Task-Specific Experience and Task-Specific Talent

Jason Cook — Cornell University
Richard Mansfield — Cornell University

PRELIMINARY AND INCOMPLETE

DO NOT CITE WITHOUT AUTHORS' PERMISSION

October 21, 2013

Abstract

In this paper, we use administrative panel data to decompose worker performance into components relating to general talent, task-specific talent, general experience, and task-specific experience. We consider the context of high school teachers, in which tasks consist of teaching particular subjects in particular tracks. Using the timing of changes in the subjects and levels to which teachers are assigned to provide identifying variation, we show that much of the productivity gains to teacher experience estimated in the literature are actually subject-specific. By contrast, very little of the variation in the permanent component of productivity among teachers is subject-specific or level-specific. *JEL Codes:* I24, I21, J45, J62.

1 Introduction

A worker's productivity at a particular job is generally assumed to depend on both the worker's innate talent at performing tasks the job requires as well as experience the worker has at performing each task (learning by doing). Knowledge of the relative importance of task-specific talent versus task-specific experience is essential for employers to maximize the productivity of their workforce. For tasks with larger potential experience gains and smaller variance in task-specific innate talent, the key to a productive workforce is employee retention: the optimal strategy is to keep employees of all talent levels at their originally assigned tasks to benefit from experience. Conversely, for tasks yielding smaller experience gains with a larger variance in task-specific talent, the optimal strategy is to fire or reassign low performing workers in an attempt to either improve general worker skill or identify superior worker-task matches.

In this paper, we use administrative panel data to decompose worker performance into components relating to general talent, task-specific talent, general experience, and task-specific experience. We consider the context of high school teachers, in which tasks consist of teaching particular subjects in particular tracks. Myriad papers have estimated education production functions featuring both teacher fixed effects and a common experience profile. The bulk of the evidence suggests that the standard deviation of permanent teacher quality is between .1 and .2 standard deviations at either the primary [1] or secondary school level[2], while teachers tend to improve with experience by around .05 test score standard deviations in their first year, another .03 to .05 over the next couple of years, and another .03 to .05 over the next several years, with the profile for mid-career teachers flattening out at between .1 and .2 standard deviations better than a novice teacher.[3]

However, this literature has generally ignored the possibility that the permanent effectiveness of a teacher and/or the gains to teaching experience might be specific to a particular classroom envi-

[1] e.g. Rockoff (2004), Hanushek *et al.* (2005), Harris and Sass (2006), Jackson (2013)

[2] e.g. Mansfield (2013), Aaronson, Barrow, and Sander (2007), Jackson (forthcoming).

[3] e.g. Rivkin *et al.* (2005), Clotfelter, Ladd, and Vigdor (2007). Wiswall (2011) finds somewhat larger returns to late-career teaching.

ronment. In such a context, models that impose homogeneity of skill across different classroom environments will return a weighted average of teacher skill across the environments each teacher actually faced (weighted by the fraction of time spent in each environment). To the extent that the teacher has faced different classroom contexts during his/her tenure, models that impose homogeneity of returns to experience across different classroom environments may underestimate the gains to context-specific experience. Similarly, such models may overestimate the returns to general experience to the extent that the teacher's classroom environment has remained somewhat stable during his/her tenure.

Of course, the distinction between context-specific and general skill or experience may not be critical if most teachers spend the bulk of their careers in a single context. Indeed, at the elementary and middle school levels, curriculum and classroom composition may change very little from year to year, particularly for teachers who remain in the same school. Even in this context, however, Ben Ost (2011) shows that teachers who always repeat elementary grade assignments improve 35% faster than teachers who never repeat grade assignments.

At the high school level, however, teachers are routinely asked to teach courses in different subjects and in different difficulty tracks. Indeed, teacher certification in most states is at the level of the field (math, science, history, etc.) rather than the subject (Biology, Chemistry, Physics), and is not specific to a level of difficulty (special education excepted). If teaching skill is specific to the subject or level, then such changes in teaching assignments may have important implications for student achievement.

On one hand, suppose teachers have innate or pre-determined comparative advantages for particular subjects or difficulty levels. Then mutually advantageous swaps among teachers could produce efficiency gains if both teachers move toward their relatively more effective subjects or levels. More generally, observing a teacher in a number of different classroom contexts early in their career might produce valuable information about his/her relative teaching strengths. Permanent subject-specific skill might exist, for example, if a teacher's undergraduate major was in a particular subject (say Physics rather than Biology). Permanent level-specific skill might exist, for example, if a teacher has strong classroom control skills due to natural charisma or sense of humor, which may be comparatively more important in remedial or basic level courses, where students may tend to be less engaged.

On the other hand, suppose task-specific skill is primarily learned through experience rather than being predetermined at the time of hire. Then rotating the classroom environments to which teachers are assigned will waste a component of each teacher's skill, and slow each teacher's progress toward his/her full potential. Subject-specific experience might be important, for example, if a teacher's knowledge of the subject content deepens over time. Level-specific experience might also be significant if methods for maintaining student attention and enthusiasm depend on the level of student ability, or if the appropriate pace at which to deliver content depends on student skill and is slowly calibrated over time. In addition, experience teaching a particular subject-level combination (e.g. honors biology) might be particularly valuable, if it allows teachers to hone particular lectures over time.

Thus, in this paper, we use administrative panel data from North Carolina to decompose effectiveness into (1) a set of permanent components capturing general talent, course-specific talent, level-specific talent, and course-level specific talent, and (2) a set of functions capturing returns from general experience, course-specific experience, level-specific experience, and course-level specific experience. The North Carolina data track teachers and students in the universe of public high schools from 1997-2009. Critically, the data feature 74,000 within-teacher changes in subject assignment and over 45,000 academic-level switches. Such rich data permit estimation of an education production function that includes general, course-specific, level-specific, and course-level-specific experience profiles as well as a full set of school-teacher-course-level fixed effects. The flexibility of our model allows us to control for many potential biases that might otherwise accompany endogenous course assignment decisions.

To preview our results, we find that much of the return to years of experience that have been estimated in the value-added literature is actually specific to the subject that the teacher taught. We find no evidence, however, of returns to either level-specific or subject-level experience. In agreement with the rest of the value-added literature, we find that the variation in innate teaching skill is comparable in magnitude to the gains to experience; in contrast to gains from experience, however, such permanent skill is mostly general. Our estimates suggest only a minor role for subject-specific or level-specific teaching talent.

Of course, the knowledge that a large fraction of the gains from experience are subject-specific may be of limited value to principals if most changes in course assignments are driven by necessity. For example, parental leave may require principals to reassign teachers to unfamiliar subjects or tracks. Thus, using our estimated experience profiles, we perform a counterfactual simulation in which we assess the potential achievement gains from maximizing the context-specific skill associated with teaching assignments. Specifically, for each year of our data, we reassign the teachers observed teaching at each school in that year to the courses that were offered at their school at the time in order to maximize student performance, given the four-dimensional profiles of experience that each teacher possesses at that point in time. To ensure that our counterfactual allocation was feasible for the principal at the time, we restrict each teacher to only teach the number of classrooms they were actually observed teaching in each year. SIMULATION RESULTS COMING SOON.

The rest of the paper proceeds as follows. Section 2 presents the education production function whose parameters we estimate. Section 3 describes how comparisons of teachers with different course assignment histories can provide joint identification of both teacher-course-level fixed effects and general, course-specific, level-specific, and course-level-specific experience profiles. Section 4 discusses the North Carolina administrative data and provides summary statistics displaying the variation in teacher course assignments. Section 5 considers estimation of the model. Section 6 presents the parameter estimates from our main specification. Section 7 discusses possible threats to our identifying assumptions and presents results from several specification tests and robustness checks. Section 8 describes the counterfactual simulation in which teachers' course assignments are made to maximize gains from context-specific experience and presents the results from the simulation. Finally, Section 9 concludes.

2 Model Specification

Because our focus is on the relative importance of various components of context-specific teacher skill and experience to test score performance, we craft our specification of the achievement production function so as to isolate the contribution of these components. Let Y_{ict} represent the standardized test score of student i in classroom c at time t. Let $r(i, c, t)$ denote the teacher that taught student i in classroom c at time t. Similarly, let $s(i, c, t)$ denote the school at which student i experienced classroom c at time t, let $j(i, c, t)$ denote the subject taught in student i's classroom c at time t, and let $l(i, c, t)$ denote the difficulty level or track associated with the classroom. $l(i, c, t) \in \{b, h\}$, where b denotes "basic" and h denotes "honors".[4] Each test score Y_{ict} is standardized so that the distribution of test scores in each subject-year combination has zero mean and unit variance.

By suppressing the dependence of s, r, j, and l on (i, c, t), we can represent the production of test score performance compactly via:

$$(1) \qquad Y_{ict} = X_{ict}\beta_{jl} + \delta_{sjl} + \mu_{srjl} + d^{tot}(exp_{rt}^{tot}) + d^{j}(exp_{rt}^{j}) + d^{l}(exp_{rt}^{l}) + d^{jl}(exp_{rt}^{jl}) + \epsilon_{ict}$$

X_{ict} represents a vector of student observable characteristics and middle school reading and math test scores, along with a vector of the average levels of observable characteristics and past test scores in classroom c. We allow the impact of student and classroom characteristics and past scores to differ by subject-level combination. This allows the students' past test scores to reveal comparative advantages in particular subjects, so that a high 8th grade math score might be a

[4]See Section 4 for a discussion of assignment of courses to difficulty levels.

stronger predictor of performance in Algebra 1 than in English 1. Similarly, classroom composition might matter more in a particular subject or level if more group work takes place in say, basic biology (labs!) than in honors math.

δ_{sjl} represents a full set of school-subject-level fixed effects. These will capture the average residual achievement at each school-subject-level combination, after removing the part of achievement that can be predicted based on observable student and classroom characteristics. The set of $\{\delta\}$ parameters will not only capture any school-level inputs such as principal quality, neighborhood quality, or quality of the facilities, they will also capture any variation in the quality of curricula or textbooks across subjects and levels within the school. Importantly, they will also capture the contribution of average unobserved inputs of the students who sort into particular school-subject-level combinations. Thus, the inclusion of δ_{sjl} acts as a control function that absorbs school inputs as well as any potential sorting biases that might otherwise be created by students' endogenous choices of school, subject, and level.

μ_{rsjl} represents a full set of school-teacher-subject-level fixed effects. The average school-teacher-subject-level will be normalized to be 0 for each school-subject-level, so that μ_{rsjl} can be thought of as the deviation of a particular teacher's performance in a particular subject-level combination from the mean (student-weighted) performance of all teachers that taught at the teacher's school-subject-level combination. This specification of the contribution of teacher quality allows the estimation of a fully non-parametric joint distribution of general teacher quality and subject-specific, level-specific, and even subject-level specific comparative advantages at particular subjects, both within and across teachers. Note that by including the identify of the school in the definition of the fixed effect, we are allowing each teacher to have a completely different average skill and set of comparative advantages for particular subjects and levels at each school at which they teach (a teacher who teaches in two schools is essentially treated as two different teachers). Variation in μ_{rsjl} around a given average $\overline{\mu}_{rsl}$ will provide evidence of subject-specific skill, while variation in $\overline{\mu}_{rsjl}$ around a given average $\overline{\mu}_{rsj}$ will provide evidence of level-specific skill. One can then average the contribution of each teacher across subjects and levels $\overline{\mu}_{rs}$ and compare these averages across teachers to examine the variation in general persistent teacher quality.

exp_{rt}^{tot} represents the total years of general teaching experience that teacher r possesses at time t. $d^{tot}(*)$ is a function that captures how additional years of total experience increases a teacher's ability to improve student performance (regardless of the subjects and levels in which this experience was earned). Analogously, exp_{rt}^{j}, exp_{rt}^{l}, and exp_{rt}^{jl} represent years of experience in the relevant subject, level, and subject-level combination respectively. The $d^{j}(*)$, $d^{l}(*)$, and $d^{jl}(*)$ functions capture how additional years of subject-specific experience, level-specific experience, and subject-level specific experience increase teachers' ability to increase student test scores. $d^{tot}(*)$, $d^{j}(*)$,, $d^{l}(*)$, and $d^{jl}(*)$ are each flexibly parametrized using indicators for narrow ranges of experience and estimated.

Finally, ϵ_{ict} represents an error component which combines time varying inputs not captured by the other components of the model. In particular, we model the error component as:

$$\epsilon_{ict} = \nu_{rt} + \phi_{st} + \zeta_{ct} + e_{ict} \tag{2}$$

ν_{rt} represents year-specific deviations in a teacher's quality from what would be expected based on his/her long run skill and level of experience in the appropriate subject-level combination. ϕ_{st} captures year-specific deviations in school inputs or student sorting relative to the sample-wide average for the school-subject-level. ζ_{ct} captures classroom level shocks, such as the archetypal dog barking outside the classroom window on test day. Finally, e_{ict} represents measurement error that captures the extent to which the student's performance on the particular exam deviates from what the student could have expected to score, given his/her accumulated knowledge in the subject. We adjust standard errors to account for the existence of each of these error components.

3 Identification

3.1 Identifying the Return to General and Context-Specific Experience

To identify the experience profiles $d^{tot}(*)$, $d^{j}(*)$, $d^{l}(*)$, and $d^{jl}(*)$, the following conditions must hold:

Assumption 1: Conditional Mean Independence of

Time-Varying Unobserved Inputs and Teacher Experience

$$E[\epsilon_{ict}|ex_{rt}^{tot} = \tilde{ex}_{rt}^{tot}, (s,r,j,l) = (\tilde{s},\tilde{r},\tilde{j},\tilde{l}), \tilde{X}_{ict}] =$$

$$E[\epsilon_{ict}|(s,r,j,l) = (\tilde{s},\tilde{r},\tilde{j},\tilde{l}), \tilde{X}_{ict}] \; \forall \; \tilde{ex} \in \mathcal{EX}, (\tilde{s},\tilde{r},\tilde{j},\tilde{l}) \in \mathcal{SRJL}, \tilde{X} \in \mathcal{X} \tag{3}$$

$$E[\epsilon_{ict}|ex_{rt}^{j} = \tilde{ex}_{rt}^{j}, (s,r,j,l) = (\tilde{s},\tilde{r},\tilde{j},\tilde{l}), \tilde{X}_{ict}] =$$

$$E[\epsilon_{ict}|(s,r,j,l) = (\tilde{s},\tilde{r},\tilde{j},\tilde{l}), \tilde{X}_{ict}] \; \forall \; \tilde{ex} \in \mathcal{EX}, (\tilde{s},\tilde{r},\tilde{j},\tilde{l}) \in \mathcal{SRJL}, \tilde{X} \in \mathcal{X} \tag{4}$$

$$E[\epsilon_{ict}|ex_{rt}^{l} = \tilde{ex}_{rt}^{l}, (s,r,j,l) = (\tilde{s},\tilde{r},\tilde{j},\tilde{l}), \tilde{X}_{ict}] =$$

$$E[\epsilon_{ict}|(s,r,j,l) = (\tilde{s},\tilde{r},\tilde{j},\tilde{l}), \tilde{X}_{ict}] \; \forall \; \tilde{ex} \in \mathcal{EX}, (\tilde{s},\tilde{r},\tilde{j},\tilde{l}) \in \mathcal{SRJL}, \tilde{X} \in \mathcal{X} \tag{5}$$

$$E[\epsilon_{ict}|ex_{rt}^{jl} = \tilde{ex}_{rt}^{jl}, (s,r,j,l) = (\tilde{s},\tilde{r},\tilde{j},\tilde{l}), \tilde{X}_{ict}] =$$

$$E[\epsilon_{ict}|(s,r,j,l) = (\tilde{s},\tilde{r},\tilde{j},\tilde{l}), \tilde{X}_{ict}] \; \forall \; \tilde{ex} \in \mathcal{EX}, (\tilde{s},\tilde{r},\tilde{j},\tilde{l}) \in \mathcal{SRJL}, \tilde{X} \in \mathcal{X} \tag{6}$$

Assumption 1 states that, for each dimension of experience, knowledge of the level of experience does not provide further information about any unobserved component of inputs, conditional on observed student inputs and the identity of the school, teacher, subject, and level. Put another way, the timing of experience accumulation in each dimension is assumed to be exogenous.

There are a number of possible threats to the validity Assumption 1, each of which relates to the exact timing of changes in experience. For example, suppose that when a school is in decline, teacher turnover begins to increase, and the teachers that remain are forced to teach both new subjects and new difficulty levels more frequently. In this case, we may observe zero subject-specific or level-specific experience more frequently when the value of ϕ_{st} is low. Since year-specific deviations in school quality from the long-run average are included in ϵ_{ict}, this scenario violates Assumption 1, and could potentially produce an overestimate of the returns to context-specific experience. Alternatively, suppose principals are reluctant to force a teacher to take on new subjects or levels when the teacher faces other short-term obstacles (such as illness, a new child, or a divorce). In that case, zero subject-specific or level-specific experience may be observed more frequently when the value of ν_{rt} is high. This scenario also violates Assumption 1, and might cause an underestimate of the returns to context-specific experience. Similarly, if teachers respond to a particularly unruly classroom by quitting teaching, or switching levels or subjects, we might underestimate the returns to exprience (since those who survive to the next year of experience will have observed above-average shocks, thereby hiding the gains to the next year of experience). In Section 7, we estimate an upper bound for the degree of bias introduced by violations of Assumption 1 triggered by correlations between our experience profiles and ϕ_{st}, which we deem the most plausible of the above scenarios. We find that endogenous responses to school-year shocks are unlikely to produce a substantial bias to any of our profiles.

Despite these concerns, however, note that Assumption 1 is still much weaker than is required to identify experience profiles in most of the literature, since it conditions on the identity of the school, teacher, level, and course. Essentially, the inclusion of school-teacher-subject-level fixed effects (μ_{srcl}) controls for any arbitrary selection of teachers into experience categories based on permanent general or context-specific skill. Conditioning on r accounts for the possibility that better teachers persist long enough to gain more experience. Similarly, conditioning on r and j accounts for the possibility that the teachers allowed to gain more subject-specific experience in a particular subject are those with comparative advantages in teaching the subject, while conditioning on r and l accounts for the possibility that persistence at teaching honors level courses might signal a comparative advantage for teaching such courses.

Even if the timing of experience accumulation is conditionally independent of the error components, the simultaneous identification of each of the four experience profiles also requires considerable variation in the history of subject and level assignments across teachers. To see how identification might be secured, consider a simple case in which there are only two subjects, chemistry (C) and physics (P), and only two difficulty levels, basic (B) and honors (H). Suppose we observe a teacher teach the following sequence: $CH \Rightarrow PB \Rightarrow CH$. Suppose her colleague teaches: $CH \Rightarrow PH \Rightarrow CH$. Under the model in 1, the difference in their performance during their first year will capture the difference in their context-specific skill: $\mu^2_{CH} - \mu^2_{CH}$. Compare their performance again during the third year, we obtain: $\mu^2_{CH} - \mu^2_{CH} + d^H(2) - d^H(1)$. The difference-in-differences isolates the return to a second year of level-specific experience. Suppose a second colleague had taught $CH \Rightarrow CB \Rightarrow CH$. The same difference-in-difference would have isolated the return to a second year of subject-specific experience, $d^C(2) - d^C(1)$. Suppose a third colleague had taught $CH \Rightarrow CH \Rightarrow CH$. The same difference-in-difference would yield $d^C(2) - d^C(1) + d^H(2) - d^H(1) + d^{CH}(2) - d^{CH}(1)$. Together with the other comparisons, this moment identifies the return to a second year of subject-level specific experience. Comparing the growth in performance between the second and third years of any teacher who stays in the same subject-level for three years yields $d^{tot}(2) - d^{tot}(1) + d^C(2) - d^C(1) + d^H(2) - d^H(1) + d^{CH}(2) - d^{CH}(1)$, which, combined with the previous moments, identifies the return to a second year of general experience.[5]

While these sample histories are stylized, note that there are many alternative moments that also provide identifying variation. For example, we could have identified $d^{tot}(2) - d^{tot}(1)$ instead by comparing teachers with the histories $CB \Rightarrow PH \Rightarrow CB$ and $CB \Rightarrow CB$ in their first year and then again using the 3rd year for the first teacher and the 2nd year for the first teacher. Alternatively, consider comparing the 2nd and 3rd years of a teacher with history $CB \Rightarrow CH \Rightarrow CH$ to one who experienced $CH \Rightarrow CH \Rightarrow CH$. The analogous difference-in-difference moment yields $d^H(2) - d^H(1)$ + $d^{CH}(2) - d^{CH}(1)$, which can be compared with the difference of the

[5]Note that since returns to experience can only be identified relative to other levels of experience, we must normalize one value for each function. We do so by setting $d^k(0) = 0$ for $k \in \{tot, j, l, lk\}$.

first and third difference-in-difference examples above. Indeed, given the frequency with which subject and level switching occurs, we frequently observe multiple teachers who have taught the same set of subjects and levels over their careers at the school, but have taught them in different orders, or in different proportions. Each such case provides a further source of identifying variation for the various context-specific experience profiles. Consequently, not only are these experience profiles precisely identified (at least for the first 10 years of experience), but there are myriad overidentifying tests that can be implemented if one worries that particular sequences may be likely to occur in conjunction with particular changes in unobserved inputs (in violation of Assumption 1).

3.2 Identification of the general and context-specific components of permanent teaching skill

Identifying permanent general and context-specific teaching skill is trickier. In particular, there is a fundamental identification problem that our model cannot overcome: we cannot distinguish average teaching quality in a school-subject-level from school or unobserved student inputs that are school-subject-level specific. If a school scores .1 student level standard deviations higher in Biology than in Chemistry, we cannot tell whether all the Biology teachers are particulary effective, or if instead the Biology textbook is superior to the Chemistry textbook (or many of the student's parents are biologists). To address this issue, we consider two polar opposite assumptions, and decompose the variance in teacher permanent skill into general, subject-specific, level-specific, and subject-level specific components under each assumption. The first assumption is that average teacher effectiveness $\overline{\mu}_{rjl}$ is uniform across all levels, subjects, and schools:

Assumption 2A: Uniform Average Teacher Quality Across Contexts

(7) $$E[\overline{\mu}_{sjl}|(s,j,l)=(\tilde{s},\tilde{j},\tilde{l})]=k \text{ for some constant k}, \forall (s,j,l) \in \mathbf{RJL}$$

This would hold if the relatively more effective teachers do not sort into particular schools, subjects, or levels. Assumption 2A implies that all the variation in average residual student performance (after removing the part predictable based on student observables) across subjects, levels, and schools can be attributed to either school inputs or unobserved student inputs. Assumption 2A can be imposed on the model by including school-subject-level fixed effects (δ_{sjl}), and normalizing the student-weighted average teacher-school-subject-level fixed effect to be zero at each school-subject-level $\frac{1}{N_{sjl}} \sum_{i \in scl} \hat{\mu}_{sjl} = 0$. Under Assumption 2A, a teacher whose Biology students perform .1 standard deviaitons better than her Chemistry students will be assumed to be equally effective at teaching both Biology and Chemistry if the school average performance difference between Biology and Chemistry is .1 standard deviations. The polar opposite approach is to assume that all the variation in average residual student performance across subjects, levels, and schools can be attributed to differences in average teacher quality:

Assumption 2B: Uniform School and Unobserved Student Quality Across Contexts

(8) $$E[\delta_{sjl} | (s, j, l) = (\tilde{s}, \tilde{j}, \tilde{l})] = k \text{ for some constant k}, \forall (s, j, l) \in \mathbf{SJL}$$

Assumption 2B would hold if students sort into high schools, subjects, and levels based only on observable characteristics and past performance, and all high schools and subject-level combinations within high schools provide the same contribution to student achievement. Assumption 2B can be imposed on the model by excluding school-subject-level fixed effects ($\delta_{sjl} = 0 \; \forall (s, j, l)$), and matching the between school-subject-level residual variation using a full set of teacher-school-subject-level fixed effects (without any normalizations). Under Assumption 2B, a teacher whose Biology students perform .1 standard deviaitons better than her Chemistry students will be assumed to be .1 standard deviations more effective at teaching both Biology and Chemistry if the school average performance difference between Biology and Chemistry is .1 standard deviations. In other words, even though the teacher is at the mean of the performance distribution in both subjects, the comparison set of Biology teachers is assumed to be .1 standard deviations superior on average to the comparison set of Chemistry teachers.

An intermediate assumption would be to assume that between-school variation in residual test scores is attributable to school quality and student sorting, but that the variation in residual performance that is within-schools but across subject-level combinations is attributable to differences in average teacher quality across these combinations:

Assumption 2C: Uniform Teacher Quality Across Schools, Uniform Student/School Quality Across Subjects and Levels

$$E[\delta_{sjl}|(s,j,l) = (\tilde{s},\tilde{j},\tilde{l})] = E[\delta_{sjl}|s = \tilde{s}]\forall(s,j,l) \in \mathbf{SJL}$$

$$E[\overline{\mu}_s|s = \tilde{s}] = k \text{ for some constant } k, \forall s \in \mathbf{S} \tag{9}$$

Estimates from such a model are useful for a principal who needs to make classroom assignments for his existing stock of teachers. She may only be interested in the breakdown of within-school teacher quality into general vs. course or level-specific components, and may believe that school inputs are divided relatively equally across subjects and levels.

While Assumptions 2A-2C allow us to separate school inputs from teacher inputs, identification of $\{\mu_{srcl}\}$ also requires that other unobserved inputs are not correlated with the observation of a particular teacher in a particular course-level combination.

Assumption 3A-3C: Conditional Mean Independence of Students' Unobserved Inputs and Teacher Experience

$$E[\epsilon_{ict}|(s,r,j,l) = (\tilde{s},\tilde{r},\tilde{j},\tilde{l}),(ex^{tot}_{rt},ex^{j}_{rt},ex^{l}_{rt},ex^{jl}_{rt}) = (\tilde{ex}^{tot}_{rt},\tilde{ex}^{j}_{rt},\tilde{ex}^{l}_{rt},\tilde{ex}^{jl}_{,rt}),X_{ict} = X'_{ict}] =$$

(10)

$$E[\epsilon_{ict}|(s,j,l) = (\tilde{s},\tilde{j},\tilde{l}),(ex^{tot}_{rt},ex^{j}_{rt},ex^{l}_{rt},ex^{jl}_{rt}) = (\tilde{ex}^{tot}_{rt},\tilde{ex}^{j}_{rt},\tilde{ex}^{l}_{rt},\tilde{ex}^{jl}_{,rt}),X_{ict} = X'_{ict}]$$

Assumption 3B: Conditional Mean Independence of

Students' Unobserved Inputs and Teacher Experience

$$3A: \quad E[\epsilon_{ict}|(s,r,j,l) = (\tilde{s},\tilde{r},\tilde{j},\tilde{l}),(ex^{tot}_{rt},ex^{j}_{rt},ex^{l}_{rt},ex^{jl}_{rt}) = (\tilde{ex}^{tot}_{rt},\tilde{ex}^{j}_{rt},\tilde{ex}^{l}_{rt},\tilde{ex}^{jl}_{,rt}),X_{ict} = X'_{ict}] =$$

$$E[\epsilon_{ict}|(s,j,l) = (\tilde{s},\tilde{j},\tilde{l}),(ex^{tot}_{rt},ex^{j}_{rt},ex^{l}_{rt},ex^{jl}_{rt}) = (\tilde{ex}^{tot}_{rt},\tilde{ex}^{j}_{rt},\tilde{ex}^{l}_{rt},\tilde{ex}^{jl}_{,rt}),X_{ict} = X'_{ict}]$$

$$3B: \quad E[\epsilon_{ict}|(s,r,j,l) = (\tilde{s},\tilde{r},\tilde{j},\tilde{l}),(ex^{tot}_{rt},ex^{j}_{rt},ex^{l}_{rt},ex^{jl}_{rt}) = (\tilde{ex}^{tot}_{rt},\tilde{ex}^{j}_{rt},\tilde{ex}^{l}_{rt},\tilde{ex}^{jl}_{,rt}),X_{ict} = X'_{ict}] =$$

$$E[\epsilon_{ict}|(ex^{tot}_{rt},ex^{j}_{rt},ex^{l}_{rt},ex^{jl}_{rt}) = (\tilde{ex}^{tot}_{rt},\tilde{ex}^{j}_{rt},\tilde{ex}^{l}_{rt},\tilde{ex}^{jl}_{,rt}),X_{ict} = X'_{ict}]$$

$$3C: \quad E[\epsilon_{ict}|(s,r,j,l) = (\tilde{s},\tilde{r},\tilde{j},\tilde{l}),(ex^{tot}_{rt},ex^{j}_{rt},ex^{l}_{rt},ex^{jl}_{rt}) = (\tilde{ex}^{tot}_{rt},\tilde{ex}^{j}_{rt},\tilde{ex}^{l}_{rt},\tilde{ex}^{jl}_{,rt}),X_{ict} = X'_{ict}] =$$

(11)

$$E[\epsilon_{ict}|s = \tilde{s},(ex^{tot}_{rt},ex^{j}_{rt},ex^{l}_{rt},ex^{jl}_{rt}) = (\tilde{ex}^{tot}_{rt},\tilde{ex}^{j}_{rt},\tilde{ex}^{l}_{rt},\tilde{ex}^{jl}_{,rt}),X_{ict} = X'_{ict}]$$

Assumption 3A states that the identify of the teacher does not provide further information about any unobserved inputs, conditional on the identities of the school, subject, and track, along with the levels of both general and context-specific experience of the teacher and the observable characteristics of the student. Note that by conditioning on all four dimensions of teacher experience, we remove the concern that a teacher will be perceived to have greater general skill because they have more general experience, or more importantly, that a teacher will be perceived to have a comparative advantage at a teaching in a particular context because many of the observations in that context are accompanied by considerable context-specific experience. Assumption 3B is much stronger, since it does not condition on the identify of the school, subject, or level, while Assumption 3C conditions on the identity of the school only.

Even in the case of Assumption 3A, there are still threats to the validity of Assumption 3. Suppose, for example, that a particular teacher R is assigned to a room with broken air conditioning each time he teaches honors physics (PH), but is assigned to functioning rooms whenever he teaches honors chemistry. If the teacher teaches the same number of years in each course, then conditioning on context-specific experience will not remove the correlation between the classroom-level error component ζ_{ct} and the fixed effect μ_{RPH}. Similarly, a teacher who happens to be assigned to basic English 1 classes during the years her kids are young (when she has little time to prepare for class) might exhibit a correlation between ν_{rt} and μ_{rEB}.

To illustrate how $\hat{\mu}_{srjl}$ can be identified given any of Assumptions 2A-2C paired with 3A-3C, consider the a teacher r' who teaches subject j' and level l' in school s' during year t. Suppose, without loss of generality, that the teacher has a vector of general and context specific experience of $(ex^{tot} = 4, ex^{j} = 3, ex^{l} = 2, ex^{jl} = 2)$. The expected residual performance of her students in subject-level combination (j, l), denoted $Z_{ict} = Y_{ict} - X_{ict}\beta$, is given by:

$$\begin{aligned} & E[Z_{ict}|(s,r,j,l) = (s',r',j',l'),(ex^{tot}_{rt}, ex^{j}_{rt}, ex^{l}_{rt}, ex^{jl}_{rt}) = (4,3,2,2)] \\ (12) \qquad & = \delta_{sjl} + \mu_{srjl} + d^{tot}(4) + d^{j}(3) + d^{l}(2) + d^{jl}(2) \end{aligned}$$

Since the experience profiles $d^{tot}(*)$, $d^{j}(*)$, $d^{l}(*)$, and $d^{jl}(*)$ were identified using variation in the sequence of course assignments in the subsection above, moments that capture the level of performance at a point in time identify $\delta_{sjl} + \mu_{srjl}$. Under Assumption 2B, $\delta_{sjl} = 0 \; \forall \; (s, j, l)$, so such moments identify μ_{srjl} directly. Under Assumption 2A, we can use the fact that the (student weighted) average teacher quality in each school-subject-level is assumed to be zero. Specifically, the average residual performance of students in a particular school-subject-level is given by:

$$(13) \qquad E[Z_{ict}|(s,r,j,l) = (s',r',j',l')] = \delta_{sjl} + \overline{d^{tot}(ex^{tot})} + \overline{d^{j}(ex^{j})} + \overline{d^{l}(ex^{l})} + \overline{d^{jl}(ex^{jl})},$$

which identifies δ_{sjl}. To identify δ_s under Assumption 2C, we simply average at the school level instead of the school-subject-level level. Thus, μ_{srjl} can be identified for each combination of school-teacher-subject-level that we actually observe in the data.

Notice that unlike the experience profiles, each μ_{srjl} fixed effect will be estimated using only a single teacher's performance during the few years in which they taught that course-level. As such, sampling error for any given $\hat{\mu}_{rjl}$ estimate will not converge to zero even with the fairly long panel we employ. Consequently, we do not focus on individual $\hat{\mu}_{rjl}$ estimates, but seek instead to decompose the variance in performance across teachers and contexts into components attributable to general teaching talent, subject-specific talent, level-specific talent, and subject-level specific talent. Specifically, note that we can first rewrite each effect μ_{srjl} using:

$$\mu_{srjl} = \overline{\mu}_{sr} + (\mu_{srjl} - \overline{\mu}_{sr}) \tag{14}$$

The first component in 14 can be interpreted as the contribution of teacher talent that may be school-specific but is general across subject-level combinations within the school. We will refer to $Var(\overline{\mu}_{sr})$ as the variance in general teaching talent. The second component consists of the teacher's persistent subject-level specific deviation in quality from her average level across all subject-level combinations. This can be interpreted as her comparative advantage or disadvantage at teaching subject-level combination (j, l). This second component can then be decomposed into three further components:

$$(\mu_{srjl} - \overline{\mu}_{sr}) \equiv \tilde{\mu}_{srjl} = \overline{\tilde{\mu}}_{srj} + \overline{\tilde{\mu}}_{srl} + (\tilde{\mu}_{srjl} - \overline{\tilde{\mu}}_{srj} - \overline{\tilde{\mu}}_{srl}) \tag{15}$$

The first component of 15 can be interpreted as the part of her comparative advantage at subject-level combination (j, l) that is common to all subjects. We will refer to $Var(\overline{\tilde{\mu}}_{srj})$ as the variance in subject-specific teaching talent. The second component of 15 can be interpreted as the part of her comparative advantage at subject-level combination (j, l) that is common to all levels. We will refer to $Var(\overline{\overline{\mu}}_{srj})$ as the variance in level-specific teaching talent. The third component of 15 is the part of a teacher's comparative advantage at (j, l) that couldn't have been predicted based on the sum of her subject-specific skill and her level-specific skill. We will refer to $Var(\tilde{\mu}_{srjl} - \overline{\tilde{\mu}}_{srj} - \overline{\overline{\mu}}_{srj})$ as the variance in subject-level specific teaching skill.

3.3 Recovering the Latent Variance Decomposition

Note that we do not observe the true variance of school-teacher-subject-level effects, but rather the sample variance, $Var(\mu_{srjl})$, which includes sampling error: $Var(\hat{\mu}_{srjl})$. To distill the true variance, we follow Aaronson, Barrow and Sander (2007) and Mansfield (2013). Specifically, we first define each estimated school-teacher-subject-level fixed effect $\hat{\mu}_{srjl}$ as the sum of the teacher's true context-specific skill and an uncorrelated error component: $\hat{\mu}_{srjl} = \mu_{srjl} + \xi_{srjl}$. Then the (student-weighted) sample variance in estimated context-specific skill can be decomposed as:

$$(16) \qquad \frac{1}{N} \sum_{(i,c,t) \in \mathbf{ICT}} (\hat{\mu}_{srjl})^2 = \frac{1}{N} \sum_{(i,c,t) \in \mathbf{ICT}} (\mu_{srjl})^2 + \frac{1}{N} \sum_{(i,c,t) \in \mathbf{ICT}} (\xi_{srjl})^2$$

where N is the number of test scores in the sample, and $\mathbf{ICT}$ is the set of (i,c,t) test score observations in the sample. As usual, the dependence of (s, r, j, l) on i, c, t has been dropped.

One would like to estimate the variance in true teacher quality as:

$$(17) \qquad \hat{Var}(\mu_{srjl}) = \frac{1}{N} \sum_{(i,c,t) \in \mathbf{ICT}} (\hat{\mu}_{srjl})^2 - \frac{1}{N} \sum_{(i,c,t) \in \mathbf{ICT}} (\xi_{srjl})^2.$$

ξ_{srjl} is not observed, but

$$(18) \qquad \frac{1}{N} \sum_{(i,c,t) \in \mathbf{ICT}} (\xi_{srjl})^2 \approx \frac{1}{N} \sum_{(i,c,t) \in \mathbf{ICT}} E[(\xi_{srjl})^2] = \frac{1}{N} \sum_{(i,c,t) \in \mathbf{ICT}} (se(\xi_{srjl}))^2,$$

so I estimate the error variance component using the standard error estimates for each school-teacher-subject-level fixed effect:

$$(19) \qquad \hat{Var}(\mu_{srjl}) = \frac{1}{N}\sum_{(i,c,t)\in\mathbf{ICT}}(\hat{\mu}_{srjl})^2 - \frac{1}{N}\sum_{(i,c,t)\in\mathbf{ICT}}(se(\xi_{srjl}))^2$$

By using the delta method to estimate standard errors for $\tilde{\hat{\mu}}_{srjl}$, denoted $se(\tilde{\xi}_{srjl})$, we can estimate $\hat{Var}(\tilde{\mu}_{srjl})$ analogously. Then, $\hat{Var}(\overline{\mu}_{sr})$ can be estimated via:

$$(20) \qquad \hat{Var}(\overline{\mu}_{sr}) = \hat{Var}(\mu_{srjl}) - \hat{Var}(\tilde{\mu}_{srjl})$$

To prevent teachers who only taught a single subject-level combination from biasing our estimate of $\hat{Var}(\overline{\mu}_{sr})$ downward, we restrict the sample of school-teacher-subject-level combinations when calculating $\hat{Var}(\tilde{\hat{\mu}}_{srjl})$ to those in which the relevant school-teacher combination was observed in at least two school-teacher-subject-level combinations.

Further use of the delta method allows the same procedure to be applied in recovering the true variance of subject-specific, level-specific, and subject-level-specific teacher talent.[6]

Finally, because we can only estimate a value of $\hat{\mu}_{srjl}$ for those combinations that we actually observe in the data, the variance in course-specific and level-specific skill that we estimate will represent the variance among the range of course and level combinations that principals actually assign. This is likely to be a selected sample: since principals may have knowledge of the relative skills of their teachers, they may avoid assigning teachers to subjects or levels at which they are likely to be particularly ineffective. For example, teaching two subjects in completely different

[6]Specifically, we calculate the true variances as follows. First, consider the alternative decomposition $\tilde{\mu}_{srjl} = \overline{\overline{\mu}}_{srj} + (\tilde{\mu}_{srjl} - \overline{\overline{\mu}}_{srj})$. We estimate the true variance of the second component by using the delta method to calculate standard errors for $\widehat{(\tilde{\mu}_{srjl} - \overline{\overline{\mu}}_{srj})}$ and applying the same method as above. We then obtain the variance in subject-specific teaching talent, $\hat{Var}(\overline{\overline{\mu}}_{srj})$, via $\hat{Var}(\overline{\overline{\mu}}_{srj}) = \hat{Var}(\tilde{\hat{\mu}}_{srjl}) - \hat{Var}((\tilde{\mu}_{srjl} - \overline{\overline{\mu}}_{srj}))$. The variance in level-specific teaching talent, $\hat{Var}(\overline{\overline{\mu}}_{srl})$, can be calculated using an identical approach. Finally, we estimate the variance in subject-level-specific teaching talent using: $\hat{Var}(\tilde{\mu}_{srjl} - \overline{\overline{\mu}}_{srj} - \overline{\overline{\mu}}_{srj}) = \hat{Var}(\tilde{\hat{\mu}}_{srjl}) - \hat{Var}(\overline{\overline{\mu}}_{srj}) - \hat{Var}(\overline{\overline{\mu}}_{srl})$.

fields (Geometry and English) may be more difficult than teaching two subjects in the same field (Algebra 1 and Geometry). While we are likely to underestimate the variance in subject-specific (or level-specific) talent across the full range of possible subjects (or levels), the estimates we do obtain may be more relevant or interesting to administrators, since nobody is proposing to assign more math teachers to English 1. The choice principals make is generally between hiring a new teacher to teach exactly the courses taught by an exiting teacher, or hiring a new teacher to teach different courses, and rotating existing teachers to new courses or levels (for example, rewarding stayers by letting them teach the honors class that was vacated by the exiting teacher). Nonetheless, in future versions of this paper (coming soon!), we will examine whether teaching skill is actually field-specific. We can compare estimates of the variance in subject-specific skill using teachers who taught multiple subjects in the same field (e.g. Algebra 1 and Geometry) to estimates that isolate the variation in subject-specific skill that derives from teachers who taught subjects in different fields (e.g. Algebra 1 and English).

4 Data

We employ data provided by the North Carolina Education Research Data Center consisting of standardized test scores of the universe of public high school students in North Carolina from 1997 - 2009 in eleven subjects and two class difficulty levels.[7]

During the sample period, North Carolina provided a standardized curriculum in which achievement was assessed via required end-of-course tests in English 1, Econ/Law/Politics, U.S. History, Algebra 1, Algebra 2, Geometry, Biology, Chemistry, Physics, and Physical Science.[8]

[7]The data originally provide nine difficulty level delineations: Special Education, Remedial, Basic, Applied/Technical, Honors, Cooperative Education, Advanced Placement, International Baccalaureate, and non-classroom. We drop student observations coming from classes labeled as special education, cooperative education, and non-classroom. We consider remedial, basic, and applied/technical classes as “basic” and advanced placement, international baccalaureate, and honors as “honors”.

[8]Testing began for Physics, Geometry, Chemistry, Physical Science, and Algebra 2 in 1999. In addition, Econ/Law/Politics was discontinued in 2004 and replaced by Civics and Economics in 2006. U.S. History was not

The data contain a large number of current student inputs[9] as well as past student inputs[10] and teacher-year inputs[11].

4.1 Generating the Experience Profile

We allow a flexible experience profile by creating indicators for eight experience cells: 0 years of experience, 1 year, 2 years, 3 years, 4 years, 5-6 years, 7-10 years, and 11 or more years of experience. Using this formulation, we track four types teacher experience: general experience, course-specific, difficulty-level-specific, and course-level-specific experience. Note that experience is measured in terms of years in which at least one classroom was taught in the relevant context, rather than by total number of classrooms ever taught. Table 1, we show how experience evolves for the second- and third- year teachers in our final sample.

4.2 Final Specification

Our empirical strategy requires that student test score observations be matched to the teachers who taught the class. Unfortunately, the teacher ID provided in the test score data corresponds to the test administrator, who may or may not be the true teacher of the class. However, personnel records contain information on the demographic composition of each class taught by each teacher, and since the student achievement data can be aggregated to the classroom level, we utilize a

tested between 2004 and 2005.

[9]Observable student inputs include classroom composition (including class size, racial composition, and number of gifted students in math and reading), as well as indicators for parental education, race, gender, gifted status, current or ever having Limited English Proficiency status, free/reduced price lunch eligibility, learning disability in math, reading, or writing, for whether the student intends to attend community college, attend four-year college, or work after high school, as well as indicators for participation in a sport, vocational club, academic club; service club, or arts club., and finally missing indicators for 7^{th} and 8^{th} grade math and reading scores.

[10]The student's 7^{th} and 8^{th} grade math and reading scores as well as the class's average 8^{th} grade math and reading scores

[11]Including the number of classes and number of different course-levels taught contemporaneously by the student's teacher

fuzzy match algorithm that matches on classroom-average demographics. See Mansfield (2013) for detailed description of the algorithm and summary statistics regarding its efficacy.

We drop from the sample test scores for students to whom we cannot match a teacher or verify a difficulty level, as well as scores from classes with fewer than 5 students. Since past test scores are critical for controlling for student sorting, we also drop observations with missing current/past test scores. Lastly, because our identification strategy relies on observing the teacher's full history of course- and level-specific experience at each point in time, we keep only test scores associated with teachers begin teaching during our sample, as indicated by an entry level paycode.

4.3 Teacher Mobility

Table 2 depicts teacher mobility in our final sample across courses. The top entry in each cell (i, j) represents the number of teachers in our sample who ever taught in subject i that also taught in subject j, while the bottom entry represents the fraction of teachers who ever taught in subject i that also taught in subject j. The table revals that there is considerable mobility acorss subjects, though the vast majority of mobility occurs within fields (e.g. math, science, etc.). This reflects the fact that certification is field-specific. Table 3 represents the corresponding transition matrix for levels. It reveals that almost all teachers who ever teach an honors class also teach at least one basic class during their career. The converse is not true; only half of teachers observed teaching at least one basic class are also observed teaching an honors class at some point during their career. This finding partly reflects the fact that there tend to be more basic courses than honors courses to staff at most schools, but is also driven by a substantial fraction of schools that do not track their classes (so that all classrooms at the school are coded as being taught at the basic level). Taken together, these tables demonstrate that teaching in multiple levels and subjects during one's career is the norm, rather than the exception.

5 Results

5.1 General and Context-Specific Experience Profiles

Table 4 displays the estimated experience profiles for each type of experience. Column 1 contains estimates of the returns to teaching experience that are general to all subject-level combinations. While there are small gains to the first two or three years of general experience, those gains do not seem to persist, although results are quite noisy for higher levels of experience (since we must observe the entire history of teacher assignments, only the cohorts of new teachers from the late 1990's are observed at the higher levels of experience in our sample). The magnitudes of these estimates are far smaller than the standard returns to experience estimated in the literature. The results in Column 2 shed light on the source of this discrepancy: much of the returns to experience generally estimated in the literature are actually specific to the subject the teacher taught. Since teachers frequently reteach the same subject many times, subject-specific experience and total years of experience are highly correlated. Thus, when returns to context-specific experience are not separated from returns to overall years of experience, the returns to subject-specific experience will generally be reflected in a larger estimated returns to general experience. Column 2 shows that teaching a subject for the second time increases the teacher's expected performance by .040 test-score standard deviations, relative to the first attempt. An extra year of subject-specific experience increases, performance by an additional .031 standard deviations, while a third year of subject-experience adds an additional .037 standard deviations. Gains begin to level off beyond the third year of experience, but overall teachers with between 3 and 10 years of subject-specific experience are about .12 student level standard deviations more effective than teachers with the same overall years of teaching experience but who are teaching the subject for the first time. While the point estimates seem to suggest that years 11-14 of subject-specific experience provide large gains relative to years 7-10, these gains are too noisily estimated to draw such an inference.

Columns 3 and 4, by contrast, show that the returns to level-specific and subject-level-specific experience seem to be virtually non-existent, once years of course-specific and total experience have been taken into account. None of the point estimates in either experience profile is statistically significant at even the 10th percentile, and most are very close to 0 in magnitude. Indeed, an F-test cannot reject the hypothesis that all of the point estimates in the level-specific and subject-level specific experience profiles are jointly equal to 0 (Table 6).

By summing across columns in Table 1, we can determine the returns to experience for a teacher who never changes the course-level he/she teaches. After 10 years, such a teacher is predicted to perform .183 standard deviations better than a novice teacher. Since many teachers teach the same course-level every year (perhaps in addition to other courses), this sum is particularly well identified. Most of the sampling error in the estimates comes from decomposing this sum into the four experience components.

Given the insignificance of the level and subject-level results from Table 4, in Table 5 we present results from a specification in which all elements of the level-specific and subject-level specific experience profiles are restricted to be zero. The results for total and subject-specific experience change very little, although the profiles are closer to increasing monotonically. Furthermore, the standard errors drop dramatically, confirming a small but non-zero payoff to the first couple of years of general experience. The returns to subject-specific experience still dwarf those of total experience, however. Due to the superior precision of the estimates, the counterfactual simulations presented in Section 7 will utilize the estimated profiles from Table 5.

5.2 The Variance of General and Context-Specific Components of Permanent Teaching Talent

Table 7 contains the results of the decomposition of the variance in permanent teacher quality into general, subject-specific, level-specific, and subject-level specific components. The first column displays the decomposition obtained from imposing Assumption 2A, in which all between school-subject-level variation in student performance is attributed to differences in school and unobserved student inputs. The row labeled “School-Course-Level-Teacher FE” provides the total estimated variance in teaching effectiveness across randomly sampled school-teacher-course-level combinations, which combines all four permanent components. The point estimate is .0184, implying that a one standard deviation increase in combined permanent teaching effectiveness is associated with a .135 increase in expected student performance. 83% of this variance in permanent teacher quality can be attributed to general teacher talent that is common to all subject-level combinations (See the row labeled “School-Teacher FE”). A student assigned to a teacher whose average effectiveness across the subject-level combinations he/she teaches is one standard deviation above average can expect a .123 standard deviation increase in test score performance relative to being assigned the average teacher at the school in the absence of knowledge about the chosen teacher’s level-specific or subject-specific skill.

Course-specific skill and level-specific skill each make up about 12% and 5% of the total variance in permanent teaching effectiveness across randomly chosen school-teacher-course-level combinations. Getting a teacher whose course-specific skill is one standard deviation above the average for a particular course increases expected performance by about .046 test score standard deviations. Note that this is still enough to move a student who would have otherwise scored at the 50th percentile to the 52nd percentile statewide. However, the variation in permanent course-specific skill is quite small relative to the returns to course-specific experience discussed above. Getting a teacher whos level-specific skill is one standard deviation above the average for a particular level increases expected performance by .03 test score standard deviations, only enough to move a student from the 50th to the 51st percentile.

Finally, the course-specific, level-specific, and general components of permanent skill combine to explain the full variance in permanent teacher skill across classroom contexts. There does not seem to be such a thing as subject-level-specific talent. In other words, a teacher’s permanent talent for teaching, say, honors biology, can be fully explained by the teacher's general teaching talent across subjects and levels, combined with his/her talent for teaching honors level courses and his/her talent for teaching biology courses, respectively.

Column 3 of Table 7 shows the alternative decomposition of permanent teacher skill that comes from imposing Assumption 2B, in which all variation in average student performance across school-subject-level combinations is attributed to differences in average teacher quality. Not surprisingly, this increases each of the variance components substantially. Note, though, that the fractions of variance explained by each component stay roughly similar to what they were under Assumption 2A. Perhaps the most compelling result from Column 3 is that the variance in level-specific skill is still only .0014, even under an assumption designed to maximize the variation attributed to teacher talent. Similarly, subject-level-specific talent does not appear to exist under Assumption 2B either. Under Assumption 2B, a one standard deviation increase in general teacher talent is associated with a .221 increase in average student performance across subject-level combinations, while a one standard deviation increase in course-specific teacher talent is associated with a .057 increase in expected student performance relative to a teacher with no comparative advantage or disadvantage at teaching the chosen subject. The results under Assumption 2C (Column 2) stem from removing only the between-school variation from the component attributed to general teacher talent. They provide a middle ground estimate of the standard deviation in general teacher talent of .178 test score standard deviations.

6 Testing the Validity of the Identifying Assumptions

6.1 Testing for Endogenous Responses to School-Year Shocks

A potential threat to the validity of the experience profile estimates raised in Section 3 stems from the possibility that reallocation of teachers across subjects and levels might be more likely when a school is enduring its relatively ineffective years (independently of the contributions of its teachers). This could occur if an inexperienced principal enters the school who has a different conception of how teachers should be allocated. It could also occur if teachers are more likely to quit during a school's relatively ineffective years, creating holes in course or level offerings that other teachers must be forced to fill. One way to test for this possibility is to examine whether schools' relatively low (or relatively high) year-specific residuals disproportionately occur with particular experience profiles. However, if we use residuals from the estimated model, any correlation between experience profiles and school-year deviations will already be reflected in biased experience profile estimates, so that the residual will have been purged of any information it might have contained about endogenous responses to school-year shocks. On the other hand, if we use residuals in which the estimated experience profiles have not been removed, then school-year average residuals will naturally be correlated with the experience profile composition of the teachers in the school-year via the causal effect of teacher context-specific experience.

The second problem can be solved, however, by re-weighting the classroom residual averages that compose school-year averages to account for differences in experience profile composition across school-years. To see how this might be done, let Z_{ict} represent student i's residual test score in classroom c at time t, where the predicted effect of all inputs in the baseline model except teacher experience have been removed:

$$\begin{aligned} Z_{ict} &= Y_{ict} - X_{ict}\hat{\beta}_{jl} - \hat{\delta}_{sjl} - \hat{\mu}_{srjl} \\ &= \hat{d}^{tot}(exp^{tot}_{rt}) + \hat{d}^{j}(exp^{j}_{rt}) + \hat{d}^{l}(exp^{l}_{rt}) + \hat{d}^{jl}(exp^{jl}_{rt}) + \hat{\epsilon}_{ict} \end{aligned} \tag{21}$$

We can form student-weighted school-year average residuals by weighting the average residuals of the classrooms in the school-year by the number of students they contained:

$$
\begin{aligned}
\overline{Z}_{st} &= \frac{1}{C_{st}} \sum_{c \in (s,t)} w_c \overline{Z}_c \\
&= \frac{1}{C_{st}} \sum_{c \in (s,t)} w_c \overline{\hat{d}}_c^{tot} + \overline{\hat{d}}_c^{j} + \overline{\hat{d}}_c^{l} + \hat{d}_c^{jl} + \overline{\hat{\epsilon}}_c \\
&\approx \frac{1}{C_{st}} \sum_{c \in (s,t)} w_c \overline{\hat{d}}_c^{tot} + \overline{\hat{d}}_c^{j} + \overline{\hat{d}}_c^{l} + \hat{d}_c^{jl} + \phi_{st} \qquad (22)
\end{aligned}
$$

where $w_c = \frac{N_c}{N_{sy}}$, and we have assumed for simplicity that classroom, teacher-year, and student error components average to approximately zero within a school-year.[12]

Equation 22 makes clear that schools which have a disproportionate fraction of classrooms taught by teachers with low stock of context-specific experience will tend to have negative school-year average residuals. However, we can remove this effect of experience composition by replacing w_c with $w_{exp(c)}$, where $w_{exp(c)}$ is the fraction of classrooms in the full sample featuring teachers with the same experience profile as the teacher who taught classroom c. Imagine for now that each school-year featured the full support of teaching profiles (though perhaps with different frequencies). Then the reweighted school-year average residual yields:

$$
\begin{aligned}
\tilde{Z}_{st} &= \sum_{c \in (s,t)} w_{exp(c)} \overline{Z}_c \approx \sum_{exp \in \mathcal{EX}} w_{exp} \overline{\hat{d}}^{tot}(exp) + \overline{\hat{d}}^{j}(exp) + \overline{\hat{d}}^{l}(exp) + \hat{d}^{jl}(exp) + \phi_{st} \\
&= K + \phi_{st} \qquad (23)
\end{aligned}
$$

where K is a constant that reflects the average contribution of teacher experience in the full sample (given the normalization chosen). Thus, the reweighted school-year average residual gives us an unbiased estimator of the school-year deviation in quality from the school's long run average.

[12] To the extent that there are relatively few teacher-years some school-years, our test statistic also will reflect endogenous responses to teacher-year shocks, another potential threat to validity.

Given the set $\{\tilde{Z}_{st}\}$, we can then examine whether particular experience levels disproportionately occur during schools' relatively ineffective years by taking a profile-specific weighted average of $\tilde{Z}_{st}$, where the weight on each school-year is the fraction of classrooms in the full sample featuring the experience profile in question that appeared within that school-year:

$$E[\phi_{st}|exp_{rt}^{k} = exp'] = \sum_{st \in \mathcal{ST}} w_{st}(exp')(\tilde{Z}_{st} - K) \tag{24}$$

Unfortunately, each four-dimensional experience profile is not observed in each school-year (a failure of common support), so that we cannot fully purge the effect of experience profile composition within a school-year by reweighting observed classroom averages. We approximate the true reweighted average as best by distributing the weight that would have been placed on unobserved profiles to observed profiles based on the L1 distance between the unobserved and observed profiles (passed through a normal kernel to smooth this distribution). While this method increases substantially the weight placed on profiles in the underrepresented region of the distribution of four-dimensional experience profiles, in school-years where all of the teachers have relatively low predicted experience component of productivity, no amount of reweighting will possibly allow the observed experience composition of the school-year to approximate the full sample distribution. However, this failure of common support biases our test against us, since our test statistic will identify spurious differences in average school-year shocks across profiles. Thus, our reweighting estimator allows us to place an upper bound on the bias produced from endogenous school-year. After forming the set of estimated school-year shocks $\{\tilde{Z}_{st}\}$, we calculate averages of these shocks for each four dimensional experience profile, then regress these averages on the design matrix for the four additively separable experience profiles to ascertain the degree to which the bias will be reflected in each of the four dimensions. Table 8 displays these upper bound estimates of bias from endogenous responses to school-year shocks. The estimates are less than .01 student-level standard deviations for nearly all levels of experience across the four profiles. While there may be a slight downward bias in the returns to general experience and a slight upward bias in the returns to subject-specific experience, the magnitudes are far too small to explain the general pattern of results. Furthermore, as emphasized, the true bias from endogenous responses to school-year shocks is likely to be even smaller, given that the test itself is biased in the direction of the estimates of returns from Table 4.

6.2 Testing for Endogenous Responses to Teacher-Year Shocks

COMING SOON

6.3 Misspecification Tests: Testing Correlation Between Teacher Talent and Teacher-Specific Returns to Experience

COMING SOON

6.4 Misspecification Tests: Testing the Additive Separability of Context-Specific Experience Profiles

COMING SOON

7 Gauging the Magnitude of Achievement Gains from Efficient Use of Context-Specific Teacher Experience

7.1 Methodology

The large gains to subject-specific experience suggest that rotating teachers across subjects can potentially result in substantial efficiency losses. To gauge the magnitude of such losses, in this section we present a counterfactual simulation in which we project the level of performance that could be achieved statewide if each principal exploited the full value of the accumulated stock of context-specific experience of the members of his or her teaching staff. To ensure that the simulation captures feasible reallocations, we hold fixed the number of classrooms of each subject-level combination at the levels that actually prevailed at each school in each year. Furthermore, we also hold fixed the total number of classrooms taught by each teacher in each year, since principals may have been constrained in the workload they could assign to their more experienced teachers.[13]. Also, because we do not observe the full teaching histories of any teacher who began teaching before the sample begins in 1995, we do not reallocate the classrooms taught by such teachers. Thus, the efficiency gains produced by our simulation will be a lower bound on the true efficiency gains available to be reaped (though this lower bound will increase toward the true predicted efficiency gain as we move through years of the sample).

Note, however, that although we only observe student test scores in the 10 tested subjects, we observe the full set of course assignments for each teacher. Hence, we can thus construct post-1995 teaching histories across all standard subjects in North Carolina (such as English 2, or Calculus). Consequently, we report two sets of results. The first contains efficiency gains from only the tested subjects, since these are the subjects on which the general and context-specific experience

[13]For example, these teachers may also have been teaching untested classes, or performing other valuable services to the school, such as lunchroom monitoring, advising student clubs, or coaching student athletic teams)

profiles were estimated. The second contains the combined efficiency gains from all subjects, under the assumption that the course-specific and general experience profiles can be extrapolated to the untested subjects. Such extrapolation beyond the support of the subject distribution used during estimation may be plausible in this context, since the tested subjects sampled from a variety of fields (science, math, social studies, and literature). We have no immediate reason to suspect that subject-specific experience would matter less in World History relative to U.S. History, or in English 2 relative to English 1.

To see how such a counterfactual would be implemented, consider the allocation of teachers to classrooms that takes place at a particular school over the set of years in our sample. In theory, we might want to solve the dynamic problem of choosing sequences of yearly allocations to maximize the average test score performance over the entire sample. However, a principal source of dynamic gains would stem from principals experimenting to learn each teacher's comparative teaching advantages. The results presented in the previous section, though, suggest that subject-specific and level-specific permanent teaching talent is quite small relative to either context-specific experience or permanent teaching skill that is general across contexts. Consequently, to highlight the potential gains from exploiting context-specific experience, we ignore any potential efficiency gains from matching teachers to their permanent comparative advantages in the simulations below[14].

Furthermore, solving the dynamic problem requires specifying principal's expectations about the probability that each teacher will remain at the school in each future year as well as expectations about the number of classrooms they will need to fill in each subject-level in each future year. This is particularly problematic for the last few years of the sample, where we cannot observe what will happen. Perhaps more importantly, under the assumption of perfectly general permanent teacher skill, the conditions under which the dynamic solution strays from the static solution at a particular point in time are peculiar, and dynamic gains are unlikely to be of first order importance.[15] Thus,

[14]specifically, we impose $\mu_{srjl} = \mu_r \forall srj \in \mathcal{SRJ}$

[15] For dynamic gains to exist, the principal needs to be able to predict both a change in the number or distribution of subject-level combinations and a change in teaching staff; otherwise the optimal allocation involves keeping all assignments the same year after year and letting all four kinds of experience accumulate. To see how such a situation

there is little loss of generality in assuming that the principal simply re-solves the static problem of maximizing the expected average test score of his/her students in each year, taking the set of classrooms and teachers to be matched in each year as exogenously given at the start of the year.

To formulate the static problem, let $\mathcal{J}$ represent the set of subjects, $\mathcal{L}$ represent the set of levels, and let $\mathcal{JL}$ be the set of subject-level combinations. Let C_{jl} be the number of classes to be staffed in subject-level combination $jl \in \mathcal{JL}$, with $N_c = \sum_{jl \in \mathcal{JL}} C_{jl}$ denoting the total number of classes to be staffed. Let $\mathcal{R}$ represent the set of teachers, with R elements. As before, ex_r^j captures the number of years of experience teaching subject j for teacher r, ex_r^l captures teacher r's years of experience teaching level l, and ex_r^{jl} captures teacher r's experience teaching the subject-level combination jl. Note that general gains from total years of experience, ex_r^{tot}, can be ignored in this simulation, since reallocations that hold each teacher's teaching load fixed do not affect aggregate student performance under the additively separable production function we have assumed. Similarly, student contributions $X_{it}\beta$ can be ignored, since they are assumed to be constant across counterfactual reallocations.

Using the estimated context-specific experience profiles ($\hat{d}^j(*)$,$\hat{d}^l(*)$,$\hat{d}^{jl}(*)$), we can predict a

might create scope for dynamic gains, suppose that there are three levels of context-experience, (low (L), moderate (M), and high(H) corresponding to 0,1 and 2+ years of experience), and suppose there are two subjects to staff for each of three years. One of the subjects will be replaced in the second year by a new subject, so that subject-specific experience will be 0 for whoever teaches the course. Suppose that in the first year the school has a novice teacher and a teacher with 1 year of subject-specific experience in the persistent course. The static solution for the first year is to assign the novice teacher to the course that's about to be eliminated, since $(L, M) \succ (L, L)$. If the discount rate is sufficiently high, the dynamic solution will involve moving the experienced teacher to the course to be eliminated, since this allows the profile $(L, L), (M, L), (H, M)$ rather than $(M, L), (L, L), (M, M)$. Similar gains might accrue if the principal can predict a decrease in enrollment in the future based on middle school enrollment, expects to eliminate honors level courses, and also expects to lay off the moderately experienced teachers in favor of the novice teachers (in which case he/she might want to allocate the novice teachers to the persistent basic level courses). Note that even these examples would not have produced dynamic gains if the principal could not predict which teachers would stay, if the teachers likely to quit had sufficiently valuable experience, and if the subject-level subject-level composition of courses wasn't predictably changing. Our judgement is that such potential gains are too rare (and too subtle!) to be worth considering. Other sources of dynamic gains can appear if we allow heterogeneity in the returns to experience by permanent teacher quality, or context-specific experience profiles that vary by subject or level, but even these dynamic gains require a change in subject-level course composition.

counterfactual performance of teacher r in course c at time $t = 0$ via:

$$\hat{\overline{Y}}^{c}_{r0} = d^{j}(exp^{j(c)}_{r0}) + d^{l}(exp^{l(c)}_{r0}) + d^{jl}(exp^{jl(c)}_{r0}) \tag{25}$$

The goal is to choose the mapping $f : \mathcal{C} \to \mathcal{R}$ from classrooms to teachers that maximizes the sum of student test scores, subject to the constraints that each teacher can only teach in as many classrooms as they were observed teaching in at time t (denoted $\overline{C}_r$), and every classroom must be taught by exactly one teacher:

$$\begin{aligned}
&\max_{f:\mathcal{C}\to\mathcal{R}} \sum_{c\in\mathcal{C}} \hat{\overline{Y}}^{c}_{f(c)} \\
&s.t. \ \sum_{r} 1(f(c) = r) = 1 \ \forall \ c \\
&s.t. \ \sum_{c} 1(f(c) = r) = \overline{C}_r \ \forall \ r
\end{aligned} \tag{26}$$

where $1(f(c) = r)$ indicates that teacher r taught course c.

This optimization problem can be recast as a binary integer programming problem:

$$\begin{aligned}
&\max_{\mathbf{x}} \mathbf{a} * \mathbf{x} \\
&s.t. \ M_c * \mathbf{x} = 1 \ \forall \ c \\
&s.t. \ N_r * \mathbf{x} = \overline{C}_r \ \forall \ r \\
&s.t. \ \mathbf{x} \in \{0, 1\}
\end{aligned} \tag{27}$$

$\mathbf{a}$ consists of a $1x(C * R)$ row vector of predicted student performances for each potential teacher-classroom combination:

$$\mathbf{a} = \begin{pmatrix} \hat{\overline{Y}}_1^1 & \dots & \hat{\overline{Y}}_C^1 & \hat{\overline{Y}}_1^2 & \dots & \hat{\overline{Y}}_C^2 & \dots & \hat{\overline{Y}}_1^R & \dots & \hat{\overline{Y}}_C^R \end{pmatrix}$$

$\mathbf{x}$ consists of a $(C * R)x1$ vector of potential teacher assignments:

$$\mathbf{x} = \begin{pmatrix} x_1^1 \\ \vdots \\ x_C^1 \\ x_1^2 \\ \vdots \\ x_C^2 \\ \vdots \\ x_1^R \\ \vdots \\ x_C^R \end{pmatrix}$$

where $x_c^r = 1(f(c) = r)$, an indicator for whether teacher r is assigned to classroom c.

M_c consists of a $1xC * R$ row vector capturing the number of teachers assigned to classroom c (restricted to be 1 $\forall$ c):

$$M_c = \begin{pmatrix} \underbrace{\overbrace{0\dots0}^{c-1}\ 1\ \overbrace{0\dots0}^{C-c}}_{\text{repeated R times}} \dots \overbrace{0\dots0}^{c-1}\ 1\ \overbrace{0\dots0}^{C-c} \end{pmatrix}$$

N_r consists of a $1xC * R$ row vector capturing the number of classrooms taught by teacher r

(restricted to be equal to $\overline{C}_r$, the number taught in the sample):

$$N_r = \left(\overbrace{0\ldots0}^{(r-1)*C} \underbrace{1\ldots1}_{C} \overbrace{0\ldots0}^{(R-r)*C} \right)$$

We solve this linear programming problem for each school in the first year of the sample using the simplex method. We then update each teacher's context-specific experience profile for the second year given the experience they gained under the optimal assignment in the first year. We repeat this process until the end of the sample so as to reap the long-run rewards associated with accumulating high levels of relevant context-specific experience.

While this procedure captures the gains that could have been reaped by the end of each year had the principal maximized the value of context-specific experience in each school starting in 1997 (the first year of the sample), note that the small payoff in the first few years conflates the fact that past switching has limited potential gains from re-optimizing with the fact that relatively few teachers are being reallocated (because we do not observe the classroom assignment histories for the vast majority of the teachers in the first few years). Thus, we also solve the linear programming problem in each year t holding fixed observed teacher assignments up through $t-1$. These results reflect the payoff to the first year of optimal reallocation. In this case,the reallocation payoff increases over the years only because we observe complete classroom histories for a greater fraction of teachers (making them eligible for reallocation). The difference between the dynamic and static simulations highlights the potential long-run payoff of maximizing context-specific experience year after year.

7.2 Results from Counterfactual Simulations

COMING SOON

8 Conclusions

This paper introduces and implements a method for decomposing worker productivity into task-specific and general components of both experience and persistent talent. For high school teachers, the bulk of productivity gains from experience are specific to the tasks (subjects) that the teacher has taught, while the bulk of permanent talent is general across all subject-level combinations.

Since the variation in general talent and the value of subject-specific experience are similar in magnitude, effective personnel management for high school administrators requires a mix of selecting/deselecting of teachers combined with retention of teachers who have considerable experience in a particular subject. Since neither level-specific skill nor level-specific experience seems to be important for teacher productivity, honors classes may be used as a non-pecuniary reward for effective teaching or other undesirable tasks (lunch duty?) without any efficiency loss, to the extent that teachers prefer to teach them. Thus, in addition to the practical importance of learning how to better manage the existing stock of public school teachers, the teacher context also represents a case in which allowing the task-specificity of worker productivity to vary across permanent and experience components turns out to be critical for correctly determining the optimal recruitment and assignment policy for an organization's workers.

Note, though, that the results of the decomposition we estimate may not generalize to other occupations or even to alternative definitions of teachers' tasks. In particular, the set of tasks we considered were still fairly similar in scope. For example, we might observe greater variation in task-specific talent among teachers if we included productivity as a high school athletic coach as one of a teacher's tasks. Similarly, developing students' cognitive and non-cognitive skills might represent two different tasks facing a teacher even within a given classroom context, and teachers good at teaching abstract concepts may not be good at handling student emotional crises. Fortunately, a similar decomposition may be estimated in any context in which worker productivity may be measured at the task level, and where the mix of tasks changes over time.

References

Aaronson, Daniel, Lisa Barrow, and William Sander (2007) 'Teachers and Student Achievement in Chicago Public Schools.' *Journal of Labor Economics* 25(1), 95–135

Clotfelter, Charles, Helen Ladd, and Jacob Vigdor (2007) 'Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects.' *Economics of Education Review* 26(6), 673–682

Hanushek, Eric, John Kain, Daniel O'Brien, and Steven Rivkin (2005) 'The Market for Teacher Quality.' NBER Working Paper 11154, National Bureau of Economic Research, Inc

Harris, Douglas, and Tim R. Sass (2006) 'Value-Added Models and the Measurement of Teacher Quality.' Unpublished Manuscript

Jackson, C. Kirabo (2012) 'Do High School Teachers Really Matter?' NBER Working Paper 17722, National Bureau of Economic Research, Inc

——— (2013) 'Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence from Teachers.' *Review of Economics and Statistics*

Mansfield, Richard (2013) 'Teacher Quality and Student Inequality.' Working Paper

Ost, Ben (2013) 'How Do Teachers Improve? The Relative Importance of Specific and General Human Capital.' *American Economic Journal - Applied Economics (Forthcoming)*

Rivkin, Steven, Eric Hanushek, and John Kain (2005) 'Teachers, Schools, and Academic Achievement.' *Econometrica* 73(2), 417–458

Rockoff, Jonah (2004) 'The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data.' *American Economic Review: Papers and Proceedings of the One Hundred Sixteenth Annual Meeting of the American Economic Association* 94(2), 247–252

Sass, Tim, Jane Hannaway, Zeyu Xu, David Figlio, and Li Feng (2010) 'Value Added of Teachers in High-Poverty Schools and Lower Poverty Schools: Implications for Research, Policy and Management.' Working Paper 52, National Center for Analysis of Longitudinal Data in Education Research

Wiswall, Matthew (2013) 'The Dynamics of Teacher Quality.' *Journal of Public Economics*

100, 61–78

9 Tables and Figures

Table 1: Experience Distribution among Classes taught by 2nd and 3rd year Teachers

	Years of Experience			Fraction of
	Course	Level	Crs-Lvl	Classes
Second-Year Teachers	1	1	1	25.0%
	1	1	0	28.6%
	1	0	0	20.6%
	0	1	0	19.1%
	0	0	0	6.7%
	Total classes taught by a 2nd-year teacher:			**280,140**
Third-Year Teachers	2	2	2	17.8%
	2	2	1	2.1%
	2	2	0	15.5%
	2	1	1	2.6%
	2	1	0	10.8%
	2	0	0	10.7%
	1	2	1	5.4%
	1	2	0	8.9%
	1	1	1	3.0%
	1	1	0	5.2%
	1	0	0	2.5%
	0	2	0	9.9%
	0	1	0	2.7%
	0	0	0	2.9%
	Total classes taught by a 3rd-year teacher:			**244,965**

Notes: The table presents classrooms in our final estimation sample having teachers in either their second or third year teaching. Notice that since multiple classes can be taught in a year, this explains seemingly impossible experience combinations such as the second row for second-year teachers. For instance, a teacher who taught both remedial chemistry and honors biology in year one would have this type of experience profile teaching remedial biology in year two.

Table 2: Teacher Mobility Across Courses: Regression Sample

Course		Course: Algebra 1	Algebra 2	Biology	Chemistry	Civics	E/L/P	English	Geometry	Physics	Physical Sciences	U.S. History
Course	Algebra 1	**1,780**	594	29	8	5	9	20	637	15	36	17
		1	0.334	0.016	0.004	0.003	0.005	0.011	0.358	0.008	0.020	0.010
	Algebra 2	594	**741**	6	3	2	3	3	380	10	8	4
		0.802	1	0.008	0.004	0.003	0.004	0.004	0.513	0.013	0.011	0.005
	Biology	29	6	**754**	133	7	20	25	7	41	339	19
		0.038	0.008	1	0.176	0.009	0.027	0.033	0.009	0.054	0.450	0.025
	Chemistry	8	3	133	**369**	1	1	2	4	82	213	1
		0.022	0.008	0.360	1	0.003	0.003	0.005	0.011	0.222	0.577	0.003
	Civics	5	2	7	1	**440**	141	9	2	0	5	245
		0.011	0.005	0.016	0.002	1	0.320	0.020	0.005	0.000	0.011	0.557
	E/L/P	9	3	20	1	141	**572**	32	3	0	22	311
		0.016	0.005	0.035	0.002	0.247	1	0.056	0.005	0.000	0.038	0.544
	English	20	3	25	2	9	32	**892**	7	0	16	27
		0.022	0.003	0.028	0.002	0.010	0.036	1	0.008	0.000	0.018	0.030
	Geometry	637	380	7	4	2	3	7	**855**	9	10	5
		0.745	0.444	0.008	0.005	0.002	0.004	0.008	1	0.011	0.012	0.006
	Physics	15	10	41	82	0	0	0	9	**169**	114	1
		0.089	0.059	0.243	0.485	0.000	0.000	0.000	0.053	1	0.675	0.006
	Physical	36	8	339	213	5	22	16	10	114	**734**	16
	Sciences	0.049	0.011	0.462	0.290	0.007	0.030	0.022	0.014	0.155	1	0.022
	U.S.	17	4	19	1	245	311	27	5	1	16	**783**
	History	0.022	0.005	0.024	0.001	0.313	0.397	0.034	0.006	0.001	0.020	1

Notes: The top entry in the (i,j)-th cell is the number of teachers who are observed teaching in both the i-th and the j-th difficulty level (not necessarily in the same year). The bottom entry of the (i,j)-th cell is the fraction of teachers ever observed teaching the i-th difficulty level who are also observed teaching the j-th difficulty level at some point during the sample.

Table 3: The Pattern of Teacher Mobility Across Difficulty Levels

		Difficulty Level	
		Low	High
Diff. Level	Low	**5,956**	2,974
		1	0.499
	High	2,974	**3,003**
		0.990	1

Notes: The top entry in the (i,j)-th cell is the number of teachers who are observed teaching in both the i-th and the j-th difficulty level (not necessarily in the same year). The bottom entry of the (i,j)-th cell is the fraction of teachers ever observed teaching the i-th difficulty level who are also observed teaching the j-th difficulty level at some point during the sample.

Table 4: The Pattern of Teacher Mobility Across Course-Levels for the Mathematics Field

Course- Level		Course-Level: Algebra 1 Low	Algebra 1 High	Algebra 2 Low	Algebra 2 High	Geometry Low	Geometry High
Algebra 1	Low Level	**1,767**	71	551	238	587	248
		1	0.040	0.312	0.135	0.332	0.140
	High Level	71	**84**	20	15	23	17
		0.845	1	0.238	0.179	0.274	0.202
Algebra 2	Low Level	551	20	**682**	236	330	138
		0.808	0.029	1	0.346	0.484	0.202
	High Level	238	15	236	**295**	128	78
		0.807	0.051	0.800	1	0.434	0.264
Geometry	Low Level	587	23	330	128	**786**	270
		0.747	0.029	0.420	0.163	1	0.344
	High Level	248	17	138	78	270	**339**
		0.732	0.050	0.407	0.230	0.796	1

Notes: The top entry in the (i,j)-th cell is the number of teachers who are observed teaching in both the i-th and the j-th difficulty level (not necessarily in the same year). The bottom entry of the (i,j)-th cell is the fraction of teachers ever observed teaching the i-th difficulty level who are also observed teaching the j-th difficulty level at some point during the sample.

Table 5: The Effect of Total, Course, Level, and Course-Level Experience on Student Test Scores

Years Experience	Total (1)	Course (2)	Level (3)	Course-Level (4)
1 yr	0.019	0.040**	0.007	0.009
	[0.017]	[0.014]	[0.016]	[0.014]
2 yrs	0.043**	0.073***	0.004	-0.002
	[0.023]	[0.020]	[0.022]	[0.019]
3 yrs	0.051**	0.110***	-0.019	-0.007
	[0.028]	[0.025]	[0.027]	[0.024]
4 yrs	0.028	0.122***	-0.007	0.000
	[0.032]	[0.029]	[0.032]	[0.028]
5-6 yrs	0.029	0.118***	0.001	-0.002
	[0.038]	[0.034]	[0.038]	[0.033]
7-10 yrs	0.024	0.122**	0.001	0.009
	[0.047]	[0.041]	[0.047]	[0.040]
11-14 yrs	0.006	0.204***	0.030	-0.057
	[0.057]	[0.052]	[0.058]	[0.052]

$N = 747,890$ student-class observations. Robust standard errors are in brackets. Significance at the 10%, 5%, and 1% levels are represented by ***, **, and * respectively.

Table 6: The Effect of Total and Course Experience on Student Test Scores (Restricted Specification)

Years Experience	Total (1)	Course (2)
1 yr	0.025***	0.049***
	[0.007]	[0.006]
2 yrs	0.046***	0.072***
	[0.010]	[0.009]
3 yrs	0.033***	0.104***
	[0.012]	[0.011]
4 yrs	0.021*	0.121***
	[0.013]	[0.013]
5-6 yrs	0.029**	0.117***
	[0.016]	[0.015]
7-10 yrs	0.025	0.129***
	[0.020]	[0.019]
11-14 yrs	0.031	0.166***
	[0.025]	[0.027]

$N = 747,890$ student-class observations. Robust standard errors are in brackets. Significance at the 10%, 5%, and 1% levels are represented by ***, **, and * respectively.

Table 7: Joint Significance of Different Experience Profiles

Experience	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total	X							X
Subject		X			X		X	X
Level			X			X	X	X
Subject-Level				X	X	X	X	X
F-ratio	1.62	5.10***	1.20	1.34	10.44***	1.57	7.59***	32.08***
p-value	0.123	0.000	0.347	0.226	0.000	0.079*	0.000	0.000

Table 8: True Variances in Fixed Effects

	True Variance Lower Bound	True Variance Middle	True Variance Upper Bound
School-Course-Level FE	0.0356	0.0170	—
School-Course FE	0.0335	0.0170	—
Level Deviations from School-Course FE	0.0020	—	—
School-Level FE	0.0182	0.0170	—
Course Deviations from School-Level FE	0.0174	—	—
School FE	0.0170	0.0170	—
Course-Level Deviations from School FE	0.0186	—	—
School-Course-Level-Teacher FE	0.0184	0.0371	0.0540
School-Course-Teacher FE	0.0175	0.0357	0.0526
Level Deviations from School-Course-Teacher FE	0.0009	0.0014	0.0014
School-Level-Teacher FE	0.0163	0.0338	0.0507
Course Deviations from School-Level-Teacher FE	0.0022	0.0033	0.0033
School-Teacher FE	0.0152	0.0321	0.0491
Course-Level Deviations from School-Teacher FE	0.0032	0.0049	0.0049

Estimates derived from the estimated variance less the student weighted average of standard errors in our fixed effect estimates.

Table 9: Estimates of the Bias in Experience Profiles Due to Endogenous Teacher Assignment Responses to School-Year Shocks

Years Experience	Total	Course	Level	Course-Level
	(1)	(2)	(3)	(4)
1 yr	-0.007	0.008**	0.003	-0.005
	[0.004]	[0.004]	[0.004]	[0.004]
2 yrs	-0.005**	0.000***	0.003	0.004
	[0.005]	[0.005]	[0.005]	[0.005]
3 yrs	-0.007**	0.007**	0.004	0.005
	[0.006]	[0.006]	[0.006]	[0.006]
4 yrs	-0.005	0.016***	-0.002	0.001
	[0.006]	[0.007]	[0.006]	[0.007]
5-6 yrs	-0.008	0.005***	0.001	0.007
	[0.006]	[0.007]	[0.006]	[0.007]
7-10 yrs	-0.008	0.006**	0.002	0.004
	[0.006]	[0.007]	[0.006]	[0.007]
11-14 yrs	—	—***	—	—
	—	—	—	—

$N = 747,890$ student-class observations. Robust standard errors are in brackets. Significance at the 10%, 5%, and 1% levels are represented by ***, **, and * respectively.