Abstract Title Page

Title: Using Subjective Teacher Evaluations to Examine Principals' Personnel Management Priorities

Authors and Affiliations:

Benjamin Master (bmaster@stanford.edu)

Stanford University

Abstract Body

Background / Context:

Teacher evaluation is at the center of current education policy reform. Most evaluation systems rely at least in part on principals' assessments of teachers, and their discretionary judgments carry substantial weight. However, we know relatively little about what they value when determining evaluations and high stakes personnel decisions. I leverage unique data from a public charter school district to explore the extent to which school administrators' formative evaluations of teachers align with teacher and school effectiveness and predict future personnel decisions. While previous research has examined administrators' subjective evaluations of teachers in surveys and in practice, this study links a detailed evaluation in practice with multiple types of personnel decisions to provide new insights into administrator decision-making.

Evaluation systems may improve the quality of teaching via two key mechanisms. First, they may identify and promote effective teaching practices that help teachers to improve (Taylor and Tyler, 2011). Second, they may facilitate personnel practices and policies that support the retention of more effective teachers and the dismissal of less effective teachers, as well as more optimal assignment of teachers to jobs in which they can have the most positive effect (Boyd et. al, 2010; Goldhaber and Theobald, 2011; Rockoff et. al, 2011). To accomplish either of these aims, educators must leverage measures of teacher effectiveness without inadvertently neglecting important contributions that occur outside the scope of measurement. That is, evaluation systems will not be as effective if the evaluation measures used miss important components of teaching that could aid in teacher improvement or more effective personnel practices.

A better understanding of the teacher contributions that administrators consider in their personnel decisions may improve the design of emerging evaluation systems. Investigations of administrator practices can help to illuminate not only what they care about, but what they can observe and how they act upon those observations. While the measures utilized in teacher evaluation system are ultimately discretionary, additional insights into administrators' perspectives and professional judgments offer two key benefits. First, they can inform the selection of measures and professional standards considered in evaluations. Second, they may identify common disconnects between desirable standards and the priorities of local school leaders who will be responsible for their implementation.

Purpose / Objective / Research Question / Focus of Study:

This study explores a variety of school personnel decisions. I link each of these to mid-year evaluative feedback provided by administrators to teachers. I also link value added measures to the evaluations teachers receive. I specifically address the following questions of interest:

1. Do overall ratings on formative mid-year teacher evaluations predict subsequent dismissal and promotion decisions by administrators?
2. Are there coherent and distinct factors within evaluative ratings that reflect different aspects of teacher performance?
3. Are different types of personnel decisions or anticipated personnel decisions predicted by different factors from the evaluative ratings?
4. Do either overall ratings or specific factors from the evaluations predict teacher value added performance in the same school year?
5. Do school aggregates of all teachers' evaluative ratings predict school-wide average value added performance in the same school year?

Setting:

My data come from a network of public charter schools that operate under a single centralized district management team. Data are available from 17 different schools who participated in a subjective, formative teacher evaluation system between SY 2008-09 and SY 2010-11. As a charter network, school and district leaders had full autonomy over teacher personnel decisions. The charter schools in this study serve a lottery-randomized population of primarily poor and minority urban students, and have been shown to be higher-performing than peer schools in lottery-randomized studies. The district's identify remains anonymous in this study.

Population / Participants / Subjects:

I examine 747 individual teacher evaluations over a three year period, and relate them to a variety of school personnel decisions affecting teachers in the same school year. Teachers taught in grades k-12, in all subject areas. For a subset of math and language arts teachers in grades 4-8, I also compare teacher evaluations to teacher value added measures available in SY 2010-11.

Intervention / Program / Practice:

Over the period of the study, teachers were evaluated and provided feedback using a set of 49 individual subjective criteria, as detailed in Figure 2. The overall distribution of teachers' evaluation ratings is quite even across the district, as shown in Figure 1. Individual within-school distributions reflect a similarly even distribution. In addition, individual school-wide average standardized teacher evaluation ratings reflect substantial variation, with a range from -0.75 and 0.64, and a standard deviation of 0.40.

The personnel decisions considered in this study are teacher dismissals, teacher resignations, and promotions to two district-specific school leadership roles: "academic assistant principal (AP)" roles and to "school culture" assistant principal roles. Academic AP roles relate primarily to leading an instructional program, as well as some general management responsibilities at district schools, while school culture AP roles are a more specialized role related to maintaining a positive school culture, including strong relationships with students and parents.

In addition to enacted personnel decisions, the district also collected some data on anticipated personnel decisions. In SY 2009-10, school leaders in the district conducted an internal census of views regarding teachers that were strong candidates for promotion in the future. This census included the AP roles previously discussed, as well as a hypothesized "Expert Teacher" role that had not yet been established in the district. This hypothetical role conceptually was focused on exemplary classroom instructional practice, rather than school management responsibilities.

The school district managed the collection of evaluation and personnel data from participating schools. Descriptive information regarding teacher characteristics and school personnel decisions in each school year is provided in Table 1. In SY 2010-11, the district contracted with an external vendor to create value added measures. Model details are not included in this abstract.

Research Design:

My research design is to use overall and component factors of evaluation ratings of teachers to predict same-year personnel decisions. I also examine how well evaluation ratings predict teacher and school value added measures.

Data Collection and Analysis:

My analysis consists of multiple distinct steps, corresponding to my research questions. To examine my first research question, I examine whether an overall evaluation rating – generated by creating a single factor from the 49 individual evaluation indicators – predicts the likelihood of various personnel decisions, as detailed in Equation 1.

$$(1)\ \ln\frac{\pi(p)}{1-\pi(p)} = \beta_0 + \beta_1 X_{it} + \beta_2 \delta_{it} + \gamma_t + \varepsilon_{it}$$

Here, the log likelihood of personnel decision p is a function of a vector *X* of teacher *i*'s characteristics in year *t*, that teacher's rating δ in year *t*, and fixed effects for individual years, γ_t. I report model results as odds ratios corresponding to my dependent variables of interest.

I then examine whether there are coherent and distinct factors within the 49 evaluation indicators using an exploratory factor analysis. Figure 2 details the factor loadings corresponding to four coherent and distinct factors present in the evaluation data, and their overlap with the district's conceptual grouping of indicators. I rotate the factors so that they are orthogonal. I label the four factors that I identify as: Student Engagement and Behavior, Instructional Specifics, Personal Organization and Planning, and Parent and Student Relationships.

To address my third research question I repeat the analysis in Equation 1 above, but in lieu of a single teacher rating δ in year *t*, *I* now include a vector of the four rating factors.

To address my fourth and fifth research questions, I predict teacher and school value added NCE percentile rankings using evaluation ratings and appropriate controls, as detailed in Equations 2 and 3 below.

$$(2)\ \mu_{is} = \beta_0 + \beta_1 X_i + \beta_2 \delta_i + \omega_i + \varepsilon_i$$

Here, teacher *i*'s SY 2010-2011 value added percentile μ in subject *s* is a function of a vector of SY 2010-2011 evaluation rating factors δ and of teacher characteristics X.

$$(3)\ \partial_{ast} = \beta_0 + \beta_1 \delta_{at} + \omega_a + \gamma_{at} + \varepsilon_i$$

Here, school a's value added percentile ∂ in subject *s* and year *t* is a function of a vector δ of one or more school-wide evaluation rating factors in the same year. I also include a fixed effect ω corresponding to school a's region, and a fixed effect γ for the 2009-2010 school year. In order to account for the clustering of data points within-schools and across years, I use a bootstrap to obtain standard errors, clustered at the school level.

Findings / Results:

I find that administrators' overall formative teacher evaluation ratings are significant and substantial predictors of future personnel decisions, as illustrated in Table 3.

Collectively, the four distinct factor ratings of a teacher explain substantially more variation in subsequent personnel decisions about that teacher than a single principal component factor. Moreover, the factors that administrators weigh in their decisions vary in accordance with the particular personnel decision in question. Table 4 provides a summary of model runs using all four orthogonally-rotated evaluation factors, in lieu of a single principal factor.

I find that lower scores in Student Engagement and Behavior and Personal Organization and Planning are strong predictors of teacher dismissal. I also find that higher scores in these two factors, particularly Student Engagement and Behavior, predict greater likelihood of teacher promotions to academic school leadership. However, promotions to student culture leadership roles are predicted by a different set of factors: Student Engagement and Behavior, and the factor for Parent and Student Relationships.

Administrators' internal assessment of teachers' potential for future promotion are summarized in Table 5. In general, the results for anticipated teacher promotions to school leadership roles reflect the same priorities as in administrators' enacted promotion decisions, but with reduced effects sizes. However, unlike other personnel decisions, administrator's identification of potential "Expert Teachers" corresponds to high teacher ratings in Instructional Specifics, as well as in Student Engagement and Behavior.

As detailed in Table 6, I find only a limited association between teacher value added and these evaluations, although Instructional Specifics predicts higher value added in language arts. However, as shown in Table 7, the average school-wide teacher rating on the factor for Student Engagement and Behavior is quite predictive of school-wide value added outcomes in math.

Conclusions:

This study offers new insights into the diversity of teacher contributions that local school administrators value in their management of school staff. I find that administrators consider a range of teacher practices when assessing teacher quality, and that their formative evaluations predict their future personnel decisions. Moreover, administrators weigh different, distinct evaluative criteria when staffing different teacher roles at their schools. The results suggest the importance of accounting for multiple aspects of teacher performance in evaluation systems that are meant to inform multiple types of personnel decisions.

Some of the criteria valued by administrators in this district could be observed in teachers' daily classroom practice, while others reflect teachers' planning, their interactions with peers, or their interactions with students and families that are more likely to be observed outside of the classroom. This suggests that evaluation systems may be more accurate if at least some of the multiple measures used to evaluate teachers address their contributions and competencies outside of instructional execution.

I also identify an association between a teacher evaluation factor that administrators valued highly across all of their personnel decisions – Student Engagement and Behavior – and school-wide average student achievement gains in math. Schools characterized by strong teacher ratings in this area were more effective at raising student achievement than those that were not. The limited sample size renders this finding somewhat tenuous. Nevertheless, the results offer suggestive evidence that evaluation systems may be more effective if they weigh teachers' coordinated efforts in addition to their individual expertise.

Appendices

Appendix A. References

Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J (2010, July 20). "Teacher layoffs: An empirical illustration of seniority v. measures of effectiveness." CALDER working paper.

Goldhaber, D., and Theobald, R.. (2011) "Managing the teacher workforce in austere times: The implications of teacher layoffs." Seattle, WA: University of Washington Center for Education Data and Research, available at http://www.cedr.us/papers/working/CEDR%20WP%202010-7%20Teacher%20Layoffs%203-15-2011.pdf

Rockoff, J. E., Staiger D. O., Kane, T. K., and Taylor, E. S. (2011). Information and employee evaluation: evidence from a randomized intervention in public schools. *American Economic Review*.

Taylor, E.S., & Tyler, J.H. (2011). "The effect of evaluation on performance: Evidence from longitudinal student achievement data of mid-career teachers." National Bureau of Economic Research working paper no. 16877.

Appendix B. Tables and Figures

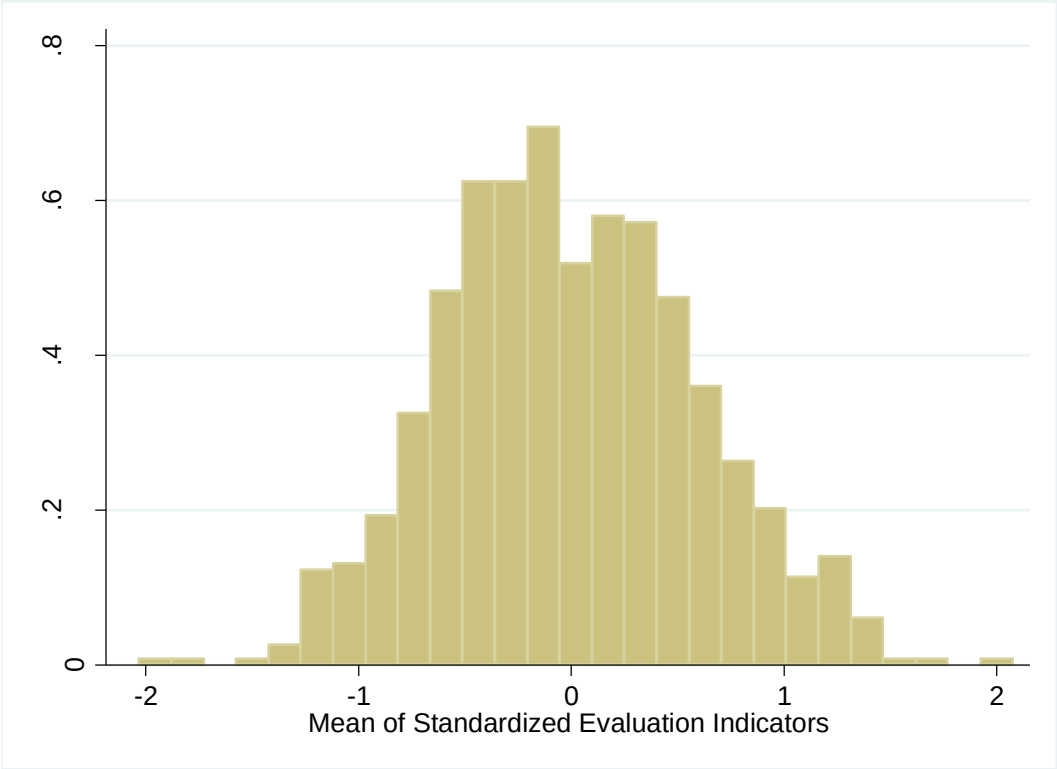

FIGURE 1. *Distribution of the Average of Individual Teachers' Standardized Indicator Ratings*

TABLE 1
Descriptive Statistics for District Teachers and Personnel Decisions, by School Year

	SY 2008-2009	SY 2009-2010	SY 2010-2011	All Years
# of Teachers Evaluations	178	263	306	747
Average Within-School Evaluation Rate	64%	78%	83%	77%
% Promoted to Academic School Leadership	3.9%	2.7%	0.7%	2.1%
% Promoted to School Culture Leadership	2.8%	1.5%	0.7%	1.5%
% Resigned	11.2%	9.1%	12.7%	11.1%
% Dismissed	7.9%	8.7%	5.6%	7.2%
% with Expert Teacher potential		14.4%		
% with Leadership potential, next 1-2 years		4.6%		
% with Leadership potential, next 3-5 years		8.4%		
% Female	77.5%	74.5%	71.2%	73.9%
% White	65.2%	64.3%	66.3%	65.3%
% Black	18.5%	16.3%	13.1%	15.5%
% Hispanic	8.4%	8.4%	10.1%	9.1%
% Asian	3.4%	4.9%	4.9%	4.6%
% Other/ Unknown	4.5%	6.1%	5.6%	5.5%

TABLE 2

Conceptual Structure and Factor Loadings of Mid-Year Formative Teacher Evaluation Indicators

Evaluation Dimension	Evaluation Indicator	Factors			
		Student Engagement and Behavior	Instructional Specifics	Personal Organization and Planning	Parent & Student Relationships
Achievement	Achievement relative to goals				
Character	Students respectful	0.77			
	Students enthusiastic	0.77			
	Students do their best	0.52			
	Students' citizenship	0.62			
	Students present/prepared	0.54			
Instruction	Clear goals for each lesson			0.53	
	Daily assessment			0.45	
	Accurate content		0.56		
	Well-planned lesson		0.64		
	Clear lesson sequence				
	Guided practice		0.61		
	Checks for understanding		0.57		
	Independent practice		0.49		
	Support during ind. Practice				
	Student work time		0.57		
	Quality responses		0.41		
	Quality questions		0.43		
	Differentiation		0.41		
Classroom Culture	All Students on-task	0.67			
	Engagement strategies	0.62			
	Classroom routines	0.75			
	High behavioral standards	0.72			
	Positive classroom environment	0.62			
	Positive student interactions	0.54			
	Character building	0.53			
	Tie character to lessons	0.44			
	Neat / orderly classroom				
	Support school culture system	0.44			
	Proper use of incentives	0.49			
Systems and Planning	Goal-setting			0.48	
	Investing students in goals				
	Knowledge of curriculum		0.50		
	Year-long instructional plan		0.44	0.49	
	Unit plans		0.44	0.48	
	Lesson plans			0.53	
	Weekly/informal data use			0.48	
	Organized data tracking			0.44	
	Periodic/formal data use			0.52	
Student and Family Relationships	Cares about students	0.41			0.40
	Relationships outside of class				0.50
	Relationships with families				0.73
	Sharing goals with parents				0.53
	Communication with parents				0.66
Personal Effectiveness	Constantly learning			0.41	
	Organized			0.52	
	Attendance				
	Communication with peers			0.41	

Note: Indicators that are not highly loaded (>0.40) on any single factor are left blank

TABLE 3

Predicting the Likelihood of Teacher Dismissals, Resignations, and Promotions with Teacher Characteristics and Overall Evaluation Ratings (Odds Ratios)

	Teacher Characteristics				Teacher Characteristics and Evaluation Ratings			
	Dismissed	Resigned	Promoted to AP of Culture	Promoted to Academic AP or Principal	Dismissed	Resigned	Promoted To AP of Culture	Promoted to Academic AP or Principal
1st Year in the District	2.174~	0.258***	0.413		0.874	0.201***	2.558	
	(0.863)	(0.087)	(0.396)		(0.373)	(0.072)	(2.849)	
2nd Year in the District	1.782	0.848	1.630		1.358	0.790	3.348	
	(0.739)	(0.241)	(1.200)		(0.570)	(0.226)	(2.739)	
Female	0.612	1.175	2.797	0.745	0.621	1.180	2.762	0.744
	(0.187)	(0.338)	(2.977)	(0.411)	(0.198)	(0.341)	(3.028)	(0.428)
Black	1.675	1.200	3.532*	0.325	1.314	1.113	8.336**	0.388
	(0.600)	(0.373)	(2.240)	(0.339)	(0.494)	(0.349)	(6.345)	(0.411)
Age	1.094***	0.925*	0.966	1.012	1.122***	0.934*	0.925	0.949
	(0.023)	(0.030)	(0.073)	(0.049)	(0.026)	(0.031)	(0.084)	(0.058)
Indicator for SY 2008-2009	1.377	0.946	3.889	6.703*	1.633	0.989	3.541	8.538*
	(0.535)	(0.287)	(3.344)	(5.433)	(0.657)	(0.301)	(3.260)	(7.178)
Indicator for SY 2009-2010	1.710	0.617~	2.033	4.328	1.842~	0.617~	2.813	5.619*
	(0.582)	(0.172)	(1.780)	(3.494)	(0.644)	(0.173)	(2.615)	(4.685)
Evaluation Rating, Single Factor					0.377***	0.756*	7.078***	3.395***
					(0.071)	(0.103)	(3.472)	(1.033)
Number of Teacher-Ratings	747	747	747	747	747	747	747	747
Psuedo R-Squared	5.907%	5.511%	11.471%	5.784%	13.721%	6.337%	30.808%	17.739%

Note: AP = Assistant Principal. As no 1st-year teachers were promoted to Academic AP or Principal roles, experience controls are omitted in those models.
~p<.1 *p < .05, **p < .01, ***p < .001.

TABLE 4

Predicting the Likelihood of Teacher Dismissals, Resignations, and Promotions with Multiple Evaluation Factors (Odds Ratios)

	Dismissed		Resigned		Promoted To AP of Culture		Promoted to Academic AP or Principal	
Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Demographic Controls		Y		Y		Y		Y
Student Engagement and Behavior	0.443***	0.373***	0.931	0.788~	3.518***	8.386***	2.628***	2.664***
	(0.076)	(0.071)	(0.110)	(0.103)	(1.233)	(4.793)	(0.753)	(0.779)
Instructional Specifics	1.171	0.935	1.019	0.992	1.354	1.939~	1.482	1.553~
	(0.185)	(0.157)	(0.122)	(0.130)	(0.410)	(0.755)	(0.359)	(0.405)
Personal Organization and Planning	0.533***	0.541***	0.809~	0.764*	0.943	1.204	1.916*	1.834*
	(0.081)	(0.088)	(0.096)	(0.096)	(0.314)	(0.503)	(0.510)	(0.501)
Parent and Student Relationships	1.013	0.954	1.151	1.044	3.857***	5.852***	0.970	0.994
	(0.147)	(0.144)	(0.133)	(0.129)	(1.468)	(2.838)	(0.272)	(0.286)
Number of Teacher-Ratings	747	747	747	747	747	747	747	747
Psuedo R-Squared	11.732%	17.558%	1.275%	7.100%	32.291%	45.684%	19.072%	20.022%

Note: AP = Assistant Principal. Demographic controls include years of experience in the district, age, race, and gender. ~p<.1 *p < .05, **p < .01, ***p < .001.

TABLE 5
Predicting the Likelihood of Teacher Identification for Future Promotions with Multiple Evaluation Factors (Odds Ratios)

	"Expert Teacher" Potential		School Leadership Potential within 1-2 Years		School Leadership Potential within 3-5 years	
Demographic Controls		Y		Y		Y
Student Engagement and Behavior	2.641***	2.492***	2.358*	2.073*	1.429	1.721*
	(0.574)	(0.589)	(0.787)	(0.742)	(0.334)	(0.454)
Instructional Specifics	2.050***	1.943**	1.360	1.198	1.332	1.453
	(0.409)	(0.406)	(0.396)	(0.363)	(0.308)	(0.363)
Personal Organization and Planning	1.393~	1.352	1.998*	1.978*	1.780*	1.830*
	(0.264)	(0.271)	(0.599)	(0.626)	(0.425)	(0.476)
Parent and Student Relationships	1.394	1.329	1.017	0.893	1.516~	1.789
	(0.298)	(0.298)	(0.354)	(0.353)	(0.378)	(0.483)
Number of Teacher-Ratings	253	253	253	253	253	253
Psuedo R-Squared	20.034%	20.793%	14.149%	17.379%	8.836%	12.813%

Note: Demographic controls include years of experience in the district, age, race, and gender. ~p<.1 *p < .05, **p < .01, ***p < .001.

TABLE 6

Predicting SY 2010-2011 Teacher Value Added NCE Percentiles with Teacher Characteristics and Evaluation Factors

	Math Value Added				Language Arts Value Added			
	Single Factor		All Factors		Single Factor		All Factors	
Region Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y
1st Year in the District		-16.910		-21.818*		4.970		7.479
		(10.347)		(10.219)		(12.422)		(11.577)
2nd Year in the District		-14.147		-24.799*		-5.845		-0.628
		(11.452)		(12.014)		(11.560)		(10.749)
Female		5.233		9.543		11.242		11.019
		(7.017)		(7.021)		(8.917)		(8.081)
Black		0.872		6.910		10.413		11.594
		(10.948)		(12.310)		(11.968)		(10.785)
Age		-1.206*		-1.062~		-0.753		-0.091
		(0.535)		(0.528)		(0.896)		(0.833)
Evaluation Rating, Single Factor	4.255	0.327			-1.745	0.647		
	(3.278)	(3.935)			(3.907)	(4.638)		
Student Engagement and Behavior			4.840	2.577			-3.017	-2.081
			(3.244)	(3.149)			(4.129)	(4.435)
Instructional Specifics			2.887	-0.471			11.948**	12.450**
			(3.388)	(3.853)			(4.001)	(4.232)
Personal Organization and Planning			5.124	5.463			1.376	0.795
			(3.452)	(4.178)			(3.733)	(3.821)
Parent and Student Relationships			-6.044	-10.528*			-10.485*	-9.596*
			(4.071)	(4.261)			(3.909)	(4.215)
Number of Teachers	38	38	38	38	53	53	53	53
Adjusted R-Squared	<0%	5.571%	1.883%	17.588%	<0%	<0%	17.776%	17.330%

Note: Teacher value added estimates are reported as normal curve equivalent percentiles, ranked relative to the mean and standard deviation of external-district teacher estimates in the same grade and subject. Sample includes 91 (out of a total of 120) district teachers with value added data for whom evaluation ratings were also available in SY 2010-2011. ~p<.1 *p < .05, **p < .01, ***p < .001.

TABLE 7

Predicting School-wide Value Added NCE Percentiles with the School-wide Average of Teachers' Mid-Year Evaluations

	Math					Language Arts				
	All Factors	Student Engagement and Behavior	Instr. Specifics	Org. and Planning	Parent and Student Relations	All Factors	Student Engagement and Behavior	Instr. Specifics	Org. and Planning	Parent and Student Relations
Region Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Student Engagement and Behavior	39.23~ (23.582)	35.475* (15.289)				3.238 (17.744)	-1.653 (13.103)			
Instructional Specifics	-8.900 (23.963)		-0.305 (15.62)			-9.812 (11.467)		-9.064 (8.031)		
Personal Org. and Planning	-2.880 (20.547)			11.263 (12.92)		-1.959 (15.371)			-0.543 (12.661)	
Parent and Student Relationships	12.615 (31.764)				13.467 (20.033)	0.243 (17.927)				-1.23 (13.108)
Number of Schools	19	19	19	19	19	19	19	19	19	19
Adjusted R-Squared	29.794%	37.350%	11.279%	16.734%	16.622%	13.428%	23.635%	30.326%	23.548%	23.611%

Note: Sample of schools with students in value added grades (4-8). School-wide averages of teacher evaluation ratings calculated as the average of all individual teacher ratings, including teachers without value added scores. School-wide average value added calculated as the average of teachers' (SY 2010-2011) or classrooms' (SY 2009-2010) value added percentiles, including teachers who were not evaluated. Bootstrap standard errors clustered at the school level. ~p<.1 *p < .05, **p < .01, ***p < .001.