Task-Specific Experience and Task-Specific Talent

Jason Cook — Cornell University

Richard Mansfield — Cornell University

PRELIMINARY AND INCOMPLETE

DO NOT CITE WITHOUT AUTHORS' PERMISSION

September 13, 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

Worker 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 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 on their given 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 assign better 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 elementary[1], middle school[2], or high school level[3], while teachers tend to improve with experience by around .05 test score standard deviations in

[1] citations here
[2] citations here
[3] citations here

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.[4]

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 environment. 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

[4] citations here

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, 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 either the subject or even the subject-level that the teacher taught. We find no evidence, however, that level-specific experience has any value if it is not applied to the same subject. 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 course-specific or level-specific teaching talent.

Of course, the knowledge that a large fraction of the gains from experience are subject and subject-level 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. While we can only observe the prior course and level histories of the subset of teachers who began teaching after 1995, our simulation from this subsample indicates that the statewide average gain from efficient use of context-specific experience (relative to the observed allocation) are as large as X student-level standard deviations by 2009, and would continue to grow over time. Since there are no changes in total teaching load for any teacher, these efficiency gains could potentially be reaped with essentially zero cost.

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. Each score is standardized so that the distribution of test scores in each classroom c and each year t has zero mean and unit variance. Let $r(i, c, t)$ denote the teacher that taught student i in course c at time t. Similarly, let $s(i, c, t)$ denote the school at which student i took course 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".[5] 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:

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

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 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

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

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^{j}_{rt}, exp^{l}_{rt}, and exp^{jl}_{rt} 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 expect 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 Students' Unobserved Inputs and Teacher Experience

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

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

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

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

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

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

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

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

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.

Nonetheless, there are still 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 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. 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. While stories might be told that suggest a correlation between experience (general or context-specific) and other unobserved inputs (such as test score measurement error or classroom shocks), they are generally less plausible. Consequently, in Section 7, we test for violations of Assumption 1 triggered by selection on ϕ_{st} and ν_{rt}. We find that ... (TBD) ....

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 with the exogenous timing of experience accumulation, 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.[6]

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 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).

[6]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 $k \in \{tot, j, l, lk\}$.

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

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

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 2 can be imposed on the

model by including school-subject-level fixed effects (δ_{sjl}), and normalizing the 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, if Honors Biology students score on average .1 standard deviations better than Chemistry students at a particular school (after controlling for past test scores and observables), the entire difference is attributed to differences in students and principals, so that if a particular teacher's Biology students score .1 standard deviations better than her Chemistry students, that teacher will be assumed to be equally effective at teaching both Biology and Chemistry. 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

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

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, if honors biology students score on average .1 standard deviations better than remedial chemistry stu-

dents at a particular school (after controlling for past test scores and observables), the entire difference is attributed to differences in the average quality of teachers teaching these courses, so that if a particular teacher's Biology students score .1 standard deviations better than her Chemistry students, that teacher will be assumed to be .1 test score standard deviations more effective at teaching Biology than Chemistry. In other words, even though the teacher is at the median 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

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

Estimates from such a model are be 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 3: Conditional Mean Independence of Students' Unobserved Inputs and Teacher Experience

$$E[\epsilon_{ict}|(s,r,j,l) = (\tilde{s},\tilde{r},\tilde{j},\tilde{l}),(ex_{rt}^{tot},ex_{rt}^{j},ex_{rt}^{l},ex_{rt}^{jl}) = (\tilde{ex}_{rt}^{tot},\tilde{ex}_{rt}^{j},\tilde{ex}_{rt}^{l},\tilde{ex},_{rt}^{jl}),X_{ict} = X'_{ict}] \quad (9)$$

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

Assumption 3 states that the identify of the school, teacher, subject, and level do not provide further information about any unobserved inputs, conditional on 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.

Nonetheless, 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 num-

ber 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_{RBH}. 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 Assumption 3 and any of Assumptions 2A-2C, 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:

$$
\begin{aligned}
& E[Z_{ict}|(s,r,j,l) = (s',r',j',l'), (ex^{tot}_{rt}, ex^{j}_{rt}, ex^{l}_{rt}, ex^{jl}_{rt}) = (4,3,2,2)] \\
& = \delta_{sjl} + \mu_{srjl} + d^{tot}(4) + d^{j}(3) + d^{l}(2) + d^{jl}(2) \qquad (10)
\end{aligned}
$$

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:

(11)

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

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:

(12)

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

The first component in 12 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:

$$(\mu_{srjl} - \overline{\mu}_{sr}) \equiv \tilde{\mu}_{srjl} = \overline{\tilde{\mu}}_{srj} + \overline{\tilde{\mu}}_{srl} + (\tilde{\mu}_{srjl} - \overline{\tilde{\mu}}_{srj} - \overline{\tilde{\mu}}_{srl}) \tag{13}$$

The first component of 13 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 13 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 $Var(\overline{\tilde{\mu}}_{srj})$ as the variance in level-specific teaching talent. The third component of 13 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 $Var(\tilde{\mu}_{srjl} - \overline{\tilde{\mu}}_{srj} - \overline{\tilde{\mu}}_{srj})$ 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:

$$(14) \qquad \frac{1}{N}\sum_{(i,c,t)\in\mathbf{ICT}}(\hat{\mu}_{srjl})^2 = \frac{1}{N}\sum_{(i,c,t)\in\mathbf{ICT}}(\mu_{srjl})^2 + \frac{1}{N}\sum_{(i,c,t)\in\mathbf{ICT}}(\xi_{srjl})^2$$

where N is the number of test scores in the sample, and **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:

$$(15) \qquad \hat{Var}(\mu_{srjl}) = \frac{1}{N}\sum_{(i,c,t)\in\mathbf{ICT}}(\hat{\mu}_{srjl})^2 - \frac{1}{N}\sum_{(i,c,t)\in\mathbf{ICT}}(\xi_{srjl})^2.$$

ξ_{srjl} is not observed, but

$$(16) \frac{1}{N}\sum_{(i,c,t)\in\mathbf{ICT}}(\xi_{srjl})^2 \approx \frac{1}{N}\sum_{(i,c,t)\in\mathbf{ICT}}E[(\xi_{srjl})^2] = \frac{1}{N}\sum_{(i,c,t)\in\mathbf{ICT}}(se(\xi_{srjl}))^2,$$

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

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

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

$$(18) \qquad \hat{Var}(\overline{\mu}_{sr}) = \hat{Var}(\mu_{srjl}) - \hat{Var}(\tilde{\hat{\mu}}_{srjl})$$

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.

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.[7]

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 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

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

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 the last section of the paper, we examine whether teaching skill is actually field-specific. We 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 form 1997 - 2009 in up to eleven subjects and three class difficulty levels[8]. During the sample period, North Carolina provided a standardized curriculum which was measured using required year-end tests in English 1, Econ/Law/Politics, U.S. History, Algebra 1, Algebra 2, Geometry, Biology, Chem-

[8]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 and international baccalaureate as "AP", leaving us with three difficulty levels: basic, honors, and AP.

The classroom description files provide a breakdown of academic level as well into: Two-year classes, regular, honors, college placement, advanced placement, and other courses

istry, Physics, and Physical Science.[9]

The data contain a large number of current student inputs[10] as well as past student inputs[11] and teacher information[12].

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, course-level-specific experience. For instance, consider a teacher who taught only Honors Chemistry in their first year, Table 1 shows the different combinations of experience the teach would have for a given second year class. Notice in row (1) that the only way to gain course-level experience is to have previously taught the exact course-level.[13]

[9]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 tested between 2004 and 2005.

[10]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.

[11]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

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

[13]Note however, that if a teacher teaches multiple classes in one year, this can explains seemingly impossible experience combinations such as a second year teacher having one year of course experience, one year of level experience and no course-level experience. For instance, a teacher who taught both honors chemistry and basic biology in year one would have this type of experience profile teaching honors biology in year two.

In Table 2, we show how experience evolves for the second- and third- year teachers in our final sample.

4.2 Difficulty Level Specfication

Since the difficulty level of the class comprises one dimension of what we consider a teacher's task-specific experience, we will provide evidence motivating our difficulty level specification. The first three columns in Table 3 show the overall distribution of the number of levels offered across our sample. For most years, a majority of schools offer only one level for each course. Of the schools offering more than one level per course, we see that depending on the year there are about three to four times as many courses available at two levels than three. Columns (4)-(6) and (7)-(9) break out the distribution of levels offered for courses offered at one and two difficulty levels respectively. We see that most schools that offer only one course, offer it at the basic level. For courses offered at two difficulty levels, a majority of the classes are offered at the basic and honors levels.

It is plausible that courses in schools only offering basic and honors difficulty levels could have similar classroom compositions for the same course offered in another school, but with basic and AP levels available. That is, the motivated students sort into the most difficult class available and vice versa. In an attempt to capture the student sorting based on the difficulty levels available for a given course, we generate our final measure of class difficulty level by recoding basic, honors, and AP classes into "low", "mid', and "high" difficulty levels. For courses only available at two difficulty levels, we recode them into two categories, "mid" and "high" and preserve

the difficulty rank ordering.[14]

4.3 Final Specification

Our empirical strategy requires that student test score observations be matched to the teachers who taught the class. Unfortunately, this connection is not explicitly given in the data. However, since we have data on the classes taught by each teacher and the classes taken by each student, we utilize a fuzzy match algorithm that matches on counts of classroom demographics.

In our final specification we drop observations for students to whom we cannot match a teacher, for whom are not in a "basic", "honors", or "AP" difficulty level class, for students in classes with fewer than 4 classmates. Further we drop observations with missing current/past test scores, or teacher experience values. Lastly, because the experience profiles must be observed from the beginning of a teacher's career, we drop students taught by teachers for whom we do not observe them entering the teaching pool,[15]

4.4 Teacher Mobility

Tables 4 and 5 depict teacher mobility in our final sample across difficulty levels and courses, respectively. While there is mobility across all three levels, there is

[14]Specifically, for courses offered only at the basic and honors levels, we consider basic as "mid" level and honors as "high" level. For basic and AP offerings, we consider these respectively as "mid" and "high". Finally, courses offered at the honors and AP levels are considered as "mid" and "high" respectively.

[15]That is we keep all teachers for whom we observe an entry level paycode.

considerably more mobility between "Low" and "Mid" difficulty level classes. As expected, most of the teacher mobility across courses occurs within different fields (e.g. math, science, etc.) reflecting the need for certification to teach within a field.

Table 6 isolates cross-course-level mobility for teachers in the field of mathematics.

5 Results

5.1 General and Context-Specific Experience Profiles

Table 7 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. In particular, the first year of experience increases teachers' effectiveness by .016 test-score-level standard deviations across all subject-levels. Third year teachers are on average .029 standard deviations more effective than they were during their first year. Teachers with 5-6 years of experience are .037 standard deviations more effective than novice teachers, while those with 7-10 years are on average .051 standard deviations more effective, and those with 11-14 years of experience tend to be .075 standard deviations more effective. While this profile of estimates matches the concave pattern usually observed in the literature, the magnitudes are smaller than are usually reported. 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 .064 test-score standard deviations, relative to the first attempt. An extra year of subject-specific experience increases, performance by an additional .025 standard deviations, while a fourth year of subject-experience adds an additional .030 standard deviations. Gains begin to level off beyond the fourth year of experience, but overall teachers with over 7 years of subject-specific experience are .136 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.

Column 3 shows that the returns to level-specific experience, by contrast, seems to be virtually non-existent, once years of course-specific and total experience have been taken into account. None of the point estimates in the level-specific expeirence profile is statistically significant at even the 10th percentile, and they are all very close to 0 in magnitude. Indeed, an F-test cannot reject the hypothesis that all of the point estimates in the level-specific experience profile are jointly equal to 0 (Table 8).

Column 4, however, shows that experience teaching the particular subject-level to which a teacher has been assigned does seem to have little bit of value. The first couple years of subject-level experience do not seem to increase performance beyond what one would expect based on the total, course-specific, and level-specific experience. However, an accumulation of four or more years of subject-level experience increases predicted performance by between .025 and .035 test score standard devi-

ations. An F-test can easily reject the hypothesis that all of the point estimates in the subject-level-specific experience profile are jointly equal to zero (See Table 8). The results suggest that the (small) payoff to teaching the same difficulty level again may come from being able to sharpen particular lectures, rather than learning how to target material more generally to students associated with a particular level of ability or preparation. By summing across columns in Table 1, we can determine the returns to experience for a teacher who never changes the course he/she teaches. After 10 years, such a teacher is predicted to perform a whopping .253 standard deviations better than a novice teacher.

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

Table 9 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 .013, implying that a one standard deviation increase in combined permanent teaching effectiveness is associated with a .114 increase in expected student performance. About 70% 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 .095 increase in test score performance relative to the average teacher at the school in the absence of knowledge about the teacher's level-specific or subject-specific skill.

Course-specific skill and level-specific skill each make up about 15% of the total variance in permanent teaching effectiveness across randomly chosen school-teacher-course-level combinations. Getting a teacher whose course-specific skill or level-specific skill is one standard deviation above the average for a particular course or level increases expected performance by about .045 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.

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 2 of Table 9 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 (specifically, by a factor of about 5). 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 2 is that the variance in level-specific skill is still only .005, 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 .212 increase in average student performance across subject-level combinations, while a one standard deviation increase in course-specific teacher talent is associated with a .110 increase in expected student performance relative to a teacher with no comparative advantage or disadvantage at teaching the chosen subject.

6 Testing the Validity of the Identifying Assumptions

COMING SOON

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

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 describe the results of 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.[16]. 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,

[16]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)

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 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[17].

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.[18] Thus, there is little loss of generality in assuming that the principal simply re-solves the static problem of maximizing the expected average

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

[18] 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 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.

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 $\mathcal{S}$ represent the set of subjects, $\mathcal{L}$ represent the set of levels, and let $\mathcal{SL}$ be the set of subject-level combinations. Let C_{sl} be the number of classes to be staffed in course-level combination $sl \in \mathcal{SL}$, with $N_c = \sum_{sl \in \mathcal{SL}} C_{sl}$ 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 counterfactual performance of teacher r in course c at time $t = 0$ via:

$$\hat{\overline{Y}}_{r0}^{c} = d^j(exp_{r0}^{j(c)}) + d^l(exp_{r0}^{l(c)}) + d^{jl}(exp_{r0}^{jl(c)}) \tag{19}$$

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:

$$\max_{f:\mathcal{C}\to\mathcal{R}} \sum_{c\in\mathcal{C}} \hat{\overline{Y}}^{c}_{f(c)}$$

$$s.t. \sum_{r} 1(f(c)=r) = 1 \ \forall \ c$$

$$s.t. \sum_{c} 1(f(c)=r) = \overline{C}_r \ \forall \ r \tag{20}$$

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

This optimization problem can be recast as a linear programming problem:

$$\max_{\mathbf{x}} \mathbf{a} * \mathbf{x}$$

$$s.t. \ M_c * \mathbf{x} = 1 \ \forall \ c$$

$$s.t. \ N_r * \mathbf{x} = \overline{C}_r \ \forall \ r \qquad s.t. \mathbf{x} \geq 0 s.t \mathbf{x} \leq 1 \tag{21}$$

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

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

$\mathbf{x}$ consists of a $C * Rx1$ 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 $1 \ \forall \ c$):

$$M_c = \left(\underbrace{\overbrace{0 \dots 0}^{c-1} \ 1 \ \overbrace{0 \dots 0}^{C-c} \dots \overbrace{0 \dots 0}^{c-1} \ 1 \ \overbrace{0 \dots 0}^{C-c}}_{\text{repeated R times}} \right)$$

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):

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

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 large long-run rewards associated with accumulating high levels of relevant context-specific experience. Table X compare the average predicted value of student test scores under optimal assignment of teachers to classrooms to the average predicted value under the actual assignment in each year of the simulation.

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, Table Y presents results from solving 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 Table Y, 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 X and Y highlights the potential long-run payoff of maximizing context-specific experience year after year.

RESULTS TO BE ANNOUNCED

8 Conclusions

COMING SOON

9 Tables and Figures

Table 1: Example of Experience Accumulation for a First-year Honors Chemistry Teacher

	Second-Year Class	General	Course	Level	Course-Level
(1)	Honors Chemistry	1	1	1	1
(2)	Honors Biology	1	0	1	0
(3)	Basic Chemistry	1	1	0	0
(4)	Basic Biology	1	0	0	0

Table 2: 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 3: Distribution of Levels by School-Course Offering

	# Levels Offered			Level Distribution: School-Courses with one Level			School-Courses with two Levels		
	1	2	3	B	H	AP	B-H	B-AP	H-AP
1997	1204	797	67	1131	69	4	723	71	3
	.58	.39	.03	.94	.06	.00	.91	.09	.00
1998	1,235	792	86	1157	73	5	722	70	0
	.58	.37	.04	.94	.06	.00	.91	.9	0
1999	2,059	1666	96	1830	218	11	1582	78	6
	.54	.44	.03	.89	.11	.01	.95	.05	.00
2000	2125	1691	554	1736	254	135	1357	242	92
	.49	.39	.13	.82	.12	.06	.80	.14	.05
2001	2165	1785	499	1739	282	144	1430	253	102
	.49	.40	.11	.80	.13	.07	.80	.14	.06
2002	2151	1795	527	1781	244	126	1501	221	73
	.48	.40	.12	.83	.11	.06	.84	.12	.04
2003	2134	1915	535	1674	332	128	1592	256	67
	.47	.42	.12	.78	.16	.06	.83	.13	.03
2004	2075	1668	394	1649	334	92	1427	176	65
	.5	.4	.1	.79	.16	.04	.86	.11	.04
2005	1876	1645	365	1516	300	60	1439	145	61
	.48	.42	.09	.81	.16	.03	.87	.09	.04
2006	2149	2259	571	1591	423	135	1976	202	81
	.43	.45	.11	.74	.20	.06	.87	.09	.04
2007	1631	1824	373	1227	363	41	1697	87	40
	.43	.48	.10	.75	.22	.03	.93	.05	.02
2008	2741	2091	307	2005	654	82	1951	91	49
	.53	.41	.06	.73	.24	.03	.93	.04	.02
2009	3028	2043	328	2114	869	45	1898	79	66
	.56	.38	.06	.70	.29	.01	.93	.04	.03

Table 4: The Pattern of Teacher Mobility Across Difficulty Levels

		Difficulty Level		
		Low	Mid	High
Difficulty Level	Low	**10,553**	6,624	2,720
		1	0.628	0.258
	Mid	6,624	**7,701**	2,503
		0.860	1	0.325
	High	2,720	2,503	**3,143**
		0.865	0.796	1

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 Pattern of Teacher Mobility Across Courses

		Course										
		Algebra 1	Algebra 2	Biology	Chemistry	Civics	E/L/P	English	Geometry	Physics	Physical Sciences	U.S. History
Course	Algebra 1	**3,422**	1,045	44	23	17	24	52	1,133	24	62	25
		1	0.305	0.013	0.007	0.005	0.007	0.015	0.331	0.007	0.018	0.007
	Algebra 2	1,045	**1,361**	6	12	5	4	2	695	19	22	5
		0.768	1	0.004	0.009	0.004	0.003	0.001	0.5110.	014	0.016	0.004
	Biology	44	6	**1,762**	234	12	20	32	10	79	654	21
		0.025	0.003	1	0.133	0.007	0.011	0.018	0.006	0.045	0.371	0.012
	Chemistry	23	12	234	**688**	2	1	2	8	135	388	1
		0.033	0.017	0.340	1	0.003	0.001	0.003	0.012	0.196	0.564	0.001
	Civics	17	5	12	2	**1,173**	374	27	4	1	9	574
		0.014	0.004	0.010	0.002	1	0.319	0.023	0.003	0.001	0.008	0.489
	E/L/P	24	4	20	1	374	**1,159**	55	2	0	22	541
		0.021	0.003	0.0170	0.001	0.3230	1	0.047	0.002	0.000	0.019	0.467
	English	52	2	32	2	27	55	**2,837**	8	0	27	46
		0.018	0.001	0.010	0.001	0.010	0.019	1	0.003	0.000	0.010	0.016
	Geometry	1,133	695	10	8	4	2	8	**1,554**	15	19	7
		0.729	0.447	0.006	0.005	0.003	0.001	0.005	1	0.010	0.012	0.005
	Physics	24	19	79	135	1	0	0	15	**297**	200	2
		0.081	0.064	0.266	0.455	0.003	0.000	0.000	0.051	1	0.673	0.007
	Physical Sciences	62	22	654	388	9	22	27	19	200	**1,455**	17
		0.043	0.015	0.449	0.267	0.006	0.015	0.019	0.013	0.137	1	0.012
	U.S. History	25	5	21	1	574	541	46	7	2	17	**1,543**
		0.016	0.003	0.014	0.001	0.372	0.351	0.030	0.005	0.001	0.011	1

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 course (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 course and difficulty level combination who are also observed teaching the j-th course and at some point during the sample.

Table 6: The Pattern of Teacher Mobility Across Course-Levels for the Mathematics Field

			Course-Level								
			Algebra 1			Algebra 2			Geometry		
			Low	Mid	High	Low	Mid	High	Low	Mid	High
Course- Level	Algebra 1	Low Level	**2,767**	1,216	489	854	551	212	884	605	234
			1	0.439	0.177	0.305	0.199	0.077	0.319	0.219	0.085
		Mid Level	1,216	**1,759**	441	524	373	149	530	417	180
			0.691	1	0.251	0.298	0.212	0.085	0.301	0.237	0.102
		High Level	489	441	**717**	215	158	91	213	166	73
			0.682	0.615	1	0.300	0.220	0.127	0.297	0.232	0.102
	Algebra 2	Low Level	854	524	215	**1,178**	618	222	537	378	150
			0.717	0.445	0.183	1	0.525	0.188	0.456	0.321	0.127
		Mid Level	551	373	158	618	**766**	191	358	257	106
			0.719	0.487	0.206	0.807	1	0.249	0.467	0.336	0.138
		High Level	212	149	91	222	191	**275**	134	95	54
			0.771	0.542	0.331	0.807	0.695	1	0.487	0.345	0.196
	Geometry	Low Level	884	530	213	537	358	134	**1,330**	729	286
			0.665	0.398	0.160	0.404	0.269	0.101	1	0.548	0.215
		Mid Level	605	417	166	378	257	95	729	**917**	265
			0.660	0.455	0.181	0.412	0.280	0.104	0.795	1	0.289
		High Level	234	180	73	150	106	54	286	265	**355**
			0.659	0.507	0.206	0.423	0.299	0.152	0.806	0.746	1

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 course and difficulty level combination (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 course and difficulty level combination who are also observed teaching the j-th course and difficulty level combination at some point during the sample.

Table 7: The Effect of Total, Course, Level, and Course-Level Experience on Student Test Scores

Years Experience	Total	Course	Level	Course-Level
	(1)	(2)	(3)	(4)
1 yr	0.016***	0.064***	-0.001	0.004
	[0.006]	[0.005]	[0.005]	[0.005]
2 yrs	0.029***	0.089***	0.005	0.009*
	[0.008]	[0.006]	[0.006]	[0.006]
3 yrs	0.027***	0.119***	-0.002	0.012**
	[0.009]	[0.008]	[0.008]	[0.007]
4 yrs	0.033***	0.129***	0.001	0.025***
	[0.011]	[0.009]	[0.009]	[0.008]
5-6 yrs	0.037***	0.134***	0.002	0.024***
	[0.013]	[0.011]	[0.010]	[0.009]
7-10 yrs	0.051***	0.137***	0.000	0.034***
	[0.015]	[0.013]	[0.011]	[0.011]
11-14 yrs	0.075***	0.136***	0.014	0.028
	[0.019]	[0.019]	[0.017]	[0.024]

Notes: $N = 1,912,870$ student-class observations. Robust standard errors are in brackets. Significance at the 10%, 5%, and 1% levels are represented by ***, **, and * respectively. Each column represents the results of an individual regression. Each regression includes as regressors: Number of classes and number of difference course-levels taught contemporaneously by the student's teacher, a cubic polynomial for both the student's 7^{th} and 8^{th} grade math and reading scores, classroom composition (including class size, the class's average 8^{th} grade math and reading scores, racial composition, and number of gifted students in math and reading), as well as dummies 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, future plans, extracurricular activities, and finally missing indicators for 7^{th} and 8^{th} grade math and reading scores. These regressions allow these covariates to have heterogeneous effects for each course-level combination.

Table 8: 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	3.64***	44.13***	0.69	2.07**	20.81***	1.63*	18.73***	107.78***
p-value	0.001	0.000	0.691	0.063	0.000	0.063	0.000	0.000

Table 9: True Variances in Fixed Effects

	True Variance Lower Bound	True Variance Middle	True Variance Upper Bound
School-Course-Level FE	0.049	0.011	—
School-Course FE	0.044	0.011	—
Level Deviations from School-Course FE	0.005	0.000	—
School-Level FE	0.013	0.011	—
Course Deviations from School-Level FE	0.036	0.000	—
School FE	0.011	0.011	—
Course-Level Deviations from School FE	0.038	0.000	—
School-Course-Level-Teacher FE	0.013	0.051	0.063
School-Course-Teacher FE	0.011	0.047	0.058
Level Deviations from School-Course-Teacher FE	0.002	0.005	0.005
School-Level-Teacher FE	0.011	0.039	0.050
Course Deviations from School-Level-Teacher FE	0.002	0.012	0.012
School-Teacher FE	0.009	0.034	0.045
Course-Level Deviations from School-Teacher FE	0.004	0.018	0.018

Notes: These come from the estimated variance less the student weighted average of standard errors in our fixed effect estimates.