whereas differences in both the lowest- and highest-FRL bins are more equally split by across- and within-school sorting.

Research Question 4: Do TQGs Across School-Level Disadvantage Vary by Specific Disability Type?

Looking across all students with disabilities may mask heterogeneity by disability type. Consequently, we disaggregate our data to examine the three largest disability subgroups: specific learning disability (SLD, ~57% of SWDs), autism (~11% of SWDs), and speech/language impairment (SLI, ~23% of SWDs). Table 5 presents our math TQG subsample estimates.[13] Within each panel, we present the TQG (for each specific disability compared to non-SWD), standard errors, and sample size for each cell. Following the format in Table 4, we also include *p*-values for disadvantage gaps, which measure if TQGs in each FRL bin are significantly different from the TQG in the most advantaged schools (<70% FRL).

Results for students with SLD follow a similar pattern to the overall sample. Across the lowest- and highest-FRL bins, SLD students have teachers with significantly lower math VAMs (–0.074 *SD* and –0.029 *SD*, respectively) compared to their peers without disabilities. Our estimates also suggest that SLD students tend to have teachers with lower evaluation scores (significant for the middle- and highest-FRL bins, –0.087 *SD* and –0.053 *SD*, respectively). Nonetheless, we find no significant differences in terms of hiring scores and novice teachers. Interestingly, we find that the VAM TQGs in the lowest-poverty bin are significantly greater than the VAM TQGs in the more disadvantaged school groups. TQGs across other teacher quality measures do not vary across FRL bins, suggesting little correlation between school-level disadvantage and SWD TQGs.

Estimates for students with autism and SLI suggest few significant differences from their non-SWD peers. If anything, our estimates suggest that, depending on the FRL bin, these subgroups may be accessing higher-quality teachers than their non-SWD peers. For example, students with SLI in the middle- and highest-FRL bins have teachers with significantly higher VAMs (0.076 *SD* and 0.058 *SD*, respectively) than their non-SWD peers. Additionally, SLI in the highest-FRL schools have teachers with significantly higher evaluation scores (0.076 *SD*) and are less likely to have novice teachers. Overall, we generally do not find evidence that TQGs vary by school-level disadvantage. The one exception is that the SLI teacher evaluation TQG between the most disadvantaged schools (0.076 *SD*) is significantly larger than the SLI TQG in the lowest-FRL schools (–0.021 *SD*).

Online Appendix Table 4 displays the estimates for ELA teachers. Like the findings in Table 5, SLD students follow the same pattern for overall ELA teacher quality differences. We find no evidence of TQGs for students with autism. Like the math results, we find evidence that SLI students are exposed to higher-quality teachers than their non-SWD peers. Of note, SLI students are consistently exposed to higher ELA VAM teachers across all FRL bins (ranging from 0.207 *SD* to 0.308 *SD*).

Discussion and Policy Implications

In this study, we provide some of the first evidence documenting the extent of TQGs between students with and without disabilities, as well as differences in these gaps by school-level disadvantage. Previous research has found few TQGs between SWDs and their peers. However, we find evidence of significant TQGs in math VAMs, ELA teacher evaluation scores, and exposure to novice ELA teachers. We also extend the current literature by showing that TQGs in general education classrooms do not generally increase with school-level disadvantage. Additionally, we find evidence that TQGs are concentrated within students with specific learning disabilities, the largest subgroup of SWDs. Students with autism or speech/language impairment are not generally placed in classrooms with teachers who are different from teachers of the average non-SWD student. This finding is partially consistent with Gilmour and Henry (2018), who also found that students with speech impairments were assigned to similar teachers as their peers without disabilities. However, the TQGs for students with learning disabilities in our study stand in contrast to their North Carolina study. One potential explanation for this may lie in local differences in how mild disabilities are categorized (e.g., Saatcioglu & Skrtic, 2019), highlighting the importance of studying special education in different contexts such as urban areas like Los Angeles.

Although we find significant math VAM gaps, the lack of significant math TQGs across other teacher quality measures suggests that principals are not actively sorting students with disabilities into classrooms with perceivably worse teacher characteristics. Instead, we believe that principals may be sorting on unobservable (to us) characteristics that are highly correlated with math VAMs. For ELA teachers, our findings suggest that although there may be some sorting across observable teacher characteristics, these do not result in exposure to lower-quality ELA teachers as measured by VAMs.

Table 5
Math Teacher Quality Gaps by Disability Type (vs. No Disability) and Free or Reduced-Price Lunch (FRL) Bins

SWD With GET vs. Non-SWD

	(1)	(2)	(3)	(4)
Disability & FRL Group	**VAM**	**Teach Eval**	**Hiring Score**	**Novice**
A. Specific learning disability				
<70% FRL TQG	–0.074***	–0.049	–0.062	0.002
SE	(0.019)	(0.047)	(0.075)	(0.004)
SWD *n*	5,071	2,110	560	7,903
Non-SWD *n*	109,112	51,285	15,280	207,006
70%–<95% FRL TQG	–0.022	–0.087*	–0.02	0.01
SE	(0.019)	(0.038)	(0.044)	(0.006)
SWD *n*	8,078	3,429	859	11,935
Non-SWD *n*	129,534	64,067	15,348	225,346
Disadv. gap [*p*-value]	[.051]	[.519]	[.627]	[.252]
≥95% FRL TQG	–0.029**	–0.053**	–0.016	0.004
SE	(0.009)	(0.020)	(0.040)	(0.002)
SWD *n*	19,997	9,930	2,446	31,959
Non-SWD *n*	328,874	196,697	45,665	653,897
Disadv. gap [*p*-value]	[.032]	[.928]	[.592]	[.661]
B. Autism				
<70% FRL TQG	0	0.012	–0.006	–0.004
SE	(0.021)	(0.035)	(0.081)	(0.004)
SWD *n*	1,405	715	199	2,834
Non-SWD *n*	109,112	51,285	15,280	207,006
70–<95% FRL TQG	0.01	0.063	0.053	0.004
SE	(0.032)	(0.041)	(0.047)	(0.005)
SWD *n*	1,126	641	139	2,090
Non-SWD *n*	129,534	64,067	15,348	225,346
Disadv. gap [*p*-value]	[.784]	[.338]	[.529]	[.226]
≥95% FRL TQG	0.005	0.033	–0.034	–0.003
SE	(0.016)	(0.026)	(0.056)	(0.003)
SWD *n*	2,267	1,405	325	4,632
Non–SWD *n*	328,874	196,697	45,665	653,897
Disadv. gap [*p*-value]	[.851]	[.619]	[.779]	[.853]
C. Speech/language impairment				
<70% FRL TQG	0.022	–0.021	0.153	–0.004
SE	(0.031)	(0.039)	(0.089)	(0.004)
SWD *n*	888	821	260	3,098
Non-SWD *n*	109,112	51,285	15,280	207,006
70–<95% FRL TQG	0.076*	0.056	–0.03	–0.002
SE	(0.029)	(0.039)	(0.070)	(0.004)
SWD *n*	1,223	1,137	231	3,647
Non-SWD *n*	129,534	64,067	15,348	225,346
Disadv. gap [*p*-value]	[.207]	[.158]	[.107]	[.752]
≥95% FRL TQG	0.058***	0.076***	0.002	–0.006**
SE	(0.014)	(0.023)	(0.042)	(0.002)
SWD *n*	3,821	4,249	855	13,701
Non-SWD *n*	328,874	196,697	45,665	653,897
Disadv. Gap [*p*-value]	[.293]	[.03]	[.125]	[.706]

Note. TQG stands for teacher quality gap. GET stands for general education teacher. SWD stands for students with disabilities. Standard errors clustered at school level. Disadv. Gap represents how similar the FRL bins are from the least disadvantaged bin (<70% FRL TQG FRL). Observations are at student-teacher cell level pooled across school years 2014–2015 to 2017–2018. A school's FRL bin is defined by taking a 4-year average of the percent of FRL-eligible students. Novice is defined by any teacher with fewer than 2 years of experience. VAM (value-added measure), Teacher Eval, and Hiring Score are z-scored measures.
*$p < .05$. **$p < .01$. ***$p < .001$.

The new evidence we provide on SWD quality gaps contributes to a growing literature addressing the contexts and needs of both students with disabilities and the educators who teach them. Our finding of math VAM quality gaps across FRL bins suggests that schools, districts, and states should be cognizant of the ways in which they distribute teachers, particularly if schools are trying to adhere to the *Endrew F.* court decision to ensure equitable outcomes for SWDs. Existing literature shows that our case is not unique; schools tend to assign novice or less-effective teachers to classes with larger proportions of low-performing students (e.g., Bruno & Strunk, 2019; Kalogrides & Loeb, 2013; Lankford et al., 2002).

Our variance decomposition suggests that the majority of SWD TQGs in LAUSD are due to within- rather than between-school factors and is in line with previous research on SWD TQGs (Gilmour & Henry, 2018). For practitioners, the implication is that solutions to SWD quality gaps do not necessarily have to come from district and state policies aimed at recruiting and retaining higher quality teachers overall—though these avenues can help schools obtain more high-quality teachers. Instead, our estimates suggest that a more immediate solution could be to shift student compositions among existing teachers within schools.

Our findings suggest several avenues for future research. This strand of research would benefit from qualitative interviews and observations to explore what, if any, factors principals take into consideration while assigning teachers to classrooms, and how this varies by subject area. It is probable that teacher characteristics beyond those in the current study are used to determine how students are matched to teachers. For example, principals may pair certain SWDs with teachers who are particularly strong at engaging their students in classroom activities or who have strong classroom management skills. Furthermore, these traits may be important for improving SWDs' academic outcomes. Similarly, future work could explore whether some general education teachers are empirically better at improving outcomes for SWD compared to other teachers and examine their instructional practices. This empirical work could help practitioners move towards the end goal of more equitable academic outcomes for SWD.

ORCID ID

Ijun Lai https://orcid.org/0000-0003-2991-3367

NOTES

We are grateful to the Los Angeles Unified School District for providing the data necessary to conduct this research, and in particular to Vivian Ekchian, Sergio Franco, Cynthia Lim, and Emily Mohr for their partnership in developing the research agenda; and Jolene Chavira, Inocencia Cordova, Kathy Hayes, Crystal Jewett, Joshua Klarin, Jonathan Lesser, and Kevon Tucker-Seeley for their assistance in obtaining and understanding the data. We also wish to thank Paul Bruno, Tasmin Dhaliwal, and Anna Terkelsen for their research assistance. The authors thank four anonymous reviewers at *Educational Researcher* for their detailed comments, and conference participants at the Association for Public Policy Analysis and Management (APPAM), Associate for Education Finance and Policy (AEFP), and National Center for the Analysis of Longitudinal Data in Education Research conferences for helpful comments and discussions. This research was supported in part by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER), which is funded by a consortium of foundations. For more information about CALDER funders, see www.caldercenter.org/about-calder. All opinions expressed in this article are those of the authors and do not necessarily reflect the views of our funders or the institutions with which the authors are affiliated.

[1]For example, we do not include students who attend home or hospital schools, special education centers, or community day schools.

[2]Although it may also be of interest to examine teacher quality gaps (TQGs) for students with disabilities (SWDs) taught in special education classrooms by special education teachers, data limitations make this problematic. In particular, we can only calculate value-added measure (VAM) scores for approximately 15% of special education teachers (SETs) in our sample because students' previous test scores are used to construct VAMs, and few SWDs with SETs have valid test scores from the previous year. Additionally, whereas SETs are evaluated on the same observation instrument and hiring criteria as general education teachers, researchers and practitioners argue that these shared measures should not be used to measure SET quality since SETs' work responsibilities and preparation programs are different from those for general education teachers (see Brownell et al., 2005, for a review). Furthermore, recent research on teacher evaluations suggest that SETs may systematically receive lower evaluation scores since effective teaching looks different for SETs than general education teachers—particularly given the individualized nature of special education (Johnson & Semmelroth, 2013; Jones & Brownell, 2014). For completeness, we calculate non-VAM TQGs for students in special education classrooms (compared to their peers in general education classrooms), broken down by school disadvantage level and disability type, in Online Appendix Tables 8 and 9 (online appendices and online appendix tables are available on the journal website). Given these concerns, we are hesitant to say whether these results are indicative of the existence or nonexistence of quality gaps among SWDs with SETs compared to their non-SWD peers.

[3]Although typically Individuals With Disabilities Education Act (IDEA) only requires districts to designate a primary disability, along with blindness and deafness as secondary disabilities, the Los Angeles Unified School District (LAUSD) operates under a consent decree that requires more detailed tracking (Weintraub et al., 2008). For our main analysis, we are focused on whether students have any disabilities listed. Consequently, we do not separately account for students with multiple disabilities. In disability-specific analysis, we include any students who have that disability subcategory listed in their individualized education plan (including those with multiple disabilities). We have also run analysis that excludes students with multiple disabilities and find little difference. Results are available upon request.

[4]In analysis not shown, we also disaggregate schools by share of students who are underrepresented minorities or have low prior test scores. The results are qualitatively similar and available upon request.

[5]As a sensitivity check, we have also split schools into four bins (<70, 70–<95, 95–<0.978, ≥0.978), and five bins (<70, 70–<80, 80–<90, 90–<95, ≥95). Results are similar to those found in our main tables and available upon request.

[6]Only Grades 3 through 8 have test scores that are usable in standard VAMs. Consequently, we calculate VAMs only for students in Grades 4 through 8 (leaving out Grade 3 to ensure there is a lagged score). All VAM scores are standardized across the full sample of LAUSD teachers in each year (instead of just our study sample). Accordingly, the overall VAM mean (displayed in Table 1) is not zero. The teacher quality literature has used multiple different ways to measure teacher value-added. As a robustness check, we also estimate 1-year teacher VAMs that use teacher fixed effects and includes student- and classroom-level demographics (see Online Appendix B for more details about the construction of these models). To address concerns that student characteristics are endogenous to teacher value-added in time t, we

use teacher's value-added score in t – 1 as a measure for teacher quality in time *t*. Additionally, we create an alternative VAM score for teachers that exclude students with disabilities from VAM calculations. These results are presented in Online Appendix Table 5 and similar to the ones we show in our main tables.

[7]We include four types of testing accommodation flags: technology (i.e., text-to-speech software), setting (i.e., small group setting), time (i.e., extended time), and format (i.e., streamlined version of text).

[8]Teacher evaluation may be deferred for employees with 10 or more years of satisfactory service, have not received a "notice of unsatisfactory act of service" in the past 4 years, and had fewer than 13 unprotected absences in the past year.

[9]We also analyzed results by teachers' final evaluation score, which only has three values: below standard performance, meets standard performance, and exceeds standard performance. Since less than 5% of teachers each year do not pass the evaluation, there is not much variation in the final score. Consequently, we focus our main results on the average score across all subcomponents. The average score across all subcomponents does not necessarily map onto the final evaluation score (although it very rarely does not match) and has the additional benefit of having more variation to distinguish between teacher scores. Results for final evaluation scores available upon request. We have also analyzed a few alternative measures for teacher evaluation scores. Following Kraft et al. (2020), we create a measure of overall performance using a graded response model for all subcomponents, *theta*, as well as a residualized *theta* measure that removes classroom- and school-level student demographic variation. However, since teachers are not all assessed on the same components, we also create a *theta* based only on the three subcomponents that are mandatory for all teachers, as well as a residualized *theta* score based on these three subcomponents. Finally, we also individually analyze the raw scores for each mandatory subcomponent. Across all these differing teacher evaluation measures, we find little evidence of SWD vs. non-SWD teacher quality gaps. All results are displayed in Online Appendix Tables 6 and 7.

[10]As a sensitivity check, we also define "novice teacher" as those with 5 or fewer years of experience. Results are qualitatively similar.

[11]We also present results on student- and school-level demographics in Online Appendix Table 1. Similar to previous literature, we find that as school-level disadvantage increases, so does the share of students who are Black, Hispanic, or labeled as English Language Learner.

[12]We do not present results for the "other disabilities" subgroup since this group encompasses a large range of disabilities from emotional disturbance to intellectual disability and interpreting any potential gaps would be difficult. However, for completeness, we include this indicator variable in our VAM calculations.

[13]We present overall SWD subgroup analysis (not broken down by school-level disadvantage) in Online Appendix Table 10.

REFERENCES

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

Angrist, J. D., & Guryan, J. (2008). Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review*, *27*(5), 483–503.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). Explaining the short careers of high achieving teachers in schools with low-performing students. *American Economic Review*, *95*(2), 166–171.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2008). Consequences? The impact of assessment and accountability on teacher recruitment and retention: Are there unintended consequences? *Public Finance Review*, *36*, 88–111. Retrieved from https://doi.org/10.1177/1091142106293446

Brownell, M. T., Ross, D. D., Colón, E. P., & McCallum, C. L. (2005). Critical features of special education teacher preparation: A comparison with general teacher education. *Journal of Special Education*, *38*(4), 242–252.

Bruno, P., & Strunk, K. O. (2019). Making the cut: The effectiveness of teacher screening and hiring in the Los Angeles Unified School District. *Educational Evaluation and Policy Analysis*, *41*(4), 426–460.

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

Child, A. E., Cirino, P. T., Fletcher, J. M., Willcutt, E. G., & Fuchs, L. S. (2019). A cognitive dimensional approach to understanding shared and unique contributions to reading, math, and attention skills. *Journal of Learning Disabilities*, *52*(1), 15–30.

Chingos, M. M., & Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, *30*(3), 449–465.

Chudowsky, N., Chudowsky, V., & Keber, N. (2009). *State test score trends through 2007–08, Part 4: Has progress been made in raising achievement for students with disabilities?* Washington, DC: Center on Education Policy.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review*, *24*, 377–392.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, *41*(4), 778–820.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673–682.

Endrew F. v. Douglas County School District RE–1, 580 U.S. ___ (2017).

Fuchs, L. S., Geary, D. C., Fuchs, D., Compton, D. L., & Hamlett, C. L. (2016). Pathways to third-grade calculation versus word-reading competence: Are they more alike or different? *Child Development*, *87*(2), 558–567.

Gilmour, A., & Henry, G. (2018). A comparison of teacher quality in math for late elementary and middle school students with and without disabilities. *Elementary School Journal*, *118*(3), 426–451.

Gilmour, A. F., & Henry, G. T. (2020). Who are the classmates of students with disabilities in elementary mathematics classrooms? *Remedial and Special Education*, *41*(1), 18–27.

Gilmour, A. F., & Wehby, J. H. (2019). The association between teaching students with disabilities and teacher turnover. *Journal of Educational Psychology*. doi:10.1037/edu0000

Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, *44*(5), 293–307.

Goldhaber, D., Quince, V., & Theobald, R. (2018). Has it always been this way? Tracing the evolution of teacher quality gaps in US public schools. *American Educational Research Journal*, *55*(1), 171–201.

Johnson, E., & Semmelroth, C. L. (2014). Special education teacher evaluation. *Assessment for Effective Intervention*, *39*(2), 71–82.

Jones, N. D., & Brownell, M. (2014). Examining the use of classroom observations in the evaluation of special education teachers. *Assessment for Effective Intervention*, *39*, 112–124.

Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher*, *42*(6), 304–316.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615–631.

Kraft, M. A., Papay, J. P., & Chi, O. L. (2020). Teacher skill development: Evidence from performance ratings by principals. *Journal of Policy Analysis and Management*, *39*(2), 315–347.

Ladd, H. F., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. *Education Finance and Policy*, *12*(2), 241–279.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Education Evaluation and Policy Analysis*, *24*(1), 37–62.

Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, *130*, 105–119.

Rice, J. K. (2010). *The impact of teacher experience: Examining the evidence and policy implications* (Brief No. 11). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.

Saatcioglu, A., & Skrtic, T. M. (2019). Categorization by organizations: Manipulation of disability categories in a racially desegregated school district. *American Journal of Sociology*, *125*(1), 184–260.

Schulte, A. C., Elliott, J. J., Tindal, S. N., & Nese, G. (2016). Achievement gaps for students with disabilities: Stable, widening, or narrowing on a state-wide reading comprehension test? *Journal of Educational Psychology*, *16*, 925–942.

Schulte, A. C., & Stevens, J. J. (2015). Once, sometimes, or always in special education mathematics growth and achievement gaps. *Exceptional Children*, *81*, 370–387.

Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives*, *24*(3), 97–118.

Swaak, T. (2020, January 22). For the first time in more than 20 years, LAUSD is in full control of its special ed system. As parents worry about accountability, the district shifts its focus. *The New York Times*. Retrieved from http://laschoolreport.com/for-the-first-time-in-20-years-lausd-is-in-full-control-of-its-special-ed-system-as-parents-worry-about-accountability-the-district-shifts-its-focus/

U.S. Department of Education. (2019). *The NCES Fast Facts Tool provides quick answers to many education questions*. Washington, DC: National Center for Education Statistics.

Weintraub, F. J., Myers, R. M., Hehir, T., & Jaque-Anton, D. (2008). A contextual overview of the modified consent decree in the Los Angeles Unified School District. *Journal of Special Education Leadership*, *21*(2), 51–57.

AUTHORS

IJUN LAI, PhD, is a human services researcher at Mathematica, 505 14th Street, Suite 800, Oakland, CA 94612; ilai@mathematica-mpr.com. Her research focuses on policies and practices that reduce or reproduce inequities.

W. JESSE WOOD, MA, is a PhD candidate at Michigan State University, 110 Marshall-Adams Hall, 486 W. Circle Dr., East Lansing, MI 48824; woodwil2@msu.edu. His research focuses on teacher efficacy and how it relates to student equity.

SCOTT A. IMBERMAN, PhD, is a professor of economics and education policy at Michigan State University, 408 W. Circle Dr., East Lansing, MI 48824; imberman@msu.edu. His research focuses on issues related to the intersection of economics and education policy with interests in special education, school choice, returns to higher education, and autism.

NATHAN D. JONES, PhD, is an associate professor in special education at Boston University's Wheelock College of Education & Human Development, 2 Silber Way, Boston, MA 02215; ndjones@bu.edu. His research focuses on teacher quality, teacher development, and school improvement, with an emphasis on the special education teacher workforce.

KATHARINE O. STRUNK is a professor of education policy and economics, the Clifford E. Erickson Distinguished Chair in Education, and the director of the Education Policy Innovation Collaborative (EPIC) at Michigan State University, 620 Farm Ln, 238 Erickson, East Lansing, MI 48824; kstrunk@msu.edu. Her research focuses on understanding and evaluating policies related to educator labor markets and education governance.

Manuscript received September 3, 2019
Revisions received February 21, 2020;
June 1, 2020; July 16, 2020
Accepted July 29, 2020