THE DISTRIBUTION AND MOBILITY OF EFFECTIVE TEACHERS:
EVIDENCE FROM A LARGE, URBAN SCHOOL DISTRICT

Economics of Education Review, *48* (2015) 86-101

Steele, J. L., Pepper, M. J., Springer, M. G., & Lockwood, J. R.

Online Technical Appendix

Appendix Section 1: Estimating teachers' value-added effects on student achievement

As noted in section 2.2.1, we generate value-added estimates for each teacher in a given year, grade, and subject area using a multivariate ANCOVA method following McCaffrey, Han, and Lockwood (2008). In this method, separately for each year, grade, and tested subject, the students' achievement scores are regressed on indicators for individual teachers teaching those students in the given year and subject, as well as multiple prior achievement scores and individual student demographics to control for variation among students. We use up to three years of prior test scores for students when available, for a total of twelve prior test scores, and require that a student has at least three prior test scores to mitigate the potential for insufficient adjustment due to test measurement error. Missing prior test score data due to mobility and missed testing are handling via a pattern-mixture approach, in which students are grouped into patterns according to which prior tests were observed. The model includes indicators for the patterns, the interaction of these indicators with the coefficients on the test scores, and student demographic variables. Student demographic variables include indicators for gender, free/reduced meal eligibility status, special education status, and ELL status, as well as a categorical variable for student race. The individual teacher indicators are parameterized using sum-to-zero constraints (Mihaly, McCaffrey, Lockwood, & Sass, 2009) so that the teacher effects are defined relative to the average teacher in the district teaching that subject, grade, and year and so that estimated standard errors are appropriate for post-hoc shrinkage. For teachers who taught across multiple grades in the same subject and year (most common in mathematics in middle school grades), effects are aggregated to the level of teacher-year-subject by a precision-weighted average of the grade-specific estimates. To contribute to a teacher's value-added estimate, the student must be linked to the teacher in both the fall and spring terms. Students linked to multiple teachers for the same subject (fewer than 10 percent) contribute fully to the estimates for each teacher.

Appendix Section 2: Stability of value-added estimates for movers versus stayers

In a supplemental analysis represented in Appendix Table A1, we examine whether teachers' value-added improves in the year after the move. Other studies have found such improvement (e.g., Jackson, 2012; Xu, Ozek, & Corritore, 2012), though the evidence has suggested that this has been driven to a larger extent by regression to the mean than by sustainable improvement (Mansfield, 2010; Xu, Ozek, & Corritore, 2012). To address this question, we regress teachers' subsequent year value-added on their current-year value-added and include a turnover indicator, along with controls for experience, age, and year fixed effects. As shown in column 3, we also interact the prior-year value-added with the turnover indicator to examine whether the slope of the relationship between subsequent and current value-added differs for those who changed schools. We find no systematic relationship between turnover and subsequent-year scores in any of the specifications, which means that teachers who change schools do not systematically raise or lower their value-added any more than teachers who do not transfer. However, in column 3, we do find a negative and significant interaction between turnover and value-added, such that current-year value-added is about half as predictive of subsequent-year value-added in those who transfer than in those who do not (with net correlations of 0.35 for movers versus 0.66 for stayers). In other words, teachers' value-added estimates do not seem to systematically improve or decline after transferring, but general performance seems less stable in the year after transferring than it would otherwise be. In estimates not shown, we find that these results are highly consistent even if we redefine teachers' pre-move value-added as a multi-year average estimate rather than a single-year estimate.

Appendix Table A1. Coefficients (and standard errors) from a multilevel linear regression of current-year on prior-year value-added estimates, as a function of whether the individual changed schools between years

	(1)	(2)	(3)
Experience: 0-3 years	0.020~	0.022*	0.021*
	(0.011)	(0.010)	(0.010)
Experience: 11-20 years	-0.002	0.001	0.001
	(0.014)	(0.013)	(0.013)
Experience: 21-42 years	-0.018	-0.005	-0.004
	(0.027)	(0.024)	(0.024)
Experience (Linear term)	0.003*	0.001	0.001
	(0.001)	(0.001)	(0.001)
Age	-0.002***	-0.001*	-0.001*
	(0.000)	(0.000)	(0.000)
Year 2004-05	-0.005	-0.018*	-0.017*
	(0.009)	(0.009)	(0.009)
Year 2005-06	-0.010	-0.018*	-0.018*
	(0.010)	(0.009)	(0.009)
Year 2006-07	-0.012	-0.017*	-0.016~
	(0.009)	(0.009)	(0.009)
Moved between Schools	-0.008	0.009	0.005
	(0.013)	(0.012)	(0.012)
Value-added (shrunken)		0.631***	0.658***
		(0.020)	(0.021)
Value-added * Moved			-0.309***
			(0.067)
Constant	0.064**	0.037*	0.039*
	(0.022)	(0.018)	(0.018)
Ln of school-level error variance	-4.470	-7.261	-7.209
Intraclass correlation	0.216	0.0198	0.0209
Teacher-by-year observations	3,801	3,801	3,801
Schools	112	112	112
Degrees of freedom in model	9	10	11
Chi-squared	23.00	1011	1033

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1
Note: Omitted experience category is 4-10 years, and omitted school year is 2007-08. Models include school random effects.

Appendix Section 3: Sensitivity of turnover analyses to misclassification of leavers

One limitation of our dataset, which includes only teachers of cores subjects in grades 4 through 8, is that we cannot distinguish teachers leaving the district from those who moved to non-tested grades or subjects in the district. As noted in our discussion of equation 3, we estimate that no more than 10 percent of teachers who appear to leave the district in fact stayed in the district but changed out of the tested grades. (Changing out of the tested subjects is unlikely for a teacher of a core content area, since all core subjects are tested.) If those stayers misidentified as leavers were representative of true leavers, then their misidentification would not affect our conclusions, but if they differed substantially from true leavers, then our conclusions about who leaves their schools (as compared to who leaves the tested grades in their schools) may be biased.

As a sensitivity test, we re-estimate the models in Tables 5 and 6 under the assumption that 10 percent of those classified as leavers actually remained in their schools but changed grades, and that they came exclusively from the top decile of leavers in terms of value added, or (separately) from the bottom decile of leavers. To bound our estimates accordingly, we recast 10 percent of the leavers as stayers from the top and bottom parts of the teacher value-added distribution and re-estimate the probability of turnover (Appendix Table A2) and the probability of changing schools (but remaining in the tested grades) versus leaving the district (Appendix Table A3).

In Appendix Table A2, we find that under both scenarios, teachers in the highest-minority quartile schools would still be more likely than those in other quartiles to leave their schools. However, if the weakest decile of leavers were actually stayers, then negative relationship of value-added to probability of leaving would disappear, and the relationship between value-added and risk of leaving would become positive in the highest minority schools.

Appendix Table A2. Re-estimation of Table 5 coefficients (and standard errors), but with top or bottom decile of leavers recast as staying in their schools in a non-tested grade or subject

	Top 10% of Leavers Recast as Stayers			**Bottom 10% of Leavers Recast as Stayers**		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
2nd quartile: % Black/Hispanic	0.073	-0.027	-0.032	-0.026	-0.022	-0.052
	(0.154)	(0.153)	(0.153)	(0.147)	(0.147)	(0.147)
3rd quartile: % Black/Hispanic	0.304*	0.207	0.201	0.153	0.156	0.127
	(0.151)	(0.150)	(0.150)	(0.144)	(0.144)	(0.145)
Highest quartile: % Black/Hispanic	0.456**	0.297~	0.325~	0.321~	0.328*	0.324~
Value-added (shrunken)		-2.722***	-3.503***		0.109	-0.642
		(0.263)	(0.580)		(0.232)	(0.506)
Value-added * 2nd quartile			0.878			0.816
			(0.779)			(0.686)
Value-added * 3nd quartile			0.840			0.806
			(0.720)			(0.625)
Value-added * Highest quartile			1.497~			1.578*
			(0.858)			(0.782)

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1
Note: Models include controls shown in Table 5, but only relevant predictors are displayed here. Models include school random effects.

In Appendix Table A3, where the top panel recasts the top 10 percent of leavers as stayers, and the bottom panel recasts the bottom 10 percent as stayers, we find that our estimates for the probability of moving between schools in the district are largely unaffected. However, we now find that value-added is more-positively related to the risk of leaving the district in the highest-minority quartile than in the other three school quartiles, and this finding holds whether the top decile or bottom decile of leavers are recast as stayers (estimated β on interaction term=1.64 [p<.1] and 2.04 [p<.05]) respectively. These shifts in the interaction effects occur because the strongest decile of leavers are underrepresented in the highest minority schools, while the weakest decile of leavers are overrepresented in the highest minority schools.

Appendix Table A3. Re-estimation of Table 6 coefficients (and standard errors), but with top or bottom deciles of leavers recast as staying in their schools in a non-tested grade or subject

	Probability of Transferring, Contingent on Staying in District			Probability of Exiting District, Contingent on Not Transferring		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Top 10% of Leavers Recast as Stayers						
2nd quartile: % Black/Hispanic	0.502~	0.442~	0.421	0.007	-0.126	-0.132
	(0.262)	(0.262)	(0.262)	(0.170)	(0.171)	(0.172)
3rd quartile: % Black/Hispanic	1.014***	0.967***	0.944***	0.123	-0.006	-0.022
	(0.242)	(0.242)	(0.242)	(0.170)	(0.171)	(0.172)
Highest quartile: % Black/Hispanic	1.131***	1.025***	0.969***	0.284	0.086	0.165
	(0.270)	(0.271)	(0.276)	(0.192)	(0.193)	(0.195)
Value-added (shrunken)		-1.417***	-3.385**		-3.291***	-3.550***
		(0.423)	(1.242)		(0.304)	(0.626)
Value-added * 2nd quartile			2.673~			0.166
			(1.503)			(0.867)
Value-added * 3nd quartile			2.338~			-0.181
			(1.383)			(0.816)
Value-added * Highest quartile			1.398			1.637~
			(1.602)			(0.947)
Bottom 10% of Leavers Recast as Stayers						
2nd quartile: % Black/Hispanic	0.482~	0.452~	0.426	-0.110	-0.096	-0.124
	(0.261)	(0.261)	(0.261)	(0.161)	(0.161)	(0.162)
3rd quartile: % Black/Hispanic	0.989***	0.963***	0.936***	-0.060	-0.046	-0.077
	(0.242)	(0.242)	(0.242)	(0.160)	(0.161)	(0.162)
Highest quartile: % Black/Hispanic	1.091***	1.037***	0.982***	0.127	0.151	0.153
	(0.270)	(0.271)	(0.276)	(0.183)	(0.184)	(0.184)
Value-added (shrunken)		-0.689~	-2.837*		0.388	-0.296
		(0.408)	(1.212)		(0.261)	(0.537)
Value-added * 2nd quartile			2.721~			0.537
			(1.466)			(0.749)
Value-added * 3nd quartile			2.538~			0.666
			(1.337)			(0.689)
Value-added * Highest quartile			1.708			2.042*
			(1.567)			(0.856)

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1
Note: Models include controls shown in Table 7, but only relevant predictors are displayed here. Models include school random effects.

In Appendix Tables A2 and A3, our estimates are most sensitive to a scenario in which the weakest leavers are actually stayers who changed grades. Since it seems unlikely that strong value-added teachers would systematically move to untested grades in the same school, these analyses suggest that if the weakest teachers are actually staying in their schools but changing grades, then it may not be the case that turnover removes the weakest teachers. In the highest-minority schools, the reverse would appear to be true. Still, these sensitivity test scenarios are fairly extreme, and we include them only to document the sensitivity of our main estimates to our decision to classify teachers who exit the dataset as leavers.

Appendix Section 4: Longitudinal rather than cross-sectional experience effects

We can examine the effects of teachers' experience in a given school over time by omitting the year fixed effects from equation 5 and replacing teacher characteristics vector $\mathbf{c}_{\mathbf{ijt}}$ with $\mathbf{c}_{\mathbf{2ijt}}$, in which experience is not measured overall, but as the number of years the teacher is observed in the dataset in the current school. This specification, shown in equation A1, incudes a teacher-level random effect, u_i, which accounts for the nesting of observations within teachers:[1]

$$y_{ijt} = \alpha + \boldsymbol{\lambda}'\mathbf{Q}_{\mathbf{jt}} + \boldsymbol{\eta}'\mathbf{c}_{\mathbf{2ijt}} + \boldsymbol{\rho}'\mathbf{Q}_{\mathbf{jt}}*\mathbf{c}_{\mathbf{2ijt}} + u_i + \varepsilon_{ijt} \qquad \text{(A1)}$$

Estimates from this equation are presented in Appendix Table A4, where our predictors of interest are the teacher's year of observation in a given school in the dataset (ranging from 1 to 5 with a mean of 2.14) and its interaction with the school's minority enrollment quartile, holding constant other teacher characteristics and a teacher's initial experience category.

In this analysis, we find no overall effect of the teacher's year of observation in the school (column 1), but testing the joint significance of the interaction terms in column 2, we do find a statistically significant interaction between year of observation and the school's minority enrollment quartile ($\chi^2_{df=3}$ =15.29, p<0.01). Specifically, the estimated effect of an additional

[1] We present a teacher random effects rather than fixed effects model for equation 6 because we have limited variation with which to identify school quartile effects in a purely within-teacher comparison. Nevertheless, models estimated with teacher fixed effects yield very similar results to those presented in Appendix Table A4, though the interactions between year of observation and minority quartiles become jointly significant at only the five-percent level ($\chi^2_{df=3}$ =2.74, p=0.042).

Appendix Table A4. Coefficients (and standard errors) from regressions estimating effects of experience by minority enrollment quartile in model that estimates change over time

	(1)	(2)
Exper. category at first observation	0.009	0.009
	(0.006)	(0.006)
Years observed in school	-0.001	-0.012*
	(0.003)	(0.005)
Years obs * 2nd quartile		0.024***
		(0.007)
Years obs * 3rd quartile		0.012~
		(0.007)
Years obs * Highest quartile		-0.003
		(0.009)
Advanced degree	-0.002	-0.004
	(0.010)	(0.010)
Competitive college	0.035**	0.035**
	(0.013)	(0.013)
Undergrad GPA	0.033**	0.032**
	(0.012)	(0.012)
2nd quartile: % Black/Hispanic	-0.024~	-0.084***
	(0.012)	(0.022)
3rd quartile: % Black/Hispanic	-0.064***	-0.093***
	(0.013)	(0.021)
Highest quartile: % Black/Hispanic	-0.088***	-0.088***
	(0.017)	(0.026)
Constant	-0.079~	-0.049
	(0.045)	(0.046)
Ln variance in teacher-level error	-2.039***	-2.044***
	(0.039)	(0.039)
Ln variance in teacher-by-year error	-1.777***	-1.779***
	(0.015)	(0.015)
Teacher-by-year observations	3,027	3,027
Teachers	818	818
Degrees of freedom in model	8	11
Chi-squared	63.01	78.74

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1
Note: Models include teacher random effects.

year in the school descends from the second quartile (0.012), to the third quartile (0), to the lowest quartile (-0.012) and the highest quartile (-0.015). Though not all of the interaction coefficients are individually significant, their joint significance indicates a low probability that the effects of an additional year observed in the school are the same across quartiles. It is not clear why the effects of years observed would be zero to positive in the middle quartiles and negative in the extreme quartiles. Given that the mean number of years teachers are observed in the same school is just over 2 (ranging from 1.95 in the highest minority enrollment quartile to 2.3 in the lowest), we view this finding as suggestive rather than conclusive. Still, a pattern in which effects of additional years in the school are lowest in the highest-minority schools would yield a gap in teacher value-added even if new hires were equally effective among quartiles, so it may warrant consideration.