The Middle School Teacher Turnover Project

A Descriptive Analysis of Teacher Turnover in New York City's Middle Schools

Technical Documentation

William H. Marinell
The Research Alliance for New York City Schools
New York University

This technical documentation provides detailed information regarding the data, methods, and variables featured in our study of New York City (NYC) middle school teacher turnover. This documentation refers to the study's five central research questions:

1. What are the characteristics of NYC middle school teachers and how have they changed over the past decade?
2. How long do middle school teachers remain in their schools?
3. How, if at all, have rates of middle school teacher turnover changed over the past decade?
4. To what extent is turnover characterized by mobility between schools or attrition from the NYC public school system?
5. Which teacher and school characteristics are associated with turnover?

This documentation addresses the following topics in the sequence which they are listed: data sources; methodologies; the variables included in the analyses; an explanation of our modeling strategy; statistical testing; and the limitations of our data and our analytical approach.

Data

This study utilizes a variety of NYC data sets to investigate the research questions stated above. Our master analytical data set was based on ten annual data files (2000-01 through 2009-10) of the NYC Department of Education's (DOE) human resources data. Each of these annual files contains information pertaining to all full-time "pedagogues" during a particular school year. The DOE's human resources files are also the source of all of the information regarding teachers' background characteristics, including their age, racial/ethnic background, gender, years of experience working in NYC schools, degree credentials, etc. To create our master analytical data set for the period of observation covered by the study (2001-10), we sorted these annual files by a unique, encrypted teacher identifier and then merged them into one data set that contained information across the ten years included in the study. Thus, the revised analytical data set contained one observation per pedagogue. From this data set, we selected out the 24,598 full-time teachers who were employed in at least one sample school for at least one school year. Descriptive information about the characteristics of these teachers is included below.

To examine the relationship between turnover and the characteristics of middle schools, we merged information from a variety of publicly-available, school-level data files onto the teacher-level master analytical data set. The DOE's *Official Audited Register, SFORM* data set was the source of information related to the composition of schools' students, such as the number and percentage of students from various racial/ethnic backgrounds.[1] School-level variables of student performance came from the *Citywide Achievement* data set.[2] The *School-Based Expenditure Reports* were the source of information regarding schools' expenditures, such as schools' annual per pupil expenditures.[3] Data about schools' borough location came from the *Official Audited Register, JFORM* data set,[4] and the *School Demographics and Accountability Snapshot* provided information about the percentage of students from high poverty backgrounds.[5] Lastly, the *New York City Progress Report Results* data set provided information about schools' environments, such as the measures from the *School Survey*.[6]

Methodology and Measures

To address the study's first research question, we described how the characteristics of the middle school teacher workforce had changed over the past decade. To address our subsequent research questions, we use a discrete time survival analysis methodology (DTSA). DTSA is a specialized type of event-history logistic regression analysis designed to address research questions that ask whether and when events occur (Singer & Willett, 1993; 2003). For this study, we are interested in determining whether middle school teachers leave their schools and, if so, how long they remain in their schools before leaving. Among the features that make DTSA superior for this type of analysis is its ability to account for the right-hand censoring of data that occurs as, in this case, teachers approach the end of the period of observation in 2010. Alternative methodologies either exclude censored cases or impute estimates of when censored cases experience the "events" in question (Singer & Willett, 1993).

Discrete time survival analysis requires that data sets are organized into a person, spell, period structure. Thus, we converted our one-observation-per-teacher master analytical data set into a hierarchical file, which contained one observation for each year a teacher was employed in a sample school between 2001 and 2010. Teachers who were employed in more than one sample school over the period of observation appear in the data set multiple times. We refer to each block of years that a teacher was employed in a sample school as a "spell." Consistent with methodological convention, our analyses refer to the individual years within each spell as "periods." Table 1 presents a hypothetical segment of the person, spell, period data set for illustrative purposes. Note that Teacher A and Teacher B each has two "spells" in NYC middle schools. For Teacher A, these spells were in two different schools. By contrast, Teacher B spent two separate spells in the same school. Also note that, since both Teacher A and Teacher B were employed in sample schools in 2010, their second spells are censored; however, the advantage of DTSA is that these cases contribute to the analysis up until the point of censoring.

Table 1
Demonstrative excerpt from a person, spell, period data set, sorted by: teacher, spell, period

Teacher	Year	SPELL	School	PERIOD (number of years in school)	Did the teacher "MOVE"?	Is this spell censored?
Teacher A	2006	1	R072	1	1	0
	2007	2	K278	1	0	1
	2008	2	K278	2	0	1
	2009	2	K278	3	0	1
	2010	2	K278	4	0	1
Teacher B	2006	1	K301	1	0	0
	2007	1	K301	2	1	0
	2010	2	K301	1	0	1

The final person, spell, period analytical data set contains 102,901 observations of 24,598 unique teachers. However, as the accompanying report explains, the bulk of our analysis focus on new-to-school teachers—in other words, teachers who entered middle schools between 2002

and 2009; it is only for this subpopulation of teachers that we can generate accurate estimates about the length of time middle school teachers remain in their schools. This smaller new-to-school teacher segment of the person, spell, period data set contains 48,912 observations of 15,628 unique new-to-school teachers. Table 2 presents descriptive information about the teachers in both the larger data set, as well as in the new-to-school data set.

The dichotomous outcome of our logistic regression model is called *MOVE,* though it refers both to teachers' decision to move between schools and to exit schools. Further, because we were ultimately interested in investigating the extent to which turnover is disruptive to schools, we employed an expansive definition of turnover, identifying a teacher as having "moved" if she either: a) left her school (due to move or exit); or b) assumed a role other than as a full-time teacher (i.e. principal or librarian), regardless of whether she remained in her school. Our rationale for considering role-changers as having "turned over" is that this transition ultimately leaves building administrators facing the same scenario: needing to fill a vacant teaching position.[7] Thus, we coded *MOVE* "1" if teachers experienced the event of moving, exiting, or changing roles, and "0" otherwise.

We fit variations of four basic models to address our second and fifth research questions. The general specifications of these models appear below, after our narrative explanation. To address our second research question about how long middle school teachers remain in their schools, we fit a time-only model (*Equation 1*). This model allowed us to estimate middle school teachers' median length of stay in schools, as well as the percentage of teachers who left the New York City public schools after each time period, on average. We fit variations of three additional models to address our fifth research question, which investigates the relationship between turnover and various teacher and school characteristics. We began by fitting a series of models that sequentially added and removed the individual baseline covariates discussed in the report to the time-only model. As described in the report, we also examined the nature and strength of the relationship between turnover and the covariates when the covariates were included in a full baseline control model (*Equation 2*).

To examine the relationship between turnover and the characteristics of teachers, we sequentially added and removed each teacher-level predictor and its two-way interaction with the continuous specification of time to the baseline control model. We then examined the nature and strength of the relationship between turnover and teacher characteristics in a full teacher-level model (*Equation 3*) that contained all of the teacher characteristics and their interactions with time, the baseline covariates, and the discrete specifications of time. *Please note: to simplify the figures and tables in the accompanying report, we did not report statistics for the interactions between teachers' characteristics and the continuous specification of time in our tables in Part II. However, we generated the turnover statistics identified in the report—the percentages of teachers who leave their schools within one, three, and five years after having begun in those schools—from models that contained these interactions.*

Table 2

Descriptive information about the teachers in the analytical data sets

		All Teachers				New-to-school Teachers			
		Across all spells		Unique teachers		Across all spells		Unique teachers	
		N	**Percent**	**N**	**Percent**	**N**	**Percent**	**N**	**Percent**
Ethnicity	White	16567	60.0	14943	60.6	10208	56.7	8924	57.1
	Black	6138	22.2	5343	21.7	4252	23.6	3639	23.3
	Hispanic	3219	11.7	2783	11.3	2180	12.1	1859	11.9
	Other ethnicity	1524	5.5	1355	5.5	1225	6.8	1063	6.8
Gender	Male	9271	33.6	8322	33.8	5675	31.5	4953	31.7
	Female	18365	66.5	16276	66.2	12344	68.5	10675	68.3
Teaches math or science	Yes	5604	20.3	4966	20.2	3883	21.6	3381	21.6
	No	22032	79.7	19550	79.5	14136	78.5	12196	78.0
	Subject changes[a]			82	0.3			51	0.4
Age	<=30	10316	37.3	9208	37.4	8361	46.4	7346	47.0
	>30-55	14110	51.1	11919	48.5	8344	46.3	6953	44.5
	>55+	3210	11.6	2846	11.6	1314	7.3	1097	7.0
	Age category changes[a]				2.5			232	1.5
Degree	B.A.	10908	39.5	9579	39.0	8265	45.9	7211	46.1
	M.A.T. or credit equivalent	8569	31.0	7144	29.0	5761	32.0	4694	30.0
	M.A.T.+ 30 additional credits	7081	25.6	6058	24.6	3055	17.0	2452	15.7
	Other	1078	3.9	820	3.3	938	5.2	745	4.8
	Degree category changes[a]				4.1			526	3.4
Experience	0-3 yrs	13278	48.1	11754	47.8	10656	59.1	9365	59.9
	>3-6 yrs	3066	11.1	2274	9.2	1988	11.0	1425	9.1
	>6-9 yrs	6025	21.8	4569	18.6	3809	21.1	2907	18.6
	>9 yrs	5267	19.1	4697	19.1	1566	8.7	1321	8.5
	Experience category changes[a]			1304	5.3			610	3.9
Total		**27636**	**100.0**	**24598**	**100.0**	**18019**	**100.0**	**15628**	**100.0**

NOTE: [a] In instances where teachers appear in multiple spells in the data set, their characteristics (such as their broad age category) occasionally change.

To examine the extent to which school characteristics predict the probability of teacher turnover, we embarked on a similar model building process that entailed adding and removing each school-level predictor to the full teacher model and examined the nature of the relationship between turnover and each of the school characteristics. In addition, we examined the nature and strength of these relationships in a full school-level model (*Equation 4*), which contained all of the school-level variables, as well as all of the previously described covariates and the teacher-level main effects and their interactions with the continuous specification of time.

An abbreviated rendition of our four basic models is as follows:

$$MOVE = \frac{1}{1 + e^{-X}}$$

Where X = :

(1) **Time-only Model:** $(\gamma_0 \times INT + \gamma_1 \times PERIOD2 + \cdots \gamma_7 \times PERIOD8)$

(2) **Baseline Control Model:**
$(\gamma_0 \times INT + (\gamma_1 \times PERIOD2) + \cdots (\gamma_7 \times PERIOD8) + (\gamma_8 \times SPELLSTR03) + \cdots (\gamma_{13} \times SPELLSTR09) + (\gamma_{14} \times ENRNUMDIFF) + (\gamma_{15} \times BRONX) + (\gamma_{16} \times BRKLN) + (\gamma_{17} \times QUEEN) + (\gamma_{18} \times STATN) + (\gamma_{19} \times NEWPRINC))$

(3) **Teacher-level Model:** $(\gamma_0 \times INT + (\gamma_1 \times PERIOD2) + \cdots (\gamma_7 \times PERIOD8) + (\gamma_8 \times SPELLSTR03) + \cdots (\gamma_{13} \times SPELLSTR09) + (\gamma_{14} \times ENRNUMDIFF) + (\gamma_{15} \times BRONX) + (\gamma_{16} \times BRKLN) + (\gamma_{17} \times QUEEN) + (\gamma_{18} \times STATN) + (\gamma_{19} \times NEWPRINC) + (\gamma_{20} \times TEACHAR) + (\gamma_{21} \times (TEACHAR * PERIOD))$

(4) **School-level Model:** $(\gamma_0 \times INT + (\gamma_1 \times PERIOD2) + \cdots (\gamma_7 \times PERIOD8) + (\gamma_8 \times SPELLSTR03) + \cdots (\gamma_{13} \times SPELLSTR09) + (\gamma_{14} \times ENRNUMDIFF) + (\gamma_{15} \times BRONX) + (\gamma_{16} \times BRKLN) + (\gamma_{17} \times QUEEN) + (\gamma_{18} \times STATN) + (\gamma_{19} \times NEWPRINC) + (\gamma_{20} \times TEACHAR) + (\gamma_{21} \times (TEACHAR * PERIOD) + (\gamma_{22} \times SCHLCHAR))$

Where *PERIOD2-PERIOD8* are a series of seven dummy variables indicating each of the discrete time periods between 2001-02 and 2008-09;[8] *SPELLSTR03-SPELLSTR09* are a series of seven dummy variables indicating the year in which new-to-school teachers began in their schools; *ENRNUMDIFF* is a continuous, time-varying measure of the annual changes in schools' student enrollment; *BRONX, BRKLN, QUEEN,* and *STATN* are dummy variables indicating a school's NYC borough location; *NEWPRINC* is an indicator of whether there was one or more incident of principal turnover in a school during the period of observation; *TEACHAR* is a vector of the teacher background characteristics of interest, including teachers' age, race, gender, etc.; *TEACHAR*PERIOD* represents the series of two-way interactions between each of the teacher background characteristics[9] and the continuous specification of time;[10] and *SCHLCHAR* represents a vector of schools' organizational characteristics of interest, including the percentage of students from low-income backgrounds, schools' peer index value, schools' per pupil expenditures, school size, schools' NYC borough location etc. Table 3 presents a complete list of the variables in this analysis. In addition, as revealed in Table 3 and discussed in the

accompanying report, we examine the two-way interaction between a teacher's race and whether her school contained a larger-than-average proportion of White students (*HIPCTWHTAVG*).

We fit a separate model to address our third research question about whether rates of NYC middle school teacher turnover have decreased over time. In this model, we used the two-way interactions between the *SPELLSTR* dummy variables, which indicate the years in which new-to-school teachers first entered their schools (which we also refer to as teachers' "cohort year") and the discrete specification of the time period dummies. This model contained a dummy variable for the first period and, as such, it did not contain an intercept term. To examine whether turnover rates have declined over time, we calculated survival probabilities from the hazard probabilities that the LSMEANS statement in SAS, version 9 generates, and then we converted the survival probabilities to our outcome metric: the percentage of new-to-school teachers who left their schools within various periods of time. We interpreted the statistical significance of the reduction in the likelihood ratio for the global null hypothesis test as an indication that the observed differences in the slopes of hazard functions across the new-to-school cohort years were statistically distinct, on average. In addition, the p-values associated with the *Wald Chi-Square* statistics (in the analysis of maximum likelihood estimates) of each of the interactions between teachers' cohort year and the discrete time periods were all less than 0.0001.

Lastly, to address our fourth research question, we simply reported the descriptive percentages of the departing new-to-school teachers who were *Movers* and *Leavers* across the entire period of observation. For the purposes of this preliminary component of the study, we did not create separate models in order to predict the probabilities of these two outcomes separately. However, as we discuss in Part III of the report, this is one of several analyses that we intend to conduct in as we extend this project.

Table 3
Variables included in the study of NYC middle school teacher turnover

Variable	Description
Outcome	
MOVE	The dichotomous outcome, coded 1 if a teacher left her school (due to move or exit) or assumed a role other than as a full-time teacher (i.e. principal or librarian), regardless of whether she remained in her school.
Discrete specification of time	
PERIOD2-PERIOD8	A series of seven dummy variables (PERIOD1 is omitted to avoid perfect collinearity) identifying the discrete time periods in the person, spell, period data set.
Baseline covariates	
SPELLSTR03-SPELLSTR09	A series of seven dummy variables (The SPELLSTR02 variable is omitted to avoid perfect collinearity) indicating the school year in which a teacher first began teaching in a sample school, coded 1 if a teacher entered in the spring of each school year (i.e. a teacher who entered her school during the 2002-03 school year is coded SPELLSTR03 = 1), 0 otherwise.
ENRNUMDIFF	A time-varying, continuous variable indicating the annual fluctuations in a school's student enrollment.
BRONX, BRKLN, QUEEN, STATN	A series of four dummy variables indicating a school's NYC borough location (Manhattan is omitted to avoid collinearity).
NEWPRINC	A dummy variable indicating whether a teacher was working in a school where there was 1 or more incident of principal turnover during the period of observation, coded NEWPRINC=1 if a teacher was working in a school where there was principal turnover, 0 otherwise.
Teacher characteristics and interactions	
MALE	A dummy variable indicating whether a teacher is male , coded MALE=1 for men; MALE=0 for women.
MALE*PERIOD	The two-way interaction between MALE and the continuous specification of time.
BLACK, HSPNC, OTHERETHNIC	A series of three dummy variables indicating teachers' race (White is omitted to avoid perfect collinearity). Coded BLACK=1 if a teacher is of African-American descent, 0 otherwise ; HSPNC = 1 if a teacher is of Hispanic descent, 0 otherwise; OTHERETHNIC=1 if a teacher is from a racial/ethnic background other than those previously identified, 0 otherwise.
BLACK*PERIOD, HSPNC*PERIOD, OTHERETHNIC*PERIOD	The two-way interactions between the dummy specifications of teachers' race and the continuous specification of time.

(continued)

Table 3 (continued)	
EXPCAT2-EXPCAT4	A series of three dummy variables (the first category, EXPCAT1, indicating that a teacher has between 0-3 years of experience, is omitted to avoid perfect collinearity) indicating a teacher's years of experience in the NYC public school system when she first entered a sample middle school. Coded EXPCAT2 = 1 if a teacher has between three and six years of experience, 0 otherwise; EXPCAT3 = 1 if a teacher has between six and nine years of experience, 0 otherwise; EXPCAT4 = 4 if a teacher has more than nine years of experience, 0 otherwise.
EXPCAT2*PERIOD-EXPCAT4*PERIOD	The two-way interactions between the dummy specifications of a teacher's years of experience in NYC schools and the continuous specification of time.
DEGCAT2-DEGCAT4	A series of three dummy variables indicating a teacher's level of degree credentials (the first category is omitted to avoid perfect collinearity). Coded DEGCAT2 = 1 if a teacher had a master's degree and/or the credit equivalent, 0 otherwise; DEGCAT3 = 1 if a teacher had a master's degree and 30 additional credit hours (the highest degree step recognized by the NYC DOE's salary scale), 0 otherwise; DEGCAT4 = 1 if a teacher had some other degree credential than those previously identified, 0 otherwise.
DEGCAT2*PERIOD-DEGCAT4*PERIOD	The two-way interactions between the dummy specifications of teacher's degree credential and the continuous specification of time.
AGECAT2-AGECAT3	A series of two dummy variables (the first age category, age 30 or younger, is omitted to avoid perfect collinearity) indicating a teacher's age when she first entered a NYC middle school. Coded AGECAT2 = 1 if a teacher is between 30- and 55-years-old, 0 otherwise; AGECAT3 = 1 if a teacher is over 55-years-old, 0 otherwise.
AGECAT2*PERIOD-AGECAT3*PERIOD	The two-way interactions between the dummy specifications of a teacher's age and the continuous specification of time.
MATHSCI	A dummy variable indicating whether a teacher's primary teaching assignment is math or science, coded MATHSCI=1 if a teacher's assignment is math or science, 0 otherwise.
School Characteristics	
SPELLSCHSZ	A continuous, numerical variable indicating the size of a teacher's school when s/he began teaching in that school.
HIPCTWHTAVG	A time-varying dichotomous variable indicating whether a teacher works in a school that has a high percentage of White students. Coded HIPCTWHTAVG = 1 if 16% of more of a school's student population is White, 0 otherwise. The weighted mean percentage of White students in NYC middle schools was 16% during the period of observation.
HIPCTWHTAVG*BLACK HIPCTWHTAVG*HISP (HIPCTWHTAVG* OTHERETHNIC)	The two-way interaction between the dummy specifications of a teacher's racial/ethnic background and the percentage of her school's student population that is White.

(continued)

Table 3 (continued)	
SQRSCORECAT1, 2-SQRSCORECAT4	A series of three dummy variables (SQRSCORECAT3 is omitted to avoid perfect collinearity) indicating the score that a school received on its 2008 quality review. Coded SQRSCORECAT1 = 1 if a school received a score of "Underdeveloped," 0 otherwise; SQRSCORECAT2 = 2 if a school received a score of "Underdeveloped with Proficient Features," 0 otherwise; SQRSCORECAT4 = 4 if a school received a score of "Well Developed" or "Outstanding," 0 otherwise.
PEERINDX	A continuous measure developed by the NYC DOE indicating how well a school's students score on NY state's standardized test scores relative to schools that serve similar student populations. The peer index measure ranges from 1-4, with lower values indicating lower performing schools that serve higher-need student populations. We use a school's peer index value from 2008 for this analysis.
SCHENRPCTPOV	A continuous measure gleaned from the NYC DOE's publicly-available Comprehensive Education Plan data set that indicates the percentage of a school's student population receiving free lunch in 2008.
PPEXPTOTL	A continuous, time-varying variable indicating a school's overall per pupil expenditures.
STMTHPCTL34ALL	A continuous, time-varying variable indicating the percentage of a school's students scoring at our above NY state's proficiency standard (i.e. scoring at or above the Level 3 performance category) on the state's standardized mathematics assessment, among all of the grades tested in a middle school.
SRCENVSCOREW	A continuous variable with values ranging from 1-15 indicating a school's weighted environment school according to the NYC DOE's 2008 *School Survey.* The DOE creates the weighted environment score by averaging teachers', students', and parents' responses to questions about a school's environment on the *School Survey*. DOE documentation indicates that, unless response rates are atypical, responses are weighted equally across teachers, students, and parents.

Practical Significance versus Statistical Significance

Readers of our report may notice that much of the discussion about the significance of reported differences revolves more around practical considerations than around tests of statistical significance. As we discuss in the report, the primary objective of this study is to describe patterns of turnover across broad types of middle schools and teachers rather than to conduct narrow investigations of the relationship between turnover and specific teacher or school characteristics. Thus, we do not conduct or report extensive statistical tests of difference between the various categories within each teacher or school measure. Rather, at each stage in our analysis, we examine whether the relationship between turnover and a particular measure—for instance, teachers' age—is statistically significant; however, we do not investigate whether differences in the rates of turnover across the three categories of teachers' age (30-years-old or younger, between age 30 and 55, older than 55) are statistically different from one another.

To determine the statistical significance of the individual baseline covariates, teacher characteristics, and school characteristics when they are individually added and removed from individual models (in the report, we refer to this process as the first analytical "step" of each of the three model-building "stages"), we examine whether the inclusion of each measure results in a statistically significant reduction in the *-2 Log Likelihood* given the degrees of freedom added by the measure. During the second analytical step, where we investigate the relationships between turnover and characteristics from various full models, we determine the statistical significance of particular variables by examining the p-values associated with the likelihood ratios in the test of the global null hypothesis for each measure. Table 4 presents the parameter estimates and significance associated with the four analytic models: 1) the time-only model, 2) the baseline control model, 3) the full teacher-level model, and 4) the full school-level model.

Table 4

The parameter estimates, statistical significance, and -2LL for the four foundational models

	Model				
	Null	**Model 1**	**Model 2**	**Model 3**	**Model 4**
		"Time only"	"Baseline control"	"Full teacher model"	"Full school model"
Intercept		-0.9949***	-0.5607***	-0.6113***	2.1581***
PERIOD1			.	.	.
PERIOD2		-0.1470***	-0.1720***	-0.0649*	-0.0056
PERIOD3		-0.4730***	-0.5389***	-0.3363***	-0.2107***
PERIOD4		-0.7462***	-0.8259***	-0.5205***	-0.3378***
PERIOD5		-0.9873***	-1.0732***	-0.6614***	-0.4108***
PERIOD6		-1.1700***	-1.2632***	-0.7519***	-0.4437***
PERIOD7		-1.1232***	-1.2213***	-0.5795***	-0.2105
PERIOD8		-1.6855***	-1.7494***	-1.1759***	-0.7607***
SPELLSTR02			.	.	.
SPELLSTR03			0.0339	-0.1139**	-0.1183**
SPELLSTR04			-0.0505	-0.0554	-0.0247
SPELLSTR05			-0.1106**	-0.1057**	-0.1023*
SPELLSTR06			-0.1376***	-0.1283**	-0.1186*
SPELLSTR07			-0.2523***	-0.2306***	-0.1747*
SPELLSTR08			-0.4233***	-0.4030***	-0.2994***
SPELLSTR09			-0.3305***	-0.2939***	-0.1414
ENRNUMDIFF			-0.0011***	-0.0011***	-0.0012***

Table 4 (continued)			
MNHTN	.	.	.
BRONX	-0.1022**	-0.0786*	-0.2451***
STATN	-0.3902***	-0.3941***	-0.3519***
BRKLN	-0.5254***	-0.5306***	-0.3430***
QUEEN	-1.0183***	-1.0041***	-0.7956***
NEWPRINC	0.0759*	0.0726*	0.0465
MALE		0.1537***	0.1229**
*MALE*PERIOD*		-0.0331*	-0.0351*
WHITE		.	.
BLACK		-0.1071*	-0.3245***
HSPNC		-0.0482	-0.2044**
OTHERETHNIC		0.1560~	0.0752
*WHITE*PERIOD*		.	.
*BLACK*PERIOD*		0.0343~	0.0213
*HSPNC*PERIOD*		-0.0316	-0.0409~
*OTHERETHNIC*PERIOD*		-0.0413	-0.0475
EXPCAT1		.	.
EXPCAT2		-0.0807	-0.0813
EXPCAT3		-0.2098**	-0.2193**
EXPCAT4		-0.5589***	-0.5530***
*EXPCAT1*PERIOD*		.	.
*EXPCAT2*PERIOD*		-0.0303	-0.03
*EXPCAT3*PERIOD*		-0.0174	-0.021
*EXPCAT4*PERIOD*		0.0733*	0.0638*
DEGCAT1		.	.
DEGCAT2		-0.0226	0.0017
DEGCAT3		0.1892*	0.2118**
DEGCAT4		1.0000***	1.0790***
*DEGCAT1*PERIOD*		.	.
*DEGCAT2*PERIOD*		0.0002	-0.0025
*DEGCAT3*PERIOD*		-0.0175	-0.0097
*DEGCAT4*PERIOD*		-0.2163***	-0.2280***
AGECAT1		.	.
AGECAT2		0.1749***	0.1707***
AGECAT3		0.4594***	0.4139***
*AGECAT1*PERIOD*			
*AGECAT2*PERIOD*		-0.1278***	-0.1323***
*AGECAT3*PERIOD*		-0.0918**	-0.0987**
MATHSCI		0.0804**	0.1010***
SPELLSCHSZ			-0.0003***
HIPCTWHTAVG			-0.2670***
*HIPCTWHTAVG*WHITE*			.
*HIPCTWHTAVG*BLACK*			0.4765***
*HIPCTWHTAVG*HSPNC*			0.5123***
*HIPCTWHTAVG*OTHER ETHNIC*			0.1475

(continued)

Table 4 (continued)					
SQRSCORECAT1					-0.0768
SQRSCORECAT2					0.0416
SQRSCORECAT3					.
SQRSCORECAT4					-0.0246
PEERINDEX					-0.4895***
SCHENRPCTPOV					-0.0030**
PPEXPTOTL					-0.0000~
STMTHPCTL34ALL					-0.0043***
SRCENVSCOREW					-0.0432***
-2LL	**50,629**	**49,494**	**48,856**	**48,371**	**47,780**
Statistical significance key: ~ p<.10; * p<.05; ** p<.01; *** p<.001					

Limitations of the content and nature of the data

There are a number of limitations to our analysis and to the master analytical data set, mostly related to information about teacher and schools that is available, but which we did not have in our data archive when we began this investigation. For instance, our data sets do not contain any information on the following topics (all of which might be related to turnover): teachers' licensure credentials; the grade(s) and subject(s) of their primary teaching assignment; transaction information that would allow us to determine whether teachers' decisions to change/leave schools were voluntary or involuntary or were intended as long-term departures or short-term, temporary departures (e.g., maternity leaves). Further, we only have information about teachers' experience in NYC schools, thus we cannot distinguish first-year novice teachers from those who are new to the NYC public schools; similarly, we are unable to determine whether teachers had any teaching experience prior to entering the NYC schools.

There are also some limitations in our school-level variables, namely that we do not have data on all of these variables across the entire period of observation. For instance, we only have information about the percentage of a school's students from low-income backgrounds from 2006-08. In these instances, we use data from the most recently-available school year to identify a school's characteristics across the period of observation. Further, there are missing data values on some of the teacher- and school-level variables, though this is only a minor problem with our various data source files. The cases of teachers with missing values are few and seem implausible and are, thus, dropped from the analysis. In cases where schools were missing values on a measure, also a rare occurrence, we use data from the most recently available year to impute values for the missing fields.

Despite these limitations, we believe that the fundamental purpose of the analysis—to identify how long middle school teachers remain in their schools—is important and unhindered by the limitations in our data set. Other related questions—such as whether turnover has decreased over time and whether turnover is the result of mobility between middle schools or attrition from middle schools—seem equally as important to address given concerns about the role that middle schools play in influencing students' performance in high school and post-secondary education. Furthermore, while there are many additional variables that we would like to include in this study, we have identified a number of important contextual and malleable school-level factors related to teacher turnover.

---

[1] See the following DOE website for more information about this database: http://schools.nyc.gov/AboutUs/data/stats/Register/SFormbyDistricts/default.htm

[2] See the following DOE website for more information about this database: http://schools.nyc.gov/Accountability/data/TestResults/ELAandMathTestResults

[3] See the following DOE website for more information about this database: https://www.nycenet.edu/offices/d_chanc_oper/budget/exp01/OLD_YEARS.asp

[4] See the following DOE website for more information about this database: http://schools.nyc.gov/AboutUs/data/stats/Register/JFormbyDistricts/default.htm

[5]See the following DOE website for more information about this database: http://schools.nyc.gov/Accountability/data/default.htm

[6] See the following DOE website for more information about this database: http://schools.nyc.gov/Accountability/tools/report/default.htm

[7] Another scenario is that teachers who change roles do so because their assignment is made redundant due to declining enrollment in a school. We do not have an ideal way for dealing with this possibility, though we attempt to control for it in our statistical models by incorporating a time-varying change in enrollment variable, *ENRNUMDIFF,* as a baseline covariate in all of our models.

[8] The intercept captures the first discrete time period, 2001-02, and thus we do not specify a dummy for PERIOD1. While we have data through the 2009-2010 school year, we cannot identify any individuals who "moved" during that final year and, thus, we cannot include a dummy for PERIOD9 in our survival analysis.

[9] The interaction, MATHSCI*PERIOD was not statistically significant and was, thus, dropped from all models.

[10] We also fit models that include interactions between the teacher characteristics and the discrete specifications of time. These interactions, as a group were statistically significant, but did not change our estimates much at all, so we opted for the more parsimonious interactions with a continuous specification of time.