Strategic Matching of Teachers and Schools with (and without) Accountability Pressure*

Tom Ahn

March 11, 2016

Abstract

Accountability systems are designed to introduce market pressures to increase efficiency in education. One potential channel by this can occur is to match with effective teachers in the transfer market. I use a smooth maximum score estimator model, North Carolina data, and the state's bonus system to analyze how teachers and schools change matching behavior in response to accountability pressure. Schools under a high degree of accountability pressure match with teachers who are better at raising test scores. Estimation with a 'control' year (when pressure was absent) sample buttresses these findings. Accountability pressure seems to motivate schools to compete for effective teachers.

Keywords: Matching, teacher transfers, accountability, ability sorting

JEL I24, J48, J62

1 Introduction

Accountability systems have a long history of being used by policy makers to introduce market forces into the education production process. The belief is that education production, due to the unobservability of teacher inputs, the noisiness of student outputs, lack of competition, and possibly different objectives pursued by school, district, and state leaderships, is inherently inefficient. The hope is that with the introduction of accountability pressures, we would observe gains in efficiency. Many studies attempt to capture the causal effect of accountability

*I am grateful to the Center for Child and Family Policy at the Sanford School of Public Policy at Duke University for access to the NCDPI dataset. All remaining errors are my own.

policy on student test scores. The implied causal effect is that teachers and principals change something about their behavior that increases education production. Examples of behavioral change examined (or implied) include inducing more or redistributing effort, changing the composition of classrooms, introducing new didactic methods, emphasizing certain subjects, and cheating the system.[1]

Accountability policies rarely prescribe the means by which the school must raise test scores, but it is widely accepted that improving the teacher-side of the production function is imperative. Nearly all education researchers and practitioners agree that teacher quality is one of the most important determinants of a student's success. The literature has focused on identifying characteristics that proxy for teacher quality, finding that experience, education level, and credentials are correlated with higher student achievement.[2] While increasing teacher effort or adopting new didactic techniques may improve test scores, school administrators have another tool they can use to raise education production: matching with more effective teachers in the teacher transfer labor market.

Some focus has been paid to the strategy of getting rid of particularly ineffective teachers.[3] Because it is nearly impossible (or at least very time consuming) to remove ineffective or undesirable teachers, increasing teacher quality through selective layoffs is unlikely to be a feasible option for most schools. However, it is also true that annual teacher turnover in an average school (in North Carolina) is higher than 10 percent. Thus, principals must replace a significant fraction of their workforce yearly.[4] This presents a challenge as well as an opportunity: successfully matching with the 'right' teacher may improve education outcomes, while matching with the 'wrong' teacher may be detrimental, compared to a random match. Schools that can recruit effective teachers may find it easier to satisfy the criterion set forth by accountability systems, while ineffective recruiters may see scores drop no matter what internal changes they institute.

I examine whether an incentive system in North Carolina pushed schools to match with different types of teachers under varying levels of accountability pressures. All teachers at the school were paid a modest $750 or $1,500 bonus if the school was able to demonstrate test

[1] See studies by Ahn (2013), Figlio (2006), Hanushek and Raymond (2006), among others.

[2] See Goldhaber and Anthony (2007), Rockoff (2004), and Clotfelter, Ladd, and Vigdor (2007) among others.

[3] See Jacob (2010).

[4] See Hanushek (2010) to explore gains by replacing the worst teachers with average ones.

score growth in-line with or exceeding the state's expectations. If the accountability system was effective, a school under more pressure should have attempted to recruit teachers who were better at raising scores. A school that was not pressured may have attempted to recruit teachers for other characteristics. At the same time, schools that were attractive to teachers should have had an easier time finding an acceptable match, while schools that were less desirable may have been forced to accept matches that did not increase academic production.

I use an empirical matching model to see where transferring teachers within North Carolina go in their next teaching assignment. There have been a few studies that have looked at teacher-school matching. Boyd et al. (2013) showed that teachers are particularly sensitive to the distance between the school and their place of birth or high school attended. In another study where the authors had detailed information about all transfer applicants, Boyd et al. (2011) determined that teachers who have better pre-service qualifications are most likely to search, and those who have better in-class performance are most likely to succeed in transferring. Ahn (2015) estimated a structural general equilibrium model of matching, finding that mid-career teachers with 4 to 10 years of experience are the most eager to search and attempt to match with a school with higher academic performance. All schools attempt to match with experienced teachers over newly minted teachers, but schools with higher academic performance are choosier about selecting teachers originating from another high performing school. While match success rate and teacher performance (measured by in-class performance or academic status of the originating school) are positively correlated in these studies, how accountability pressure impacts matching is not directly explored and is an under-studied topic. Clotfelter et al. (2004) find that in North Carolina, accountability pressure made it difficult to retain good teachers at under-performing schools. In New York, Boyd et al. (2008) find that the introduction of standardized testing for accountability actually increased retention rates.

As I only observe successful matches, I utilize a matching framework and maximum score estimator from Horowitz (2002) and Fox (2009). The next section presents a model of matching between teachers and schools. Section 3 details the North Carolina accountability system. Section 4 describes the data, and section 5 introduces the econometric model. I present results in section 6, discuss in more detail the impact of accountability in section 7 by estimating the model in a year when accountability pressure was absent, and conclude in section 8.

2 Theoretical Model

Consider a set of teachers A and schools S. Utility gain of teacher a moving from his or her original school s_o to the new school s is $u_a(s, s_o) + \epsilon_{a,s}$, where vector s contains wage and non-wage characteristics, and $\epsilon_{a,s}$ is a match-specific error. Similarly, marginal utility gain of a school (or principal) s matching with teacher a compared to the outside option 0 (a new teacher) is $g_s(a) + \epsilon_{s,a}$.[5] The joint utility function of the match is:[6]

$$f(s, s_o, a) + \eta_{a,s} = u_a(s, s_o) + g_s(a) + \eta_{a,s}$$

where $\eta_{a,s} = \epsilon_{a,s} + \epsilon_{s,a}$.[7]

Following Fox (2009), if a feasible teacher-school assignment is defined as $A - S_1$ and the same assignment with *any pair* of teachers (a, a') and schools (s, s') matches flipped defined as $A - S_2$, with some data of teachers and schools X, I assume rank-ordering as:[8]

$$f(s, s_o, a) + f(s', s'_o, a') \geq f(s, s_o, a') + f(s', s'_o, a)$$

if and only if:

$$Pr(A - S_1|X) \geq Pr(A - S_2|X)$$

This condition defines pairwise stable matches.

One limitation to note is that portions of the joint utility function are not exclusively attributable to the teacher or the principal. For instance, highly experienced teachers may match more frequently with high performing schools because older teachers prefer better performing

[5] I remain agnostic about the principal's utility function, by *not* assuming that he or she seeks to maximize test scores. A principal may seek to minimize classroom disruption through recruitment of experienced teachers or minimize effort exertion by recruiting less ambitious teachers. It may also make sense to think of the principal and the teachers at the school as a coalition with a common utility function.

[6] For a match to be stable, I assume: 1) it is mutually agreed upon, but a break-up may be unilateral. 2) A school (teacher) may offer inducements to a teacher (school) to break up a match.

[7] The joint error term η can be interpreted as match-specific error, which could be, for instance, personal/professional compatibility between the principal and the teacher during the recruiting process.

[8] With this assumption, results from Fox (2009) can be used to show that the probability of any hypothetical market-wide teacher-school assignment is equal to the integration of an indicator function of the particular assignment maximizing the joint output of all matches over the error distribution, given an initial parameter guess.

students, these schools may believe that experienced teachers will increase performance, or both. Because of this, the observed match (and the estimated utility production) is assumed to be joint, inseparable, and specific to the match.

If all portions of $f(\cdot)$ is attributable to the teacher or the principal separately, incentive constraints sum to the joint production framework and pairwise stability.[9] That is:

$$u_{a'}(s', s'_o) + g_s(a) \geq u_{a'}(s, s'_o) +_s (a') \quad \text{and}$$
$$u_a(s, s_o) + g_{s'}(a') \geq u_a(s', s_o) + g_{s'}(a)$$

As abstract as the theory model is, the actual transfer process mirrors the model. It seems that principals and teachers actively search for each other. The hiring of new teachers and transfers in North Carolina is handled at both the school and district levels. While vacancies are advertised at the school level, employment lists are maintained centrally by the district. From discussions with principals and policy practitioners, it is clear that principals motivated to find the 'best' candidates aggressively search through district lists. Ambitious applicants are known to contact principals directly or through other means, such as seeking exposure while working as a student-teacher or lobbying through colleagues.[10]

It should be noted that the initial decision to search by teachers is taken as given, and I do not model how the expected matching outcome in turn influences the search decision. While this is a limitation of the model, in this paper, I am interested in what the stable distribution of teachers across schools with different accountability pressures looks like, conditional on teachers searching.[11] In addition, the modest bonus amount seems unlikely to spur a major decision to move to a new school.[12]

[9]This is equivalent to saying that the inducements that must be offered to the teacher (by the school) or the principal (by the teacher) who may want to 'trade up' must be so large that all parties find it optimal to maintain the original match. See online appendix for complete details.

[10]Some districts have attempted to prevent aggressive principals from recruiting. For instance, Wake County limits the number of teachers principals can take with them, as they moves to new assignments. The policy was implemented in response to complaints that exiting principals were poaching effective teachers.

[11]See Ahn (2015) for a model that accounts for the initial search decision by a teacher (and how it affects all other teachers' search decisions).

[12]The bonus may be large enough to induce a short-distance move, perhaps within the same district. An analysis of bonus receipt history of schools from Ahn and Vigdor (2014) shows that there are very few schools that are almost always out of the running for the bonus. In fact, across 5 years, the average school receives the

3 The North Carolina Accountability System

The North Carolina accountability program (also known as ABC[13]) began in 1995/96. With the exception of minor changes,[14] the mechanism of offering cash bonuses for student achievement gains remained stable for more than a decade. All North Carolina students in the public education system in grades 3 through 8 are required to take End-of-Grade (EOG) exams in reading and mathematics. The tests are on a developmental scale, allowing comparison of scores from consecutive grades. Students entering grade 3 are administered a 'pre-test' within the first three weeks of the school year to serve as the baseline performance measure. Using the formula below, North Carolina Department of Public Instruction (NCDPI) determines the expected achievement gains threshold for each school's bonus eligibility based on the school's students' performance last year.

$$\Delta y_{mgst} = \Delta y_{gs94} + b_1 ITP_{mgt} + b_2 IRM_{mgst}$$

Δy_{mgst} is the expected change in the score for subject s for students in grade g in year t in school m compared to the score on the same subject from year $t-1$. Δy_{gs94} is average gain in scores for students from 1992/93 to 1993/94, the first two years of exam administration. The second and third terms are 'correction factors.' 'Index of True Proficiency' (ITP_{mgt}) and 'Index for the Regression to the Mean' (IRM_{mgst}) adjust test score goals for shocks in students' performances last year. For a complete description, see Vigdor (2008).

For a school with G tested grades, $2G$ thresholds are produced, which are compared to the actual average test score improvement. The school scores and threshold scores are differenced, standardized[15], and weighted by the number of students in each grade. If this average, termed 'expected growth' score is greater than zero, all teachers in the school receive \$750. The procedure is repeated after increasing the growth threshold by 10 percent, to generate the 'high growth' score. Teachers in schools that make high growth receive an additional \$750.

bonus 3.77 times. A teacher would have to move from one of the worst schools to one of the best, for expected bonus amount to make a substantial difference. Within the district, differences in bonus receipt history of schools is relatively small. See the online appendix.

[13]The acronym stands for strong Accountability, teaching the Basics, and emphasis on local Control.

[14]For instance, middle and high schools achievement gains were measured starting in 1997/98. In addition to academic achievement, drop out rates are considered for high schools.

[15]The difference is divided by the standard deviation of the difference across all schools in the state.

While the cash bonus amount may not be enough to induce a teacher to transfer to a different school, the accountability system may induce a principal to recruit effective teachers in two ways. First, the teachers currently at the school collectively have an interest in bringing an effective teacher on board, as this will increase the expected bonus of all teachers. If teachers have formal or informal influence in hiring decisions, the bonus can influence who the principal recruits.[16] Second, schools are recognized for making (or failing to make) expected/high growth, as such records are published on-line as part of 'school report cards.'[17] If academic performance impacts the principals' professional evaluations, recruiting effective teachers to attain higher growth will be important to principals.[18]

It is also worth noting that schools that 'fail' under the ABC system are not necessarily undesirable schools. A high *achieving* school that maintains its level of excellence could be labeled as failing, and a school with low test scores (and other undesirable characteristics) could see a modest increase in test scores and be labeled as succeeding.[19]

4 Data

I use an administrative dataset for the North Carolina public school system from the academic years 1998/99 to 2002/03. I use 2009/10 data for a 'control' group.[20] The dataset contains information on all public schools, students, and teachers in the state. The data is collected

[16]The Working Conditions Survey in 2004 asked elementary school teachers (with 3+ years of experience) whether they felt that they had influence in selecting new teachers. About 45 percent of teachers agreed that they had a role in selecting new teachers, while about 37 percent disagreed.

[17]See www.ncreportcards.org/src.

[18]On the other hand, if schools are also judged on standards that are largely unrelated to growth scores, say the percentage of staff that is fully certified, this may drive recruiting in a different direction.

[19]There is very low correlation between growth and achievement in North Carolina. Ahn and Vigdor (2014) showed in the years 2005-07 (when the No Child Left Behind (NCLB) system was evaluating achievement), that over 40 percent of schools that made expected growth failed to make adequate yearly progress (AYP), and 30 percent of schools that made AYP failed to make expected growth.

[20]The data, which is collected by the North Carolina Department of Public Instruction (NCDPI), was made available by North Carolina Education Research Data Center (NCERDC: www.pubpol.duke.edu/centers/child/nceddatacenter.html) at the Center for Child and Family Policy. While student and teacher level data are confidential, aggregate data and summary statistics are publicly available at the NCDPI web site (www.ncpublicschools.org/reportstats.html).

annually and can link students and teachers across years.

In the matching decision, student data is aggregated to class level, since a teacher is offered a 'bundle' of students at the school. However, in the education production function, I allow a teacher to impact students individually. Table 1 summarizes the student, class, and school characteristics. I focus on grades 3, 4, and 5 for two reasons: one, only grades 3 to 8 are tested. There is no good way to measure teacher effectiveness for lower grades. Middle school students in grades above 5 have several teachers throughout the day, and it is difficult to assign credit to a particular teacher for education production due to spill-over between subjects.

I exclude teachers with no previous teaching experience. By excluding new teachers in the matching model, I am assuming that new teachers do not compete with experienced teachers and are considered the outside option for schools that fail to match with an experienced teacher. Previous research has shown that new teachers are consistently matched with high minority, high poverty schools, and within the school, they are matched with higher-than-school-average minority and disadvantaged classrooms.[21] In addition, new teachers are disadvantaged in the matching process because they are not as likely to be informed about which schools are desirable nor will they have as extensive a social network of teachers and administrators who can lobby on their behalf. For schools, new teachers are virtually indistinguishable among each other. While I can 'look ahead' to observe a teacher's fixed effect, the principal who is hiring the teacher at year zero cannot 'see' this value. Given that over 99 % of new teachers do not have certification, the only observable differences among new teachers are gender and race.[22] This implies that a school is indifferent among all new teachers (unless the principal has a preference for a gender or a race). The summary statistics show that new teachers are more likely to match with schools that have more students in traditionally disadvantaged groups. Furthermore, within these schools they are assigned to classrooms that have a higher concentration of minority and low parental education students. New teachers seem to be outside options for schools and experienced teachers appear to have more bargaining power.[23]

[21] See Clotfelter, Ladd, and Vigdor (2006).

[22] See Ballou (1996) for further evidence that a new teacher is not recruited based on observable criteria, such as the selectivity of her undergraduate institution.

[23] Some principals may prefer new teachers over transfers that have shown themselves to be ineffective. Unfortunately, the matching model is unable to account for such preferences.

The average teacher who does and does not transfer are not drastically different, except for years of experience. New teachers start in classrooms and schools that have more traditionally disadvantaged children. These teachers then transfer to better environments later in their careers. New teachers' fixed effects are observable ex-post from subsequent years performances. Again, note that new teacher characteristics are tabulated to show that they are outside options and non-competitive with transferring teachers. They are not used in the analysis.

5 Empirical Specification

I assume that teacher's *gain* in utility is linear in the percent differences in 'amenities' offered by the origin and destination schools. Implicit in this setup is the notion that the destination school must be at least as desirable as the origin school. I allow distance, urbanicity, and classroom and school minority percent to affect a teacher's decision when transferring.[24]

The portion of the joint production that can reasonably be attributed to teacher utility is constructed in the following manner. Abusing notation from the theory, I define the school and class characteristics *vectors* at the origin and destination schools as s^o and s, respectively. These vectors are each weighted by a matching coefficient vector β.[25] Therefore, if a teacher a is offered positions at different schools, he or she is deciding (considering a total of $K+1$ school or class characteristics) between:

$$\tilde{u_a}(s*, s^o) = \sum_{k=1}^{K} \beta_k \frac{s*_k - s^o_k}{s^o_k} + \beta_{urb}(s*_{urb} - s^o_{urb}) \quad \forall s* \in S$$

$s*_{urb}$ is the NCES metro-centric locale variable, which runs from 1 (large city) to 8 (rural, inside MSA).[26] Therefore, a positive sign for the difference between the old and new school urbanicity indicates that the teacher moved to a more urban location. I take the natural log of miles between the origin school and possible destination schools as the measure of distance

[24]The distance measure should be the distance from the teacher's home to the destination school. Because the dataset does not contain this information, I use the address of the teacher's origin school. Alternatively, the distance measure could proxy as an information acquisition cost.

[25]Distance between schools is excluded from this vector.

[26]For the full category definitions, refer to nces.ed.gov/ccd/rurallocales.asp.

Table 1: Summary Statistics

	Stationary Teachers	(Mobile Teachers =	New Entrants	+ Transfers)	Transfers (2009)
FE Reading	0.0024 (0.1100)	-0.0028 (0.0968)	-0.0018 (0.0944)	-0.0047 (0.1010)	0.0189 (0.1165)
FE Mathematics	0.0056 (0.1718)	-0.0027 (0.1582)	-0.0015 (0.1559)	-0.0050 (0.1623)	0.0170 (0.1696)
Minority	0.2211 (0.4150)	0.2198 (0.4141)	0.2028 (0.4021)	0.2517 (0.4341)	0.1425 (0.3499)
Female	0.9252 (0.2630)	0.8845 (0.3196)	0.8728 (0.3333)	0.9066 (0.2911)	0.9035 (0.2955)
W/in district transfer				0.6808 (0.4663)	0.6162 (0.4868)
Move to Rural				-0.0088 (0.4128)	-0.0175 (0.4541)
Yrs. of Experience	15.1550 (12.00)	3.6539 (8.02)	0.0000 (0.00)	10.5092 (10.62)	11.8599 (10.06)
Certified	0.0436 (0.2042)	0.0211 (0.1439)	0.0043 (0.0656)	0.0527 (0.2236)	0.0723 (0.2593)
Cls Size	19.6430 (8.03)	18.5067 (8.45)	17.9590 (8.72)	19.5342 (7.82)	19.6383 (4.79)
Cls Minority %	0.4665 (0.2805)	0.5210 (0.2908)	0.5373 (0.2904)	0.4632 (0.2891)	0.3870 (0.2879)
Cls Female %	0.4650 (0.1639)	0.4495 (0.1879)	0.4422 (0.1973)	0.4632 (0.1679)	0.4988 (0.1154)
Cls Low Parent Ed. %	0.4707 (0.2957)	0.5148 (0.3139)	0.5335 (0.3164)	0.4707 (0.3061)	
Schl Minority %	0.4506 (0.2283)	0.4917 (0.2369)	0.5039 (0.2328)	0.4684 (0.2428)	0.3844 (0.2656)
Schl Female %	0.4896 (0.0337)	0.4854 (0.0371)	0.4896 (0.0368)	0.4895 (0.0375)	0.4993 (0.0416)
Schl Low Parent Ed. %	0.4492 (0.2028)	0.4740 (0.2127)	0.4862 (0.2104)	0.4510 (0.2151)	
Expected Bonus	787.67 (171.01)	761.68 (184.95)	750.06 (184.29)	783.49 (184.26)	
Observations	17,177	6,434	4,197	2,237	456

NCERDC data from 1998/99 to 2002/03 for first 4 columns. NCERDC data from 2009/10 for the last column. FE is fixed effect, calculated from an education production function (see Table 2). Low parental education is defined as high school degree or below, not available for 2009 data.

and set the parameter equal to -1.[27]

In this matching model, it is usually impossible to disentangle how much of the joint match utility should be assigned to the teacher or the school (principal). For the percent change characteristics described above, it seems logical to assign this portion of the joint production to teachers. The percentage change in characteristics from the origin to the destination school will impact the teacher's utility, but it is unlikely to impact the school's objective function. One possible way a principal *might* care is if he or she wants to attract a teacher who has taught previously at a similar school, perhaps because the expectation is that the teacher will be more self-sufficient. In this case, the teacher utility portion is indistinguishable from the principal's objective function.[28]

I use a teacher's certification status, experience, ethnicity, and math fixed-effect to distinguish among teachers. Rothstein (2008) showed that estimates of the fixed-effects may be biased due to nonrandom sorting of students into classroom. The fixed-effect estimate may not be the 'true' measure of teacher value-added. However, the fixed-effect is what is observable to the principal, and in this sense, it may be the more appropriate measure to use.

The parts of the joint production function that contains interaction terms between the school's status in the ABC system and characteristics of the teacher that may or may not drive education achievement (such as experience and certification status) are tentatively assigned to the principal. As shown in Ahn (2013), the incentive pressure in the NC system is non-monotonic in the school's academic performance. For very low probability of qualifying under the ABC standard (when the bonus is essentially unattainable and the school will be labeled as under-performing) and for very high probability (when the bonus is all-but-assured and the school will be labeled as high-achieving), principals may not have a strong incentive to recruit highly effective teachers. It is only in the middle of the probability CDF, when the bonus outcome is in doubt, that a principal may look to match with candidates to maximize

[27]SMSE framework requires one of the parameters to be normalized at $+1$ or -1. Setting the distance parameter to -1 maximized the number of satisfied inequalities.

[28]Some studies have shown that students paired with a teacher of the same race seem to perform better. See Dee (2004). If pairing of teacher to students have differential impact (based on race, say), and the principal's utility is maximized by matching the 'effective' teacher with an 'undesirable' class, a principal will have to weigh the utility gain from such a match against the decreased ability to lure the effective teacher to her school. Unfortunately, the matching model is unable to discern between these motives.

academic outcome, because hiring one or more effective teachers may be the difference in being recognized as a good school and attaining the bonus. Because of this non-monotonicity, there is no obvious connection between school quality (such as non-monetary amenities and student body quality) and accountability pressure. Abusing notation from above, I define the *vector* of L total characteristics of the teacher that are relevant to maximizing the school objective function as a and the vector of characteristics of an alternative teacher as a'. The portion of the joint production that may be attributable to the school side is defined as:

$$\tilde{g_s}(a*) = \sum_{l=1}^{L} \delta_l P_s \cdot a *_l \quad \forall a* \in A$$

While accountability pressure may spur principals to look for efficient teachers, the absolute academic performance of the school may also drive matching results. A school that has a highly regarded reading program for example, may itself be a draw for teachers, and schools that have such high absolute performance may match with effective teachers purely due to their characteristics. This portion of utility is truly joint, and it is impossible to attribute the production exclusively to one side of the match. If this joint production aspect is *not* accounted for, and if academic achievement and accountability pressure are strongly correlated, the matching parameter may not be consistent with the preferences of the principal and teacher, much akin to an omitted variable bias problem. To account for this possibility, I include an interaction term between the level achievement of the school and teacher characteristics. If a school s that has achievement level A_s under accountability pressure level P_s and a teacher a is evaluating the gain in utility from the match, the teacher's total utility and principal's total utility are, respectively:

$$\begin{aligned} u_a(s) &= \tilde{u_a}(s, s^o) + \sum_{l=1}^{L} \theta_l A_s \cdot a_l \\ g_s(a) &= \sum_{l=1}^{L} (\delta_l P_s \cdot a_l + (1 - \theta_l) A_s \cdot a_l) \end{aligned}$$

Because the two utilities are summed, θ_l is unidentified. An alternative interpretation of A_s and P_s is the difference between a growth criterion and a level criterion. A school s can be under different levels of pressure with different accountability systems. In North

Carolina, schools are categorized by growth and achievement,[29] and a school can be considered performing well under one system while being deficient in another. I use the pdf of the likelihood of a school to make expected growth, according to the ABC system as the measure of accountability pressure (P_s),[30] and the percentile ordering of schools according to school-wide performance as the measure of achievement (A_s).[31]

Following Fox (2009) and Horowitz (1992, 2002), I estimate coefficients of the matching model using a smooth maximum score estimation (SMSE) procedure:[32]

$$\max_{\beta,\delta,\gamma,h} S(\beta,\delta,\gamma,h) = \frac{1}{T \cdot \sum_{t=1}^{T} N_d} \sum_{t=1}^{T} \sum_{n=1}^{N_t} (2 \cdot 1(v_{nt} > 0) - 1) K\left(\frac{v_{nt}}{h}\right)$$

where $(2 \cdot 1(v_{nt} > 0) - 1)$ term is an indicator function that equals 1 when $v_{nt} > 0$ where:

$$v_{nt} = f(s, s_o, a) - f(s, s_o, a') + f(s', s'_o, a') - f(s', s'_o, a) > 0$$

and where $K(\cdot)$ is a continuous function in the real line such that:

$$\lim_{v \to -\infty} K(v) = 0 \text{ and } \lim_{v \to \infty} K(v) = 1.$$

The $K(\cdot)$ function is:

$$K(v) = \begin{cases} 0 & \text{if } v < -1 \\ 0.5 + (105/64)((v/5) - (5/3)(v/5)^3 + (7/5)(v/5)^5 - (3/7)(v/5)^7) & \text{if } -1 \leq v \leq 1 \\ 1 & \text{if } z > 1 \end{cases}$$

The K function is analogous to a kernel in non-parametric regressions, except that K behaves like a distribution. The indicator function (with some coefficients $\tilde{\beta}$, $\tilde{\delta}$, $\tilde{\gamma}$, and $\tilde{h}$)) equals 1

[29] The analysis sample runs through the 2002/03 academic year, the inaugural year for NCLB. Because schools are not sanctioned until their second consecutive failure to make adequate yearly progress under NCLB, no accountability sanctions are associated with deficient achievement, although a school can be publicly labeled as failing. To test whether the 2002/03 sample teachers and schools face differing incentives due to NCLB, the matching model is estimated with this year excluded. Qualitative results remain largely unchanged. See online appendix.

[30] I convert the pdf value that each school has (since ABC status is assigned at the school level) to a percentile ordering of schools by accountability pressure. Therefore, the school under the greatest amount of pressure (peak of pdf) is assigned a 1, and schools that are under the least amount of pressure (*both* tails of the pdf) are assigned a 0. See Ahn(2013) for complete details.

[31] School-wide achievement is defined as the grade-level proficiency rate across both math and reading.

[32] An alternative maximum score estimation (MSE) following Fox (2009) directly was also attempted, which yielded qualitatively similar results. For the MSE, the estimator would lose the kernel $K(\cdot)$.

when the actual match ($\{a, s\}$ and $\{a', s'\}$) yields higher benefits than the proposed alternative match ($\{a, s'\}$ and $\{a', s\}$). Standard errors are generated using a sandwich estimator-like asymptotic covariance matrix using first and second-derivative matrices of the S objective function. The bandwidth, h is chosen by a method analogous to the plug-in method in kernel density estimation.

The market is defined as the entire state in each year t. As shown in Fox (2009), I do not need to generate counter-factuals for the entire sample. Since there are over 2,000 teachers who transfer, generating all counter-factuals would create over 4 million inequalities (for each parameter vector guess). Instead, I sample 10 percent of the observations in each year as possible counter-factuals to evaluate the SMSE.[33] In the choice of the counter-factual (a' and s'), I note that for a 'market' (year labeled t) with N_t matched teachers, there are $N_t - 1$ potential counter-factuals teachers (each tied to a particular school s').

6 Econometric Results

Table 2 presents the results of the SMSE estimation.[34] The top portion, labeled 'Teacher-side,' represents matching preferences that could reasonably be attributed to teachers. To generate fixed effect estimates, an education production function was estimated. A one standard deviation increase in teacher fixed effect for reading (mathematics) results in approximately 0.17 (0.24) standard deviation increase in test scores. A school motivated to raise academic achievement by hiring more 'productive' teachers should select, all else equal, a teacher with a higher fixed effect value.[35]

The parameter on urbanicity is weakly negative, indicating that moving to a more urban location is not an important consideration in teacher transfers. The statistical insignificance on the parameter implies that teachers are not avoiding rural or urban schools due to geography, but monetary or non-monetary characteristics of the position.

Class-level and school-level minority percent affects teacher transfer outcomes for white

[33] Restricting the sample in this way decreases the number of inequalities to be evaluated to under 20,000 (for each guess at the parameter vector), making estimation feasible.

[34] The SMSE estimation of the matching function is duplicated at half and double the 'ideal' bandwidth. The results are qualitatively unchanged. Tables are presented in the online appendix.

[35] See online appendix for complete results.

and minority teachers in different ways. White teachers often move to schools that have a lower proportion of minority students compared to their origin schools. The parameter on classroom minority percent is more negative compared to the parameter on school minority percent, implying that the composition of the teacher's classroom is more important. Somewhat surprisingly, the parameter on difference in achievement levels between origin and destination schools is relatively weak and negative. 'Moving up' academically does not seem to be a strong driver in matching. Therefore, the negative parameter on minority student percent for white teachers is most likely attributable to differences in work environment or other unobservable factors that are correlated with the student demographic makeup.[36]

Minority teachers move toward schools with a *higher* proportion of minority students. While the pattern of moving toward schools with less traditionally disadvantaged students is as expected for white teachers, the opposite observed for minority teachers is not. There are two hypotheses for these patterns across teacher ethnicity. The first is that teachers prefer to move to schools more heavily populated with students that 'look' similar to them. Teachers may prefer to teach students with whom they can more readily identify or to live in and contribute to neighborhoods that mirror their own background and ethnicity. Alternatively, the pattern could be due to accountability pressure. Minority-heavy schools may actively court minority teachers who have demonstrated the ability to 'reach' and motivate minority students.[37] In essence, these school may be induced into seeking out complementary productive capacity in teachers due to incentive pressure. As we will see in joint portion of utility as well as the 'control' sample analysis, the latter hypothesis is more consistent with the evidence.

The bottom portion of the table represents joint match utility based on interaction of teacher characteristics with school level of accountability pressure and level achievement. If all schools were equally motivated to raise education production, every school would have a cardinal preference ordering over teachers, and teachers would then choose their most preferred

[36]Estimation restricting the sample to only intra-district transfers yields qualitatively similar results. There is some corroborating evidence for this result in Ahn (2015). In that study I find that both low- and high-achieving schools have a preference for matching with teachers from high-achieving schools. However, low-achieving schools tend to be less successful luring these teachers. High-achieving schools, on the other hand, match more often with teachers originating from comparable schools. Then, what is observed here may be the expression of strong preferences from the school-side, desiring to match with teachers from better schools.

[37]See Jackson (2009). Schools with few minority students may court white teachers for similar reasons.

school in order. That is, the teacher that yields the highest academic achievement gain would select the closest school that offered the best teaching environment, and the teacher with the next highest value would choose the next best school, etc. Analysis of the joint portion of the matching function reveals that accountability pressure plays an important role in matching certain teacher characteristics to certain schools.

Ignoring the level achievement terms for the moment, the matching coefficients on the interaction between ABC pressure and teacher characteristics can be interpreted in the following way. For illustration, assume there exists two levels of accountability pressure, $\overline{P}$ (high) and $\underline{P}$ (low), and two levels of some teacher characteristic a_k, $\overline{a_k}$ (high) and $\underline{a_k}$ (low).

If schools under high pressure match with teachers with high a_k values, we will observe combinations of ($\overline{P}$, $\overline{a_k}$) and ($\underline{P}$, $\underline{a_k}$) more often, as opposed to ($\overline{P}$, $\underline{a_k}$) and ($\underline{P}$, $\overline{a_k}$). Then:

$$\delta(\overline{P} \cdot \overline{a_k} + \underline{P} \cdot \underline{a_k}) > \delta(\overline{P} \cdot \underline{a_k} + \underline{P} \cdot \overline{a_k}) \text{ iff } \delta > 0$$

In fact, the higher the association between $\overline{P}$ and $\overline{a_k}$, the greater δ should be. In contrast, if $\overline{P}$ schools actually match more often with $\underline{a_k}$ teachers, then:

$$\delta(\overline{P} \cdot \overline{a_k} + \underline{P} \cdot \underline{a_k}) < \delta(\overline{P} \cdot \underline{a_k} + \underline{P} \cdot \overline{a_k}) \text{ iff } \delta < 0$$

This time, the stronger the association between $\overline{P}$ and $\underline{a_k}$, the more negative δ should be. If there is no relationship between P and a_k:

$$\delta(\overline{P} \cdot \overline{a_k} + \underline{P} \cdot \underline{a_k}) = \delta(\overline{P} \cdot \underline{a_k} + \underline{P} \cdot \overline{a_k}) \text{ iff } \delta = 0 \text{ or } \overline{P} = \underline{P}$$

Further, for a vector of normalized teacher characteristics a_1, a_2, ... , a_L (say between 0 and 1) interacted with the accountability pressure P, the corresponding δ_1, δ_2, ... , δ_L parameters will also be scaled such that they are directly comparable to determine how much more (or less) a particular characteristic is desirable compared to another characteristic.

If schools with high levels of achievement (A_s) but low levels of P_s are most often associated with a particular characteristic a_k (that is, the parameter on $A_s \cdot a_k$ is greater than that on $P_s \cdot a_k$), a teacher with a high a_k value is more likely to match with a school with a high level achievement as opposed to a school under ABC accountability pressure.

The 'triple' interaction among ABC pressure, level achievement, and teacher characteristics pays special attention to schools that are doing well in absolute terms that are also

under accountability pressure, compared to schools doing poorly in level terms and facing low accountability pressure.

The econometric results show that experience, fixed effect values, and certification status are all positive predictors of matching with schools under accountability pressure. While teachers with high experience and fixed effects are equally likely to match with pressured schools, certified teachers are only approximately 60 percent as attractive. As certification is often discussed in the literature as being relatively ineffective in increasing test scores, this may be indicative of principals prioritizing stronger predictors of educational production. Alternatively, certified teachers may be averse to matching with pressured schools. It is especially impressive that the parameter on matching fixed effects is so substantial. As fixed effect (or even raw year over year average test score gains) is difficult to observe, principals seem to be putting in the effort to identify and match with these teachers.

Like pressured schools, schools with high academic achievement positively match with teacher who have high experience and fixed effects, as well as certification. However, unlike pressured schools, the parameters on the three characteristics are similar in size. The parameter on certification is about 80 percent of the parameter on fixed effect. Schools with high achievement do not have to compromise in selecting teachers with desirable characteristics.[38]

The difference in magnitudes of the matching parameters of accountability pressure and level achievement for experience and fixed effects seems to show that accountability is a modest force in driving matching outcomes. While accountability legislation in North Carolina may be incentivizing principals correctly, in absolute magnitudes, high achieving schools are much more likely to match with teachers with desirable characteristics. Schools under pressure may not have resources available to them to attract their most desired targets when in competition with schools with high achievement.

How pressured schools compete at all against high achieving schools becomes clear with the relative differences in the parameters for certification and minority. Pressured schools are less likely to match with certified teachers. While ratio of parameters for experience and fixed effects for pressured to high achieving school is approximately 0.67 and 0.57, respectively, the

[38]That certification status is such a strong matching component is puzzling, since the literature has shown that it does not strongly impact achievement. Preference for these teachers may be reflective of the fact that the percentage of certified teachers is a publicly-released number, as part of the school's annual report card.

Table 2: Estimates of Matching Parameters[†]

Variable	Point Estimate	95% CI
Teacher-side		
Δ distance %	-1	super-consistent
Δ urbanicity	-4.1645	(-9.9742 , 1.6452)
Δ classroom minority %	-16.0079	(-21.7881 , -10.2277)
Δ school minority %	-12.4768	(-14.3188 , -10.6348)
Δ classroom minority % X teacher minority	20.0798	(13.8278 , 26.3318)
Δ school minority % X teacher minority	22.2767	(20.4347 , 24.1187)
Δ level achievement	-3.785	(-6.9693 , -0.6007)
Level achievement X experience	24.1429	(20.7237 , 27.5621)
Level achievement X math FE	27.6966	(24.3399 , 31.0533)
Level achievement X certification	21.8939	(18.7164 , 25.0714)
Level achievement X teacher minority	-18.5917	(-21.8512 , -15.3322)
ABC pressure X experience	15.7851	(11.9718 , 19.5984)
ABC pressure X math FE	15.8741	(11.2621 , 20.4861)
ABC pressure X certification	9.5156	(6.4938 , 12.5374)
ABC pressure X teacher minority	11.7704	(7.4591 , 16.0817)
Level X ABC X experience	20.6388	(17.5234 , 23.7542)
Level X ABC X math FE	19.9136	(16.6636 , 23.1636)
Level X ABC X certification	9.8859	(6.7801 , 12.9917)
Level X ABC X teacher minority	-17.9003	(-21.0763 , -14.7243)
h (bandwidth)	1.127	

[†]Experience and fixed effects are converted to percentile values. Level achievement is average math proficiency rate at the school-level.

ratio of parameters for certification is about 0.43. Most strikingly, while the high achieving schools are likely to match with white teachers, pressured schools are more likely to match with minority teachers. Then, the strategy for pressured schools becomes clear: these schools match most often with minority teachers without certification, who nonetheless have the experience and fixed effects necessary to help the school qualify for the bonus. In this sense, these schools display a significant amount of sophistication (and effort) in matching with teachers who have high academic impact.

The triple interaction term shows that when high achieving schools are also pressured, their matches prioritize experience and fixed effects over certification. Parameters for experience and fixed effects are double in size compared to the parameter for certification. Interestingly, the parameter on minority remains negative and large. High achieving, pressured schools, which on average have a higher fraction of the student body that is white, may be courting white teachers because of the complementarity in education production. The positive parameter on minority for pressured schools in general and the negative parameter for minority specifically for pressured schools that are high achieving supports the hypothesis that teacher ethnicity sorting may be driven by accountability pressure.

Although the model is useful in its ability to elucidate potential strategic matching behavior by teachers and schools, there are limitations imposed by structure. It is a possibility that teacher transfer policies are made at the district level, with teachers and principals having little role in the process (other than requesting a transfer and posting a vacancy). District superintendents may be able to compel schools to accept particular candidates or veto their hire. Within the context of the matching model, this is a difficult issue to address. Because over 60 percent of transfers happen within the district, most job characteristic differences (from the current district to the future district) that define the teacher's marginal utility gain from transfer would be zero. In addition, the district objective function is unclear, since shifting teachers from one school to another within the same district does not decrease overall achievement (although it may ameliorate or exacerbate inequality across schools).[39]

[39]In 2003/04, the superintendent of Charlotte Mecklenburg district (CMS) instituted a policy forbidding intra-district transfers of teachers into highly desirable schools (as defined by academic achievement and other measures). The plan was for this list to be updated yearly. The design was clearly intended to stem the bleeding of experienced teachers from bad to good schools within the district. However, just one year later, the

Alternatively, the abundance of intra-district transfers, may indicate that there exists formalized policies that prioritize these hires over outside transfers. As far as I was able to learn, there are no explicit rules in place that force principals to consider intra-district transfers ahead of inter-district (or out-of-state) transfers. However, principals have confided that they tend to accept more intra-district teacher transfers because they have better information on them. I interpreted this to mean that distance (or more precisely, social connections) were driving at least some of these transfers. To explore further, I ran a version of the model with the sample restricted to intra-district transfers, where if there is any competition for teachers, it is occurring at the school level. Results are available in the online appendix. Caution should be exercised in interpreting these results, as restricting the sample in this way is equivalent to saying that new teachers *and* experienced teachers from outside the district are both outside options. More troublesome is the implied assumption that teachers would consider inter-district matches as equivalent to the outside option. Because the decision to search nor the impact of failure to match is not modeled in this paper, it is difficult to reconcile over 30 percent of successfully matching teachers 'choosing' to match with an outside option, instead of opting to remain at the current school. Results remained qualitatively similar to the full sample estimates, with the exception of smaller estimations on the interaction of ABC pressure and teacher characteristics. This may be due to the fact that there is relatively little difference in ABC pressure among schools within a given district.

Finally, teachers may move from (to) a high-stakes grade where EOG exams are administered, to (from) a low-stakes grade after a transfer. This has the potential to affect the matching results. From the teacher-side, this may not be a significant issue, as teachers, whether they are in a tested or non-tested grade, are rewarded as a group in the bonus system. From the school-side, however, placing a teacher in a high-stakes grade, as opposed to 'hiding' the teacher in a low-stakes grade, may impact the probability of bonus receipt. For this study, the sample is restricted to teachers that are associated with test scores before and after the transfer. There is a small fraction of teachers who move to different-level stakes grades from one year to the next. Demographic comparisons show that they are broadly similar to teachers who remain in high-stakes grades. However, one distinct difference may indicate that these teachers comprise a separate market. Among those that transition to dif-

superintendent resigned, and the policy was immediately scrapped by his successor.

ferent level of stakes (regardless of transfer status), there are many more teachers that are newly minted or out-of-state transfers in the current year, who cannot reveal (or themselves do not know) their fixed effect impact. This may be showing queuing behavior by teachers or the principal learning about the quality of the teacher in the initial year, and moving the teacher to a better-suited position in the subsequent year. See the online appendix.

7 Matching in the Absence of Pressure

The econometric results showed that schools under accountability pressure match with teachers who possess the ability to increase standardized test scores. Schools where the bonus outcome is in doubt are motivated to seek out effective teachers, as the marginal teacher's performance may determine whether the school receives the bonus.

While we may speculate that schools not under accountability pressure, whether because they are assured of the bonus or are completely out of the running, would not seek to pursue these effective teachers (and thus remain unaffected by accountability), it remains unclear what matching across the entire distribution of schools would look like in the absence of the bonus program. This is important for policy, as the introduction of market pressure impacts strategies of schools under pressure, which in turn may impact strategies (and outcomes) of other schools not under pressure.

Normally, a 'pre-policy' sample would be used to compare differences in parameter estimates across the two regimes. Lack of quality data before the bonus system implementation precludes this possibility. However, North Carolina completely discontinued its bonus system in 2009, while keeping the standardized tests that were used to generate the score to determine bonus receipt in place. Thus, it is possible to use a 'post-post-policy' sample to analyze changes in matching behavior of teachers and schools.[40] The last column in Table 1 presents summary statistics for the 2009 sample, and Table 3 presents the matching estimation results.

Interpreting the results should be approached with caution, as several years separate the

[40]In 2008, the state unexpectedly cut bonus payouts by 30 percent. The next year, the program was discontinued. I do not use the 2007/08 or 2008/09 samples as transfers would have occurred under uncertainty about the bonus system. By 2009, the writing was on the wall, and it was clear there would be no more bonuses. At the same time, because upwards of $100 million per year were saved by discontinuing the program, districts did not have to cut positions during the beginning of the Great Recession, keeping the transfer market stable.

'treatment' and 'control' samples. The bonus system was discontinued due to building financial pressure from the Great Recession. Thus, the economic environment faced by teachers were different, which may have impacted the decision to seek transfers. In addition, by 2009, schools had accumulated various sanctions under the NCLB regime, with some schools facing restructuring (leadership turnover) unless they were able to sharply increase proficiency rates of their student body. Teachers may have shunned these schools undergoing turmoil.

Overall transferring teachers are similar, compared to the 1989/99 to 2002/03 sample. The two key differences are that fewer teachers of minority status move to new schools, and when any teacher does transfer, they move to schools and classrooms that have proportionally fewer minority students. Teachers now appear to seek out schools with more traditionally advantaged students.

While absolute magnitudes differ from Table 2, the signs and relative magnitudes for class and school minority percent is similar; however, the interaction effect between minority percent and teacher minority is very different from the 'treatment' years. The parameters on the interactions terms are now close to zero. The positive matching between teacher and school/class ethnicity that was observed when the accountability system was in place has largely disappeared. This provides further support to the hypothesis that minority-heavy schools were previously courting minority teachers for accountability reasons.

Schools with high achievement now focus more on experience and fixed effects, with the parameter on experience and fixed effects about twice the size of the parameter on certification. These relative parameter sizes differences are similar to the triple interaction results. Once the accountability pressure from the bonus system disappears, in theory, the differences between the level interaction and triple interaction terms should disappear. That the differences are not completely gone may be indicate that principals still care about the school's performance, perhaps because these results still mattered for NCLB and the school report card.

The largest change in the matching function estimates is the 'hypothetical' accountability pressure interaction terms.[41] While these schools still match positively with experienced teachers, the parameter is only about 1/3 the size of the parameter for high achieving schools. Recall that these schools were much more competitive with high achieving schools when ac-

[41] Again, note that accountability pressure from the bonus system no longer exists. The hypothetical pressure that schools would have been under is calculated with the standardized EOG test scores.

Table 3: Estimates of Matching Parameters (Control)[†]

Variable	Point Estimate	95% CI
Teacher-side		
Δ distance %	-1	super-consistent
Δ urbanicity	0.1356	(-4.1356 , 4.4068)
Δ classroom minority %	-24.3885	(-30.4305 , -18.3465)
Δ school minority %	-23.8055	(-29.6121 , -17.9989)
Δ classroom minority % X teacher minority	-0.3647	(-2.0261 , 1.2967)
Δ school minority % X teacher minority	4.3775	(2.4769 , 6.2781)
Δ level achievement	-7.8197	(-13.1034 , -2.536)
Level achievement X experience	24.937	(21.7359 , 28.1381)
Level achievement X math FE	25.3869	(21.814 , 28.9598)
Level achievement X certification	12.3977	(9.0765 , 15.7189)
Level achievement X teacher minority	-7.126	(-10.6009 , -3.6511)
ABC pressure X experience	8.0599	(5.4071 , 10.7127)
ABC pressure X math FE	-5.8735	(-8.3833 , -3.3637)
ABC pressure X certification	-0.1077	(-2.6142 , 2.3988)
ABC pressure X teacher minority	3.184	(0.3469 , 6.0211)
Level X ABC X experience	24.3723	(21.9194 , 26.8252)
Level X ABC X math FE	23.9108	(21.505 , 26.3166)
Level X ABC X certification	9.0297	(6.6493 , 11.4101)
Level X ABC X teacher minority	-12.5573	(-15.225 , -9.8896)
h (bandwidth)	1.127	

[†]Experience and fixed effects are converted to percentile values. Level achievement is average math proficiency rate at the school-level.

countability pressure existed. Certification is uncorrelated with hypothetical accountability pressure, and most strikingly, fixed effect is *negatively* associated with 'hypothetical' accountability pressure. Because it takes effort to identify teachers with high fixed effects, in the absence of market pressure, the 'haves' (students at high achieving schools) end up with better teachers, as schools educating the 'have-nots' stop searching and competing for effective teachers.

8 Conclusion

Using a SMSE framework and the North Carolina administrative education panel dataset, this study analyzed the teacher transfer market, with particular attention paid to the role of accountability pressures in affecting what types of teachers and schools match. The econometric results showed that white (minority) teachers transfer to schools with a higher proportion of white (minority) students. This sorting seems to be due to accountability pressure leading schools to seek out complementary education production by matching teachers to student population who 'look more like them.'

The joint portion of the estimates of the match production reveals that schools under pressure are more likely to match with teachers that have high experience and fixed effects. The size of the parameter on fixed effect is particularly encouraging, as it takes some amount of effort to identify teachers with high fixed effects. Accountability pressure may induce principals to seek out teachers that will have a strong impact on test scores or to change the school environment to emphasize testing. Teachers who dislike testing may try to match with schools that are not under accountability pressure, and teachers who excel at raising test scores may seek out schools where their strengths will be more appreciated.[42]

Estimates on achievement showed that experience, fixed effects, and certification status matter for recruiting. Comparing the magnitudes of the parameters between accountability pressure and achievement, high achieving schools are more likely to match with teachers with desirable characteristics, compared to pressured schools. While accountability pressure is more associated with teacher characteristics that produce education output, schools under

[42]See Ahn and Vigdor (2016) to see how principals change the school environment in response to the North Carolina incentive system.

accountability pressure often lose out to schools with high achievement (and low pressure) for these teachers. While high achieving schools seem to have the edge in matching with teachers with high experience, fixed effects, and certification, pressured schools relinquish matching with certified teachers to pursue minority teachers with experience and high fixed effect.

The same matching estimation done with data on a year during which the accountability system was discontinued shows that in the absence of market pressure, schools that would have been under accountability pressure cease to aggressively compete for teachers with high fixed effects. In addition, the relatively sophisticated strategy of matching with minority teachers who have the characteristics that drive education production is largely abandoned. As a result, high achieving schools are more likely to find themselves matched with experienced teachers with higher fixed effects. Ultimately, the results provide evidence that market-based reform can have a positive impact on the decision making processes of schools, but more resources to attract teachers must be provided for the policy to be effective.

References

[1] Ahn, T. (2013) "The Missing Link: Estimating the Impact of Incentives on Effort and Effort on Production Using Teacher Accountability Legislation" Journal of Human Capital, Vol. 7(3), pp. 230-273.

[2] Ahn, T. (2015) "Matching Strategies of Teachers and Schools in General Equilibrium." IZA Journal of Labor Economics, Vol. 4: 5.

[3] Ahn, T. and J. Vigdor (2014) "When Incentives Matter Too Much: Explaining Significant Responses to Irrelevant Information." NBER Working Paper 20321.

[4] Ahn, T. and J. Vigdor (2016) "Opening the Black Box: Behavioral Responses of Teachers and Principals to Pay-for-Performance Incentive Programs." Working Paper.

[5] Boyd, D., Lankford, H., Loeb, S., Wyckoff, J. (2008). "The Impact of Assessment and Accountability on Teacher Recruitment and Retention Are There Unintended Consequences?" Public Finance Review, 36(1), 88-111.

[6] Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J. (2013). "Analyzing Determinants of the Matching of Public School Teachers to Jobs: Estimating Compensating Differentials in Imperfect Labor Markets." Journal of Labor Economics, Vol. 31(1), pp. 83-117.

[7] Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J. (2011). "The Role of Teacher Quality in Retention and Hiring: Using Applications-to-Transfer to Uncover Preferences of Teachers and Schools." Journal of Policy Analysis and Management, Vol. 31(1), pp. 88-110.

[8] Clotfelter, C., Ladd, H., Vigdor, J., Diaz, R. (2004). "Do school accountability systems make it more difficult for lowperforming schools to attract and retain highquality teachers?" Journal of Policy Analysis and Management, 23(2), pp. 251-271.

[9] Clotfelter, C., H. Ladd, J. Vigdor (2006) "Teacher-Student Matching and the Assessment of Teacher Effectiveness."Journal of Human Resources, vol. 41(4), pp 778-820.

[10] Clotfelter, C., H. Ladd, J. Vigdor (2007) "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects." NBER Working Paper 13617.

[11] Dee , T. (2004) "Teachers, Race and Student Achievement in a Randomized Experiment." Review of Economics and Statistics, vol. 86(1), pp. 195-210.

[12] Figlio, D., L. Getzler (2002) "Accountability, Ability, and Disability: Gaming the System?" NBER Working Paper 9307.

[13] Figlio, D. (2006) "Testing, Crime, and Punishment." Journal of Public Economics 90: 837-851.

[14] Figlio, D., L. Kenny (2007) "Individual teacher incentives and student performance," Journal of Public Economics, Elsevier, vol. 91(5-6), pp 901-914.

[15] Fox, J. (2009) "Estimating Matching Games with Transfers" Working Paper

[16] Hanushek E., M. Raymond (2005) "Does School Accountability Lead to Student Improvements?" Journal of Policy Analysis and Management, 24, pp 297-327.

[17] Horowitz, J. (1992) "A Smoothed Maximum Score Estimator for the Binary Response Model" Econometrica v. 60 pp. 505-531

[18] Horowitz, J. (2002) "Bootstrap critical values for tests based on the smoothed maximum score estimator" Journal of Econometrics v. 111 pp. 141-167

[19] Jacob B., S. Levitt (2003) "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." Quarterly Journal of Economics, 118(3), pp. 843-77.

[20] Jacob B., (2005) "Accountability, Incentives and Behavior: Evidence from School Reform in Chicago." Journal of Public Economics 89, pp. 761-796.

[21] Jacob B., (2007) "Test-Based Accountability and Student Achievement: An Investigation of Differential Performance on NAEP and State Assessments." NBER Working Paper 12817.

[22] Jackson C., (2009) "Student Demographics, Teacher Sorting, and Teacher Quality: Evidence from the End of School Desegregation." Journal of Labor Economics, Vol. 27, No. 2, pp. 213-256.

[23] Kalogrides, D., Loeb, S., and T. Beteille (2013) "Systematic Sorting: Teacher Characteristics and Class Assignments" Sociology of Education, Vol. 86(2), pp. 103-123.

[24] Kelso, A. and V. Crawford. (1982) "Job matching coalition formation and gross substitutes." Econometrica 50, pp. 1483-1504.

[25] Lavy, V. (2002) "Evaluating the Effect of Teachers Group Performance Incentives on Pupil Achievement." Journal of Political Economy, 110, 1286-1317.

[26] Rivkin S., Hanushek, E., and J. Kain (2005) "Teachers, Schools, and Academic Achievement." Econometrica (73), 417-458.

[27] Rockoff, J. (2004) "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." American Economic Review 94(2), 247-52.

[28] Rothstein J. (2008) "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." Quarterly Journal of Economics, Forthcoming.

[29] Sanders, W. and J. Rivers (1996) "Cumulative and residual effects of teachers on future student academic achievement." Research Progress Report. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

[30] Vigdor, J. (2008) "Teacher Salary Bonuses in North Carolina." National Center for Performance Incentives Working Paper 2008-03.