Estimating the Effects of Principal Quality and Experience
by
Stuart Buck
August 2012

I. Introduction

Recent public policy debates have focused heavily on measuring and improving teacher quality. Teachers are said to be the most important in-school variable that affects student achievement (Rivkin, Hanushek, and Kain 2005; Wright, Horn, and Sanders 1997). Yet teacher compensation and hiring systems are geared towards paper credentials such as certification or masters' degrees that have little (if any) effect on student achievement (Goldhaber and Brewer 1999). A better way, so goes the claim, would be to employ value-added measurement of teacher quality, so that we could properly assess which teachers are producing the greatest student achievement. In turn, we could get a better idea of which teachers to recruit and possibly reward with higher salaries or bonuses, which teachers should receive further professional development, and which teachers should be fired.

Value-added measurement of teachers suffers from some serious drawbacks, however. Most notably, an individual teacher is responsible for a relatively tiny number of children each year. With a small *n*, the variance in student outcomes can be large, and can be heavily influenced in either direction by random or non-random variation, which in turn means that an individual teacher's "value-added" can vary widely from year to year

(McCaffrey, Sass, and Lockwood 2008). Assignment of students to classrooms is rarely random, and if students who are more (or less) likely than average to make gains are assigned to the same teacher, that teacher's true effectiveness will be mismeasured. In all events, an individual teacher mostly affects the students in her own classroom,[1] and then only within the constraints provided by the rest of the system (such as the choice of curriculum).

Principals, by contrast, can affect student achievement much more widely than in a single classroom. Principals have a school-wide effect in evaluating teachers and in deciding who to hire and retain as teachers in the first place. Principals have the ability to fire bad teachers or to make their lives so miserable that they voluntarily quit or transfer. Harris and Sass (2009) and Jacob and Lefgren (2008) both find that principals' subjective evaluations can do a good job at predicting teacher value-added measurements, better than teacher experience and masters' degrees, and thus it matters how seriously principals take that task.

But principal selection and evaluation of teachers is not all. Principals can decide how teachers should be assigned to various classrooms and grades. Principals can encourage teachers to collaborate with each other across classrooms in order to ensure consistency of coverage (such collaboration may have a large effect size, see Saunders, Goldenberg, and Gallimore 2009), or they could stifle collaboration and discourage

[1] I say "mostly" because teachers *can* influence students outside their own classrooms to some extent. For example, high-quality teachers may, through the power of peer effects, improve the quality of other teachers in the same school, thereby indirectly affecting other students. Jackson and Bruegmann (2009).

teachers from leaving their own classrooms. Principals can affect what kind of professional development is available and how often. To be sure, principals' authority may depend on how decentralized the school system is, but more and more school systems are granting principals greater flexibility (Ouchi 2009).

As instructional leaders, principals can influence what pedagogical techniques teachers use and in some cases even the curriculum itself. Indeed, American teaching can be inconsistent and arguably of low quality: one large-scale observational study of classroom behavior by teachers found that "opportunities to learn for this sample of mostly middle-class students proved highly variable and did not appear congruent with the high performance standards expected for students or for teachers as described by most state teacher certification and licensure documents" (Pianta et al. 2007). In other words, there may be plenty of room for principals to push teachers in the direction of more successful teaching.

In short, even if teachers are the "most important" in-school factor, principals choose teachers in the first place, and then have numerous opportunities to either facilitate or discourage their success. Given that principals take all these actions on a *school*-wide basis, not just as to a single classroom, it may be more important for us to measure principal quality more objectively.

It is also important for us to design social policies that give principals the right incentive to take all these actions with student achievement in mind. For example, we could have a merit pay system with the objective of providing teachers an incentive to maximize student achievement, but some teachers may do better in certain classrooms or grades than others. If the principal assigns teachers to classrooms and grades not based on their aptitude but on who is popular within the school, one or more teachers could have much less value-added than if they had been assigned to the best classroom or grade for their own particular aptitudes. The same is true for all the other achievement-related decisions that principals make – if principals are basing decisions on politics, popularity, or any other irrelevant factor, they could be undermining *all* of the teachers' ability to affect student achievement in the right direction. Moreover, evaluating principals by value-added analysis should not be subject to all the same objections as are teacher value-added scores. For instance, principals affect a much larger number of students. A larger *n* will mean fewer random swings from year to year.[2]

Using three different models, I estimate principal value-added for Arkansas schools between 2004-05 and 2010-11. I find that there is no systematic relationship between years of experience at a given school and principal quality. Results as to the match between a student's race and the principal's race are inconsistent across different races and different model specifications.

[2] To say that evaluating principal performance empirically is a nice idea is not, of course, to say that it is easy or even possible. I discuss the methodological difficulties further in Section II.

II. Methodology

The scholarly literature has historically lacked a solid basis for evaluating principal performance. Much of the early literature on principal performance suffered from glaring deficiencies, such as a small sample size, a lack of controls for student demographics, and a reliance on raw test scores or even on measures that are more subjective and harder to define, such as student engagement or the teachers' own evaluations of principal performance. (See, for example, Ballou and Podgursky 1993; Eberts and Stone 1988; Waters, Marzano, and McNulty 2003; Clark, Martorell and Rockoff 2009, p. 2). Much of the "educational leadership" literature is even worse, full of cheery exhortations and somber advice that seems more like airport bookstore material than serious empirical inquiry into principal quality.

How, then, might we measure principal performance more objectively? The first complication is that principals do not teach students directly. Whatever impact they have on students is mediated through teachers. So do we just estimate teacher quality and impute that to the principal? No: teachers may have many skills and characteristics that are independent of the principal. On the other hand, we do not need to separate principal performance from teacher performance *entirely* – part of teacher performance may be due to the principal's influence over hiring, training, classroom assignment, and other working conditions. Thus, we need to separate out *only* that part of teacher performance that pre-exists a particular principal, while crediting the principal for whatever part of teacher performance he or she facilitated in some way.

A second complication is that there are unmeasured characteristics of schools that are partly or fully outside of the principal's control. Most worrisome, families often select into attendance zones based on their beliefs (true or erroneous) about the quality of schools. These selection effects could easily confound the true effectiveness of principals. Nor is relying on school-level value-added measures likely to be a good substitute for measuring principal effectiveness (Chiang, Lipscomb, and Gill 2012).

Yet a third complication is that principals themselves may sort over time into different schools, and high-quality principals may be preferentially able to migrate into schools that already have advantaged student bodies and/or high-quality teachers. Thus, we need a way to separate out principal selection effects as well.

Recent empirical work has attempted to solve the separability and selection effect problems by using both school and principal fixed effects (Branch, Hanushek, and Rivkin 2012; Li 2011; Dhuey and Smith 2011). Using both levels of fixed effects at once means that principal effects are being identified from switches between different principals at the same school. With school fixed effects in place, the model is hopefully correcting for all of the time-invariant qualities of the school, including family selection into the attendance zone and pre-existing teacher quality. Alternatively, Clark, Martorell and Rockoff (2009) use school-level fixed effects, but with additional controls for principal characteristics such as graduation from particular preparation programs, which means that an effect would be identified only where different principals with differing characteristics have taught at the same school.[3]

Relying on principal switchers obviously limits the number of schools and principals that can be analyzed at all, because some principals will remain at a particular school for the entirety of a given dataset. The loss of datapoints limits precision and increases standard errors, of course, but it can add bias as well: perhaps principals who stay at one school for long periods of time are systematically better or worse than principals who switch schools, either because of permanent characteristics of the principal or because of how well the principal's personality and management style match with the particular school. Nonetheless, most researchers are willing to accept this possible bias due to the even greater selection bias that seems possible if one leaves out school fixed effects.

One endemic problem with fixed effects estimations is that we may want to inspect the actual fixed effect parameters (rather than sweeping them out through the within transformation), but the estimates for each individual or group are often made relative to an arbitrary holdout from the sample. Although the relative positioning of each individual or group should remain the same, the actual parameters can be negative or positive depending on which person or group the software chooses to exclude; worse, the

[3] Clark, Martorell and Rockoff (2009) find no correlation between school performance and the principal's type of prior work experience or the prestige of the principal's undergraduate degree (p. 3). They do find a link, however, between the principal's years of experience and student test scores in math, and this gain to principal experience is "especially steep over the first few years" (p. 3).

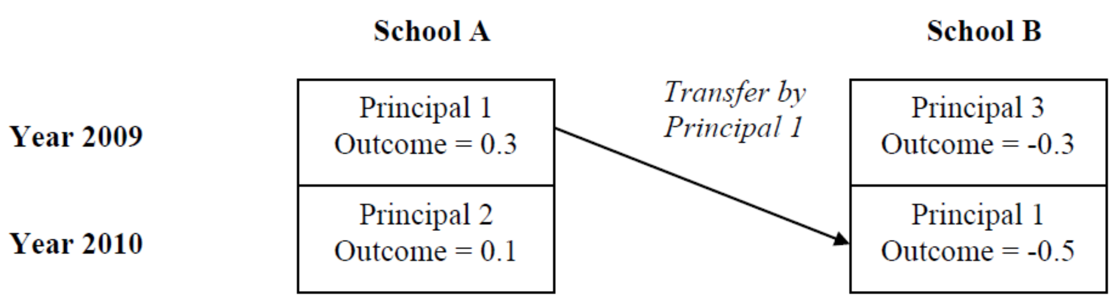
standard errors of each individual's estimate can vary depending on what holdout is chosen, and can vary even more if you try to recenter the fixed effects coefficients (or combine them with the constant) in order to create a more interpretable result. An even further problem in the principal/school context is that *every* comparison is being made between two or occasionally three principals who served at the same school.

Labor economists Abowd, Creecy, and Kramarz (2002) initially developed a statistical algorithm for estimating fixed effects parameters in high-dimensional matrices. In their algorithm, estimates are made within "groups" of workers and firms that are connected in a network of switchers. As they put it, "When a group of persons and firms is connected, the group contains all the workers who ever worked for any of the firms in the group and all the firms at which any of the workers were ever employed. In contrast, when a group of persons and firms is not connected to a second group, no firm in the first group has ever employed a person in the second group, nor has any person in the first group ever been employed by a firm in the second group" (p. 3). Andrews, Shank, and Upward (2005) extend the analysis by showing Stata code to implement what they call a "FEiLSDVj" estimator that involves fixed effects that are swept out for each worker *i* along with least squares dummy variables for each firm *j*. Yet another labor economist then wrote a Stata routine that implements such an algorithm in a way that saves considerable memory (Cornelissen 2008).

Finally, Mihaly et al. (2010) published a further Stata routine: felsdvregdm. The added "dm" is for "deviations from means," which means that the estimates for the least squares dummy variables are calculated in terms of deviations from each group's mean, which is centered at zero. Just as with the original labor economics paper, Mihaly et al.'s routine classifies observations into connected "groups": if Principals A and B have both worked at School 1, while Principals A and C and D have worked at School 2 and Principals B and E have worked at School 3, then Principals A through E and Schools 1 through 3 will all be in the same group.

The following figure from Chiang, Lipscomb, and Gill (2012) explains in simplified form how these groups work:

Figure 1.
Hypothetical Connected Network



Note: "Outcome" denotes the mean outcome of students at the given school in the given year after adjusting for all covariates in the principal VAM.

In that figure, we can deduce that Principal 3 is better than Principal 2, even though neither worked at the same school; we can do this because of a sort of transitivity principle, in that Principal 3 is better than Principal 1, who in turn is better than Principal 2.

Within each comparison group, the principal effects found by the felsdvregdm routine will be interpretable as deviations from the mean effect of zero. The standard errors do not vary arbitrarily, and can be more reliably used for empirical Bayesian shrinkage (although this is a more natural fit for random effects estimation).

III. Data

My dataset includes all Arkansas 3rd through 8th grade standardized test scores in reading and math from 2004-05 to 2010-11—some 1,457,375 student-year observations. Initially, the dataset identified these Arkansas students by student ID, school number, grade, year of observation, economically disadvantaged status, an indicator for ELL status, an indicator for an IEP being present, gender, ethnicity, and reading and math scale scores and z-scores. I then merged in a separate dataset provided by the Arkansas Department of Education that included information about the schools and principals for each relevant school year,[4] including school name, school type (Arkansas codes these as elementary, middle/jr, and high school), principal's name, principal's race and gender, principal's undergraduate and graduate institutions, principal's hire date within the particular school at issue, principal's licensure type, and principal's salary (for a few years). I then separately merged in two further datasets that provided school racial demographics and school status for federal Title I funding purposes (no assistance, targeted assistance, or school-wide assistance).

Ideally, I would be able to control for years of experience as principal, but no one possesses that data on a statewide basis in Arkansas. As the next best thing, the Arkansas Department of Education was able to provide data on two experience-related variables. First, I know when each school's current principal began working in that school in any

[4] The original test score dataset included scores from 2002-03 and 2003-04, but the Arkansas Department of Education did not identify principals for those years.

capacity, including as teacher. This variable can give us information about experience and stability: perhaps working one's way up within a given school, or having teaching experience in a given school before becoming principal, can all contribute to principal effectiveness. The second variable shows how many years the principal has been in the Arkansas education workforce, including service at other schools. Again, while this variable does not separate out experience as principal from experience as a teacher, it can give us information about whether the level of experience as *either* teacher or principal is useful (some make the argument that experience as a teacher should improve a principal's effectiveness).

There are 1,913 principals observed in the dataset, compared to 1,275 schools. In Table 1, I show descriptive statistics on how many principals switch schools during the observation periods.

Table 1
Principal Switchers

Principal Duplicates → Principal-School Duplicates ↓	0	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
0	363	276	79	8	0										726
1		1274	362	67	17	2	1								1723
2			1109	316	209	57	22	5	3	2					1723
3				864	111	141	74	22	4	3	4	1			1224
4					670	75	67	50	17	5	1	1			886
5						519	50	47	8	9	5	0			638
6							397	32	9	5	4	3	1		451
7								294	4	3	5	1		1	308
Total	363	1550	1550	1255	1007	794	611	450	45	27	19	6	1	1	7679

This table was generated by first searching for all duplicates of each principal's name. Overall, there were between 0 duplicates, for principals who appear only once in the database for one school in one year, and 13 duplicates, for principals who appear 14 times.[5] I then searched for duplicates based on both principal name *and* school id number. To the extent that a given principal's name appeared more often as a duplicate than did his or her name *in combination* with a particular school, that principal must have served in more than one school. Note that there are 308 cases where the school-principal combination is the same for all seven years of the dataset.

[5] Even though the dataset extends only from 2004-05 to 2010-11, a few rural Arkansas principals appear up to 14 times because they serve as principal for both a small elementary school and for a junior/high school at the same time.

IV. Analysis

In Model 1, I use the felsdvregdm routine with student and principal fixed effects. Formally, the model is:

$$Y_{iqt} = \alpha_1 + \beta_1 X_{it} + \beta_2 Z_{qt} + \beta_3 P_{jt} + \beta_4 Yr + \beta_5 SchoolType + \tau_i + \delta_j + \varepsilon_{ijqt}$$

where Y_{ijt} means student *i*'s math z-score in school q at time t, X includes time-varying student characteristics, Z includes school characteristics, P_q includes time-varying principal characteristics, Yr includes year indicators, SchoolType includes indicators for middle/junior and high schools, and τ_i and δ_j are student and principal fixed effect terms respectively. Given that I wanted to compare minority students with minority principals directly and only to minority students with white principals (and the same for white students and white principals), I ran two separate student/principal fixed effects analyses.

In the first, limited to minority students, the grouping portion of the felsdvregdm algorithm found a total of 8 groups across Arkansas, spanning 121,357 minority students and 1,348 principals. 199 minority students are left ungrouped, and therefore do not form part of the analysis. The vast majority of the students and principals (121,203 and 1,330, respectively) are in the same group. With so many students and principals being part of one statewide "group" or network, one might worry that this is largely due to students switching between different *types* of schools – going from elementary to middle school, for example. The inclusion of indicators for middle/junior high school and for high school hopefully captures the effect of being in those types of schools (compared to elementary schools), and the indicator for a student having switched schools in a given year will hopefully capture any dip experienced in the first year in a new school.[6]

The following table shows the results:

[6] I originally attempted to code for structural moves separately from non-structural moves, but this proved somewhat slippery given different grade configurations in Arkansas schools and given that some students appear in the database in a given year with no indication whether they are transferring from a school of the same type in another state or whether they graduated from a private school that stopped at the previous grade. (In some cases, that is, there is no way to tell whether a given student is making a structural move or a non-structural move.) In any event, we have reason to believe that both structural and non-structural moves cause a downward dip in student achievement, and it should not be necessary to identify such moves separately in analyzing principal effects.

Table 3 Minority Students and Principal Fixed Effects						
	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Indicator for Student Poverty	0.032985	0.004691	7.03	0	0.023791	0.042179
ELL	0.050602	0.009854	5.13	0	0.031288	0.069916
IEP	0.044483	0.007899	5.63	0	0.029	0.059965
Principal's Years at School	-0.00323	0.001137	-2.84	0.005	-0.00546	-0.001
Principal's Years in Workforce	-0.00158	0.000409	-3.86	0	-0.00238	-0.00078
Middle School	-0.14659	0.014402	-10.18	0	-0.17482	-0.11837
High School	-0.16901	0.025532	-6.62	0	-0.21905	-0.11896
Percent Black	-0.02531	0.043869	-0.58	0.564	-0.11129	0.060676
Percent Hispanic	0.281192	0.059711	4.71	0	0.164161	0.398224
Targeted Title I Aid	0.03693	0.012365	2.99	0.003	0.012694	0.061165
Schoolwide Title I Aid	0.028079	0.00935	3	0.003	0.009753	0.046406
Switched Schools	-0.01015	0.002453	-4.14	0	-0.01496	-0.00534
2006	0.16019	0.02334	6.86	0	0.114445	0.205936
2007	0.155767	0.024283	6.41	0	0.108172	0.203362
2008	0.261015	0.02457	10.62	0	0.212858	0.309171
2009	0.307752	0.02522	12.2	0	0.258322	0.357183
2010	0.338338	0.025802	13.11	0	0.287767	0.388909
2011	0.337275	0.026412	12.77	0	0.285509	0.389041
Race Match	-0.04453	0.027016	-1.65	0.099	-0.09748	0.008419
N=291789						
F-test that person and firm effects are zero: F(122695,169074)=8.703 Prob > F = 0						
F-test that person effects are equal to zero: F(121348,169074)=8.135 Prob > F = 0						
F-test that firm effects are equal to zero: F(1339,169074)=7.759 Prob > F = 0						

With only a few exceptions, all coefficients are highly significant at the .001 level or below. We must be careful in interpreting the student-level variables, however. With student fixed effects in place, the variables for IEP status, ELL status, and poverty status (which all appear to have a slight *positive* effect) really represent the fact that a given student *switched* into that status over the period of observation. In other words, given that a student is someone who needs an IEP or ELL status, getting that status has a small benefit.

The indicators for switching schools, for going to a middle/junior high school, and for going to a high school all have a negative effect. The percent black in a school has a tiny negative effect, while the percent Hispanic has a surprisingly positive effect, although this variable may be highly collinear with ELL status (most Arkansas Hispanics are from families that immigrated within the past 20 years).

As for principal characteristics, the point estimates on principal experience at the particular school and in the Arkansas teacher workforce are both negative and highly significant, albeit very small. As for race-matching between the student and the principal, Model 1 suggests that for minority students, having a minority principal had a small negative effect, although this was only marginally significant at the 0.10 level.

In Model 2, I perform the same analysis for white Arkansas students. The grouping portion of the analysis put 231,460 Arkansas students into the same group, along with 1,390 principals.

Table 4: White Students and Principal Fixed Effects						
	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Indicator for Student Poverty	0.001576	0.002985	0.53	0.598	-0.00427	0.007426
ELL	-0.09695	0.058041	-1.67	0.095	-0.21071	0.016804
IEP	0.020518	0.00541	3.79	0	0.009914	0.031122
Principal's Years at School	-0.00336	0.000745	-4.52	0	-0.00482	-0.0019
Principal's Years in Workforce	-0.00163	0.000502	-3.25	0.001	-0.00262	-0.00065
Middle School	-0.06046	0.010111	-5.98	0	-0.08028	-0.04064
High School	-0.11664	0.01292	-9.03	0	-0.14196	-0.09132
Percent Black	-0.04914	0.043199	-1.14	0.255	-0.13381	0.035524
Percent Hispanic	0.252775	0.057438	4.4	0	0.140198	0.365352
Targeted Title I Aid	-0.00696	0.007273	-0.96	0.339	-0.02122	0.007294
Schoolwide Title I Aid	-0.00531	0.007065	-0.75	0.452	-0.01915	0.008539
Switched Schools	-0.02851	0.001831	-15.57	0	-0.0321	-0.02492
2006	0.101752	0.02286	4.45	0	0.056946	0.146557
2007	0.10473	0.02379	4.4	0	0.058103	0.151358
2008	0.132604	0.024096	5.5	0	0.085378	0.17983
2009	0.145919	0.024614	5.93	0	0.097676	0.194162
2010	0.153618	0.025104	6.12	0	0.104415	0.202822
2011	0.172761	0.025567	6.76	0	0.122651	0.222871
Race Match	0.117582	0.020328	5.78	0	0.07774	0.157424
N=577908						
F-test that person and firm effects are zero: F(232848,345040)=8.698 Prob > F = 0						
F-test that person effects are equal to zero: F(231459,345040)=8.198 Prob > F = 0						
F-test that firm effects are equal to zero: F(1389,345040)=10.721 Prob > F = 0						

Just as in Model 1, being in a middle or high school, and switching schools generally, has a negative and significant effect. A school's level of federal aid seems to make no significant difference, and neither does adding poverty or ELL status (which is probably a rare event for white Arkansas students anyway).

As for principal characteristics, the point estimates on principal experience at the particular school and in the Arkansas teacher workforce are both negative and highly significant, albeit very small. As for race-matching between the student and the principal, Model 2 suggests that for white students, having a white principal had a positive and highly significant effect.

Model 3 is similar to Models 2 and 3, except that 1) it uses school fixed effects instead of student fixed effects; 2) it uses student lagged math scores; and 3) for obvious reasons, it omits the Schooltype indicator. Formally, the model is:

$$Y_{iqt} = \alpha_1 + \beta_1 Y_{ijt-1} + \beta_2 X_{it} + \beta_3 Z_{qt} + \beta_4 P_{jt} + \beta_5 Yr + \tau_i + \delta_j + \varepsilon_{ijqt}$$

where Y_{iqt} means student *i*'s math z-score in school q at time *t*, X includes student characteristics, Z includes time-varying school characteristics, Pj includes time-varying principal characteristics, Yr includes year indicators, and τ_i and δ_j are school and principal fixed effects respectively.

The grouping portion of the analysis created 349 groups with 397,313 student-year observations, 579 schools, and 1024 principals. 216 of the groups have only two principals, 63 groups have three principals, 31 groups have four principals, and 39 groups have five or more principals. We thus have the opposite worry from Model 1: instead of worrying that much of the state was placed into a single group, we might now worry that there are too many disconnected "groups" of only two or three principals being compared to each other. (In the case of two principals being compared only to each other, the point estimates for those specific principals will be precisely the same, except that one will be negative and one will be positive.) At worst, though, Model 3 is in the same situation as a model with the usual school and principal fixed effects, wherein *all* comparisons are being made between the two or three principals who actually presided over the same school. The felsdvregdm algorithm at least allows one to make comparisons within the several dozen "groups" that do involve principals moving across multiple schools.

The results can be seen in Table 5:

Table 5 School and Principal Fixed Effects						
	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Lagged Math Score	0.74	0.001	717.32	0	0.74	0.743
Indicator for Student Minority Status	-0.064	0.003	-23.2	0	-0.07	-0.06
Indicator for Student Poverty	-0.102	0.002	-50.38	0	-0.106	-0.098
ELL	0.006	0.005	1.19	0.236	-0.004	0.015
IEP	-0.27	0.003	-88.44	0	-0.28	-0.27
Principal's Years at School	0.017	0.003	5	0	0.01	0.024
Principal's Years in Workforce	-0.001	0.0006	-1.65	0.099	-0.002	0.0002
Percent Black	0.101	0.084	1.21	0.226	-0.063	0.27
Percent Hispanic	0.296	0.109	2.72	0.006	0.08	0.51
Targeted Title I Aid	0.04	0.01	3.99	0	0.02	0.06
Schoolwide Title I Aid	0.07	0.01	7.39	0	0.05	0.09
Switched Schools	-0.05197	0.002	-26.05	0	-0.05588	-0.04806
2006	0.085703	0.0125	6.85	0	0.061196	0.110209
2008	0.027167	0.007	3.92	0	0.013576	0.040758
2009	0.008492	0.0094	0.91	0.365	-0.00986	0.026847
2010	-0.01847	0.0122	-1.51	0.131	-0.04243	0.005494
2011	-0.02488	0.0154	-1.62	0.105	-0.05501	0.005242
Minority Student/Minority Principal	-0.091	0.0227	-3.99	0	-0.135	-0.046
White Student/Minority Principal	-0.05051	0.0225	-2.24	0.025	-0.09467	-0.00636
N=397313						
F-test that person and firm effects are zero: F(1253,396040)=12.56 Prob > F = 0						
F-test that person effects are equal to zero: F(230,396040)=6.769 Prob > F = 0						
F-test that firm effects are equal to zero: F(675,396040)=4.905 Prob > F = 0						

In Model 4, IEP status and poverty status for an individual student have significant negative effects, as does switching schools. A school becoming eligible for greater levels of federal aid has a small positive impact. Increases in the percent black in a given school has no significant impact, while changes in the percent Hispanic seem to have a significant positive impact around 0.3 standard deviations (again, this may be misleading due to collinearity with ELL status).

As for a principal's years of experience at a given school, there is a slight positive and significant impact for the number of years the principal has served at the school in any capacity, but a marginally significant (at the 0.10 level) and negative impact for the principal's number of years in the workforce. With race-matching, I coded for white-student-minority-principal and minority-student-minority-principal. Both categories of student-principal race matches have a highly significant and negative effect. These results are consistent with what I found in the student fixed effects models, where white students matched with white principals fared better than when matched with minority principals, while minority students also fared better with white principals than with minority principals.

V. Conclusion

The value of a principal's tenure at a given is uncertain in this dataset. In student fixed effects models, a principal's years at the school and years in the workforce both have a negative and significant impact, while in the school fixed effects model, the principal's years at the school have a positive and significant impact. For both white and minority students, having a white principal seems to make a slight positive difference, a finding that was contrary to my expectation given previous literature finding that minority students benefit slightly from having a minority teacher.

REFERENCES

Abowd, John M., Robert H. Creecy, and Francis Kramarz. 2002. "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data." Working paper, available at https://courses.cit.cornell.edu/jma7/abowd-creecy-kramarz-computation.pdf.

Andrews, Martyn, Thosten Schank, and Richard Upward. 2005. "Practical fixed effects estimation methods for the three-way error components model." Working paper, available at http://staff.bath.ac.uk/ecsjgs/Research/Papers%20to%20Read/Martyn%20Andrews/leed2.pdf.

Ballou, Dale, and Michael J. Podgursky. 1993. "What Makes a Good Principal? How Teachers Assess the Performance of Principals." *Economics of Education Review* 14(3): 243–52.

Ballou, Dale, William Sanders, and Paul Wright. 2004. "Controlling for Student Background in Value-Added Assessment of Teachers." *Journal of Educational and Behavioral Statistics* 29 no.1: 37-65.

Branch, Gregory, Eric Hanushek, and Steven Rivkin. 2009. "Estimating Principal Effectiveness." CALDER Working Paper 32.

Chiang, Hanley, Stephen Lipscomb, and Brian Gill. 2012. "Is School Value-Added Indicative of Principal Quality?" Paper for AEFP Conference, March 2012, available at http://www.aefpweb.org/sites/default/files/webform/School%20VA%20and%20Principal%20Quality.pdf.

Clark, Damon, Paco Martorell, and Jonah Rockoff. 2009. "School Principals and School Performance." CALDER Working Paper 38.

Cornelissen, Thomas. 2008. "The Stata command felsdvreg to fit a linear model with two high-dimensional fixed effects." *Stata Journal* 8 no. 2: 170-189.

Dhuey, Elizabeth & Justin Smith. 2011. "How Important Are School Principals in the Production of Student Achievement." Working paper, available at http://homes.chass.utoronto.ca/~edhuey/datastore/files/docs/dhuey_smith_princ_nov2011.pdf.

Eberts, Randall W., and Joe A. Stone. 1988. "Student Achievement in Public Schools: Do Principals Make a Difference?" *Economics of Education Review* 7(3): 291–99.

Goldhaber, Dan, and Dominic Brewer. 1999. Teacher Licensing and Student Achievement. In M. Kanstoroom & C. E. Finn, Jr (Ed.), *Better Teachers, Better Schools* (pp. 83-102). Washington, DC: The Thomas B. Fordham Foundation.

Grissom, Jason A., and Susanna Loeb. 2011. "Triangulating Principal Effectiveness: How Perspectives of Parents, Teachers, and Assistant Principals Identify the Central Importance of Managerial Skills." *American Educational Research Journal* 48(5): 1091-1123.

Guarino, Cassandra M., Mark D. Reckase, and Jeffrey M. Wooldridge. 2011. "Can Value-Added Measures of Teacher Performance Be Trusted?" Working paper.

Guimaraes, Paulo, and Pedro Portugal. 2009. "A Simple Feasible Alternative Procedure to Estimate Models with High-Dimensional Fixed Effects." IZA Working Paper, available at http://ftp.iza.org/dp3935.pdf.

Harris, Douglas N., and Tim R. Sass. 2006. "Value-Added Models and the Measurement of Teacher Quality." Working paper retrieved from http://myweb.fsu.edu/tsass/Papers/IES%20Harris%20Sass%20EPF%20Value-added%2014.pdf.

Harris, Douglas N., and Tim R. Sass. 2009. "What Makes for a Good Teacher and Who Can Tell?" CALDER Working Paper No. 30. Retrieved from http://www.urban.org/uploadedpdf/1001431-what-makes-for-a-good-teacher.pdf.

Jackson, C. Kirabo, and Elias Bruegmann. 2009. "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers." *American Economic Journal: Applied Economics* 1(4): 85–108.

Jacob, Brian A., and Lars Lefgren. 2008. "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education." *Journal of Labor Economics* 26:101–36.

Li, Danielle. 2011. "School Accountability and Principal Mobility: How No Child Left Behind Affects the Allocation of School Leaders." Working paper, available at http://econ-www.mit.edu/files/5437.

McCaffrey, Daniel F., Tim R. Sass, and J. R. Lockwood. 2008. "The Intertemporal Stability of Teacher Effect Estimates." Working paper.

McCaffrey, D., J. R. Lockwood, T. Louis, and L. Hamilton. 2004. "Models for Value-Added Models of Teacher Effects." *Journal of Educational and Behavioral Statistics* 29 (1): 67-101.

Mihaly, Kata, Daniel F. McCaffrey, J. R. Lockwood, and Tim R. Sass. 2010. "Centering and reference groups for estimates of fixed effects: Modifications to felsdvreg."

Stata Journal 10 no. 1: 82-103. Available at http://www.econ.uzh.ch/department/library/research/statajournal/sj10-1.pdf.

Ouchi, William G. 2009. *The Secret of TSL:The Revolutionary Discovery That Raises School Performance*. New York: Simon and Schuster.

Pianta, Robert C., Jay Belsky, Renate Houts, and Fred Morrison. 2007. "Opportunities to Learn in America's Elementary Classrooms." *Science* 315 (March 30): 1795-96.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, School, and Academic Achievement." *Econometrica* 73 no. 2: 417-58. Retrieved from http://www.econ.ucsb.edu/~jon/Econ230C/HanushekRivkin.pdf.

Rivkin, Steven G. 2007. "Value-Added Analysis and Education Policy." CALDER Brief 1. Washington, D.C.: The Urban Institute.

Saunders, William M., Claude N. Goldenberg, and Ronald Gallimore. 2009. "Increasing Achievement by Focusing Grade-Level Teams on Improving Classroom Learning: A Prospective, Quasi-Experimental Study of Title I Schools." *American Educational Research Journal* 46(4): 1006-33.

Waters, Tim, Robert J. Marzano, and Brian McNulty. 2003. "Balanced Leadership: What 30 Years of Research Tells Us About the Effect of Leadership on Student Achievement." McREL Working Paper. Denver, CO.: Mid-continent Research for Education and Learning.

Wright, S. Paul, Sandra P. Horn, and William L. Sanders. 1997. "Teacher and Classroom Context Effects on Student Achievement: Implications for Teacher Evaluation." *Journal of Personnel Evaluation in Education* 11:57-67.