Task-Specific Experience and Task-Specific Talent: Decomposing the Productivity of High School Teachers

Jason B. Cook[a], Richard K. Mansfield[1b]

[a]*Cornell University, B30 Ives Hall, Ithaca, NY 14853; E-mail: jbc266@cornell.edu*

[b]*Cornell University, 266 Ives Hall, Ithaca, NY 14853; E-mail: rm743@cornell.edu*

Abstract

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 difficulty levels to which teachers are assigned to provide identifying variation, we show that a substantial part of the productivity gains to teacher experience are actually subject-specific. Similarly, while three-quarters of the variance in the permanent component of productivity among teachers is portable across subjects and levels, there exist non-trivial subject-specific and level-specific components. Counterfactual simulations suggest that maximizing the test-score contribution of task-specific experience and task-specific talent can increase student performance by as much as .04 test score standard deviations relative to random assignment of teachers to classrooms. *JEL Codes:* I21, I28, J24, J45, L23.

Keywords: Task-Specific Human Capital, Teacher Experience, Teacher Value-Added, Teacher Quality

---

[1]Corresponding Author. Telephone: 781-724-1418. Fax: 607-255-4496.

1. Introduction

How should principals allocate teachers to courses so as to maximize the teachers' contribution to student achievement?

The optimal course assignment depends on teachers' existing comparative advantages in different courses or classroom environments, as well as the extent to which current assignments will increase teachers' future productivity (or the principal's information about such productivity). However, the large literature on teacher value-added and the returns to teaching experience (discussed below) has focused primarily on estimating variation in teacher productivity that is assumed (or restricted) to be common to all course or grade assignments. If this assumption is true, then any allocation of existing teachers with fixed course loads will feature the same distribution of value-added contributions to achievement. However, if this assumption is false, then improving the mechanism by which teachers are assigned to courses may yield significant gains at potentially low cost (Jacob and Rockoff (2011)).

To see this, suppose first that teachers have pre-determined comparative advantages for particular subjects or difficulty levels. Then course swaps among teachers could produce efficiency gains if both teachers move toward their relatively more effective courses. Furthermore, if principals cannot ascertain teachers' relative strengths at the time of hire, then the optimal assignment strategy might involve rotating teachers across several different classroom contexts early in their careers so as to produce information about the courses the teachers will be particularly effective at teaching. Permanent subject-specific skill might exist, for example, if a teacher's choice of undergraduate major leads to a deeper understanding of the content in a particular subject (e.g. Physics rather than Biology). Permanent level-specific skill might exist, for example, if a teacher's natural charisma or sense of humor leads to strong classroom control skills that are comparatively more important in the remedial or basic level courses where students may tend to be less engaged.

Now suppose instead that task-specific skill is primarily learned through experience rather than predetermined prior to 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 with each opportunity to teach it. Track- or level-specific experience might also be significant if the appropriate pace at which to deliver content depends on student skill and is slowly calibrated with practice. In addition, experience teaching a certain subject-level combination (e.g. honors biology) might be particularly valuable if it allows teachers to hone particular lectures over time that would be inappropriate for either a different level or a different subject.

More generally, knowledge of the importance of task-specific talent and task-specific experience is essential for any employer wishing to maximize the productivity of his/her 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 featuring smaller experience gains and a larger variance in task-specific talent, the optimal strategy is to lay off or reassign low performing workers in an attempt to either improve general worker skill or identify superior worker-task matches.

Thus, in this paper we introduce a method for decomposing worker productivity into components relating to general talent, task-specific talent, general experience, and task-specific experience. Our decomposition requires data featuring (1) signals (possibly noisy) of individual workers' task-specific output, (2) histories of worker task assignments, and (3) frequent rotation of workers across tasks. We implement our method using the context of high school teachers, in which tasks consist of teaching particular subjects in particular tracks (difficulty levels).[3]

Specifically, we use administrative panel data from the North Carolina Education Research Data Center (NCERDC) to decompose teacher effectiveness at improving student achievement into (1) a set of permanent components capturing general talent, subject-specific talent, level-specific talent, and subject-level specific talent, and (2) a set of functions capturing returns from general experience, subject-specific experience, level-specific experience, and subject-level-specific experience. The data track teachers and students in the universe of public high schools in North Carolina from 1997-2009. Critically, the data feature over 24,000 within-teacher changes in subject assignments and over 18,000 changes in academic-level assignments. Such rich data permit estimation of an education production function that features general, subject-specific, level-specific, and subject-level-specific experience profiles as well as a full set of school-teacher-subject-level fixed effects. The flexibility of our model allows us to control for many potential biases that might otherwise accompany endogenous course assignment decisions. We then use our results to project the potential student achievement gains that could be reaped by better utilizing knowledge about course-specific experience and skill relative to the course assignment patterns observed in the data.

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 test score standard deviations at both the primary or secondary school levels.[4] Similarly, the existing literature suggests that 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.[5] More recent studies relax the functional form assumptions imposed in these early studies and find somewhat larger returns to high levels of teaching experience.[6]

However, this literature has generally ignored the possibility that the baseline 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 productivity across different classroom environments will return a weighted average of teacher productivity across the environments each teacher actually faced (weighted by the fraction of time spent in each environment). To the extent that teachers face different classroom contexts over their careers, models that impose homogeneity of returns to experience across different classroom environments may underestimate the gains to context-specific experience. Similarly, to the extent that teachers' classroom environment remain somewhat stable during their career, such models may overestimate the returns to general experience.

A few papers, though, have addressed various aspects of the context-specificity of teacher productivity, mostly using elementary or middle school data. Jackson (2013) shows that a substantial portion of the

[3]Throughout the paper below, we use the term "task" to refer to a subject-level combination, while we use the term "context" more generally to refer to particular characteristics or features of the classroom environment, which include but are not limited to the subject and level.

[4]For primary school estimates, see, for example, Rockoff (2004), Hanushek et al. (2005), Clotfelter et al. (2006), Sass et al. (2014), Boyd et al. (2008), Jackson and Bruegmann (2009), Harris (2009), Harris and Sass (2011), and Jackson (2013). For secondary school estimates, see, for example, Aaronson et al. (2007), Jackson (2014), and Mansfield (Forthcoming). Harris (2009), by contrast, finds little evidence of returns to experience using high school data from Florida.

[5]e.g. Rivkin et al. (2005), Clotfelter et al. (2007).

[6]Wiswall (2013) and Papay and Kraft (2015).

variation in teacher contributions to student achievement is specific to the school in which a teacher has taught. Lockwood and McCaffrey (2009) and Aucejo (2011) examine the degree to which teachers have comparative advantages at teaching relatively high versus low ability students, and find evidence that a small component of teaching productivity is specific to student ability level. Perhaps more closely related to our paper is work by Ost (2014) showing that teachers who always repeat elementary grade assignments improve 35% faster than teachers who never repeat grade assignments. Similarly, Master et al. (2012) show that the efficacy of a teacher teaching English-language learners (ELL) depends on his/her experience teaching the ELL population. The paper most closely related to ours is Condie et al. (2014), who also consider subjects as tasks. They demonstrate the existence of meaningful comparative advantages of elementary teachers at teaching English vs. math. These papers, however, focus either on context-specific experience or context-specific skill, rather than providing a unified treatment of both factors.

Given the applicability of our methodology to the broader worker-to-task assignment problem, our paper also contributes to a growing literature on task-specific human capital, brought to the forefront by Gibbons and Waldman (2004), which considers the possibility that a considerable portion of a worker's human capital might be specific to the particular tasks the worker has performed at the jobs the worker has held.[7] Part of the literature on task-specific human capital either has assumed that only the experience component of human capital is task-specific (e.g. Gibbons and Waldman (2004), Clement et al. (2007), and DeAngelo and Owens (2012)). Alternatively, Polataev and Robinson (2008) assume that the degree of task-specificity is common between the talent and experience components of human capital, while Gathmann and Schoenberg (2010) instrument to remove the influence of the task-specific talent component in order to focus on task-specific experience. Yamaguchi (2012) allows for both task-specific talent and gains to task-specific experience, but does not have productivity data, and thus must infer the size of each component indirectly from observed sequences of occupational choices.

To preview our results, we find that about a quarter to a third of the returns to years of experience are actually specific to the subject that the teacher taught. We find little evidence of returns to level-specific experience and no evidence of returns to subject-level experience. In agreement with the rest of the value-added literature, we find that the variation in fixed teaching skill is comparable in magnitude to the gains to experience. While 74% of the variance in permanent skill is general to all subjects and levels, we find a small but meaningful role for subject-specific (17%) and level-specific (9%) teaching talent: receiving a teacher whose subject-specific (level-specific) talent is one standard deviation above his/her average among all subjects (levels) he/she teaches increases a student's expected test score by .063 (.046) standard deviations.

We test for and fail to find convincing evidence of estimation biases driven by dynamic assignment responses to teacher-year or school-year shocks or unmodeled teacher-specific heterogeneity in gains from experience. Backcasting tests for bias from non-random student sorting to teachers suggest that, if anything, the significant gains to both general and subject-specific experience that we estimate may be understated. Split-sample forecast tests suggest that our estimates of teachers' combined general and task-specific talent have considerable out-of-sample predictive power, though admittedly slightly less than what a model with no bias or misspecification would imply. While similar split-sample forecast tests for teachers' estimated task-specific comparative advantages (more important for teacher assignment) are underpowered, they do not find evidence of any forecast bias in subject-specific talent estimates, though level-specific talent estimates do not seem to predict out-of-sample comparative advantages nearly as well.

[7] See, for example, Yamaguchi (2012), Clement et al. (2007), Polataev and Robinson (2008), Gathmann and Schoenberg (2010), DeAngelo and Owens (2012).

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 or teacher retirements may require principals to reassign teachers to unfamiliar subjects or tracks. Using our estimated experience profiles, we address this possibility by performing counterfactual simulations in which we reassign the teachers observed teaching in each school-field combination in the chosen year to the courses that were offered at their school at the time in order to maximize student performance, given posterior beliefs about the teachers' course-specific talent as well as the four-dimensional stocks of general and context-specific experience that these teachers possessed at the beginning of the year.

Our simulations indicate that efficient use of task-specific experience and talent can, in principle, increase student achievement non-trivially: relative to random assignment of teachers to classrooms, the efficient allocation raises mean test scores by as much as .04 student-level standard deviations for school-field combinations with seven or more teachers. The degree to which principals' classroom assignments already effectively incorporate information about teacher comparative advantages is difficult to discern; however, under the strong assumption that the information about teachers' subject-specific and level-specific talent reflected in our statewide panel of 1997-2009 test scores is a superset of the information available to principals, our simulations suggest that efficient use of context-specific experience might increase mean test scores in larger high schools by as much as .02-.03 student-level standard deviations relative to the observed patterns of teachers' classroom assignments. Furthermore, since we hold the teaching load fixed for each teacher, these efficiency gains could potentially be reaped with near zero cost.[8] These simulated gains are comparable in magnitude to the gains from subject-specialization in elementary school projected by Condie et al. (2014). We also show that they are comparable in magnitude to the gains administrators could expect to reap from a policy in which the least effective 10% of teachers are removed and replaced by average teachers.

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 school-teacher-subject-level fixed effects and general, subject-specific, level-specific, and subject-level-specific experience profiles. Section 4 discusses the North Carolina administrative data and provides summary statistics illustrating the variation in teacher course assignments. Section 5 presents the parameter estimates from our main specifications. Section 6 presents tests for possible threats to our identifying assumptions and demonstrates the robustness of our results to alternative choices regarding sample selection, variable definition, and model specification. Section 7 describes and presents results from the counterfactual simulations that gauge the test score gains that might be achievable through effective use of a teaching staff's context-specific talent and experience. Section 8 concludes.

2. Model Specification

Because our goal is to determine the relative importance of context-specific teacher skill and experience for test score performance, we craft our specification of the achievement production function in a fashion that permits us 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

[8]Note that we cannot address the possibility that proposed reallocations would either detract from competing non-test score objectives or carry compensating differential costs (e.g. if teachers have strong preferences for teaching courses in their comparative disadvantages).

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 (Basic or Honors).[9] Since North Carolina used different test forms for each subject in each year, we standardize each test score Y_{ict} 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:

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

Because we estimate the model at the classroom level, we aggregate this production function and focus our attention on the classroom-mean of test score performance, denoted Y_{ct}:

$$Y_{ct} = X_{ct}\beta_{jl} + \delta_{sjl} + \mu_{srjl} + d^{gen}(exp^{gen}_{rt}) + d^{j}(exp^{j}_{rt}) + d^{l}(exp^{l}_{rt}) + d^{jl}(exp^{jl}_{rt}) + \epsilon_{ct} \tag{2}$$

X_{ct} represents a vector of class-averages of student observable characteristics and middle school reading and math test scores, along with other classroom characteristics (e.g. class size $|I_c|$) and a full set of calendar year indicators. We allow the coefficients on X_{ct}, β_{jl}, to differ across subject-level combinations.[10] This allows for the possibility that a high class average of 8th grade math scores might be a stronger predictor of class 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 (e.g. labs) than in honors math. X_{ct} is included to control for non-random sorting of students to particular teachers within school-subject-level cells (discussed further in Section 3.2).[11]

δ_{sjl} represents inputs provided by the school-subject-level combination. The set of $\{\delta\}$ parameters will not only capture the contribution of any school-level inputs such as principal quality, neighborhood quality, or quality of the school facilities, they will also capture any variation in the quality of curricula or textbooks across subjects and levels within the school. δ_{sjl} will be estimated via a full set of school-subject-level fixed effects, $\hat{\delta}_{sjl}$. These fixed effects 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 classroom characteristics. Importantly, in practice they will also reflect the contribution of average unobserved inputs of the students who sort into particular school-subject-level combinations. Thus, the school-subject-level design matrix also 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_{srjl} captures the experience-invariant component of teacher r's ability to increase student achievement in subject-level (j,l) at school s. μ_{srjl} will be estimated via a full set of school-teacher-subject-level fixed effects, $\hat{\mu}_{srjl}$. The average school-teacher-subject-level will be normalized to 0 for each school-subject-level in our baseline specification (see Section 3.2 for further discussion), so that $\hat{\mu}_{srjl}$ can be interpreted as the deviation of a particular teacher's performance in a particular subject-level combination from the

[9]Section 4.2 describes how we assign courses to difficulty levels.

[10]The coefficients on the calendar year indicators are restricted to be the same across all subject-levels to improve efficiency.

[11]Given that we include classroom averages of student inputs to better control for sorting on unobservable student characteristics (Altonji and Mansfield, 2014), aggregation of our outcome test scores to the classroom level is essentially without loss of generality. This is because the student-level observables are orthogonal to all the inputs of interest once class averages of these student observables have been conditioned on, since the inputs of interest display no within-class variation.

mean (student-weighted) performance of all teachers that taught in the chosen teacher's school-subject-level combination during the sample (e.g. how a particular honors biology teacher's students performed relative to the honors biology students of his/her colleagues). This specification of the contribution of teacher quality allows the estimation of a fully non-parametric joint distribution of general teacher talent and subject-specific, level-specific, and even subject-level-specific permanent comparative advantages within and across teachers. Note that by including the identity of the school in the definition of the fixed effect, we are allowing each teacher's mean contribution and comparative advantages for particular subjects and levels to be different at each school at which they teach (a teacher who teaches in two schools is essentially treated as two different teachers).

exp_{rt}^{gen} represents the total number of years of general teaching experience that teacher r possessed at the beginning of year t, defined as the number of previous years in which the teacher taught at least one course. Analogously, exp_{rt}^{j}, exp_{rt}^{l}, and exp_{rt}^{jl} represent previous years of experience teaching at least one course in the subject, level, and subject-level combination associated with classroom c, respectively. $d^{gen}(*)$ is a function that captures the part of the gains from years of teacher experience that are portable ("general") across all subjects and levels. 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 affect a teacher's ability to increase student test scores. $d^{gen}(*)$, $d^{j}(*)$, $d^{l}(*)$, and $d^{jl}(*)$ are each flexibly parameterized using indicators for narrow ranges of experience.

Because the estimates from the "baseline" specification in (2) prove to be somewhat sensitive to choice of controls and the exact parametrization of the experience profiles, we also devote considerable attention to a less volatile "restricted" specification in which we constrain $\mu_{srjl} = \overline{\mu}_{sr}\ \forall(j,l)$ and (s,r), allowing us to replace the school-teacher-subject-level fixed effects with school-teacher fixed effects only:

$$Y_{ct} = X_{ct}\beta_{jl} + \delta_{sjl} + \overline{\mu}_{sr} + d^{gen}(exp_{rt}^{gen}) + d^{j}(exp_{rt}^{j}) + d^{l}(exp_{rt}^{l}) + d^{jl}(exp_{rt}^{jl}) + \epsilon_{ct} \quad (3)$$

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

$$\epsilon_{ct} = \phi_{st} + \nu_{rt} + \zeta_{ct} + \frac{1}{|I_c|}\sum_{i \in c} e_{ict} \quad (4)$$

ϕ_{st} captures year-specific deviations in school inputs relative to the sample-wide average for the school-subject-level (e.g. due to school renovation). ν_{rt} represents year-specific deviations in a teacher's quality from what would be expected based on the teacher's time-invariant skill and context-specific experience (e.g. due to teacher illness). ζ_{ct} captures classroom level shocks, such as an uncontrollably disruptive student or the archetypal dog barking outside the classroom window on test day. Finally, e_{ict} represents the contributions of residual student-level inputs that are unpredictable based on observables as well as measurement error reflecting the extent to which the student's exam performance deviates from what the student could have expected to score, given his/her accumulated knowledge in the subject.

3. Identification

3.1. Identifying the Returns to General and Task-Specific Experience

Let $Exp = [Exp^{gen}, Exp^{j}, Exp^{l}, Exp^{jl}]$ represent the random vector of general and context-specific experience stocks for classroom teachers accumulated as of year t, of which each observed combination $[exp^{gen}_{rt}, exp^{j}_{rt}, exp^{l}_{rt}, exp^{jl}_{rt}]$ is a draw. Similarly, let M and D represent random vectors of school-teacher-subject-level and school-subject-level cell indicators, respectively. Each draw from M and D represents a row of the design matrices corresponding to the fixed effects capturing $\{\mu_{srjl}\}$ and $\{\delta_{sjl}\}$, respectively. Finally, let X represent the random vector of observed classroom characteristics, and let ϵ represent the random variable of which ϵ_{ct} is a draw. To identify the functions mapping experience stocks to productivity, $\{d^{gen}(*), d^{j}(*), d^{l}(*), d^{jl}(*)\}$, we assume that the following condition holds:

Assumption 1: Conditional Mean Independence of Time-Varying Unobserved Inputs and Teacher Experience

$$E[\,\epsilon\,|Exp, M, D, X] = E[\,\epsilon\,|M, D, X] \tag{5}$$

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

Recall from (4) that the regression error contains school-year, teacher-year, and classroom shocks (along with class-averages of individual-level unobserved inputs): $\epsilon_{ct} = \phi_{st} + \nu_{rt} + \zeta_{ct} + \frac{1}{|I_c|}\sum_{i \in c} e_{ict}$. Thus, there are a number of sources of possible threats to the validity of 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 be more likely to observe zero subject-specific or level-specific experience when the contribution of time-varying school inputs ϕ_{st} is low. Since year-specific deviations in school quality from the sample-wide average are included in ϵ_{ct}, this scenario violates Assumption 1 and could potentially produce an overestimate of the returns to task-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 or pregnancy). In that case, zero subject-specific or level-specific experience may be observed more frequently when the value of the teacher-year shock ν_{rt} is high. This scenario also violates Assumption 1, and might cause an underestimate of the returns to task-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 experience (since those who survive to the next year of experience will have experienced above-average classroom shocks the previous year, thereby hiding the gains to the next year of experience). Finally, returns to experience could also be overestimated if students with superior unobserved inputs systematically sort into classes within subject-levels taught by teachers with high general or context-specific experience. We address the possibility of such violations of Assumption 1 in Section 6 and find little evidence of violations of sufficient magnitude to produce a substantial bias to any of our profiles.

Despite these concerns, however, note that Assumption 1 is still much weaker than the assumptions required to identify experience profiles in most of the literature, since it conditions on the combined identities of the school, teacher, subject and level. Essentially, the inclusion of school-teacher-subject-level fixed effects (μ_{srjl}) controls for any arbitrary selection of teachers into experience categories based on the fixed component of general or context-specific productivity. Conditioning on the identity of the teacher accounts for the possibility that better teachers persist long enough to gain more experience. Similarly, conditioning on the teacher-subject combination 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 the teacher-level combination accounts for the possibility that persistence at teaching honors 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 and estimation of each of the four experience profiles also requires considerable variation in the history of subject and level assignments across teachers. Such variation is necessary to satisfy the OLS rank condition and, more importantly, to produce sufficiently precise estimates. Appendix A illustrates how identification of the context-specific experience profile in each context dimension might be secured for our baseline model, and provides insight into the patterns of student performance in the data that inform estimates of the experience profile parameters.

The examples in Appendix A reveal that the experience profiles are fully identified from comparisons of different teachers' rates of performance growth (divergence/convergence of average student residuals) across years in which the same subject-level combination was taught. Because the average performance of each teacher in each school-subject-level combination is perfectly fit by the unrestricted school-teacher-subject-level and school-subject-level fixed effects, such cell averages provide no identifying variation for the experience profiles. Put another way, the inclusion of these fixed effects forces the identification of the experience profiles to be delivered exclusively from the path of productivity growth within school-teacher-subject-level combinations.

3.2. Identification of the General and Context-Specific Components of Fixed Teaching Skill

Identifying fixed or pre-determined general and context-specific teaching skill is more difficult. In particular, there is a fundamental identification problem that our model cannot overcome: we cannot distinguish average teaching quality in a particular school-subject-level from school or unobserved student inputs that vary across school-subject-level cells. For example, suppose a school's students score 0.1 student-level standard deviations higher in Biology than in Chemistry. In the absence of restrictions on the distribution of subject-specific teacher skill, we cannot tell whether all the teachers at the school are particularly effective at teaching Biology relative to Chemistry, 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 perform a sensitivity analysis in which we introduce two polar opposite assumptions and one moderate assumption for apportioning the between school-subject-level achievement variation between teachers and other inputs. We decompose the variance in teacher time-invariant productivity into general, subject-specific, level-specific, and subject-level-specific components under each assumption.

The first extreme assumption is that average teacher effectiveness is uniform across all levels, subjects, and schools:

Assumption 2A: Uniform Average Teacher Quality Across Contexts

$$\frac{1}{|I_{sjl}|}\sum_{(i,c,t)\in(s,j,l)}\hat{\mu}_{srjl} = k \text{ for some constant } k,\ \forall\ (s,j,l)\in\mathcal{SJL} \tag{6}$$

where $|I_{sjl}|$ is the number of students observed taking subject j in level l at school s and $\mathcal{SJL}$ is the set of all school-subject-level combinations. This assumption will hold (with a sufficiently large pools of teachers) 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 across subjects, levels, and schools (after removing the part that is predictable based on classroom observables) can be attributed to either school inputs or unobserved student inputs. Assumption 2A can be imposed on the model by estimating a full set of school-subject-level fixed effects ($\hat{\delta}_{sjl}$), and normalizing the student-weighted average teacher-school-subject-level fixed effect to be zero at each school-subject-level: $\frac{1}{|I_{sjl}|}\sum_{(i,c,t)\in(s,j,l)}\hat{\mu}_{srjl} = 0$ (note that the common mean k does not contribute to variance estimates). Under Assumption 2A, if the school average performance difference between Biology and Chemistry is 0.1 standard deviations then a teacher whose Biology students perform 0.1 standard deviations better than her Chemistry students will be assumed to be equally effective at teaching both Biology and Chemistry.

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

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

$$\begin{aligned}&\delta_{sjl} = \overline{\delta}_s\ \forall\ (s,j,l)\in\mathcal{SJL}\\ &\frac{1}{|I_s|}\sum_{(i,c,t)\in s}\mu_{srjl} = k \text{ for some constant } k,\ \forall\ s\in\mathcal{S}\end{aligned} \tag{7}$$

Estimates from such a model are useful for a principal who needs to determine classroom assignments for his/her existing stock of teachers. The principal will only require the decomposition of the within-school variance in time-invariant teacher productivity, and may believe that school inputs are divided relatively equally across subjects and levels. Assumption 2B is implemented by replacing the school-subject-level effects with school fixed effects only, and restricting the average school-teacher-subject-level effect to be 0 at each school.

Finally, the opposite extreme 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 2C: Uniform School and Unobserved Student Quality Across Contexts

$$\delta_{sjl} = k \text{ for some constant } k,\ \forall\ (s,j,l)\in\mathcal{SJL} \tag{8}$$

Assumption 2C will 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 2C can be imposed on the model by excluding school-subject-level fixed effects entirely ($\hat{\delta}_{sjl} = 0\ \forall\ (s,j,l)$), 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 2C, a teacher whose Biology students perform 0.1 standard deviations better than her Chemistry students will be assumed to be 0.1 standard deviations more effective at teaching Biology than Chemistry if the school average performance difference between Biology and Chemistry is 0.1 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 0.1 standard deviations superior on average to the comparison set of Chemistry teachers.

While Assumptions 2A-2C allow us to separate school inputs from teacher inputs, identification of $\{\mu_{srjl}\}$ also requires that other unobserved inputs are not correlated with the observation of a particular teacher in a particular subject-level combination. As before, M and D represent the random vectors of school-teacher-subject-level and school-subject-level cell indicators, while Exp represents the random vector of teacher experience stocks and X represents the random vector of observed classroom characteristics. Similarly, let S represent the random vector of school indicators (draws of which would represent a row of a design matrix for a set of school fixed effects). Then assumptions 3A-3C capture this additional condition for each of the three cases considered:

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

$$\begin{aligned} 3A: \quad & E[\epsilon|M, Exp, X] = E[\epsilon|D, Exp, X] \\ 3B: \quad & E[\epsilon|M, Exp, X] = E[\epsilon|S, Exp, X] \\ 3C: \quad & E[\epsilon|M, Exp, X] = E[\epsilon|Exp, X] \end{aligned} \tag{9}$$

Assumption 3A states that the identity 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 general and context-specific experience of the teacher and the observable classroom characteristics. 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 he/she has more general experience, or that a teacher will be perceived to have a comparative advantage at teaching in a particular context because many of the test-score observations from that context are accompanied by considerable context-specific experience. Assumption 3B is much stronger, since it conditions on the school rather than the school-subject-level, while Assumption 3C, which conditions only on teacher experience stocks and observed classroom inputs, is the strongest assumption of all.

There remain several potential threats to the validity of the fixed effect estimates even in the case of Assumption 3A. Suppose, for example, that a particular teacher is assigned to a room with broken air conditioning each time the teacher teaches honors physics, but is assigned to functioning rooms whenever the teacher teaches honors chemistry. In this case, conditioning on context-specific experience will not remove the correlation between the classroom-level error component ζ_{ct} and the indicator for the school-teacher-subject-level combination associated with the chosen teacher teaching honors physics. 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 the unobserved teacher-year shock ν_{rt} and the indicator for the school-teacher-subject-level combination associated her basic English 1 course.

Perhaps the most serious concern stems from the possibility that unobservably superior students are able to disproportionately select a particular teacher.[12] This possibility is somewhat less likely at the high school

[12] Rothstein (2010) documents non-random student sorting into particular classrooms within North Carolina elementary schools. However, Kinsler (2012) retests the same data, accounting for small sample sizes, and fails to reject such non-random sorting.

level, since class assignments are frequently generated by scheduling algorithms (given students' subject-level choices), making it difficult for students to select into particular classrooms within a subject-level. We rely on classroom averages of student covariates to absorb any within-subject-level sorting based on student inputs. Altonji and Mansfield (2014) show that classroom averages of observable characteristics can in theory absorb all the bias from sorting on both observables *and* unobservables, if the set of observables is diverse enough to span the set of classroom amenities that are driving students to sort. Furthermore, Mansfield (Forthcoming) and Jackson (2014), using the same NCERDC dataset we employ here, find little evidence of remaining student sorting after controlling for track. In Section 6.2, though, we investigate further the possibility that sorting of students to teachers could bias our estimated production function.

Appendix B provides a concrete example that illustrates the kinds of moments in the data that identify time-invariant teaching skill. The example in Appendix B reveals that each school-teacher-subject-level fixed effect $\hat{\mu}_{srjl}$ will be estimated using only a single teacher's performance during the few years in which they taught the subject-level associated with the fixed effect. As such, sampling error for any given fixed effect estimate $\hat{\mu}_{srjl}$ will not converge to zero even with the fairly long panel we employ. Consequently, we do not focus on individual $\hat{\mu}_{srjl}$ estimates, but instead seek to characterize the joint distribution of the components of time-invariant teaching skill. Specifically, we 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.

To see how this may be done, note first that we can rewrite the true value of teacher r's context-specific productivity μ_{srjl} via:

$$\mu_{srjl} = \overline{\mu}_{sr} + (\mu_{srjl} - \overline{\mu}_{sr}) \tag{10}$$

The first component in (10) can be interpreted as the contribution of teacher talent that may be school-specific, but is general or portable across tasks (subject-level combinations) within the school. We will refer to $Var(\overline{\mu}_{sr})$ as the variance in general teaching talent. The second component contains the teacher's persistent subject-level-specific deviation in quality from the teacher's average across all subject-level combinations. This can be interpreted as the teacher's comparative advantage or disadvantage at teaching subject-level combination (j, l). This second component can then be decomposed into three further components:

$$(\mu_{srjl} - \overline{\mu}_{sr}) \equiv \tilde{\mu}_{srjl} = \overline{\tilde{\mu}}_{srj} + \overline{\tilde{\mu}}_{srl} + (\tilde{\mu}_{srjl} - \overline{\tilde{\mu}}_{srj} - \overline{\tilde{\mu}}_{srl}) \tag{11}$$

The first component of (11) can be interpreted as the part of the teacher's comparative advantage at subject-level combination (j, l) that is portable across levels but not subjects. We will refer to $Var(\overline{\tilde{\mu}}_{srj})$ as the variance in subject-specific teaching talent. The second component of (11) can be interpreted as the part of the teacher's comparative advantage at subject-level combination (j, l) that is portable across subjects but not levels. We will refer to $Var(\overline{\tilde{\mu}}_{srl})$ as the variance in level-specific teaching talent. The third component of (11) is the part of a teacher's comparative advantage at (j, l) that is not portable across levels or subjects, and thus could not have been predicted based on the sum of the teacher's subject-specific skill and the teacher's 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.

Note that we do not observe the true variance of school-teacher-subject-level effects, $Var(\mu_{srjl})$, but rather the sample variance, which contains sampling error: $Var(\hat{\mu}_{srjl})$. To recover the true latent variance decomposition, we follow the method of Aaronson et al. (2007) and Mansfield (Forthcoming). Appendix C describes this sampling error correction in detail.

Because we can only estimate a value of $\hat{\mu}_{srjl}$ for those school-teacher-subject-level combinations that we actually observe in the data, the variance in subject-specific and level-specific skill that we estimate will represent the variance among the range of subject and level combinations that principals actually assign.

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 are more relevant and interesting to principals and administrators. Much of the missing variance stems from variation in the strength of teacher's comparative disadvantages among classroom assignments that are never seriously considered by principals (i.e. variation among English teachers in their ability to teach physics). Rather, the choice principals generally face is between hiring a new teacher to teach exactly the courses taught by an exiting teacher and hiring a new teacher to teach different courses while rotating existing teachers who are certified in the chosen field to new subjects or levels within that field (for example, rewarding stayers by letting them teach the honors class that was vacated by the exiting teacher).[13] Given the limited support for the distribution of comparative advantages that underlies our estimates, in our simulations in Section 7 we only reallocate teachers across classrooms within fields.

4. Data

4.1. Overview

The decomposition of worker productivity developed in Sections 2 and 3 requires that the data 1) contain signals of worker output in each task, 2) allow the construction of accurate measures of general and task-specific experience, and 3) exhibit considerable worker rotation among tasks. We employ administrative data provided by the North Carolina Education Research Data Center (NCERDC) that satisfies each of these three conditions for the context of high school teaching.

4.2. Task-Specific Output and Sample Restrictions

The NCERDC data consists of standardized test scores for the universe of public high school students in North Carolina from 1997 - 2009 in eleven subjects and two course difficulty levels.[14]

During the sample period, North Carolina provided a standardized curriculum in each subject and assessed achievement via statewide end-of-course tests.[15] The eleven subjects, which can be grouped into four fields

[13]Teacher certification in North Carolina, as 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).

[14] The student-level End-of-Course test data provide a set of four difficulty level categories (honors, AP, college placement, and other) that do not perfectly match the difficulty level categories provided with the beginning-of-year classroom data (Special Education, Remedial, Basic, Applied/Technical, Honors, Cooperative Education, Advanced Placement, International Baccalaureate, and Non-Classroom), which contain the correct teacher ID (on which the level-specific experience stocks are based). In order to minimize the probability that the relevant level-specific experience of the teacher is mismeasured, we drop student observations coming from classes labeled as Special Education, Cooperative Education, and Non-Classroom. We classify Remedial, Basic, and Applied/Technical classes as "basic" and Advanced Placement, International Baccalaureate, and Honors as "honors". In the rare cases where schools offer distinct Advanced Placement and Honors courses in the same tested subject we drop observations from classrooms where the teacher's relevant level-specific experience depends on whether these two difficulty levels are combined during the construction of level-specific experience stocks.

[15]Note that these tests are subject-specific but not level-specific.

based on common certification requirements, are as follows: *Math*: Algebra 1, Algebra 2, Geometry; *Science*: Biology, Chemistry, Physical Science, Physics; *Social Studies*: Econ/Law/Politics, Civics and Economics, U.S. History; *English*: English 1.[16] Because statewide achievement tests were administered immediately at the conclusion of each year-long course, and the subjects are (largely) distinct from one another, average student performance in each course represents a signal (albeit a noisy, possibly biased one) of the task-specific output of the teacher. [17]

In our framework, accurately distilling the signal of a teacher's task-specific productivity from student sorting requires rich data on student inputs. Fortunately, the NCERDC data contain information about a variety of current student inputs (or proxies for such inputs), as well as past student inputs.[18] In addition, we also include in X_{ct} the number of classes and number of distinct courses taught contemporaneously by the student's teacher in order to capture teacher workload, and include indicators for whether the student's teacher taught the current subject, level, and subject-level (and whether he/she taught at all) in the previous year to capture depreciation of human capital. We also include in X_{ct} a full set of dummies for the calendar year t in which the test was taken, as emphasized by Papay and Kraft (2015).

Properly measuring teacher contributions to achievement also requires that each student test score observation be matched to the teacher who taught the class in which the student's test score was generated. We utilize the fuzzy matching algorithm developed by Mansfield (Forthcoming), which exploits the fact that classroom-average demographics can be constructed and compared for both the test-score-level data and the classroom-level data (which contains the valid teacher ID).[19]

Our original dataset consists of 8,407,382 test scores from 460,792 classrooms, 28,347 teachers, and 1,307 high schools. We drop from the sample 2,878,254 test score observations for which we cannot match a teacher and 794,541 for which we cannot verify a difficulty level (discussed in footnote (14) above).

In addition, recall that the second key data requirement is that measures of both general and task-specific experience can be accurately constructed. The NCERDC data contain all classroom assignments (subject and level) for each teacher for the years 1995-2009, even in non-tested subjects. However, complete histories of classroom assignments, necessary to construct subject-specific, level-specific, and subject-level-specific experience, can only be assembled for teachers who began teaching after the data collection commences in 1995 (as indicated by an entry level paycode). Because our identification strategy relies on observing each teacher's full history of subject- and level-specific experience at each point in time, we drop an additional 2,364,544 test scores associated with teachers for whom we cannot properly construct context-specific experience stocks. Note that we cannot distinguish novice teachers from teachers who previously taught outside of North Carolina unless such transferring teachers are given partial credit for their prior experience (and

[16] 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.

[17] In principle, one might worry that differences in teacher performance may be reflecting the extent to which teachers adhere to the state curriculum rather than differences in ability to foster learning. Fortunately, several features of the North Carolina context mitigate such concerns. First, in recent years No Child Left Behind legislation has put pressure on principals to ensure that teachers teach the standard curriculum, since schools that fail to meet state standards are subject to sanctions and possible closure. Second, the North Carolina end-of-course exam scores we use as outcome measures must comprise 25% of the student's year-end grade in a given subject, so that parents are likely to complain about teachers that ignore the standard curriculum. Finally, during the sample period, teacher bonuses of up to $1,500 were linked to average test scores of the students in the school at which they teach.

[18] Observable current student inputs include indicators for parental education, race, gender, gifted status, current grade, and current Limited English Proficiency status. Observable past student inputs include the student's 7^{th} and 8^{th} grade math and reading scores (though to reduce the influence of missing data we only include 7th grade test scores as a robustness check in Section 6.4).

[19] See Mansfield (Forthcoming) for a full description of the algorithm and summary statistics regarding its efficacy.

thus would not have an entry level paycode). Nonetheless, the problem of accurately constructing stocks of context-specific experience would be considerably more severe in contexts where data exist only for a single school district (even a large one).

After several other sample restrictions, our final sample consists of 1,126,300 test scores aggregated to 61,993 classroom-level observations, from 8,750 teachers, and 596 high schools.[20] Basic summary statistics comparing the original and final samples are presented in Table 1.

4.3. Generating the Experience Profile

For the baseline specification we construct flexible experience profiles by creating indicators for eight experience categories: 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. In our featured specifications, experience is measured as the number of prior years in which at least one classroom was taught in the relevant context for the chosen experience dimension. We posit that teaching a second classroom in the same year, when there is no opportunity to alter the lesson plan or assignments, is likely to provide negligible experience value relative to teaching a classroom in a different year. However, as a robustness check we also present results from specifications in which experience is measured using the total number of classrooms taught prior to the year of the observation.

We also assume that teachers' general and context-specific experience is fully portable across schools. This assumption is partly driven by the existence of a statewide testing regime that is tied to students' course grades, so that curriculum differences between schools should be minimal. Further, since only 16% of our classrooms are taught by a teacher who in his/her second (or greater) school, we also have limited variation with which to test this assumption. However, in Section 6.4 we examine the sensitivity of our results to this assumption by estimating our general and context-specific experience profiles for only the subset of classrooms featuring teachers in their initial schools.

To capture depreciation in teachers' "experience capital", we include in X_{ct} a set of indicators for whether the teacher of the classroom taught the subject, level, and subject-level in the previous year, as well as an indicator for whether the teacher taught at all in the previous year. We also include a second set of analogous indicators for whether the teacher of the classroom taught in the relevant contexts two years prior.

Finally, to account for possible decreases in teacher effort prior to an assignment change (explained further in Section 6.1) we also include four indicators that equal one if the observation is from a classroom that represents the teacher's last year teaching the school-subject combination, the school-level combination, the school-subject-level combination, and at the school in any classroom, respectively.

[20] We restrict the sample in several additional ways. First, we drop 21,915 scores from classes with fewer than 5 students (since these are likely to represent data entry errors). Given our focus on high schools, we also drop 263,893 test scores from students in grades 6-8. We also drop test scores with invalid or outlier values, as well as all scores from 1997 and from Physical Science in 1999 due to concerns about data quality (270,395 scores). Since past test scores are critical for controlling for student sorting, we also drop 685,116 observations for students with missing 8th grade math or reading test scores. Finally, identification of experience cell fixed effects (estimated in our "full specification" discussed in section 6.4 below) requires that four-dimensional experience cells and school-teacher-subject-level cells form a connected graph, with the experience cells as vertices and school-teacher-subject-level cells as edges (or vice versa). We drop 2,424 test scores that are associated with school-teacher-subject-level combinations not contained within the largest connected component of the graph.

4.4. The Frequency of Teacher Assignment Rotations

The third data requirement for our decomposition is that we observe considerable worker rotation across tasks. Table 2 depicts teacher rotation across subjects in our final sample. The top (bottom) entry in each cell (i, j) represents the number (fraction) of teachers in our sample who ever taught in subject i that also taught in subject j. The table reveals that there is considerable rotation across subjects, though the vast majority of rotations occur within fields. This reflects the fact that certification is field-specific. Teacher rotation across levels is also substantial. The vast majority (87%) of teachers who ever teach an honors class also teach at least one basic class during their career. The converse is not true; only 43% of teachers observed teaching at least one basic class are also observed teaching an honors class at some point during their careers. 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).

Table 3 displays the pattern of rotation across subject-level combinations for teachers in the field of mathematics. The table illustrates that teachers do not merely teach either multiple levels of the same subject or multiple subjects at the same level, but rather are frequently observed teaching at the basic level in one subject and at the honors level in a different subject. It is this variation that allows us to distinguish the returns to subject-level-specific experience from the returns to subject-specific and level-specific experience, respectively. Taken together, these tables demonstrate that rotating across multiple subjects, levels, and subject-levels during one's career is the norm, rather than the exception.

As the example in Appendix A makes clear, identification of all four dimensions of experience relies on teachers continuing to introduce new subjects and levels into their repertoire after their career is already underway, as well as taking single year or multi-year breaks from teaching particular subjects before returning to them later.[21]

Figure 1 shows, for each level of general experience, the fraction of teachers who teach a subject, level, or subject-level for either the first time (1a) or last time (1b), as well as the fraction of teachers who leave teaching in North Carolina altogether ("General" in 1b). Figure 1 reveals that introducing new courses is quite common even in mid-career: 19% of teachers with seven prior years of experience teach a new subject for the first time in their eighth year, while 11% teach a new level and 29% teach a new subject-level combination.

Gap years in which teachers fail to teach (and then return to) a particular subject or level are also quite common. 22.5% of unique teacher-subject-level combinations exhibit one or more gap years at some point during our sample, while 19.9% and 14.1% of teacher-subject and teacher-level combinations exhibit at least one gap year. By contrast, 10.2% of observed teachers leave public school teaching entirely for at least a year before returning. These statistics reveal that there is more variation available to identify the returns to subject- or level-specific experience than there is to identify gains that are portable across all contexts (Wiswall (2013), Papay and Kraft (2015)).

Finally, Figures 2a and 2b display the distributions of subject-specific, level-specific, and subject-level spe-

[21] To see this clearly, note that if every teacher taught the same exact subject/level combinations each year for their entire career, level-specific, subject-specific, and subject-level-specific experience would all increment by one every year, and would thus be perfectly collinear with general experience. By contrast, the relative within-teacher performance among multiple courses taught simultaneously provides an important source of variation in identifying the variance in permanent task-specific talent.

cific experience for classrooms taught by second and third year teachers in our final sample. The data underlying these figures are presented in Appendix Table H.1. About 71% of classrooms taught by 2nd year teachers are in subject-level combinations that these teachers taught in their first years, while 55% of classrooms taught by 3rd year teachers are in subject-level combinations that these teachers taught in both of their first two years.

4.5. Estimation and Calculation of Standard Errors

We estimate the model at the classroom level via weighted OLS by exploiting the sparsity of the design matrices for the school-subject-level and school-teacher-subject-level fixed effects. Weights for each classroom observation are proportional to the number of students in the classroom, so that the variances in teacher productivity presented below capture the variation in teacher contributions across student-course combinations. Cluster-robust standard errors are calculated for each parameter. We cluster at the teacher level in order to accommodate the possibility of autocorrelated teacher-year shocks.

5. Results

5.1. Variation in the General and Context-Specific Components of Time-Invariant Teacher Productivity

Table 4 contains the results of the decomposition of the variance in time-invariant teacher productivity ("talent") into general, subject-specific, level-specific, and subject-level-specific components using the baseline specification (2). 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-Teacher-Subject-Level Combos" provides the total estimated variance (and corresponding standard deviation) in time-invariant teacher contributions to test scores across randomly sampled student-course combinations, which combines all four components of time-invariant teacher productivity. A one standard deviation increase in combined permanent teaching effectiveness is associated with a .154 standard deviation increase in expected student performance. 74% of this variance in permanent teacher quality can be attributed to general teacher talent that is portable across all subject-level combinations (see the row labeled "General Talent"). A student assigned to a teacher whose average effectiveness across the subject-level combinations he/she teaches is one standard deviation above the school average can expect a .132 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 experience or level-specific and subject-specific skill.

Subject-specific skill and level-specific skill make up about 17% and 9%, respectively, of the total variance in permanent teaching effectiveness across randomly chosen student-course combinations (tests). Receiving a teacher whose subject-specific skill in the selected subject is one standard deviation above the teacher's subjectwide average increases expected student achievement by about .063 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 53rd percentile statewide. Getting a teacher whose level-specific skill is one standard deviation above his/her levelwide average increases expected performance by .045 test score standard deviations, enough to move a student from the 50th to the 52nd percentile.

Finally, the subject-specific, level-specific, and general components of time-invariant teacher productivity combine to explain nearly the full variance in time-invariant teacher productivity across classroom contexts. Subject-level-specific talent does not seem to exist. 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 the teacher's talent for teaching honors-level courses and the teacher's talent for teaching biology courses, respectively.

Columns 3 and 4 of Table 4 display the alternative decomposition of permanent teacher skill that comes from imposing Assumption 2B, in which all the variation in average student performance across subject-level combinations within schools is also attributed to differences in average teacher quality. Not surprisingly, this increases each of the variance components substantially. Note, though, that the fractions of variance in teacher productivity explained by each component stay roughly similar to what they were under Assumption 2A. Under Assumption 2B, a one standard deviation increase in general teacher talent is associated with a .192 increase in average student performance, while a one standard deviation increase in subject-specific (level-specific) teacher talent is associated with a .077 (.058) increase in expected student performance relative to a subject (level) in which the teacher has no comparative advantage or disadvantage. Subject-level-specific talent does not appear to exist under Assumption 2B either. These results are roughly in line with those of Mansfield (Forthcoming).

The results under Assumption 2C (Columns 5 and 6) assign all the between school-subject-level variation in student performance to differences in teacher inputs rather than school or student inputs. They provide an upper bound estimate of the standard deviation in general teacher talent of .225 test score standard deviations.

Overall, we conclude that most of the time-invariant variation in teacher productivity is portable across all subjects and levels, but that there is a non-negligible achievement gain from being taught by a teacher who is relatively well-matched to the level and particularly the subject associated with the classroom.

5.2. General and Context-Specific Experience Profiles

Table 5 presents the estimated experience profiles for each type of experience from the baseline specification (2).[22] Panel A of Figure 3 displays these experience profiles graphically. Column 1 of Table 5 contains estimates of the part of the returns to teaching experience that are portable to all subject-level combinations, while Columns 2-4 contain estimates of the part of the returns to teaching experience that are subject-, level-, and subject-level-specific, respectively. There are considerable gains from the first two years of general experience, such that teachers teaching in their third year can expect to improve student performance by .085 test score standard deviations more than a novice teacher, even if they are teaching at a new level in a new subject. These gains grow to .113 by 7 years of experience, but seem to plateau thereafter. However, the results become 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.

Row 1 of Column 2 indicates that teaching a subject for the second time increases the teacher's expected performance by .014 test-score standard deviations within that subject, relative to the first attempt. An additional year of subject-specific experience increases performance by an additional .019 standard deviations,

[22] The coefficients on our controls for teacher workload and depreciation of experience capital for this specification are presented in Appendix Table H.2.

while a third year of subject-experience adds an additional .016 standard deviations. Gains seem to slow beyond the third year of subject experience. Overall, teachers with more than 7 years of subject-specific experience are between .046 and .067 student level standard deviations more effective than teachers with the same total years of general teaching experience but who are teaching the subject for the first time.

The results in Column 2 suggest that part of the returns to experience generally estimated in the literature are actually specific to the subject taught. Since teachers frequently reteach the same subject many times, subject-specific experience and overall (general) years of experience are highly correlated. Thus, when returns to subject-specific experience are not separated from returns to general experience, the returns to subject-specific experience will generally be reflected in larger estimated returns to general experience.

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 subject-specific and general experience have been taken into account. In fact, the returns to subject-level-specific experience seem to be negative. Note that such negative returns are not implausible in principle: teaching the exact same course again and again could cause teachers to lose enthusiasm or to stop updating course materials (even as the state curriculum drifts slightly).

That said, this negative profile might also be spurious if it is merely the product of overfitting; while including a full set of school-teacher-subject-level fixed effects removes potential bias from teachers systematically repeating the courses at which they are relatively effective more frequently, it also considerably limits the remaining variation in experience stocks that can be used to identify gains from experience.[23] Given that subject-specific and subject-level-specific experience are very highly correlated, OLS may be able to reduce squared residuals more by fitting sampling error than by fitting true productivity gains.

To address concerns about overfitting, we turn attention to our "restricted" specification (3) that replaces the school-teacher-subject-level effects μ_{srjl} with school-teacher fixed effects only.[24] Because the results from the restricted specification have proven to be more robust to alternative sample restrictions and the inclusion of additional controls, we focus primarily on experience profiles that maintain these restrictions for the remainder of the paper.

Table 6 displays the estimated general and context-specific experience profiles for the restricted specification (with Panel B of Figure 3 providing a graphical depiction). The results for general and level experience are essentially unaffected by the restrictions, but the negative effects of subject-level-specific experience disappear, while the gains to subject-specific experience are somewhat diminished. Specifically, a teacher with two (four) prior years of subject-specific experience could be expected to increase achievement by .023 (.041) test score standard deviations relative to the teacher's expected performance when teaching the subject for the first time (holding the other experience components fixed).

Column 5 in Table 6 sums across the first four columns to provide the returns to experience for a teacher who never changes the subject-level he/she teaches. After two (four) years, such a teacher is predicted to perform .118 (.138) standard deviations better than a novice teacher. Since many teachers teach the same subject-level every year (perhaps in addition to other courses), this sum is particularly well identified. Most of the

[23]Specifically, the inclusion of these fixed effects implies that only relative growth rates in performance within a school-teacher-subject-level cell provide identifying variation.

[24]Consistent estimation of experience profiles in the restricted specification requires that teachers do not systematically gain more general or context-specific experience in the subjects or levels in which they have experience-invariant comparative advantages. However, given that the previous sub-section revealed relatively small variances in subject- and level-specific permanent talent, even substantially elevated rates of re-assignment of teachers to their more effective subjects and levels would produce minimal bias.

sampling error in the estimates comes from decomposing this sum into the four experience components.

Given the failure to observe meaningful level-specific and subject-level-specific experience effects, the first two columns of Table 7 display results from a yet more parsimonious specification in which the level-specific and subject-level specific experience profiles are constrained to be zero everywhere. The basic pattern of results for total and subject-specific experience exhibit little change; there are still meaningful gains from the first several years of both total experience and subject-specific experience. Imposing these further restrictions increases the precision of the estimates considerably, however, so that experienced teachers are statistically significantly more effective than novice teachers for all categories of general experience and for all but the highest experience category of subject-specific experience.

The fourth column of Table 7 presents estimates from the standard specification in the literature, in which only a single “general” experience term enters the production function. This standard experience profile, which is driven by both general and specific returns, matches fairly closely those found in the literature.

Overall, the relative magnitudes of the coefficients for the different dimensions of experience parallel the results for context-specific talent presented in Section 5.1: a large role for the general component, with a moderate role for the subject-specific component and small-to-nonexistent roles for the level-specific and subject-level-specific components.

6. Tests of Identifying Assumptions and Robustness Checks

6.1. Testing for Dynamic Classroom Assignment Responses to Unobserved Shocks

Assumption 1, which is necessary for consistent estimates of experience profiles, will be violated if particular experience profiles are more likely to be observed during years in which either teachers or their schools are experiencing positive or negative year-specific deviations in productivity relative to what could be predicted given their full sample performance and teachers’ observed levels of each dimension of experience.

There are a variety of scenarios that could bring about such a correlation. Some involve endogenous allocation responses to idiosyncratic shocks, and may not exhibit any pre-trend. For example, a teacher who is less effective while pregnant may quit teaching after the baby arrives. Scenarios such as these would imply that the set of teachers who make it to the next year of teaching (or perhaps teaching in a particular context) are those whose teacher-year (or perhaps classroom) shocks were not too negative. Thus, the expected change in the teacher-year error component would be negative among those who persist, creating a potential downward bias in our estimate of the return to general experience.

We address endogenous responses to idiosyncratic shocks by including in all our specifications four indicator variables that are set to one if the observation is from a classroom that represents the teacher’s last year teaching at the school in any classroom, in the current school-subject combination, in the current school-level combination, and in the current school-subject-level combination, respectively. These indicators capture the extent to which the year before an assignment change tends to exhibit particularly low performance, thereby preventing such dips from being fit by the experience profile parameters of interest. In addition to controlling for the most plausible dynamic response to health shocks, these dummies also control for the possibility that teachers who anticipate quitting put forth less effort in their final year (which could also bias downward the estimated general experience profile). Indeed, the coefficients on the dummies corresponding to the last year in the school-subject and school-subject-level (Appendix Table H.3) are negative and statistically significant.

However, other scenarios that produce violations of Assumption 1 might involve trends over time in error components rather than merely single-year idiosyncratic shocks, so that our "last year" indicators are inadequate controls. One particularly plausible mechanism stems from the possibility of heterogeneity in the gains to experience among teachers.[25] Since both our baseline and restricted specifications constrain the gains from general experience to be common to all teachers, any heterogeneity in rates of growth among teachers in the sample will be reflected in the teacher-year error component, ν_{rt}.

Thus, our context-specific experience profiles could be biased upward if teachers with faster than average growth rates are more likely to stay in the courses and levels they are teaching: the average value of the teacher-year error component ν_{rt} would be higher for higher values of subject-specific or level-specific experience. This might occur if rapidly improving teachers are rewarded with the opportunity to continue teaching their courses (while forcing others to adjust to changing classroom demand created by, say, teacher turnover or variation in student cohort size).

An analogous bias could be created by endogenous responses to school-year shocks. For example, teachers may be more likely to quit a declining school, thereby creating holes in subject or level offerings that other teachers must be forced to fill. In this case, the school-year error component ϕ_{st} would be positively correlated with levels of context-specific experience, leading to overestimates of the gains to experience.

We can test both of these hypotheses jointly by examining whether the trend in a teacher's performance (relative to the estimated experience profile) predicts the teacher's future teaching assignments. Indeed, such a test will also reveal the potential bias from *any* other sources of dynamic assignment patterns that involve a time trend in the composite error ϵ_{ct} within a teacher.

Specifically, we first identify all teacher-year combinations in which a teacher fails to teach any classroom in the following year. We then calculate and plot in Appendix Figure H.1a the average test score residuals across all classrooms of students taught by the teachers from these teacher-year combinations in the years leading up to their breaks from teaching (denoted t in event time). We see no evidence of any trend in teacher-year residuals in advance of the break from teaching. In order to distinguish quits/retirements from parental leave, Figure H.1e plots the same time path of teacher-year residuals leading up to the smaller sample of teacher-year combinations in which a teacher fails to teach in any classroom in *any* future year in the sample. No obvious trend is observed.

We then perform the analogous exercise for changes in subject, level, and subject-level assignments. Specifically, for Figure H.1b (H.1f) we identify all teacher-subject-year combinations in which the teacher fails to teach any classrooms in the chosen subject in the following year (any future year), and plot the time path of average teacher-subject-year residuals leading up the change in subject assignment. Figures H.1c - H.1h plot the analogous trends in teacher-level-year and teacher-subject-level-year residuals leading up to breaks from teaching a given difficulty level or subject-level combination. None of the Figures H.1a-H.1h show any evidence of a significant trend in residuals preceding an assignment change that might suggest biases from dynamic reallocations of teacher assignments in response to unobserved shocks/input trends.[26]

[25] Atteberry et al. (2013) finds evidence of heterogeneous teacher growth in New York City.

[26] The point estimates that underlie these figures are presented in Appendix Table H.4.

6.2. Testing for Dynamic Student Sorting

In this subsection, we focus our attention specifically on violations of Assumptions 1 and 2 that are caused by nonrandom student sorting. To gauge the possible severity of the problem, we implement a "backcasting" test in the spirit of Rothstein (2010) in which we replace class averages of students' contemporaneous test scores with class averages of their math standardized test scores from 7th grade.[27] The intuition behind the test is that if students were randomly assigned to teachers conditional on controls, current teacher identity or experience should not predict past student performance. To the extent that it does, part of the estimated gains to teacher experience could simply be capturing the ability of more experienced teachers to attract/be assigned to unobservably superior students.

The results of this exercise for the restricted specification 3 are presented in Appendix Table H.5. While the estimates are generally relatively small in magnitude, a number of estimates are statistically significantly different from zero, creating some cause for concern. A closer look, though, reveals that teachers with more general and subject-specific experience seem to be attracting students with inferior 7th grade math scores (conditional on 8th grade scores and the other controls), while teachers with more level-specific and subject-level specific experience seem to be attracting students with superior past test scores. Thus, the backcasting test suggests that the substantial gains to general and subject-specific experience reported above are, if anything, understated. By contrast, the gains to level-specific and subject-level specific experience could be slightly negative. Thus, these results do not undermine the qualitative conclusions of Section 5.2.

Furthermore, while such backcasting tests are well known in the literature and are valuable for flagging potential selection and sorting biases, recent research by Kinsler (2012) and Goldhaber and Chaplin (2015) suggest that these tests may find evidence of significant dynamic student sorting even where none exists. For example, suppose that classroom assignment in 9th or 10th grade is partially based on 7th grade test scores (perhaps because these test scores still affect principal or student beliefs about student ability), but that the part of persistent student inputs captured by 7th grade test scores is fully reflected in the included controls. In this case, current teacher assignments could significantly predict past test score noise or transitory student inputs, yet estimates of teacher value-added and gains from experience would nonetheless be unbiased.[28] Indeed, when we add 7th grade math and reading scores as controls as a robustness check in the next section, we find negligible changes in estimated gains from general and context-specific experience.

6.3. Evaluating Forecast Bias in Estimates of Context-Specific Teacher Talent

While the previous subsections have investigated several sources of potential bias in our estimated experience profiles, in this section we seek to determine the degree to which our estimates of teacher talent, the estimated fixed effects $\{\hat{\mu}_{srjl}\}$, properly capture the true talent contributions $\{\mu_{srjl}\}$. Following Chetty et al. (2014), we do this by measuring forecast bias: the degree to which teachers' context-specific talent estimates from

[27] These test scores are not included in our baseline specification because 7th test scores are missing for our first cohort (since they had already reached 8th grade the first year the statewide database was constructed). We wanted a consistent set of controls for all cohorts in our sample, and did not want to exclude our earliest student cohort, since their 1997 performance creates a baseline of productivity for the 1997 cohort of new teachers, which permits estimation of the gains to the 14th year of teacher experience and generally increases the precision of estimates of mid-career teaching (to which few cohorts of teachers contribute).

[28] Similarly, Chetty et al. (2016) point out that track-level, field-level, or school system-level shocks that are correlated across years could produce sampling error that is correlated across students' current and past classroom observations. This represents an additional mechanism by which backcasting tests could yield spurious "evidence" of bias.

one partition of our data predict mean residual achievement in the same context in a second, left out partition. The implementation of our tests for forecast bias, which mirrors Chetty et al. (2014), is described in detail in Section Appendix F.

We first test for forecast bias in our estimates of combined general and task-specific talent $\{\hat{\mu}_{srjl}\}$. This involves regressing differences in the performance of pairs of teachers within the same school-subject-level context from a left-out sample of classrooms on our posterior mean belief about the difference in the two teachers' talent in the chosen context. This empirical Bayes (EB) posterior belief is formed by multiplying the difference in estimated fixed effects $\hat{\mu}_{srjl} - \hat{\mu}_{sr'jl}$ from the primary sample by a reliability ratio that shrinks the estimated difference toward zero. If the estimated variance in teachers' talent contributions across randomly chosen test scores presented in Table 4 is valid, multiplying by this reliability ratio removes the attenuation bias created by sampling error in the fixed effect estimates $\{\hat{\mu}_{srjl}\}$ that would otherwise occur in a forecast regression of outcome differences from one sample on outcome differences in a second disjoint sample. Consequently, under the null hypothesis that the estimated (lower bound) teacher talent variance across tests is valid, the coefficient on the EB posterior mean from the forecast regression should converge in probability to 1.

The actual estimated regression coefficient (Appendix Table H.6, Column 1) is 0.825, with a standard error of 0.019. While this exercise reveals that our estimates of teacher talent can be used to forecast contributions to student performance out-of-sample fairly accurately, our estimator does not seem to be "forecast unbiased". The most straightforward explanation for a coefficient below 1 is that our estimate $\hat{Var}(\mu_{stcl})$ slightly overstates the true variance in teacher talent contributions $\hat{Var}(\mu_{stcl})$, so that the reliability ratio we use in shrinkage overstates the degree of signal in the fixed effect differences. However, a couple of alternative explanations exist. First, the reliability ratio could also be overstated if we are underestimating the standard errors used to construct the estimated "noise". Second, the subsample of school-teacher-subject-level combinations that satisfy the criteria for eligibility for the forecast sample (See Section Appendix F) might feature a slightly lower true variance in teacher talent contributions than the population.

While this test captures the model's ability to consistently estimate the combined general and context-specific talent that a teacher contributes to a given context, the ability to improve the efficiency of teachers' classroom assignments only depends on the model's success in isolating and consistently estimating the context-specific components of teacher talent. Thus, we also construct two additional forecast tests that measure the degree to which our estimates of subject-specific and level-specific talent can forecast out-of-sample teachers' subject-specific and level-specific comparative advantages, respectively.

Unlike our tests of the consistency of our combined talent estimates, which could be performed using differences among teachers who taught in the same school-subject-level context, evaluating our comparative advantage estimates requires measuring the degree to which difference-in-differences between teachers who taught the same two courses at the same school can be forecast. This necessitates restricting the forecasting sample to pairs of teachers who each taught multiple classes in the same two subjects within the same school-level combination (or, for the second test, both basic and honors in the same school-subject combination). Only 205 and 289 difference-in-differences exist on which to perform the forecast test for subject-specific and level-specific talent estimates, respectively. In essence, there is far less overidentifying variation available to test the model's ability to detect true variation in subject-specific and level-specific talent.

The methodology for the context-specific forecast tests is otherwise perfectly analogous to the forecast test for combined teacher talent. Difference-in-differences in residual mean test scores from among the left-out classrooms in the forecasted sample across teachers and either subjects or levels (conditioning on the same school-level or school-course environment as appropriate) are regressed on empirical Bayes estimates of difference-in-differences in the teachers' context-specific talent from the forecasting sample.

The regression coefficient on the forecasted difference-in-difference from the subject-specific forecast sample is 1.013, with a standard error of 0.242. Thus, while the point estimate suggests negligible forecast bias in estimates of subject-specific talent, the confidence interval is quite wide: only values below 0.539 can be ruled out with 95% confidence. Nonetheless, the test provides some reassurance that the kind of achievement data available to principals can provide some meaningful signal of subject-specific skill that might be used to guide classroom assignments.

The regression coefficient from the level-specific forecast regression is 0.456, with a standard error of 0.333. The point estimate indicates that our ability to infer level-specific talent is less strong than what our estimate of the true variance in level-specific talent would suggest. However, the test is severely underpowered: both 0 and 1 are within the 95% confidence interval. The large standard errors are partly due to the limited overidentifying variation just discussed, but are also attributable to the small estimated variance in level-specific skill: each classroom provides an extremely weak signal of level-specific skill relative to the "noise" stemming from the contributions of general teacher talent and other student and school inputs.

6.4. Further Robustness Checks

This subsection aims to provide a broader sense of the robustness of the main results to the array of difficult choices regarding specifications, variable definitions, and sample restrictions described in sections 2 and 4.

First, so far we have defined experience in a given context as the number of previous years in which the teacher taught at least one classroom in that context. This assumes that additional classes taught simultaneously in a context within a year (e.g. two periods of honors Biology classes) do not provide additional productivity value, which is based on the idea that teachers often have little time to alter materials between classes in a given day. However, Appendix Table H.7 presents estimated experience profiles in which experience in each context is defined as the total number of classrooms taught in the chosen context in prior years. While the scales are difficult to compare, the results based on the classroom-based definition of experience are qualitatively very similar to those based on the year-based definition: substantial gains to general experience, moderate gains to the first few years of subject-specific experience, and negligible gains to level-specific and subject-level specific experience. Due to the near perfect correlation between year-based and classroom-based measures of experience, we are unable to determine which measure better captures the true accumulation of productivity gains from experience. Appendix Table H.8 shows that the decomposition of the variance in teacher talent is insensitive to the definition of experience.

Second, while we allow the permanent component of teacher productivity to be school-specific, to this point we have assumed that gains from general experience and from each dimension of context-specific experience retain their full value at new schools. Appendix Table H.9 presents estimated experience profiles based on the subsample of classrooms associated with teachers teaching in their first schools, where there is no concern about mismeasurement of experience stocks due to imperfect portability across schools. This subsample comprises 83.5 percent of our full sample of classrooms. The experience profiles remain essentially unchanged. Similarly, the decomposition of teacher talent for this subsample (Appendix Table H.10) is nearly identical to its full sample counterpart.

Third, Appendix Tables H.11 and H.12 present results from a specification in which we alter our controls for depreciation in experience-based human capital. Specifically, we replace indicators for whether the teacher taught the chosen subject, level, and subject-level (and whether the teacher taught at all) in the last year with linear controls for the number of years since having taught the relevant subject, level, or subject-level (or taught in any classroom). The estimates of the returns to experience are not sensitive to our handling of depreciation in experience-based human capital, and our depreciation controls are generally close to zero and statistically insignificant.

Fourth, Section 6.2 revealed that the identities of students' high school teachers can partially predict their prior 7th grade test scores, suggesting that 7th grade math and reading test scores might be valuable controls for student sorting. Thus, Appendix Table H.13 reports estimated experience profiles from a specification that includes class-averages of 7th grade math and reading test scores as controls (and sets missing 7th grade test scores to the samplewide mean of zero). Inclusion of 7th grade math and reading scores has almost no impact on the estimated profiles. These results reinforce the idea that failure of a backcasting test need not imply substantive bias in estimates.

Fifth, the baseline and restricted specifications presented in Tables 5 and 6 impose that the returns to general and subject-specific experience are the same across fields. In Appendix Table H.14, we present separate estimates of general and subject-specific experience profiles for math, science, social studies, and English subjects. Comparing across columns, we see that general and subject-specific returns to experience are fairly similar across all four fields, providing support for the pooled specifications above. However, there is some variation in experience gains across fields. In particular, the gains to general experience appear highest in math, and the gains to subject-specific experience appear to be highest in science.

Sixth, up to this point we have combined years of experience 5 and 6, 7 through 10, and 11 and beyond into bins rather than introducing separate indicator variables for each year of experience. We did this because we expected gains from experience to slow down at higher levels of experience (as our estimates suggest they do), and combining multiple years into bins allows us to reap additional efficiency gains and identifying power (necessitated by the need to observe teachers' full teaching histories in order to construct their stocks of general and task-specific experience, which removes most well-experienced teachers from the sample). However, Wiswall (2013) points out that grouping experience levels into broad bins imposes arguably unrealistic restrictions on the experience profile that can potentially produce substantial bias. To address this concern, Appendix Table H.15 presents estimated general and context-specific experience profiles from a version of the restricted specification in which indicators are included for years 1 through 14 of experience.[29] While the estimates become prohibitively noisy beyond six or so years of experience, the results for the first several years of experience are extremely similar to those presented in Table 6 across all four dimensions of experience, suggesting that pooling multiple years of experience into experience category indicators is not generating substantial bias, at least for the unpooled experience categories.

Along the same lines, Appendix Table H.17 displays predicted values for the first ten years of experience in each dimension from a specification in which the set of indicators for each number of years of experience is replaced by a quartic in each experience dimension (general, subject-specific, level-specific, and subject-level specific). The gains to each dimension of experience are again similar, illustrating that a smoother, more parsimonious specification can still capture the basic qualitative results.

Finally, both the baseline and restricted specifications impose that the separate components of experience

[29] When the full set of dummies is introduced into the baseline specification, the results become nonsensical, with enormous offsetting positive and negative effects across dimensions. This is not surprising, as combining multiple years into bins was helping to break the collinearity between the various dimensions of experience, so that the overfitting/collinearity problem discussed in Section 5.2 now becomes even more severe.

are additively separable in the education production function:

$$d(exp^{gen}, exp^j, exp^l, exp^{jl}) = d^{gen}(exp^{gen}) + d^j(exp^j) + d^l(exp^l) + d^{jl}(exp^{jl}) \tag{12}$$

However, general experience and different dimensions of context-specific experience may interact with one another. For example, perhaps students only learn if the teacher has developed effective ways to both explain a subject's content *and* maintain control of the classroom. Lectures that deliver content effectively may require subject-specific experience, whereas classroom control skills may be learned through general or level-specific experience. Alternatively, perhaps a teacher can keep student attention by either having exceptional command of the content *or* by having excellent classroom control skills, in which case the different components of experience would be substitutable rather than complementary.

We relax the additive separability assumption by estimating a "full" specification that captures the contribution of experience to teacher productivity via a non-parametric function of the four experience components:

$$Y_{ct} = X_{ct}\beta_{jl} + \delta_{sjl} + \mu_{srjl} + d(exp^{gen}_{rt}, exp^j_{rt}, exp^l_{rt}, exp^{jl}_{rt}) + \epsilon_{ct} \tag{13}$$

We implement this specification by replacing the four dimension-specific experience profiles with a full set of four-dimensional experience cell fixed effects.[30] This specification is isomorphic in structure to a model with worker and firm fixed effects. Since the estimated experience cell fixed effects are measured with considerable sampling error, to better reveal the underlying structure of the experience contributions we smooth estimates for each experience cell by using a normal kernel to give weight to "nearby" estimates.[31]

We then take partial derivatives of this smoothed non-parametric experience production function with respect to each dimension of experience and integrate over these dimension-specific partial derivative functions to construct a set of standard experience profiles analogous to those from our additively separable baseline and restricted specifications. Section Appendix E describes this procedure in further detail.

The results of this exercise are displayed in Table H.18, while Table H.19 displays the corresponding marginal effect estimates from a "restricted" version that a replaces school-teacher-subject-level fixed effects with school-teacher fixed effects. Compared to the additively separable results from Table 6, the results in Table H.19 feature quite similar general and subject-level experience profiles, but somewhat larger gains to both subject-specific and level-specific experience. Overall, though, accounting for possible misspecification from ignoring interactions among experience components does not change the basic qualitative conclusion that the bulk of the gains from experience stem from general and subject-specific experience.[32]

[30] For example, the vector of experience stocks $(exp^{gen}, exp^j, exp^l, exp^{jl}) = (2, 1, 1, 1)$ is captured by a different indicator variable than $(2, 1, 2, 1)$.

[31] Appendix D.1 provides a more detailed explanation of this smoothing procedure.

[32] Previous versions of this paper attempted to use the full specification to characterize the nature of complementarity present in the experience production function. While such attempts produced suggestive evidence that general, subject, and level experience are substitutes rather than complements, identification and estimation of the degree of complementarity places extremely strong demands on the data, so that the results were both noisy and fragile. Thus, we have chosen to remove the full discussion of sub/supermodularity of the production function in this version.

7. Projecting the Achievement Gains from Efficient Use of Context-Specific Teacher Experience and Talent

7.1. Methodology

The moderate variance in subject- and level-specific time-invariant productivity differences among teachers, combined with the estimated gains to subject-specific experience, suggest that fully exploiting a teaching staff's task-specific human capital could potentially generate non-trivial efficiency gains. In this section we develop a set of counterfactual simulations to gauge the magnitude of the performance gains that could be achieved statewide if each principal exploited the full value of the stocks of task-specific experience and talent of the members of his or her teaching staff.

To see how such simulations might be implemented, consider the allocation of teachers to classrooms that takes place at a particular school in a particular field over the set of years in our sample. Ideally, we would solve the dynamic problem of choosing sequences of yearly allocations to maximize the average test score performance over the entire sample (and perhaps beyond). However, the state space of such a dynamic problem is prohibitively large: it must include, for each teacher in the school, both the teacher's stock of task-specific experience as well as posterior beliefs (with corresponding precisions) about the teacher's talent in each subject and level.

Consequently, we instead simulate the dynamic effects of re-solving each year the static optimization problem in which the expected average test score for the year is maximized, taking the set of classrooms and teachers to be matched in the chosen year as exogenously given at the start of the year. Four-dimensional experience stocks are then updated for the next year based on the efficient static allocation. While this approach necessarily understates the true gains to dynamic optimization, it represents an allocation rule that principals can automatically implement each fall with minimal computational burden and without making any projections about enrollment and teacher attrition. By evaluating the dynamic implications of static optimization, we can ensure that the short-run efficiency gains from implementing the statically optimal allocation are not undermined by long-run efficiency losses.

Even static optimization, however, requires specifying the principal's belief about each teacher's time-invariant task-specific productivity for each subject-level combination to which the teacher could potentially be assigned. Thus, we calculate empirical Bayes posterior beliefs about each teacher's task-specific talent based on our school-teacher-subject-level fixed effect estimates and their standard errors, and use these for any school-teacher-subject-level combinations that are observed in our sample. We assign task-specific productivities of 0 (the population mean) to any school-teacher-subject-level combination that we do not observe.

These posterior beliefs are designed to make efficient use of the information about teacher comparative advantages contained in the student test score data, given our assumed achievement production function. However, a major concern is that the principal may have information about teachers' subject-specific or level-specific talent that is not reflected in the test score data, and is therefore unobserved by the econometrician. Such information might be derived from classroom observations or from knowledge of the teacher's college preparation (e.g. a biology major might be likely to have a comparative advantage in biology relative to chemistry). If such additional sources of principal information exist, then allocations that are optimal based on the posterior beliefs we calculate may identify spurious efficiency "gains" in which teacher assignments that were driven by the unobserved component of principal's information are altered to better fit the noise in our empirical Bayes estimates of task-specific talent, and thus would in fact represent achievement *losses*.

In light of this concern, we compare our simulated "optimal" allocations to two different baselines representing different informational assumptions. The first baseline consists of the achievement contribution of actual teacher assignments in our sample under the assumption that the information available to the principal at the time of allocation is a subset of the information contained in our entire sample of test scores for all public school teachers in North Carolina. Under this assumption, the gains we identify from teacher reallocations should be correct in expectation; we are as likely to understate as to overstate the gains from our alternative allocation.[33]

However, since this assumption may cause us to overstate (possibly dramatically) the potential efficiency gains from effective use of test-score-based information about task-specific experience and talent, we also compare the achievement gains from our "optimal" allocations to a baseline in which teachers are randomly allocated to classrooms within field. This random baseline allows the reader to gauge the potential importance of utilizing information about task-specific experience and talent contained in test score data without making any assumption about the degree to which this and other sources of information are already being used by principals. Principals and other administrators may simply wish to know whether it is worth the time and effort to track task-specific experience and generate beliefs about task-specific talent and to incorporate these pieces of information into classroom assignments, or if instead they should allocate classrooms based on other objectives that may be nearly orthogonal to the short run maximization of test scores (e.g. minimizing parent dissatisfaction).

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.[34] This counterfactual simulation can be rewritten as a binary integer programming problem. The formal presentation of the problem is located in Appendix G.

Since our estimated gains from general and task-specific experience are based on only the 11 tested subjects, our simulations only consider efficiency gains from reallocating classrooms in which the tested subjects were taught. We also do not reallocate classrooms in which English 1 was taught, since this is the only tested subject in English. In addition, because we do not observe the full teaching histories of any teacher who began teaching before the sample begins in 1995, for some of our simulations, we do not reallocate the classrooms taught by such teachers; for other simulations, we impute the full teaching histories for such teachers. We use the estimates from the full specification in equation (13) for both the posterior beliefs

[33] Given that the fixed effect estimates are based on the entire sample, one could argue that they are partially based on information (teachers' average test score performance from future years) that principals cannot possibly have observed at the time of allocation, so that the gains we compute overstate gains from a feasible allocation algorithm even if principals do not have other sources of information on teachers' task-specific talent. However, for highly experienced teachers that are included in only a subset of our simulations (but would make up a substantial fraction of the principals actual staff), the principal will generally have many years of test-score data on which to base posterior beliefs, so that their posterior beliefs may closely correspond to our empirical Bayes posterior beliefs. Another possible approach would have been to calculate posterior beliefs for each teacher in each year based on their performance record up to that date. This would be quite computationally costly for us, since it requires re-estimating the model (and calculating standard errors) for each year in our sample, but would likely be feasible for an actual school that is allocating only a handful of teachers. Thus, even our richer static optimization program could be fairly easily implemented by any school, given accurate records on the teachers' past course assignments and student performance.

[34] 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.

about context specific talent and the predicted contributions of each four-dimensional vector of experience stocks.[35]

Our simulation 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 and context-specific talent in each school starting in 1995 (the first year of the sample). However, estimates in the first few years of the sample conflate the fact that past rotations have limited potential gains from re-optimizing with the fact that relatively few teachers are being reallocated.[36] Thus, we focus on efficiency gains among classrooms assigned in the last 5 years of the sample, when a substantial fraction of teachers are eligible for reassignment. We do not extend our simulations beyond the last year of the sample, so that the gains we report may not fully capture the very long run (steady state) gains from repeated optimization. We hold the allocation of teachers to schools fixed (thus ignoring any possible effects of classroom reassignments on teacher turnover), and we continue to assume that context-specific experience is fully portable across schools.

We also compare the results of the "dynamic" simulation to a fully static simulation that solves the binary integer 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 static reallocation. The static simulation serves to illustrate the decomposition of gains into the part stemming from initial reassignment to better match teachers' context-specific experience and talent to the courses they teach and the part stemming from longer run gains associated with the specialization of the teacher work force.

7.2. Results from Counterfactual Simulations

The bar charts in Figure 4 present the student-weighted average expected test score gain from optimal reallocation among all school-year combinations for both the single-year "static" simulation (Panels C and D) and the "dynamic" simulation in which static re-assignments affect the following year's experience stocks (Panels A and B). The numerical values that correspond to the bars in Figure 4 are presented in Table 8.

The results in Figure 4 are reported separately by whether the baseline is the actual allocation observed in the sample (Panels A and C) or an allocation in which teachers are randomly assigned to classrooms within field (Panels B and D). In addition, because the scope for efficiency gains from matching and specialization increases in the size of the teaching force, achievement gains are also presented separately by number of teachers in the school-field-year combination eligible to be reallocated (i.e. the number who taught at least one classroom in that school-field-year combination in the actual data for whom the full teaching history is observed).[37]

While optimal reallocations were implemented separately by field, the results displayed in Figure 4 pool the

[35] We smooth the nonparametrically-estimated experience function to a greater degree for the simulations to ensure that our simulated efficiency gains do not stem from better exploiting the sampling error portion of the estimated returns to experience. We use a bandwidth (variance on a normal PDF) of 5 to smooth the estimates used for the simulation. In theory, the appropriate smoothing represents a delicate balance: smoothing too little creates the possibility of spurious gains from better fitting sampling error in estimates, while smoothing too much also removes the signal. Indeed, complete smoothing would make the productivity of each experience cell identical, and would therefore eliminate the possibility of any gains from better use of teacher experience stocks. In practice, however, we have found that bandwidth choices between 2 and 10 yield very similar estimates.

[36] This is because we do not observe the classroom assignment histories for the vast majority of the teachers in the first few years.

[37] In the case where only one teacher is observed teaching all of the courses in the field, there can be no gains from teacher reallocation. Thus, school-field-years featuring only one teacher are omitted from the simulations presented in Figure 4.

classroom-level gains across the three fields (math, science, social studies). We pool the results because there was surprisingly little heterogeneity in simulated gains from reallocation across fields (See Appendix Table H.20 for the disaggregated results). In each panel of Figure 4, the height of the rightmost bar in each set of three bars represents the total simulated per-student standardized test score gain from optimally allocating teachers to classrooms, while the heights of the leftmost and middle bars decompose this total per-student gain into the components stemming from better (or worse) use of teachers' context-specific experience and context-specific talent, respectively.

We focus first on Panel A, which displays results from our "dynamic" simulations in which the actual allocation observed in North Carolina is used as a baseline. These results indicate that better use of context-specific talent in particular has the potential to reap non-trivial efficiency gains. Specifically, the total gains relative to the actual allocation from better use of context-specific teacher productivity grow from .017 test score standard deviations for school-year-fields in which only two teachers are eligible to be reallocated to .033 for four-teacher fields and .044 standard deviations for school-year-fields featuring eleven or more eligible teachers.[38] Moreover, these total gains derive almost entirely from more efficient use of teachers' task-specific talent, while gains from better use of teacher-specific experience are negligible and in some cases slightly negative.

If instead the random allocation is used as a baseline (Panel B), two-teacher fields reap efficiency gains of .025 standard deviations, while four-teacher fields produce gains of .044 standard deviations and fields with eleven or more teachers produce gains .054 standard deviations. Generally speaking, about 20% of the gains relative to the random allocation comes from effective use of context-specific experience rather than talent; .05, .010 and .015 of the total per-student test score gain can be attributed to better exploiting teacher subject-specific and level-specific experience for school-year-fields with two teachers, four teachers, and eleven or more teachers eligible for reallocation, respectively. The combined results in Panels A and B suggest that while effective use of teachers' stocks of context-specific experience could be an important source of efficiency gains in some contexts, North Carolina principals already seem to be effectively exploiting the context-specific experience of their staffs, possibly even at the expense of subject-specific and level-specific teacher talent.

Panels C and D display the corresponding results for the "static" simulations, in which teacher assignments up until time $t - 1$ are held fixed when choosing simulated classroom allocations at t. They reveal that nearly all of the long-run gains from optimal reallocation of teachers are reaped in the first period of reallocation. This is not surprising given the small fraction of the total efficiency gains in the dynamic simulations attributable to better use of task-specific experience.

Note that if principals have very precise information about task-specific talent at the time of hire, then there is no tension between maximizing the contributions of task-specific experience versus task-specific talent: teachers can be assigned when hired to the courses in which they have the strongest comparative advantages, and then can continue to teach these courses, building up the relevant task-specific experience. However, for principals that have minimal information about teachers' context-specific talent at the time of hire, our estimates suggest that the degree to which teachers should be rotated among courses is likely to depend strongly on a school's teacher turnover rate. For schools with very low turnover rates, the variance

[38]This pattern is mirrored in the fraction of classrooms whose assigned teacher in the simulation differs from the one observed in the data. 30.3, 44.9, and 51.9 percent of classrooms in the math field with two, four, and 11+ teachers have their original teachers reassigned in the dynamic simulation. The corresponding percentages are 25.3, 40.4, and 45.6 for the static simulation. Appendix Table H.21 presents the full set of reallocation rates from our simulations.

in context-specific talent is sufficiently large that principals might find it worthwhile to rotate teachers for several years in order to learn the set of course assignments that best utilize task-specific talent. However, for schools with high turnover rates, the signal about task-specific talent received from a small number of classrooms is sufficiently coarse that the knowledge necessary to benefit from superior allocation of task-specific talent cannot be gathered in time for it to be valuable; by contrast, the productivity gains from the first two years of subject-specific experience are reasonably large, and can be reaped even among teachers who are only likely to stay for three or four years. This logic suggests that high turnover schools are likely to be better off minimizing the degree to which teachers are rotated among courses.

In the absence of an analytical solution to the full dynamic problem, a more precise characterization of the optimal amount of teacher rotation requires simulating test score contributions from alternative rotation strategies for a variety of parameter combinations governing, for example, turnover rates, principal information, teaching loads, and the number of distinct subjects, levels, and courses. We leave such an extensive simulation exercise for future work. However, we wish to emphasize that each individual school likely faces fixed and known values of many of these remaining parameters, so that the estimated task-specific experience profiles and underlying variances in subject-specific and level-specific talent presented in this paper provide the information necessary for school administrators to perform their own customized simulations to guide their classroom assignment decisions.

In Appendix Table H.22, we also display results from simulations in which all observed teachers who taught the tested courses are eligible for reallocation. We impute context-specific experience stocks for those teachers whose full teaching history is not observed based on the distribution of context-specific experience among the most experienced teachers whose full histories are observed. Adding in the full roster of teachers reduces dynamic gains relative to the actual allocation to .005 standard deviations for two-teacher fields, .014 standard deviations for four-teacher fields, and .025 for fields with eleven or more teachers (though note that 29% of the school-year fields in the full sample feature 11+ teachers, relative to 2% for the complete history subsample). These smaller simulated gains indicate that principals might make better use of their experienced teachers' context-specific talent, suggesting that they may learn teachers' comparative advantages slowly. When the random baseline is used instead, the corresponding gains are .011 for two-teacher fields, .027 standard deviations for four-teacher fields and .042 for fields with eleven or more teachers.

On one hand, these magnitudes are clearly not large enough to dramatically shift the distribution of student achievement; a .025 standard deviation test score gain is only enough to move an average student from the 50th to the 51st percentile of the state test score distribution. However, a number of other considerations suggest a more optimistic interpretation of these efficiency gains.

First, note that these gains are virtually costless: no change in existing staff is required, and all teaching loads are held fixed. It is rare to find the potential for across-the-board gains from policy changes that require so little upheaval.

Second, given that the vast majority of the test-score variation is within classes, most other school-level policies are likely to have a similarly-sized impact. For example, consider a policy that aims to identify and replace the worst 10 percent of teachers with new hires. Using the estimates from Table 4, the expected contribution of a randomly chosen teacher below the 10th percentile of general skill is -.22 test score standard deviations, so that if such teachers teach only 10 percent of students, average test scores would increase by 0.022 standard deviations even under the optimistic assumption that replacement teachers were of average quality.

Third, note that the vast majority of students are taught in high schools that feature seven or more teachers in a field. Furthermore, classrooms were only reallocated in tested courses, so that, for example, teachers who only taught calculus were not available for reallocation. Thus, the largest efficiency gains from our simulations are probably the relevant gains in most situations, and in fact may still be underestimates for most large schools.

Finally, these average gains conceal considerable heterogeneity in potential gains among schools. Consider the specification that incorporates task-specific talent, reallocates only teachers with fully observed teaching histories, and uses the observed allocation as the baseline. Focusing on schools with fields that generally feature seven or more teachers and averaging across fields and years, the mean dynamic gain from optimal reallocation among the 10 percent of schools featuring the smallest gains is only .004 standard deviations, while schools among the top decile of the distribution of dynamic gains are predicted to enjoy test score increases of .047 standard deviations on average. Thus, there seem to be a non-trivial subset of schools that might be able to reap substantial gains simply from changing their teacher assignment mechanism.

On the other hand, several additional caveats and limitations of our simulations should be noted. First, recall that the projected gains relative to the actual allocation rely on the questionable assumption that the principal does not have alternative sources of information beyond what is reflected in the full sample of test scores. Second, we are unable to evaluate the extent to which any achievement gains from an alternative teacher assignment mechanism would also contribute to or detract from other important non-test score student or school outcomes. Furthermore, because we do not allow our simulated classroom assignments to affect teacher turnover, the simulated efficiency gains could overstate even the true achievement gains if, for example, good teachers have a taste for variety, and quit more frequently if they are forced to teach the same subject-level combination repeatedly.[39] Similarly, our data do not permit us to estimate gains (or losses) to high levels of general and context-specific experience. It may be that some of the excess rotation of teachers away from their comparative advantages is necessary to prevent burnout or human capital depreciation among the most senior teachers. This might also lead us to overstate potential gains from optimal reallocation.[40]

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, about a third of the productivity gains from experience are specific to the subjects to which a teacher has been assigned, while about 74% of the variance in experience-invariant talent is portable across all courses. Nonetheless, our simulations provide suggestive evidence that existing allocations of teachers to classrooms in public high schools might be failing to exploit the variation in subject-specific and level-specific human capital that does exist, suggesting the potential for efficiency gains of around .02-.03 student test score standard deviations on average, with larger gains for some schools.

Note, however, that the results of the decomposition we estimate may not generalize to other occupations or

[39] However, Ost and Schiman (2015) suggests that the opposite is true in the elementary school context: teachers who rotate more frequently among grades exit schools at a higher rate.

[40] Note, though, that this scenario could also cause us to understate the gains to reallocation, since we have essentially assumed away any experience-based gains from reallocation among very experienced teachers by assigning the same experience productivity values to all levels of experience beyond 10 years.

even to alternative definitions of teachers' tasks. In particular, the tasks we consider are still fairly similar in scope. For example, we might observe greater variation in task-specific talent among teachers if we included serving 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.[41]

The methodology, however, does generalize: a similar decomposition may be estimated in any context in which worker productivity may be measured at the task level and where the blend of tasks changes over time. Indeed, there are many other organizational contexts in which we might also expect productivity to reflect a mix of general and task-specific talent as well as general and task-specific experience, and in which the nature of this production function may not be easily observable by employers or managers. A company employing a sales team to sell different products to different types of clientele, for example, might have both the wherewithal and the need to implement our decomposition.

9. Acknowledgments

We wish to thank Ron Ehrenberg, Michael Lovenheim, Maria Fitzpatrick, Damon Clark, Jordan Matsudaira, John Bishop, Jonah Rockoff, Karthik Muralidharan, and seminar participants at Cornell University, Queen's University, McMaster University, University of Rochester, as well as those at the NBER Economics of Education Fall Meeting, the Society of Labor Economists Annual Meeting and the AEA LERA session for immensely valuable input. This research is based on data from the North Carolina Education Research Data Center at Duke University. We acknowledge the North Carolina Department of Public Instruction for collecting and providing this information.

10. References

Aaronson, D., Barrow, L., Sander, W., 2007. Teachers and Student Achievement in Chicago Public Schools. Journal of Labor Economics 25 (1), 95–135. 2, 11

Altonji, J. G., Mansfield, R. K., 2014. Group-average observables as controls for sorting on unobservables when estimating group treatment effects: the case of school and neighborhood effects. Tech. rep., National Bureau of Economic Research. 5, 11

Atteberry, A., Loeb, S., Wyckoff, J., 2013. Do first impressions matter? improvement in early career teacher effectiveness. Tech. rep., National Bureau of Economic Research. 20

Aucejo, E., 2011. Assessing the role of teacher and student interactions. Working paper. 3

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., Wyckoff, J., 2008. The narrowing gap in new york city teacher qualifications and its implications for student achievement in high-poverty schools. Journal of Policy Analysis and Management 27 (4), 793–818. 2

[41] For instance, teachers who are effective at teaching abstract concepts may not be effective at handling student emotional crises.

Chetty, R., Friedman, J. N., Rockoff, J., 2016. Using lagged outcomes to evaluate bias in value-added models. Nber working paper 21961, National Bureau of Economic Research. 21

Chetty, R., Friedman, J. N., Rockoff, J. E., 2014. Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. The American Economic Review 104 (9), 2593–2632. 21, 22, 51, 52

Clement, M. B., Koonce, L., Lopez, T. J., 2007. The roles of task-specific forecasting experience and innate ability in understanding analyst forecasting performance. Journal of Accounting and Economics 44 (3), 378–398. 3

Clotfelter, C., Ladd, H., Vigdor, J., 2007. Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects. Economics of Education Review 26 (6), 673–682. 2

Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., 2006. Teacher-student matching and the assessment of teacher effectiveness. Journal of Human Resources 41 (4), 778–820. 2

Condie, S., Lefgren, L., Sims, D., 2014. Teacher heterogeneity, value-added and education policy. Economics of Education Review 40, 76–92. 3, 4

DeAngelo, G., Owens, E., 2012. Learning the Ropes: Task-Specific Experience and the Output of Idaho State Troopers. Working paper. 3

Gathmann, C., Schoenberg, U., 2010. How General Is Human Capital? A Task-Based Approach. Journal of Labor Economics 28 (1), 1–49. 3

Gibbons, R., Waldman, M., 2004. Task-specific human capital. The American Economic Review 94 (2), 203–207. 3

Goldhaber, D., Chaplin, D. D., 2015. Assessing the rothstein falsification test: Does it really show teacher value-added models are biased? Journal of Research on Educational Effectiveness 8 (1), 8–34. 21

Hanushek, E., Kain, J., O'Brien, D., Rivkin, S., 2005. The Market for Teacher Quality. Nber working paper 11154, National Bureau of Economic Research, Inc. 2

Harris, D. N., 2009. Would accountability based on teacher value added be smart policy? an examination of the statistical properties and policy alternatives. Education Finance and Policy 4 (4), 319–350. 2

Harris, D. N., Sass, T. R., 2011. Teacher training, teacher quality and student achievement. Journal of public economics 95 (7), 798–812. 2

Jackson, C. K., 2013. Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence from Teachers. Review of Economics and Statistics 95, 1096–1113. 2

Jackson, C. K., 2014. Do High School Teachers Really Matter? Journal of Labor Economics 32 (4). 2, 11

Jackson, C. K., Bruegmann, E., 2009. Teaching students and teaching each other: The importance of peer learning for teachers. American Economic Journal: Applied Economics 1 (4), 85–108. 2

Jacob, B., Rockoff, J., 2011. Organizing schools to improve student achievement: start times, grade configurations, and teacher assignments. Brookings Institution, Hamilton Project. 1

Kinsler, J., 2012. Assessing rothstein's critique of teacher value-added models. Quantitative Economics 3 (2), 333–362. 10, 21

Lockwood, J., McCaffrey, D., 2009. Exploring Student-Teacher Interactions in Longitudinal Achievement Data. Education Finance and Policy 4 (4), 439–467. 3

Mansfield, R., Forthcoming. Teacher Quality and Student Inequality. Journal of Labor Economics. 2, 11, 13, 17

Master, B., Loeb, S., Whitney, C., Wyckoff, J., 2012. Different skills: Identifying differentially effective teachers of english language learners. Working paper. 3

Ost, B., 2014. How Do Teachers Improve? The Relative Importance of Specific and General Human Capital. American Economic Journal: Applied Economics 6 (2), 127–51. 3

Ost, B., Schiman, J. C., 2015. Grade-specific experience, grade reassignments, and teacher turnover. Economics of Education Review 46, 112–126. 31

Papay, J. P., Kraft, M. A., 2015. Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. Journal of Public Economics. 2, 13, 15

Polataev, M., Robinson, C., 2008. Teachers, Schools, and Academic Achievement. Journal of Labor Economics 26 (3), 387–420. 3

Rivkin, S., Hanushek, E., Kain, J., 2005. Teachers, Schools, and Academic Achievement. Econometrica 73 (2), 417–458. 2

Rockoff, J., 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. 2

Rothstein, J., 2010. Teacher quality in educational production: Tracking, decay, and student achievement. The Quarterly Journal of Economics 125 (1), 175–214. 10, 21

Sass, T. R., Semykina, A., Harris, D. N., 2014. Value-added models and the measurement of teacher productivity. Economics of Education Review 38, 9–23. 2

Wiswall, M., 2013. The Dynamics of Teacher Quality. Journal of Public Economics 100, 61–78. 2, 15, 24

Yamaguchi, S., 2012. Tasks and heterogeneous human capital. Journal of Labor Economics 30 (1), 1–53. 3

11. Tables and Figures

Table 1: Effect of Sample Restrictions on Sample Composition

	Full Sample	Regression Sample
	(1)	(2)
School-Year Averages		
Enrollment	1,346.0	1,362.3
	(654.5)	(646.4)
# Teachers	23.1	23.3
	(8.9)	(8.7)
Teacher-Year Averages		
General Experience	4.96	3.28
	(3.65)	(3.01)
Subject Experience	3.78	2.37
	(3.30)	(2.48)
Level Experience	4.70	3.11
	(3.55)	(2.91)
Subject-Level Experience	3.42	2.17
	(3.11)	(2.35)
Classes Taught Per Year	3.44	3.34
	(1.52)	(1.50)
Unique Subj./Lvl. Taught Per Year	1.67	1.63
	(0.70)	(0.67)
Student-Year Averages		
Standardized Subject Test	0.041	-0.024
	(0.662)	(0.636)
Fraction of White Students	0.667	0.641
	(0.270)	(0.276)
Fraction of Black Students	0.259	0.278
	(0.252)	(0.258)
Fraction of Other Students	0.074	0.081
	(0.100)	(0.104)
8th Grade Standardized Reading Scores	0.095	-0.024
	(0.975)	(0.614)
8th Grade Standardized Math Scores	0.075	-0.030
	(0.976)	(0.651)
N (Aggregated Classroom Observations)	207,951	61,993

Notes: Student-test-weighted means and standard deviations (in parentheses) of classroom observations are reported for each sample. *Full Sample* includes all classroom observations with valid values for the variables in this table (i.e. current and 8th grade test scores, subject and level designation, race variables, teacher experience, class size, and grade). *Regression Sample* includes only classroom observations that satisfy the more extensive set of sample restrictions described in Section 4. The most important restriction is that the full history of course assignments must be observed for the teacher of the classroom. The *School-Year Averages* for the Regression Sample in Column (2) present the school-average student enrollment and teaching staff size from the *Full Sample*, but for the classrooms represented in the *Regression Sample* (now weighted by the number of student-tests in the regression sample). If we only count the subset of students and teachers that actually contribute an observation to our Regression Sample, student-test-weighted school means of enrollment and number of teachers are 504.9 and 8.4, respectively.

Table 2: Teacher Mobility Across Subjects: Regression Sample

		Math			Science				Social Studies			English
		Algebra 1	Algebra 2	Geometry	Biology	Chemistry	Physics	Physical Sciences	Civics	E/L/P	U.S. History	English
Math	Algebra 1	**1,860**	749	742	26	18	16	37	11	12	13	33
		1.000	0.403	0.399	0.014	0.010	0.009	0.020	0.006	0.006	0.007	0.018
	Algebra 2	749	**1,078**	533	4	9	14	14	1	3	3	0
		0.695	1.000	0.494	0.004	0.008	0.013	0.013	0.001	0.003	0.003	0.000
	Geometry	742	533	**1,142**	8	3	10	6	1	3	3	4
		0.650	0.467	1.000	0.007	0.003	0.009	0.005	0.001	0.003	0.003	0.004
Science	Biology	26	4	8	**1,472**	185	69	525	7	24	20	26
		0.018	0.003	0.005	1.000	0.126	0.047	0.357	0.005	0.016	0.014	0.018
	Chemistry	18	9	3	185	**554**	112	307	0	0	1	1
		0.032	0.016	0.005	0.334	1.000	0.202	0.554	0.000	0.000	0.002	0.002
	Physics	16	14	10	69	112	**243**	165	0	0	2	0
		0.066	0.058	0.041	0.284	0.461	1.000	0.679	0.000	0.000	0.008	0.000
	Physcial Sciences	37	14	6	525	307	165	**1,151**	6	24	15	21
		0.032	0.012	0.005	0.456	0.267	0.143	1.000	0.005	0.021	0.013	0.018
Soc. Stu.	Civics	11	1	1	7	0	0	6	**904**	279	412	12
		0.012	0.001	0.001	0.008	0.000	0.000	0.007	1.000	0.309	0.456	0.013
	E/L/P	12	3	3	24	0	0	24	279	**952**	414	52
		0.013	0.003	0.003	0.025	0.000	0.000	0.025	0.293	1.000	0.435	0.055
	U.S. History	13	3	3	20	1	2	15	412	414	**1,235**	36
		0.011	0.002	0.002	0.016	0.001	0.002	0.012	0.334	0.335	1.000	0.029
English	English	33	0	4	26	1	0	21	12	52	36	**2,162**
		0.015	0.000	0.002	0.012	0.000	0.000	0.010	0.006	0.024	0.017	1.000

Notes: E/L/P denotes Econ/Law/Politics. 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 subject (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 subject who are also observed teaching the j-th subject at some point during the sample.

Table 3: Teacher Mobility Across Math Subject-Level Combinations: Regression Sample

		Algebra 1		Algebra 2		Geometry	
		Low	High	Low	High	Low	High
Algebra 1	Low Level	**1,855** 1.000	27 0.015	676 0.364	331 0.178	678 0.365	315 0.170
	High Level	27 0.844	**32** 1.000	18 0.563	10 0.313	14 0.438	7 0.219
Algebra 2	Low Level	676 0.700	18 0.019	**966** 1.000	341 0.353	451 0.467	194 0.201
	High Level	331 0.731	10 0.022	341 0.753	**453** 1.000	202 0.446	118 0.260
Geometry	Low Level	678 0.644	14 0.013	451 0.428	202 0.192	**1,053** 1.000	368 0.349
	High Level	315 0.689	7 0.015	194 0.425	118 0.258	368 0.805	**457** 1.000

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 subject- 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 subject-level who are also observed teaching the j-th subject-level at some point during the sample.

Table 4: True Variances in Fixed Effects (Using Year-Based Measure of Teacher Experience with the Baseline Specification)

	Lower Bound		**Intermediate**		**Upper Bound**	
	Var.	SD	Var.	SD	Var.	SD
	(1)	(2)	(3)	(4)	(5)	(6)
Sch-Subj-Lvl-Tch Combos	0.0236	0.154	0.0467	0.216	0.0605	0.246
General Talent	0.0175	0.132	0.0368	0.192	0.0506	0.225
Subj-Lvl Combos	0.0061	0.078	0.0099	0.099	0.0099	0.099
Sch-Lvl-Tch Combos	0.0197	0.140	0.0407	0.202	0.0545	0.234
Subject Talent	0.0039	0.063	0.0060	0.077	0.0060	0.077
Sch-Subj-Tch Combos	0.0215	0.147	0.0433	0.208	0.0571	0.239
Level Talent	0.0021	0.045	0.0034	0.058	0.0034	0.058
Subject-Level Talent	0.0001	0.011	0.0005	0.023	0.0005	0.023

Notes: Standard errors are clustered at the teacher level. *Lower Bound* estimates allocate all of the between school-subject-level variance in residual test scores to school and student inputs (Assumption 2A). This is implemented by including school-subject-level fixed effects and normalizing the mean among school-teacher-subject-level fixed effects to be 0 in each school-subject-level. *Intermediate* estimates allocate the between school variance in residual test scores to school and student inputs, and the within-school/between subject-level variance to teachers (Assumption 2B). This is implemented by replacing the school-subject-level fixed effects with school fixed effects only. *Upper Bound* estimates allocate all of the between school-subject-level variance in residual test scores to teachers (Assumption 2C). This is implemented by removing all school-level controls. See Section 3.2 for details.

Table 5: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Baseline Specification)

Years Experience	General	Subject	Level	Subj.-Level	Combined
	(1)	(2)	(3)	(4)	(5)
1 yr	0.066***	0.014	-0.006	0.000	0.074***
	[0.017]	[0.015]	[0.015]	[0.013]	[0.006]
2 yrs	0.085***	0.033*	-0.006	-0.013	0.099***
	[0.025]	[0.023]	[0.022]	[0.020]	[0.010]
3 yrs	0.090***	0.049**	-0.001	-0.027	0.110***
	[0.030]	[0.029]	[0.028]	[0.026]	[0.014]
4 yrs	0.097***	0.053*	-0.004	-0.033	0.113***
	[0.035]	[0.033]	[0.033]	[0.031]	[0.018]
5-6 yrs	0.097***	0.056*	0.008	-0.044	0.116***
	[0.040]	[0.039]	[0.037]	[0.036]	[0.023]
7-10 yrs	0.113***	0.046	-0.006	-0.054	0.098***
	[0.046]	[0.046]	[0.044]	[0.043]	[0.031]
11-14 yrs	0.093**	0.067	0.024	-0.099**	0.085**
	[0.054]	[0.059]	[0.055]	[0.060]	[0.046]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. The regression includes school-teacher-subject-level fixed effects, calendar year fixed effects, and a vector of classroom observable characteristics with subject-level-specific coefficients. The regression also includes controls for teacher workload (number of current class periods and number of distinct subject-levels taught) and depreciation of experience capital (indicators for whether the teacher taught a class in the current subject, level, subject-level, or taught at all last year, as well as analogous indicators for teaching in each context two years ago). Finally, the regression also includes controls for decreasing effort/productivity shocks in the year prior to an assignment change (indicators for whether the current year is the final time the teacher taught the subject, level, subject-level associated with the observation, as well as whether the current year is the teacher's final year of teaching high school in North Carolina. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every year of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

Table 6: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification)

Years Experience	General	Subject	Level	Subj.-Level	Combined
	(1)	(2)	(3)	(4)	(5)
1 yr	0.065***	0.014*	-0.003	0.013*	0.089***
	[0.011]	[0.009]	[0.010]	[0.008]	[0.004]
2 yrs	0.085***	0.023**	-0.004	0.014*	0.118***
	[0.014]	[0.012]	[0.012]	[0.010]	[0.006]
3 yrs	0.093***	0.036***	-0.007	0.008	0.131***
	[0.016]	[0.014]	[0.014]	[0.012]	[0.007]
4 yrs	0.101***	0.041***	-0.011	0.007	0.138***
	[0.018]	[0.015]	[0.016]	[0.014]	[0.008]
5-6 yrs	0.103***	0.041***	-0.002	0.009	0.152***
	[0.019]	[0.017]	[0.017]	[0.015]	[0.009]
7-10 yrs	0.114***	0.025	-0.008	0.006	0.138***
	[0.022]	[0.021]	[0.020]	[0.019]	[0.012]
11-14 yrs	0.107***	0.027	0.027	-0.019	0.141***
	[0.028]	[0.038]	[0.028]	[0.041]	[0.026]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. *Restricted Specification* refers to the specification in equation (3) in which the school-teacher-subject-level fixed effects $\hat{\mu}_{srjl}$ from Equation (2) are restricted to be common across subject-levels (i.e. replaced by school-teacher effects). Refer to notes below Table 5 for a full description of the control variables. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every year of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

Table 7: Effect of Years of General and Subject-Specific Experience on Student Test Scores (Restricted Specification with Level and Subject-Level Experience Additionally Constrained to 0)

	Restricted Specification w/ Lvl. & Subj.-Lvl. Exp. Gains Constrained to 0			**Standard Specification**
	General	Subject	Combined	"General"
	(1)	(2)	(3)	(4)
1 yr	0.063***	0.025***	0.088***	0.084***
	[0.007]	[0.006]	[0.004]	[0.004]
2 yrs	0.081***	0.036***	0.118***	0.113***
	[0.009]	[0.008]	[0.005]	[0.005]
3 yrs	0.087***	0.045***	0.133***	0.127***
	[0.010]	[0.009]	[0.007]	[0.006]
4 yrs	0.092***	0.048***	0.140***	0.136***
	[0.011]	[0.010]	[0.007]	[0.007]
5-6 yrs	0.100***	0.050***	0.151***	0.148***
	[0.012]	[0.011]	[0.009]	[0.008]
7-10 yrs	0.107***	0.032**	0.139***	0.148***
	[0.014]	[0.014]	[0.011]	[0.010]
11-14 yrs	0.124***	0.020	0.143***	0.157***
	[0.020]	[0.025]	[0.023]	[0.016]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. *Restricted Specification* refers to the specification in equation (3). Columns (1-3) report results from imposing on the Restricted Specification the additional restrictions that gains from level-specific and subject-level-specific experience are constrained to be 0: $d^l(exp) = 0$ and $d^{jl}(exp) = 0 \; \forall \; exp$. Column 4 reports results from imposing the further restriction that $d^j(exp) = 0 \; \forall \; exp$, for ease of comparison with with standard experience profiles estimated in the literature. Refer to notes below Table 5 for a full description of the control variables. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1 & 4) or in the subject (Col. 2) associated with the current classroom observation. Column 3, entitled *Combined*, captures the combined predicted contribution of both dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every year of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

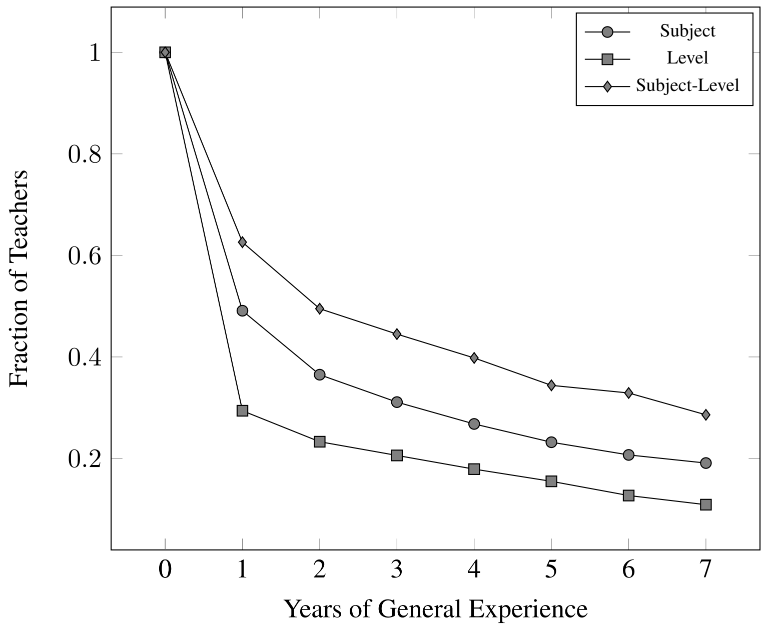
Table 8: Counterfactual Simulations: Achievement Gains from Optimal Allocation Relative to Actual and Random Allocations (Year-Based Measure of Experience, Excluding Teachers Without Full Histories)

Eligible Teach.		**Static**		**Dynamic**	
		Actual (1)	Random (2)	Actual (3)	Random (4)
2	**Total**	**.017**	**.025**	**.017**	**.025**
	Talent	.018	.021	.018	.020
	Exper.	-.001	.004	-.001	.005
3	**Total**	**.026**	**.035**	**.027**	**.039**
	Talent	.027	.030	.028	.031
	Exper.	-.001	.005	-.000	.008
4	**Total**	**.031**	**.041**	**.033**	**.044**
	Talent	.032	.035	.033	.035
	Exper.	-.001	.006	.001	.010
5-6	**Total**	**.038**	**.048**	**.040**	**.053**
	Talent	.039	.041	.039	.042
	Exper.	-.001	.007	.001	.011
7-10	**Total**	**.039**	**.050**	**.041**	**.052**
	Talent	.039	.042	.039	.041
	Exper.	.000	.008	.002	.012
11+	**Total**	**.042**	**.053**	**.044**	**.054**
	Talent	.039	.044	.039	.039
	Exper.	.003	.009	.005	.015

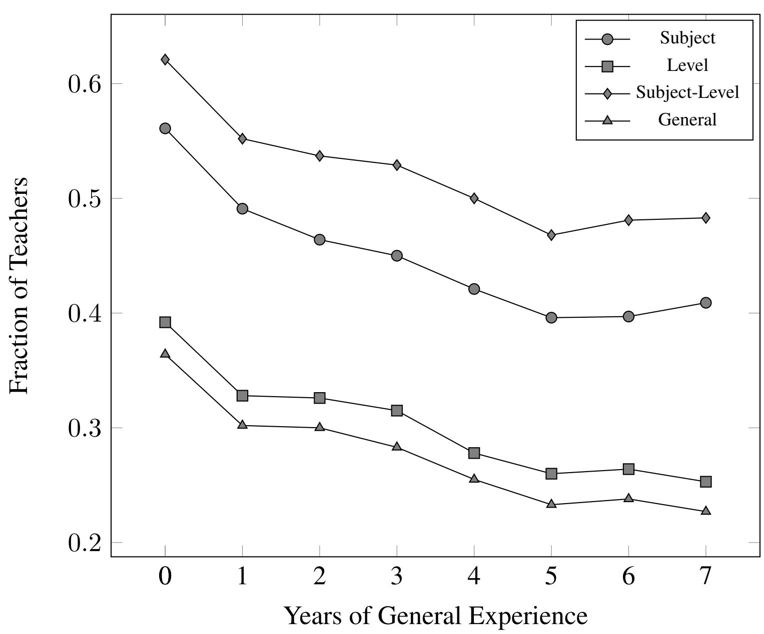
Notes: Each cell presents simulated achievement gains from the optimal allocation of teachers to classrooms relative to either the observed allocation (in columns labeled "Actual") or a randomly selected feasible allocation (columns labeled "Random") among all school-year-field combinations with the number of eligible teachers specified by the row label. Classroom-level gains are pooled across the three fields (math, science, and social studies). The top entry in each cell displays the total achievement gains, while the middle and bottom entries display the components of the gains attributable to task-specific experience and task-specific talent, respectively. *Static* refers to simulations in which teacher experience stocks are held fixed as they were in the actual sample through year $t-1$ prior to simulated reassignment in year t. *Dynamic* refers to simulations in which teacher experience stocks used as the basis for simulated reassignment in year t are based on simulated assignments from 1995 through year $t-1$. See Section 7.1 and Appendix Section Appendix G for further detail about simulation methodology. A teacher is eligible for reassignment if their full teaching history is observed in the data. Estimates of gains from task-specific experience and of teachers' task-specific talent are derived from the Full Specification (equation (13)). The principal incorporates information from empirical Bayes posterior beliefs about each teacher's task-specific talent based on our school-teacher-subject-level fixed effect estimates for any school-teacher-subject-level combination that is observed in our sample. We assign task-specific productivities of 0 for any school-teacher-subject-level combination that we do not observe.

Figure 1: Fraction of Teachers Starting New or Discontinuing Existing Courses by Year of General Experience

(a) First Time Teaching Course



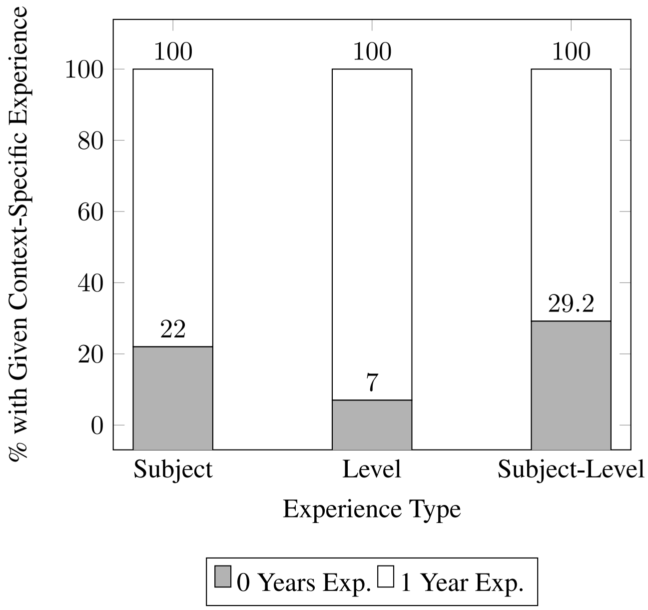
(b) Last Time Teaching Course



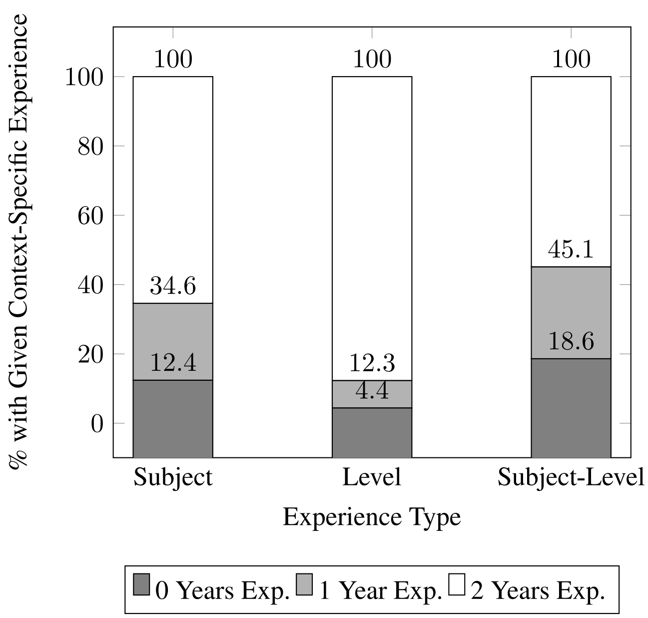
Notes: Panel A plots the fraction of teachers with the given number of years of general experience that teach a new subject, level, and subject-level combination, respectively, in that year that they have not previously taught. Panel B plots the fraction of teachers with the given number of years of general experience that discontinue teaching at least one subject, level, and subject-level combination, respectively, after the chosen year. Counts for the number of school-teacher-year observations associated with each general experience level are: 0 (5,294), 1 (4,249), 2 (3,545), 3 (2,901), 4 (2,322), 5 (1,792), 6 (1,385), 7 (1,106).

Figure 2: The Distribution of Context-Specific Experience among Second- and Third-Year Teachers

(a) Second-Year Teachers

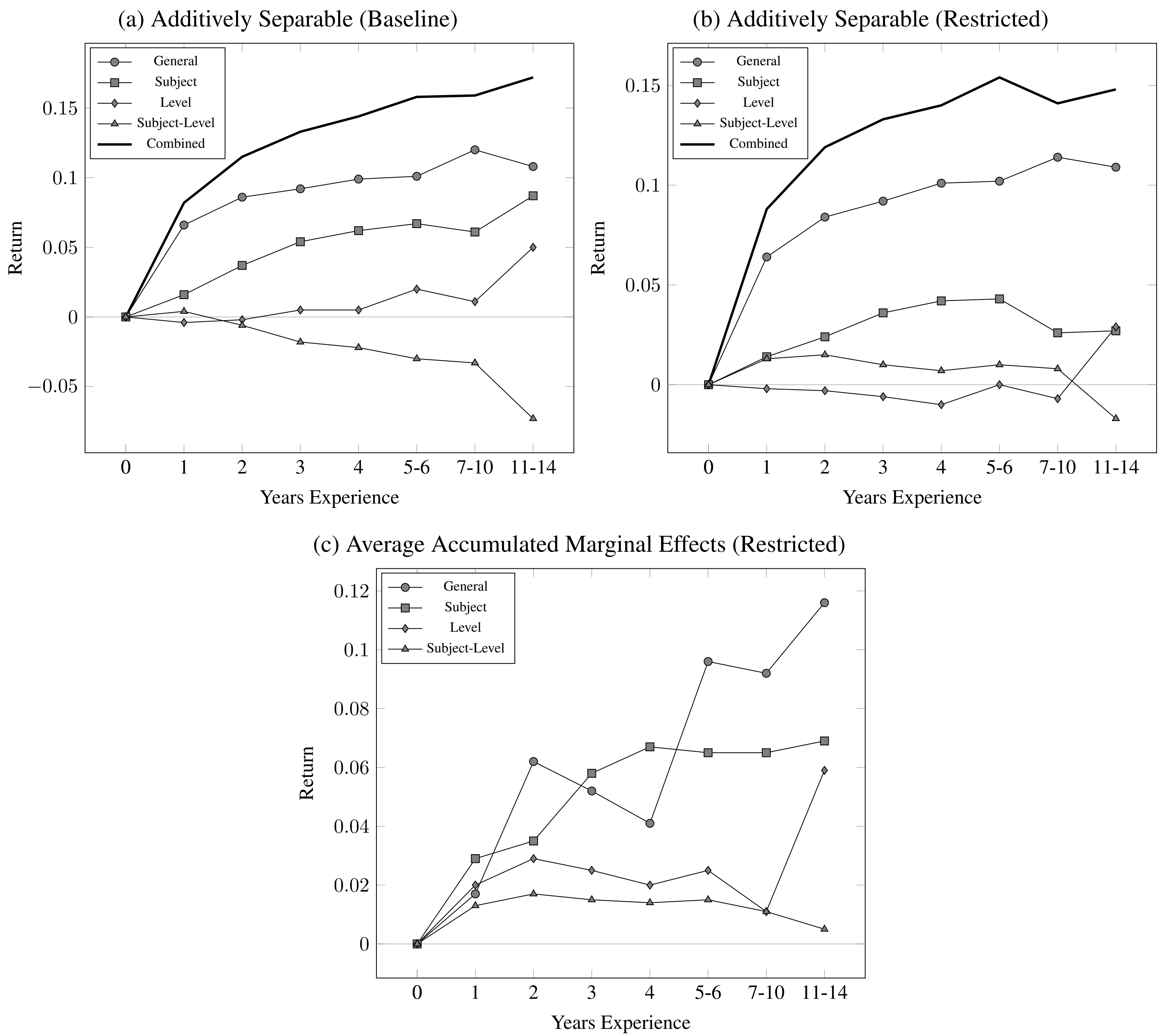


(b) Third-Year Teachers



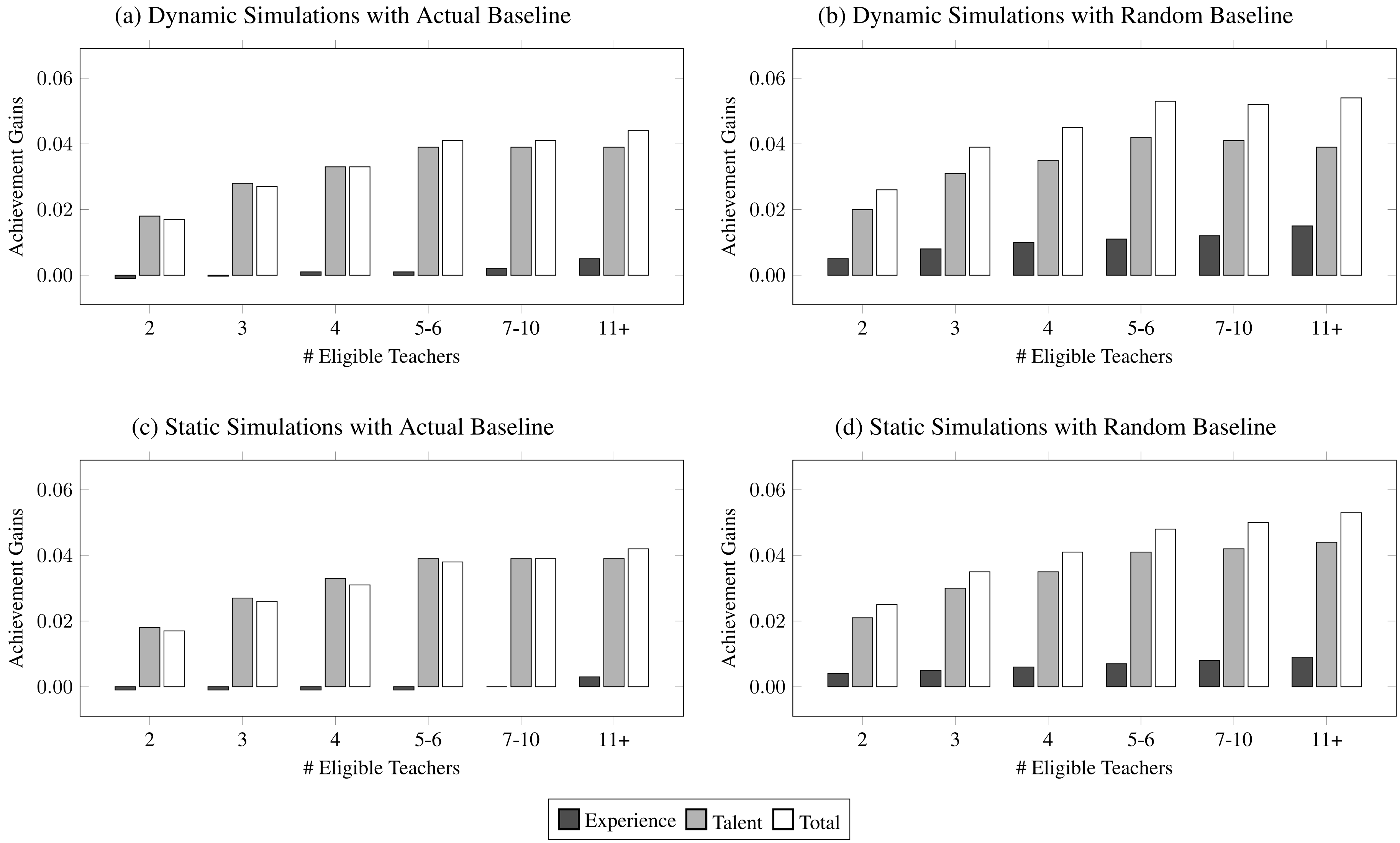
Notes: The figure displays the classroom-weighted distribution of four-dimensional experience stocks among second- and third-year teachers in our final sample. The sample includes 10,270 and 8,665 total classes taught by a second-year and third-year teacher respectively. Panel A displays the fractions of classrooms taught by second-year teachers in which the teacher has 0 versus 1 prior years of the relevant subject-, level-, and subject-level-specific experience, respectively. Panel B displays the fractions of classrooms taught by third-year teachers in which the teacher has 0, 1, and 2 prior years of the relevant subject-, level-, and subject-level-specific experience, respectively. Note that multiple subject-level combinations can be taught in a year. The full joint distribution of four dimensional experience profiles for second- and third-year teachers can be found in Appendix table H.1.

Figure 3: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Various Specifications)



Notes: Figures 3a, 3b, and 3c plot the entries from Tables 5, 6, and H.19, respectively. Refer to the notes from these tables for further detail concerning these specifications.

Figure 4: Counterfactual Simulations: Achievement Gains from Optimal Allocation Relative to Actual and Random Allocations (Year-Based Measure of Experience, Excluding Teachers Without Full Histories)



Notes: Each cell presents simulated achievement gains from the optimal allocation of teachers to classrooms relative to either the observed allocation (in sub-figures labeled "Actual") or a randomly selected feasible allocation (sub-figures labeled "Random") among all school-year-field combinations with the number of eligible teachers specified on the x-axis. The gains reported are averages of classroom-level gains across all classrooms in math, science, and social studies from the final 5 years of simulated allocations (2005-2009). The white cells display the total achievement gains, while the grey and black cells display the components of the gains attributable to more efficient use of task-specific talent and task-specific experience, respectively. Subfigures labeled *Static* refer to simulations in which teacher experience stocks are held fixed as they were in the actual sample through year $t-1$ prior to simulated reassignment in year t. Subfigures labeled *Dynamic* refer to simulations in which simulated classroom assignments from 1995 through year $t-1$ are used construct the teacher experience stocks that determine the simulated reassignment in year t . See Section 7.1 and Appendix Section Appendix G for further detail about simulation methodology. A teacher is eligible for reassignment if their full teaching history is observed in the data. Estimates of gains from task-specific experience and of teachers' task-specific talent are derived from the Full Specification (equation (13)). The principal incorporates information from empirical Bayes posterior beliefs about each teacher's task-specific talent based on our school-teacher-subject-level fixed effect estimates for any school-teacher-subject-level combination that is observed in our sample. We assign task-specific productivities of 0 for any school-teacher-subject-level combination that we do not observe.

Technical Appendices

Appendix A. Identification of Experience Profiles

To see how identification of the returns to both general and three dimensions of context-specific experience 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 that four different teachers (not necessarily at the same school) each teach different subject-level combinations in their first years: Teacher 1 teaches basic physics (BP) in her first year, while teacher 2 teaches honors physics (HP), teacher 3 teaches basic chemistry (BC) and teacher 4 teaches honors chemistry (HC). Suppose then that all four teach honors chemistry (HC) every year thereafter. To keep the example simple, suppose further that the gains from each of the components of experience are fully persistent (no depreciation), and that each teacher only teaches classes in one subject-level per year. Panel A of Table H.23 displays the course assignment paths taken by each teacher, along with the observed stocks of general, subject-specific, level-specific, and subject-level specific experience that teachers will possess at the beginning of each of their school years.

Consider a difference-in-difference estimator that compares the change in teacher 1's average student test scores between years 2 and 3 with the corresponding change for teacher 2. Since each teacher teaches the same subject-level (HC) in both year 2 and year 3, focusing on changes over time differences out the permanent general and context-specific skills of the two teachers (as well as any differences in time-invariant school quality). Furthermore, comparing across teachers removes the common gains from the second year of (previous) general experience and the first year of subject-specific and subject-level specific experience. Because teacher 2 taught at the honors level in her first year, the extent to which teacher 1's performance converges to or diverges from teacher 2's performance between years 2 and 3 will reflect the relative value of the 2nd year of level-specific experience compared to the 1st year: $(d^l(2) - d^l(1)) - (d^l(1) - 0)$.[42] If instead we compare the change in student performance between years 3 and 4 for the same two teachers (1 and 2), we recover the relative value of the 3rd year of level-specific experience compared to the 2nd year: $(d^l(3) - d^l(2)) - (d^l(2) - d^l(1))$. Indeed, conditional on knowing the value of the first year of level-specific experience, $d^l(1)$, we can trace out the entire path of returns to level-specific experience by comparing the divergence/convergence in the performance of teachers 1 and 2 as they progress through their careers. If we replace teacher 2 with teacher 3 in the comparisons above, we instead trace out the path of returns to subject-specific experience. Now that the returns to subject-specific and level-specific experience have been identified, replacing teacher 3 with teacher 4 identifies the path of returns to subject-level-specific experience. Finally, the growth path of teacher 4, who never switched subjects or levels, identifies the path of returns to general experience.

To see how the value of the first year of experience might be identified for each component of experience, consider a second scenario in which teacher 1 teaches the following sequence of courses in her first four years: $BC \rightarrow HC \rightarrow BP \rightarrow HC$. Teacher 2 teaches the same set of courses, but in a different sequence: $BP \rightarrow HC \rightarrow BC \rightarrow HC$. Panel B of Table H.23 illustrates the stocks of general and context-specific experience each teacher possesses at the beginning of each year of teaching. Since both teachers teach honors chemistry with the same accumulated experience profile in year 4, comparing the performance of the two

[42] 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 $k \in \{gen, j, l, lk\}$.

teachers identifies the difference in permanent teaching skill between the two teachers (part of which may be honors-chemistry specific): $\mu_{2CH} - \mu_{1CH}$. Once relative permanent skill has been identified, comparing the same two teachers' average student residuals in year 2 (when both were teaching honors chemistry) identifies the return to the 1st year of subject-specific experience, $d^j(1)$. Replacing basic chemistry with honors physics in this example would instead identify the return to the 1st year of level-specific experience ($d^l(1)$), while replacing it with honors chemistry would identify the return to the 1st year of subject-level specific experience ($d^{jl}(1)$). The return to the first year of general experience ($d^t(1)$) can then be identified via the growth in student average residuals from the 1st to the 2nd year from teachers who teach the same subject-level in each of their first two years.

While the sample histories used in these scenarios are stylized, note that there are many alternative moments that also provide identifying variation. 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. Since each different sequence also implies a different pattern of potential depreciation for a given model of depreciation, such comparisons allow us to simultaneously estimate the rates at which different experience components depreciate.[43]

Furthermore, each subject or level switch, regardless of the point in the career, provides a further source of identifying variation for the various context-specific experience profiles. Consequently, not only are these experience profiles estimable with reasonable precision (at least for the first several 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). Indeed, in Section 6 we show that the function linking four-dimensional stocks of general and context-specific teacher experience to student performance is non-parametrically identified, and we present estimates from a more flexible (though noisily estimated) specification.

Appendix B. Identification of Permanent Teaching Skill

To illustrate how $\hat{\mu}_{srjl}$ can be identified given any of Assumptions 2A-2C paired with 3A-3C, consider a teacher r' who teaches subject j' and level l' in school s' during years t_1 to t_2. Let $Z_{ct} = Y_{ct} - X_{ct}\beta$ represent the average test score residual in classroom c at time t, after removing the component predictable based on classroom inputs. Then the average residual performance of students taught by teacher r' in school-subject-level combination (s', j', l') is given by:

$$\begin{aligned} &E[Z_{ct}|(s,r,j,l) = (s',r',j',l')] \\ &= \delta_{s'j'l'} + \mu_{s'r'j'l'} + \sum_{t'=t_1}^{t_2} w_{t'}[d^{gen}(exp^{gen}_{r't'}) + d^j(exp^j_{r't'}) + d^l(exp^l_{r't'}) + d^{jl}(exp^{jl}_{r't'})] \end{aligned} \tag{B.1}$$

where the weight $w_{t'}$ captures the fraction of all the students teacher r' taught in combination (s', j', l') that were taught in year t'. Since the experience profiles $d^{gen}(*)$, $d^j(*)$, $d^l(*)$, and $d^{jl}(*)$ were identified

[43] In practice, after some experimentation, we include in our estimated specifications four dummy variables indicating whether the teacher taught the current subject last year, the current level last year, the current subject-level last year, and whether the teacher taught any class last year.

using comparisons of changes in performance across years in Section Appendix A, the average level of performance of teacher r' while teaching in school-subject level combination (s', j', l') identifies $\delta_{s'j'l'} + \mu_{s'r'j'l'}$. Under Assumption 2C, $\delta_{sjl} = 0 \; \forall \; (s, j, l)$, so this moment identifies $\mu_{s'r'j'l'}$ 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:

$$\begin{aligned} & E[Z_{ct}|(s,j,l) = (s',j',l')] \\ & = \delta_{s'j'l'} + E[d^{gen}(exp^{gen}) + d^j(exp^j) + d^l(exp^l) + d^{jl}(exp^{jl})|(s,j,l) = (s',j',l')], \end{aligned} \tag{B.2}$$

which identifies $\delta_{s'j'l'}$, leaving the teacher-specific average to identify $\mu_{s'r'j'l'}$. To identify $\delta_{s'}$ under Assumption 2B, 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.

Appendix C. Recovering the Latent Variance Decomposition

This section shows how to distill the true decomposition of time-invariant skill into general, subject-specific, level-specific, and subject-level specific components from the estimated cell fixed effects $\{\hat{\mu}_{srjl}\}$. We first assume that each estimated school-teacher-subject-level fixed effect $\hat{\mu}_{srjl}$ can be written as the sum of the teacher's true context-specific skill and an uncorrelated error component: $\hat{\mu}_{srjl} = \mu_{srjl} + \xi_{srjl}$. Let $\mathcal{C}$ and C represent the set of classrooms and the total number of classrooms in the sample, respectively. In addition, let $\mu_{srjl(c)}$ represent the context-specific skill of the teacher that taught classroom c. Then the (student-weighted) sample variance in estimated context-specific skill can be decomposed as:

$$\frac{1}{C}\sum_{c\in\mathcal{C}} w_c(\hat{\mu}_{srjl(c)})^2 = \frac{1}{C}\sum_{c\in\mathcal{C}} w_c(\mu_{srjl(c)})^2 + \frac{1}{C}\sum_{c\in\mathcal{C}} w_c(\xi_{srjl(c)})^2 \tag{C.1}$$

where each w_c is a weight capturing the fraction of all student test scores in the sample that were associated with classroom c.

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

$$\hat{Var}(\mu_{srjl}) = \frac{1}{C}\sum_{c\in\mathcal{C}} w_c(\hat{\mu}_{srjl(c)})^2 - \frac{1}{C}\sum_{c\in\mathcal{C}} w_c(\xi_{srjl(c)})^2. \tag{C.2}$$

The sampling error components $\{\xi_{srjl}\}$ are not observed, but

$$\frac{1}{C}\sum_{c\in\mathcal{C}} w_c(\xi_{srjl(c)})^2 \approx \frac{1}{C}\sum_{c\in\mathcal{C}} w_c E[(\xi_{srjl(c)})^2] = \frac{1}{C}\sum_{c\in\mathcal{C}} w_c se(\xi_{srjl(c)})^2, \tag{C.3}$$

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

$$\hat{Var}(\mu_{srjl}) = \frac{1}{C}\sum_{c\in\mathcal{C}} w_c(\hat{\mu}_{srjl(c)})^2 - \frac{1}{C}\sum_{c\in\mathcal{C}} w_c se(\xi_{srjl(c)})^2. \tag{C.4}$$

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

$$\hat{Var}(\overline{\mu}_{sr}) = \hat{Var}(\mu_{srjl}) - \hat{Var}(\tilde{\mu}_{srjl}) \tag{C.5}$$

To prevent teachers who only taught a single subject-level combination from biasing our estimate of $\hat{Var}(\overline{\mu}_{sr})$ downward, when calculating $\hat{Var}(\tilde{\hat{\mu}}_{srjl})$ we restrict the sample of school-teacher-subject-level combinations 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. [44]

Appendix D. Testing the Additive Separability of Context-Specific Experience Profiles

Appendix D.1. Smoothing the Nonparametric Experience Contribution Function

To address the volatility of our experience cell fixed effect estimates, we assume that the true structural function $d(*, *, *, *)$ is differentiable everywhere, and smooth our estimates using a kernel function featuring the normal PDF with zero mean and standard deviation 0.5 (denoted ϕ despite the non-unity standard deviation):

$$\tilde{d}(\mathbf{exp}) = \frac{\sum_{\mathbf{exp'} \in \mathcal{EX}} w_{exp'} \phi(|\mathbf{exp'} - \mathbf{exp}|) \hat{d}(\mathbf{exp'})}{\sum_{\mathbf{exp'} \in \mathcal{EX}} w_{exp'} \phi(|\mathbf{exp'} - \mathbf{exp}|)}, \tag{D.1}$$

where $\hat{d}(\mathbf{exp'})$ is the estimate on the given experience profile from equation (13). The argument to the normal density $|\mathbf{exp'} - \mathbf{exp}|$ is the L1 norm or taxicab distance between the two experience profiles: $|\mathbf{exp'} - \mathbf{exp}| = |exp^{gen'} - exp^{gen}| + |exp^{j'} - exp^{j}| + |exp^{l'} - exp^{l}| + |exp^{jl'} - exp^{jl}|$. The weight $w_{exp'}$ represents the fraction of observations in the sample in which the experience profile exp' is observed. Thus, the impact of $\hat{d}(1, 1, 1, 1)$ on $\tilde{d}(1, 1, 1, 0)$ will be greater than that of $\hat{d}(1, 1, 0, 0)$, despite equal L1 distances, because $\hat{d}(1, 1, 1, 1)$ is much more precisely estimated than $\hat{d}(1, 1, 0, 0)$. The chosen bandwidth yields a four-dimensional function $\tilde{d}(*, *, *, *)$ that is smooth enough to remove considerable sampling error, yet is still flexible enough to reveal true complementarities where they may occur.

[44] 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 first using the delta method to calculate standard errors for $\widehat{(\tilde{\mu}_{srjl} - \overline{\tilde{\mu}}_{srj})}$ and then applying the same method as above. We then obtain the variance in subject-specific teaching talent, $\hat{Var}(\overline{\tilde{\mu}}_{srj})$, via $\hat{Var}(\overline{\tilde{\mu}}_{srj}) = \hat{Var}(\tilde{\hat{\mu}}_{srjl}) - \hat{Var}((\tilde{\mu}_{srjl} - \overline{\tilde{\mu}}_{srj}))$. 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: $\hat{Var}(\tilde{\mu}_{srjl} - \overline{\tilde{\mu}}_{srj} - \overline{\tilde{\mu}}_{srl}) = \hat{Var}(\tilde{\hat{\mu}}_{srjl}) - \hat{Var}(\overline{\tilde{\mu}}_{srj}) - \hat{Var}(\overline{\tilde{\mu}}_{srl})$.

Appendix E. Marginal Effects Example

This subsection shows how we estimate profiles of returns to single dimensions of experience from the smoothed nonparametric experience cell estimates. These profiles can then be compared with the corresponding dimension-specific profiles from our additively-separable baseline specification. For each initial value v of each component of experience, we approximate the average partial derivative of the experience production function at v ($E[\frac{\partial d(exp^{gen}, exp^{j}, exp^{l}, exp^{sl})}{\partial exp^{dim}} | exp^{dim} = v]$ for $dim \in \{gen, j, l, sl\}$) by calculating a weighted average marginal effect of an extra year of the chosen component of experience (holding the other experience components fixed). The weighted average is taken over all combinations of the other three experience dimensions that are observed among experience cells that feature the chosen initial value v in the chosen dimension $dim \in \{gen, j, l, sl\}$.

For example, let $\mathcal{Q}^{j,v}$ denote the set of experience cells at which a partial derivative for subject-specific experience at initial value v may be calculated:

$$\begin{aligned}&\mathcal{Q}^{j,v} = \{(exp^{t}, exp^{j}, exp^{l}, exp^{sl}) : \\ &exp^{j} = v, (exp^{t}, exp^{j}, exp^{l}, exp^{sl}) \in \mathcal{D}, (exp^{t}, exp^{j}+1, exp^{l}, exp^{sl}) \in \mathcal{D}\}.\end{aligned}$$

Then the average marginal effect of subject-specific experience among cells featuring $exp^{j} = v$ can be calculated via:

$$\begin{aligned}&\frac{\overline{\partial} d(exp^{t}, v, exp^{l}, exp^{sl})}{\partial exp^{j}} = \\ &\sum_{k \in \mathcal{Q}^{j,v}} w_k [\hat{d}(exp^{t}_{k}, v+1, exp^{l}_{k}, exp^{sl}_{k}) - \hat{d}(exp^{t}_{k}, v, exp^{l}_{k}, exp^{sl}_{k})]\end{aligned}$$

The weight w_k is composed of the product of two sub-weights associated with the two experience cells included in the partial derivative estimate. Each sub-weight represents the fraction of all teacher-school-subject-level-year cells that feature the chosen experience combination. The w_k are then re-scaled to sum to 1.

Appendix F. Methodology for Measuring Forecast Bias

This appendix describes the implementation of the test for forecast bias in our estimates of teacher talent discussed in Section 6.3. The intuition for the test is that if the estimated fixed effects $\{\hat{\mu}_{srjl}\}$ properly capture the true talent contributions $\{\mu_{srjl}\}$, then differences in fixed effects among teachers in a chosen context estimated from one partition of our data should predict differences in mean residual achievement in the same context in a second, left out partition. Our methodology closely mirrors that of Chetty et al. (2014).

The first step is to construct an appropriate sample of school-teacher-subject-level (hereafter SRJL) combinations. Two conditions must be met for a given SRJL combination to enter the test sample. First, the teacher must have taught at least three classes in the chosen school-subject-level (hereafter SJL). At least one class must be available for the forecasted partition, and at least two classes must be available for the partition that forms the basis for the forecast (so that a valid standard error for the estimate $\hat{\mu}_{srjl}$ can be formed).[45] Second, the SJL context associated with the chosen SRJL combination must be shared by at least one other teacher. This ensures that persistent school inputs that are specific to (or differentially important in) the chosen course-level can be differenced out. 2,285 out of an original 18,257 SRJL combinations satisfy these two restrictions.

Let $Z^F_{ct} \equiv Y_{ct} - X_{ct}\hat{\beta}_{jl} - \hat{d}^{gen}(exp^{gen}_{rt}) - \hat{d}^j(exp^j_{rt}) - \hat{d}^l(exp^l_{rt}) - \hat{d}^{jl}(exp^{jl}_{rt})$ represent classroom level residuals, where the values $\hat{\beta}_{jl}$, $\hat{d}^{gen}$, $\hat{d}^j$, $\hat{d}^l$, and $\hat{d}^{jl}$ are used to form the residuals represent the estimated coefficients on the student demographics, year dummies, and context-specific experience profiles from our restricted specification (3).

We randomly select one classroom from each eligible SRJL to form the forecasted partition (Partition 2), and assign the remaining classrooms to Partition 1. Let $Z^F_{ct} \equiv [Z^{1F}_{ct}, Z^{2F}_{ct}]$ capture the corresponding partition of classroom-level residuals. We fit the following model to the class-level achievement data from Partition 1:

$$Z^1_{ct} = \mu^F_{srjl} + \epsilon^F_{ct} \tag{F.1}$$

Since the estimated fixed effects from this model, $\hat{\mu}^F_{srjl}$, still contain the contributions of school-subject-level inputs δ_{sjl}, we choose a teacher r' from each SJL environment and subtract this teacher's context-specific fixed effect from those of the other teachers in the SJL environment to form the differences $\hat{\mu}^F_{srjl} - \hat{\mu}^F_{sr'jl}$.

Because these differences still contain sampling error, the coefficient in a regression of differences in residuals $Z^{2F}_{c(s,r,j,l)t} - Z^{2F}_{c(s,r',j,l)t}$ from the forecasted partition on the estimated fixed effect differences $(\hat{\mu}^F_{srjl} - \hat{\mu}^F_{sr'jl})$ will be attenuated toward zero even when these fixed effect differences are unbiased estimates of true differences in task-specific talent. Still following Chetty et al. (2014), we therefore shrink the estimated fixed effect differences by a pair-specific reliability ratio to form empirical-Bayes difference estimates: $Diff^{EB}_{srjl} = (\frac{\hat{Var}(\mu_{srcl})}{\hat{Var}(\mu_{srcl}) + (\hat{\sigma}^F_{srjl})^2})(\hat{\mu}^F_{srjl} - \hat{\mu}^F_{sr'jl})$, where $(\hat{\sigma}^F_{srjl})^2$ is the squared standard error of the estimated fixed effect difference (obtained from the component fixed effect estimates via the delta method), and $\hat{Var}(\mu_{srjl})$ is the estimated true variance in teacher talent contributions across classrooms (.154) from our lower bound decomposition presented in Section 5.1.

We then regress the vector of differences in mean classroom residuals in the forecasted sample (Partition 2) on the shrunken fixed effect differences $Diff^{EB}_{srjl}$:

$$Z^{2F}_{c(s,r,j,l)t} - Z^{2F}_{c(s,r',j,l)t} = \beta^F Diff^{EB}_{srjl} + e^{2F}_{ct} \tag{F.2}$$

If the estimates of both the true variance in teacher talent contributions across classrooms (the "signal") and the standard errors of the fixed effect differences (the "noise") are correct, the coefficient β^F should converge in probability to 1, so that the talent estimates are "forecast unbiased" (Chetty et al. (2014)).

While this test captures the ability of our specification to consistently estimate the combined general and context-specific talent of a given teacher teaching in a given context, the ability to choose teachers' classroom

[45]If only one class is available for given SRJL combination for the partition within which production function is estimated, the estimated value $\hat{\mu}_{srjl}$ will be chosen to perfectly fit the classroom mean test score residual, and there will be no regression error with which to form heteroskedasticity-robust standard errors. While other approaches to estimating standard errors (estimating at the test score level, imposing homoskedastic standard errors) would not require this second classroom in the forecasting partition, we want the method for constructing standard errors in the model used for the forecast test to mimic as closely as possible the one employed for our main estimates.

assignments in a way that maximizes achievement contributions depends critically on the ability to isolate and consistently estimate only the context-specific components of teacher fixed contributions to achievement. Thus, we also construct two additional forecast tests that capture the degree to which our estimates of subject-specific and level-specific talent can forecast out-of-sample subject-specific and level-specific comparative advantages, respectively.

Unfortunately, unlike our tests of the consistency of our combined talent estimates, which could be performed using differences among teachers who taught in the same SJL context, testing the consistency of comparative advantage estimates requires evaluating the degree to which difference-in-differences between teachers who taught the same two courses at the same school can be forecast. Thus, a given SRJL combination only enters the subject (level) forecast sample if (1) the teacher taught at least three classrooms in both the chosen SJL *and* a second course that shares the same school-level (school-subject) environment, and (2) there exists a second teacher who also taught at least three classrooms in the same two school-subject-level contexts. These criteria are far more stringent. Much of the variation that identified the estimated true variances in subject-specific and level-specific talent came from difference-in-differences in which at least one of the teachers taught fewer than three subjects in at least one of the school-subject-level contexts. Indeed, applying these criteria leaves us with 205 and 289 difference-in-differences on which to perform the forecast test for subject-specific and level-specific talent estimates, respectively.

The methodology for the context-specific forecast tests is otherwise perfectly analogous to the forecast test for combined teacher talent. Difference-in-differences in residual mean test scores from among the left-out classrooms in Partition 2 across teachers and either subjects or levels (conditioning on the same school-level or school-course environment as appropriate) are regressed on empirical Bayes estimates of difference-in-differences in teacher context-specific talent from the forecasting sample:

$$\begin{aligned}(Z^{2F}_{c(s,r,j,l)t} - Z^{2F}_{c(s,r,j',l)t}) - (Z^{2F}_{c(s,r',j,l)t} - Z^{2F}_{c(s,r',j',l)t}) &= \beta^F Diff\_in\_Diff^{EB}_{srjl} + e^{F,j}_{ct} \\ (Z^{2F}_{c(s,r,j,l)t} - Z^{2F}_{c(s,r,j,l')t}) - (Z^{2F}_{c(s,r',j,l)t} - Z^{2F}_{c(s,r',j,l')t}) &= \beta^F Diff\_in\_Diff^{EB}_{srjl} + e^{F,l}_{ct}\end{aligned} \tag{F.3}$$

The results of the test for forecast bias the combined talent estimates as well as the corresponding tests for the subject-specific and level-specific talent estimates are presented in Table H.6.

Appendix G. Formulation of the Counterfactual Simulation

To formulate the static problem, first let $\mathcal{J}$ represent the set of subjects offered within a given school-field combination. Similarly, let $\mathcal{L}$ represent the set of levels, and let $\mathcal{JL}$ represent the set of subject-level combinations. Let C_{jl} represent 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, exp^{gen}_r captures the number of prior years in which teacher r has taught any classroom, and exp^j_r, exp^l_r, and exp^{jl}_r capture the number of prior years in which teacher r has taught at least one classroom in subject j, level l, and subject-level combination jl, respectively. Student and classroom contributions $X_{ct}\beta_{jl}$ can be ignored, since they are assumed to be constant across counterfactual reallocations (and are assumed to be additively separable from teacher inputs).

Using the estimated smoothed non-parametric experience production function from the "full" specification (13) introduced in Section 6.4, we can predict the contribution of context-specific experience to the counterfactual performance of teacher r in classroom c in a given year t via:

$$\hat{\overline{Y}}^{c}_{rt} = \hat{d}(exp^{gen}_{rt}, exp^{j(c)}_{rt}, exp^{l(c)}_{rt}, exp^{jl(c)}_{rt}) \tag{G.1}$$

For simulations in which we incorporate principal beliefs about teachers' task-specific talent, teacher r's predicted contribution to test scores in classroom c in school s in year t becomes:

$$\hat{\overline{Y}}^{c}_{rt} = \mu^{EB}_{srj(c)l(c)} + \hat{d}(exp^{gen}_{rt}, exp^{j(c)}_{rt}, exp^{l(c)}_{rt}, exp^{jl(c)}_{rt}) \tag{G.2}$$

where $\mu^{EB}_{srj(c)l(c)}$ is an empirical Bayes estimated posterior belief about teacher r's context-specific talent for increasing test scores in school s in subject-level combination $(j(c), l(c))$. [46]

The goal is to choose the mapping $f : \mathcal{C} \rightarrow \mathcal{R}$ from classrooms to teachers that maximizes the sum of student test scores, subject to the constraints that each teacher can only teach as many classrooms as they were observed teaching at time t (denoted $\overline{C}_r$), and every classroom must be taught by exactly one teacher[47]:

$$\begin{aligned} &\max_{f:\mathcal{C}\rightarrow\mathcal{R}} \sum_{c\in\mathcal{C}} \hat{\overline{Y}}^{c}_{f(c)} \\ &s.t. \sum_{r} \mathbb{1}(f(c) = r) = 1 \; \forall \, c \\ &s.t. \sum_{c} \mathbb{1}(f(c) = r) = \overline{C}_r \; \forall \, r \end{aligned} \tag{G.4}$$

where $\mathbb{1}(f(c) = r)$ indicates that teacher r is assigned to classroom c.

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

$$\begin{aligned} &\max_{\mathbf{x}} \mathbf{a} * \mathbf{x} \\ &s.t. \; M_c * \mathbf{x} = 1 \; \forall \, c \\ &s.t. \; N_r * \mathbf{x} = \overline{C}_r \; \forall \, r \\ &s.t. \; \mathbf{x} \in \{0, 1\} \end{aligned} \tag{G.5}$$

[46] $\mu^{EB}_{srj(c)l(c)}$ is calculated by shrinking the fixed effect estimate $\hat{\mu}_{srj(c)l(c)}$ toward zero (the global mean contribution) by multiplying it by the reliability ratio:

$$\mu^{EB}_{srjl} = (\hat{\mu}_{srjl})\left(\frac{Var(\mu_{srjl} - \overline{\mu}_{sr})}{\hat{\sigma}^2_{\hat{\mu}_{srjl}} + Var(\mu_{srjl} - \overline{\mu}_{sr})}\right). \tag{G.3}$$

$Var(\mu_{srjl} - \overline{\mu}_{sr})$ is the estimated true variance in subject-level deviations from school-teacher general talent taken from Row 3 of Column 1 of Table 4. $\hat{\sigma}^2_{\hat{\mu}_{srjl}}$ is the estimated squared standard error from the fixed effect estimate $\hat{\mu}_{srjl}$, which captures the contribution of noise (sampling error) to the variance in the school-teacher-subject-level fixed effect estimates $\{\mu_{srjl}\}$. We set $\mu^{EB}_{srjl} = 0$ for school-teacher-subject-level combinations that are feasible in the simulated assignment but are never observed in our sample.

[47] We suppress dependence on the year (t) in what follows.

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

$$\mathbf{a} = \begin{pmatrix} \hat{\bar{Y}}_1^1 & \dots & \hat{\bar{Y}}_1^C & \hat{\bar{Y}}_2^1 & \dots & \hat{\bar{Y}}_2^C & \dots & \hat{\bar{Y}}_R^1 & \dots & \hat{\bar{Y}}_R^C \end{pmatrix}$$

$\mathbf{x}$ consists of a $(C * R) \times 1$ vector of potential teacher assignments:

$$\mathbf{x} = \begin{pmatrix} x_1^1 \\ \vdots \\ x_1^C \\ x_2^1 \\ \vdots \\ x_2^C \\ \vdots \\ x_R^1 \\ \vdots \\ x_R^C \end{pmatrix}$$

where $x_r^c = \mathbb{1}(f(c) = r)$ is an indicator for whether teacher r is assigned to classroom c.

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

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

N_r consists of a $1 \times C * R$ 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):

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

We solve this binary integer programming problem for each school-field combination in the first year of the sample. 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.[48] 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

[48]Since non-tested subjects are not reallocated, any general or level-specific experience teachers accumulated in those subjects under the true allocation is also included in the update.

context-specific experience. The "static" version of the simulation does not update each teacher's context-specific experience profile for the next year after allocating teachers in a given year, but instead treats every year in the sample as if it were the first year.

Some of our simulations exploit the full sample of teachers, rather than restricting attention to those teachers with fully observed teaching histories. For teachers who begin teaching after 1995, when our sample begins, we impute their teaching history as of 1995 by randomly assigning them the teaching history of a full history teacher who 1) was observed (later in the sample) at the 1995 general experience level of the imputed teacher, and 2) who shares the same most commonly taught subject-level across all the years of our sample as the imputed teacher. Some teachers are sufficiently experienced in 1995 so that there is no teacher with a fully observed teaching history who is ever observed at such a high level of general experience in our sample. These teachers are randomly assigned a 1995 teaching history from among the full history teachers who are observed at 12+ years of general experience who share the same most commonly taught subject-level. Once a 1995 teaching history has been imputed for all teachers with missing histories, we accumulate their post-1995 stocks of general and context-specific experience as it existed in the data (if constructing actual stocks) or as it was optimally assigned (if constructing simulated stocks).

Finally, some of our simulation results tables use a random allocation of teachers to classrooms as a baseline rather than the actual allocation of teachers. To enure that the random allocations are feasible, we construct them by selecting, for each school-field, a random permutation of the allocation identified as the solution to the binary assignment problem.

Appendix H. Appendix Tables

Table H.1: The Distribution of Years of Experience among Classes Taught by 2nd and 3rd Year Teachers

Panel A: Second-Year Teachers

General	Subject	Level	Subj.-Lvl	%
1	1	1	1	70.8%
1	1	1	0	2.7%
1	1	0	0	4.5%
1	0	1	0	19.4%
1	0	0	0	2.5%

Panel B: Third-Year Teachers

General	Subject	Level	Subj.-Lvl	%
2	2	2	2	54.9%
2	2	2	1	3.0%
2	2	2	0	0.6%
2	2	1	1	4.1%
2	2	1	0	0.6%
2	2	0	0	2.2%
2	1	2	1	17.7%
2	1	2	0	1.0%
2	1	1	1	1.7%
2	1	1	0	0.8%
2	1	0	0	1.1%
2	0	2	0	10.5%
2	0	1	0	0.8%
2	0	0	0	1.1%

Notes: The table presents the classroom-weighted distribution of four-dimensional experience stocks among 2nd and 3rd year teachers in our final sample. 10,270 and 8,665 total classes were taught by 2nd-year and third-year teachers respectively. Note that multiple subject-level combinations can be taught in a year.

Table H.2: Coefficient Estimates Associated with Control Variables Capturing Teacher Workload, Depreciation of Experience Capital, and Productivity Declines in the Last Year Teaching Any Class or Teaching a Class in the Chosen Subject, Level, or Subject-Level Combination (Baseline Specification)

	(1)
# of Concurrent Classes Taught	0.000
	[0.000]
# of Concurrent Subject-Level Combinations Taught	0.000
	[0.002]
1(Did Not Teach Last Year)	0.004
	[0.020]
1(Did Not Teach Subject Last Year)	-0.003
	[0.013]
1(Did Not Teach Level Last Year)	0.006
	[0.015]
1(Did Not Teach Subject-Level Last Year)	-0.001
	[0.012]
1(Did Not Teach in Last 2 Years)	-0.006
	[0.036]
1(Did Not Teach Subject in Last 2 Years)	-0.014
	[0.026]
1(Did Not Teach Level in Last 2 Years)	0.003
	[0.028]
1(Did Not Teach Subject-Level in Last 2 Years)	0.002
	[0.022]
1(Final Year Teaching)	-0.005
	[0.019]
1(Final Year Teaching Subject)	0.001
	[0.012]
1(Final Year Teaching Level)	-0.004
	[0.018]
1(Final Year Teaching Subject-Level)	-0.010
	[0.011]

Notes: Regression also contains a full set of school-subject-level and school-teacher-subject-level fixed effects, calendar year effects, a set of observable classroom covariates, and a set of four additively separable flexibly parameterized profiles capturing productivity gains from years of general, subject-specific, level-specific, and subject-level-specific experience. See Table 5 for estimates of these experience profiles. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

Table H.3: Coefficient Estimates Associated with Control Variables Capturing Teacher Workload, Depreciation of Experience Capital, and Productivity Declines in the Last Year Teaching Any Class or Teaching a Class in the Chosen Subject, Level, or Subject-Level Combination (Restricted Specification)

	(1)
# of Concurrent Classes Taught	-0.000
	[0.000]
# of Concurrent Subject-Level Combinations Taught	-0.004*
	[0.002]
1(Did Not Teach Last Year)	0.014
	[0.017]
1(Did Not Teach Subject Last Year)	-0.020*
	[0.011]
1(Did Not Teach Level Last Year)	0.004
	[0.012]
1(Did Not Teach Subject-Level Last Year)	0.005
	[0.010]
1(Did Not Teach in Last 2 Years)	0.005
	[0.031]
1(Did Not Teach Subject in Last 2 Years)	-0.014
	[0.019]
1(Did Not Teach Level in Last 2 Years)	-0.001
	[0.022]
1(Did Not Teach Subject-Level in Last 2 Years)	0.000
	[0.017]
1(Final Year Teaching)	-0.004
	[0.014]
1(Final Year Teaching Subject)	-0.029***
	[0.008]
1(Final Year Teaching Level)	0.008
	[0.012]
1(Final Year Teaching Subject-Level)	-0.017**
	[0.007]

Notes: Regression also contains a full set of school-subject-level and school-teacher fixed effects, calendar year effects, a set of observable classroom covariates, and a set of four additively separable flexibly parameterized profiles capturing productivity gains from years of general, subject-specific, level-specific, and subject-level-specific experience. See Table 6 for estimates of these experience profiles. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

Table H.4: Tests for Dynamic Course Assignment Responses to Unobserved Time-Varying Endogenous Inputs

Year Relative to Change	**Permanent Teacher Changes**				**Temporary Teacher Changes**			
	General	Subject	Level	SL	General	Subject	Level	SL
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$t-1$	0.010	0.005	0.039**	-0.016	-0.000	-0.001	0.000	0.001
	[0.008]	[0.007]	[0.022]	[0.016]	[0.001]	[0.001]	[0.001]	[0.001]
$t-2$	-0.012	-0.016*	-0.021	0.022	0.001	0.003*	0.001	0.002
	[0.012]	[0.011]	[0.017]	[0.019]	[0.002]	[0.002]	[0.002]	[0.002]
$t-3$	-0.003	0.032**	0.037**	0.035*	0.005**	0.003	-0.001	-0.001
	[0.012]	[0.016]	[0.016]	[0.024]	[0.003]	[0.003]	[0.003]	[0.003]
$t-4$	0.017*	-0.010	0.014	-0.009	0.002	0.001	0.006*	0.004
	[0.013]	[0.014]	[0.038]	[0.017]	[0.004]	[0.003]	[0.004]	[0.004]
$t-5$	-0.003	-0.014	0.017	0.024	-0.006*	-0.002	0.002	0.004
	[0.029]	[0.013]	[0.049]	[0.037]	[0.005]	[0.004]	[0.004]	[0.005]
$t-6$	-0.021	0.002	-0.083***	0.032	0.008*	0.005	-0.001	0.005
	[0.018]	[0.019]	[0.031]	[0.067]	[0.006]	[0.005]	[0.006]	[0.006]
$t-7$	-0.040	-0.012	-0.182	-0.115**	0.005	-0.001	0.015***	-0.006
	[0.038]	[0.022]	[0.160]	[0.055]	[0.008]	[0.005]	[0.006]	[0.007]

Notes: Table entries display average school-teacher-year residuals (Columns 1 and 5), school-teacher-subject-year residuals (Columns 2 and 6), school-teacher-level-year residuals (Columns 3 and 7), and school-teacher-subject-level-year residuals (Columns 4 and 8), respectively, in the years leading up to a change in classroom assignment (using residuals from the *Restricted Specification* in equation (3). A permanent change in general course assignment (Column 1) is defined as a teacher-year combination in which the teacher is not observed teaching any course in a subsequent sample year. A permanent change in subject assignment (Column 2) is defined as a teacher-subject-year combination in which the teacher teaches the chosen subject, but is not observed teaching the chosen subject again in subsequent sample years. Permanent changes in level (Column 3) and subject-level (Column 4) assignments are defined analogously to permanent subject changes. Temporary changes in assignment (Columns 5-8) are defined in a similar manner as permanent course assignment changes, except the teacher is observed returning to teach (Column 5) or teach in the chosen subject (Column 6), level (Column 7), or subject-level (Column 8) in a subsequent sample year. Bootstrap standard errors (in brackets) are computed using 1,000 iterations. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively.

Table H.5: Backcasting Test for Non-Random Student Sorting (Restricted Specification with Classroom Average 7th Grade Math Scores as the Outcome Variable)

Years Experience	General (1)	Subject (2)	Level (3)	Subj.-Level (4)	Combined (5)
1 yr	-0.027***	-0.005	0.024***	0.002	-0.005*
	[0.009]	[0.007]	[0.009]	[0.006]	[0.004]
2 yrs	-0.014*	-0.022***	0.016**	0.016**	-0.004
	[0.010]	[0.009]	[0.010]	[0.008]	[0.004]
3 yrs	-0.026**	-0.021**	0.032***	0.024***	0.009**
	[0.013]	[0.011]	[0.011]	[0.010]	[0.005]
4 yrs	-0.019*	-0.030***	0.022**	0.025***	-0.002
	[0.013]	[0.011]	[0.012]	[0.010]	[0.005]
5-6 yrs	-0.016	-0.020*	0.020*	0.026**	0.010**
	[0.014]	[0.013]	[0.013]	[0.012]	[0.005]
7-10 yrs	0.008	-0.028**	-0.009	0.039***	0.011*
	[0.015]	[0.015]	[0.014]	[0.014]	[0.006]
11-14 yrs	0.007	0.009	-0.011	0.012	0.017*
	[0.020]	[0.025]	[0.020]	[0.026]	[0.013]

Notes: $N = 61,993$ test-score-weighted classroom observations. Results are based on an altered version of the *Restricted Specification* in equation (3) in which the actual classroom average of students current test scores from the chosen class are replaced by the classroom average of the 7th grade math scores of these students. Refer to notes below Table 5 for a full description of the control variables. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every year of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

Table H.6: Testing for Forecast Bias in Estimates of Time-Invariant Task-Specific Teacher Talent

	Outcome (Forecasted Sample)		
	Diff (Sch-Tea-Subj-Lvl)	Diff-in-Diff (Subj)	Diff-in-Diff (Lvl)
Forecasting Sample	(1)	(2)	(3)
Diff^{EB} (Sch-Tea-Subj-Lvl) $[\hat{\mu}_{srjl} - \hat{\mu}_{sr'jl}]^{EB}$	0.825 (0.019)		
Diff-in-Diff (Subj) $[(\hat{\mu}_{srjl} - \hat{\mu}_{srj'l}) - (\hat{\mu}_{sr'jl} - \hat{\mu}_{sr'j'l})]^{EB}$		1.013 (0.242)	
Diff-in-Diff (Lvl) $[(\hat{\mu}_{srjl} - \hat{\mu}_{srjl'}) - (\hat{\mu}_{sr'jl} - \hat{\mu}_{sr'jl'})]^{EB}$			0.456 (0.333)
Observations	7,246	205	289

Notes: The entries in this table are coefficients (with standard errors in brackets) capturing the degree of forecast bias in estimates of combined (general and task-specific) talent, subject-specific talent, and level-specific talent, respectively, from a set of split sample tests. See Section 6.3 for a detailed description of the methodology. In each specification, the estimator should yield a forecast coefficient that converges in probability to 1 if our achievement production function is correctly specified. Specifically, the outcome in Column 1 is the difference in average test score residuals among a pair of classes from the same school-subject-level taught by two different teachers from a partition of our main sample. The entry in Row (1), Column (1) captures the coefficient on a vector of empirical Bayes forecasts of the expected difference in achievement among these pairs of teachers based on (appropriately shrunken) estimates of the difference in their combined general and course-specific productivity from a second, mutually exclusive partition used to construct the forecast. The entries in Column 2 and Column 3 replace these pair-specific differences on both sides of the equation with differences-in-differences among pairs of teachers across common pairs of courses that differ only in subject (Column 2) or level (Column 3). These coefficients capture the degree of forecast bias in the model's ability to estimate subject-specific and level-specific comparative advantages. Heteroskedasticity-robust (White) standard errors (in brackets) are computed for each coefficient.

Table H.7: Effect of Number of Courses of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification)

Course Experience	General	Subject	Level	Subj.-Level	Combined
	(1)	(2)	(3)	(4)	(5)
1 crs	0.019	0.011	-0.006	0.008	0.031
	[0.036]	[0.012]	[0.012]	[0.009]	[0.035]
2 crs	0.051**	0.007	0.002	0.002	0.062***
	[0.027]	[0.011]	[0.011]	[0.009]	[0.025]
3 crs	0.075***	0.021**	-0.005	0.002	0.094***
	[0.017]	[0.011]	[0.012]	[0.010]	[0.015]
4-5 crs	0.066***	0.032***	0.001	0.004	0.104***
	[0.013]	[0.011]	[0.011]	[0.010]	[0.008]
6-9 crs	0.054***	0.049***	0.009	0.002	0.114***
	[0.012]	[0.012]	[0.012]	[0.011]	[0.006]
10-20 crs	0.070***	0.062***	0.005	-0.006	0.131***
	[0.014]	[0.013]	[0.013]	[0.013]	[0.007]
21+ crs	0.082***	0.071***	0.003	-0.009	0.146***
	[0.015]	[0.015]	[0.015]	[0.016]	[0.010]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. *Restricted Specification* refers to equation (3). Refer to notes below Table 5 for a full description of the control variables. Experience is measured as the number of classes taught in prior years by the classroom's teacher in total (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every classroom of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

Table H.8: True Variances in Fixed Effects (Using Course-Based Measure of Teacher Experience with the Baseline Specification)

	Lower Bound		**Intermediate**		**Upper Bound**	
	Var.	SD	Var.	SD	Var.	SD
	(1)	(2)	(3)	(4)	(5)	(6)
Sch-Subj-Lvl-Tch Combos	0.0237	0.154	0.0467	0.216	0.0604	0.246
General Talent	0.0176	0.133	0.0368	0.192	0.0505	0.225
Subj-Lvl Combos	0.0061	0.078	0.0099	0.099	0.0099	0.099
Sch-Lvl-Tch Combos	0.0198	0.141	0.0407	0.202	0.0544	0.233
Subject Talent	0.0039	0.063	0.0060	0.077	0.0060	0.077
Sch-Subj-Tch Combos	0.0216	0.147	0.0433	0.208	0.0569	0.239
Level Talent	0.0021	0.045	0.0034	0.059	0.0034	0.059
Subject-Level Talent	0.0001	0.011	0.0005	0.022	0.0005	0.022

Notes: This variance decomposition is based on a version of the baseline specification (equation 2 in which experience in each context dimension (general, subject-specific, level-specific, and subject-level-specific) is measured as the total number of previous classrooms taught in the relevant context. *Lower Bound* estimates allocate all of the between school-subject-level variance in residual test scores to school and student inputs (Assumption 2A). This is implemented by including school-subject-level fixed effects and normalizing the mean among school-teacher-subject-level fixed effects to be 0 in each school-subject-level. *Intermediate* estimates allocate the between school variance in residual test scores to school and student inputs, and the within-school/between subject-level variance to teachers (Assumption 2B). This is implemented by replacing the school-subject-level fixed effects with school fixed effects only. *Upper Bound* estimates allocate all of the between school-subject-level variance in residual test scores to teachers (Assumption 2C). This is implemented by removing all school-level controls. See Section 3.2 for details. See Section 3.2 for details.

Table H.9: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification with the Sample Restricted to Classrooms Featuring Teachers in their First School)

Years Experience	General (1)	Subject (2)	Level (3)	Subj.-Level (4)	Combined (5)
1 yr	0.062***	0.018**	-0.003	0.011	0.088***
	[0.012]	[0.011]	[0.011]	[0.009]	[0.005]
2 yrs	0.087***	0.022**	-0.011	0.018*	0.117***
	[0.016]	[0.014]	[0.014]	[0.012]	[0.006]
3 yrs	0.093***	0.035**	-0.016	0.014	0.126***
	[0.018]	[0.016]	[0.016]	[0.014]	[0.007]
4 yrs	0.105***	0.035**	-0.022	0.012	0.130***
	[0.020]	[0.017]	[0.018]	[0.015]	[0.008]
5-6 yrs	0.109***	0.034**	-0.019	0.023*	0.147***
	[0.021]	[0.020]	[0.019]	[0.017]	[0.010]
7-10 yrs	0.116***	0.020	-0.031*	0.027	0.131***
	[0.025]	[0.025]	[0.023]	[0.023]	[0.013]
11-14 yrs	0.101***	0.019	0.016	-0.014	0.122***
	[0.034]	[0.044]	[0.035]	[0.049]	[0.033]

Notes: $N = 51,773$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. *Restricted Specification* refers to equation (3). The sample is restricted to classrooms featuring a teacher that is teaching in his/her first school (i.e. the teacher's entire teaching history was acquired at the current school). Refer to notes below Table 5 for a full description of the control variables. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every year of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

Table H.10: True Variances in Fixed Effects (Using the Baseline Specification with the Year-Based Measure of Teacher Experience and a Sample Restricted to Classrooms Featuring Teachers in Their First Schools)

	Lower Bound		**Intermediate**		**Upper Bound**	
	Var.	SD	Var.	SD	Var.	SD
	(1)	(2)	(3)	(4)	(5)	(6)
Sch-Subj-Lvl-Tch Combos	0.0225	0.150	0.0454	0.213	0.0590	0.243
General Talent	0.0167	0.129	0.0356	0.189	0.0491	0.222
Subj-Lvl Combos	0.0058	0.076	0.0098	0.099	0.0098	0.099
Sch-Lvl-Tch Combos	0.0188	0.137	0.0395	0.199	0.0530	0.230
Subject Talent	0.0037	0.061	0.0060	0.077	0.0060	0.077
Sch-Subj-Tch Combos	0.0205	0.143	0.0421	0.205	0.0556	0.236
Level Talent	0.0019	0.044	0.0033	0.058	0.0033	0.058
Subject-Level Talent	0.0001	0.011	0.0005	0.023	0.0005	0.023

Notes: Lower Bound estimates allocate all of the between school-subject-level variance in residual test scores to school and student inputs (Assumption 2A). This is implemented by including school-subject-level fixed effects and normalizing the mean among school-teacher-subject-level fixed effects to be 0 in each school-subject-level. *Intermediate* estimates allocate the between school variance in residual test scores to school and student inputs, and the within-school/between subject-level variance to teachers (Assumption 2B). This is implemented by replacing the school-subject-level fixed effects with school fixed effects only. *Upper Bound* estimates allocate all of the between school-subject-level variance in residual test scores to teachers (Assumption 2C). This is implemented by removing all school-level controls. See Section 3.2 for details.

Table H.11: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification with Linear Depreciation)

Years Experience	General	Subject	Level	Subj.-Level	Combined
	(1)	(2)	(3)	(4)	(5)
1 yr	0.066***	0.013*	-0.004	0.015**	0.089***
	[0.011]	[0.009]	[0.010]	[0.008]	[0.004]
2 yrs	0.086***	0.023**	-0.005	0.014*	0.118***
	[0.014]	[0.012]	[0.012]	[0.010]	[0.006]
3 yrs	0.095***	0.036***	-0.006	0.006	0.131***
	[0.016]	[0.014]	[0.014]	[0.012]	[0.007]
4 yrs	0.103***	0.040***	-0.009	0.004	0.137***
	[0.018]	[0.015]	[0.016]	[0.014]	[0.008]
5-6 yrs	0.106***	0.040**	-0.000	0.005	0.151***
	[0.019]	[0.018]	[0.017]	[0.015]	[0.009]
7-10 yrs	0.117***	0.025	-0.004	-0.001	0.136***
	[0.022]	[0.021]	[0.020]	[0.019]	[0.012]
11-14 yrs	0.111***	0.027	0.031	-0.029	0.140***
	[0.028]	[0.038]	[0.028]	[0.041]	[0.026]

Notes: Regression specification mimics the *Restricted Specification* (see Equation (3) and the notes to Table 6), except that the indicator sets for whether the teacher failed to teach a course in each of the relevant dimensions of context (general, subject, level, and subject-level) last year or in the last two years are replaced by linear controls for the number of years since the teacher taught any course and since the teacher taught in the subject, level, and subject-level associated with the classroom observation. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every classroom of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

Table H.12: Coefficient Estimates Associated with Control Variables Capturing Teacher Workload, Depreciation of Experience Capital, and Productivity Declines in the Last Year Teaching Any Class or Teaching a Class in the Chosen Subject, Level, or Subject-Level Combination (Restricted Specification with Linear Depreciation)

	(1)
# of Concurrent Classes Taught	-0.000
	[0.000]
# of Concurrent Subject-Level Combinations Taught	-0.003
	[0.002]
# of Years Since Last Taught	0.004
	[0.006]
# of Years Since Last Taught Subject	-0.005
	[0.004]
# of Years Since Last Taught Level	0.001
	[0.004]
# of Years Since Last Taught Subject-Level	-0.005
	[0.004]
1(Final Year Teaching)	-0.005
	[0.014]
1(Final Year Teaching Subject)	-0.028***
	[0.008]
1(Final Year Teaching Level)	0.008
	[0.012]
1(Final Year Teaching Subject-Level)	-0.018**
	[0.007]

Notes: Regression specification mimics the ***Restricted Specification*** (see Equation (3) and the notes to Table 6), except that the indicator sets for whether the teacher failed to teach a course in each of the relevant dimensions of context (general, subject, level, and subject-level) last year or in the last two years are replaced by linear controls for the number of years since the teacher taught any course and since the teacher taught in the subject, level, and subject-level associated with the classroom observation. Regression also contains a full set of school-subject-level and school-teacher fixed effects, calendar year effects, a set of observable classroom covariates, and a set of four additively separable flexibly parameterized profiles capturing productivity gains from years of general, subject-specific, level-specific, and subject-level-specific experience. See Table H.11 for estimates of these experience profiles. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

Table H.13: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification with 7th Grade Math and Reading Test Scores Added as Controls)

Years Experience	General	Subject	Level	Subj.-Level	Combined
	(1)	(2)	(3)	(4)	(5)
1 yr	0.069***	0.014*	-0.007	0.014**	0.090***
	[0.011]	[0.009]	[0.010]	[0.008]	[0.004]
2 yrs	0.088***	0.025**	-0.007	0.013*	0.119***
	[0.014]	[0.012]	[0.012]	[0.010]	[0.006]
3 yrs	0.099***	0.037***	-0.013	0.007	0.131***
	[0.016]	[0.014]	[0.014]	[0.012]	[0.007]
4 yrs	0.106***	0.042***	-0.016	0.005	0.138***
	[0.018]	[0.015]	[0.015]	[0.014]	[0.008]
5-6 yrs	0.108***	0.042***	-0.006	0.007	0.151***
	[0.019]	[0.017]	[0.017]	[0.015]	[0.009]
7-10 yrs	0.115***	0.026	-0.009	0.002	0.135***
	[0.022]	[0.021]	[0.019]	[0.019]	[0.011]
11-14 yrs	0.108***	0.018	0.025	-0.014	0.137***
	[0.028]	[0.037]	[0.028]	[0.041]	[0.026]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. *Restricted Specification* refers to equation (3). Refer to notes below Table 5 for a full description of the control variables. This regression also includes 7th grade math and reading standardized test scores as additional controls. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every classroom of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

Table H.14: Heterogeneity across Subject Fields in the Effects of Years of General and Subject-Specific Experience on Student Test Scores (Restricted Specification with Level and Subject-Level Experience Additionally Constrained to 0)

Years Exp.	**Math**		**Science**		**Social Studies**		**English**	
	General	Subject	General	Subject	General	Subject	General	Subject
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 yr	0.072***	0.021**	0.070***	0.046***	0.072***	0.017	0.020	0.013
	[0.012]	[0.010]	[0.013]	[0.011]	[0.014]	[0.012]	[0.016]	[0.015]
2 yrs	0.090***	0.038***	0.098***	0.047***	0.091***	0.027*	0.017	0.035*
	[0.014]	[0.012]	[0.017]	[0.015]	[0.017]	[0.015]	[0.022]	[0.021]
3 yrs	0.102***	0.048***	0.101***	0.062***	0.091***	0.032*	0.020	0.036
	[0.016]	[0.014]	[0.019]	[0.018]	[0.018]	[0.017]	[0.026]	[0.026]
4 yrs	0.095***	0.053***	0.089***	0.078***	0.119***	0.026	0.036	0.022
	[0.017]	[0.015]	[0.022]	[0.021]	[0.020]	[0.019]	[0.031]	[0.031]
5-6 yrs	0.120***	0.057***	0.089***	0.075***	0.119***	0.042**	0.023	0.025
	[0.018]	[0.017]	[0.025]	[0.024]	[0.021]	[0.022]	[0.037]	[0.037]
7-10 yrs	0.134***	0.038*	0.062**	0.075**	0.133***	0.005	0.038	0.009
	[0.020]	[0.020]	[0.032]	[0.031]	[0.022]	[0.026]	[0.045]	[0.045]
11-14 yrs	0.154***	0.008	0.073	0.097*	0.125***	-0.081	0.066	0.042
	[0.028]	[0.042]	[0.047]	[0.051]	[0.036]	[0.058]	[0.055]	[0.058]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. Results are based on an altered version of the *Restricted Specification* in equation (3) in which 1) we impose the additional restrictions that gains from level-specific and subject-level-specific experience are constrained to be 0: $d^l(exp) = 0$ and $d^{jl}(exp) = 0 \; \forall \; exp$, and 2) we generalize the gains from years of general and subject-specific experience to be field-specific: $d^{gen}(exp) \rightarrow d^{gen}_{field}(exp)$, $d^j(exp) \rightarrow d^j_{field}(exp)$, $field \in \{\text{Math}, \text{Science}, \text{Social Studies}, \text{English}\}$. Refer to notes below Table 5 for a full description of the control variables. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (odd Columns) or in the subject (even Columns) associated with the current classroom observation. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

Table H.15: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification with Full Set of Indicator Variables for Each Observed Years of Experience)

Years Experience	General	Subject	Level	Subj.-Level	Combined
	(1)	(2)	(3)	(4)	(5)
1 yr	0.064***	0.013*	-0.002	0.013**	0.089***
	[0.011]	[0.009]	[0.010]	[0.008]	[0.004]
2 yrs	0.084***	0.021**	-0.003	0.016*	0.118***
	[0.014]	[0.012]	[0.012]	[0.010]	[0.006]
3 yrs	0.093***	0.033***	-0.005	0.010	0.131***
	[0.016]	[0.014]	[0.014]	[0.012]	[0.007]
4 yrs	0.100***	0.036***	-0.008	0.010	0.137***
	[0.018]	[0.015]	[0.016]	[0.014]	[0.008]
5 yrs	0.101***	0.041***	0.003	0.012	0.156***
	[0.019]	[0.017]	[0.017]	[0.016]	[0.009]
6 yrs	0.101***	0.021	-0.001	0.026*	0.146***
	[0.021]	[0.019]	[0.018]	[0.018]	[0.011]
7 yrs	0.117***	0.018	-0.017	0.019	0.137***
	[0.023]	[0.022]	[0.020]	[0.020]	[0.012]
8 yrs	0.110***	-0.017	0.002	0.028	0.123***
	[0.026]	[0.026]	[0.023]	[0.024]	[0.014]
9 yrs	0.120***	-0.001	0.002	0.027	0.148***
	[0.028]	[0.031]	[0.026]	[0.029]	[0.017]
10 yrs	0.120***	-0.020	0.006	0.021	0.127***
	[0.032]	[0.036]	[0.029]	[0.037]	[0.022]
11 yrs	0.104***	0.006	0.056**	-0.025	0.140***
	[0.035]	[0.047]	[0.032]	[0.050]	[0.029]
12 yrs	0.153***	-0.086*	-0.014	0.147**	0.199***
	[0.044]	[0.062]	[0.042]	[0.065]	[0.039]
13 yrs	0.087*	-0.063	0.010	0.176*	0.210***
	[0.053]	[0.105]	[0.055]	[0.113]	[0.044]
14 yrs	0.127**	-0.156	-0.041	0.231*	0.162
	[0.060]	[0.131]	[0.065]	[0.163]	[0.141]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. Refer to notes below Table 5 for a full description of the control variables. Results are based on an altered version of the *Restricted Specification* from equation (3) in which bins for years of experience 5-6, 7-11, and 11+, respectively, are replaced by indicator variables for each individual year of experience (general, subject, level, and subject-level combination). Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every year of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

Table H.16: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification Featuring Quartics in Each Dimension of Context-Specific Experience)

Years Experience	General	Subject	Level	Subj.-Level
	(1)	(2)	(3)	(4)
Year Exp.	0.063***	0.019*	0.006	0.007
	[0.012]	[0.011]	[0.010]	[0.010]
(Year Exp.)^2	-0.014***	-0.003	-0.004	-0.001
	[0.004]	[0.004]	[0.003]	[0.004]
(Year Exp.)^3	0.001***	0.000	0.001	-0.000
	[0.000]	[0.001]	[0.000]	[0.001]
(Year Exp.)^4	-0.000***	-0.000	-0.000	0.000
	[0.000]	[0.000]	[0.000]	[0.000]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. *Restricted Specification* refers to equation (3). Refer to notes below Table 5 for a full description of the control variables. Experience profiles in this regression are generated by replacing year-of-experience dummy variables from the restricted specification with a quartic in each of the four dimensions of experience (general, subject, level, and subject-level). Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

Table H.17: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification Featuring Quartics in Each Dimension of Context-Specific Experience: Predicted Values for First 10 Years of Experience)

Years Experience	General (1)	Subject (2)	Level (3)	Subj.-Level (4)
Pred. Exp. in Years 1	0.050***	0.016**	0.003	0.007
	[0.009]	[0.008]	[0.008]	[0.007]
Pred. Exp. in Years 2	0.078***	0.027**	0.001	0.011
	[0.013]	[0.011]	[0.012]	[0.010]
Pred. Exp. in Years 3	0.092***	0.032**	-0.002	0.014
	[0.016]	[0.013]	[0.014]	[0.011]
Pred. Exp. in Years 4	0.098***	0.034**	-0.005	0.015
	[0.017]	[0.015]	[0.015]	[0.013]
Pred. Exp. in Years 5	0.099***	0.032*	-0.006	0.014
	[0.018]	[0.016]	[0.016]	[0.014]
Pred. Exp. in Years 6	0.100***	0.027	-0.005	0.013
	[0.020]	[0.018]	[0.017]	[0.016]
Pred. Exp. in Years 7	0.101***	0.019	-0.002	0.013
	[0.022]	[0.021]	[0.019]	[0.019]
Pred. Exp. in Years 8	0.105***	0.009	0.004	0.014
	[0.024]	[0.025]	[0.021]	[0.022]
Pred. Exp. in Years 9	0.110***	-0.004	0.010	0.019
	[0.027]	[0.029]	[0.024]	[0.027]
Pred. Exp. in Years 10	0.116***	-0.018	0.016	0.030
	[0.030]	[0.034]	[0.027]	[0.032]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. *Restricted Specification* refers to equation (3). Refer to notes below Table 5 for a full description of the control variables. Experience profiles in this regression are generated by replacing year-of-experience dummy variables from the restricted specification with a quartic in each of the four dimensions of experience (general, subject, level, and subject-level). Predicted values are used for the first 10 years of experience in each dimension. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Standard errors (in brackets) are calculated by applying the delta method to the cluster-robust standard errors for the experience estimates from Table H.16, which were clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 2 for methodological details.

Table H.18: Average Accumulated Marginal Effects Derived from Non-Parametric Experience Production Function (Full Specification with Year-Based Definition of Experience)

Years Experience	General	Subject	Level	Subj.-Level
	(1)	(2)	(3)	(4)
1 yr	0.021***	0.024**	0.006	0.014**
	[0.007]	[0.013]	[0.012]	[0.007]
2 yrs	0.073**	0.043**	0.006	0.016*
	[0.033]	[0.019]	[0.018]	[0.012]
3 yrs	0.076**	0.080***	0.016	0.016
	[0.044]	[0.026]	[0.023]	[0.017]
4 yrs	0.073**	0.094***	0.030	0.019
	[0.044]	[0.032]	[0.028]	[0.021]
5-6 yrs	0.097**	0.098***	0.051**	0.018
	[0.047]	[0.039]	[0.031]	[0.024]
7-10 yrs	0.114**	0.117***	0.035	0.016
	[0.050]	[0.049]	[0.034]	[0.028]
11-14 yrs	0.128***	0.135***	0.072**	0.007
	[0.053]	[0.056]	[0.041]	[0.037]

Notes: Refer to the notes below Table 5 for a full description of the control variables. Experience profiles are generated by integrating partial derivatives of extra experience in each experience dimension, evaluated at each level of experience, over all the levels of experience. These partial derivatives are derived from a smoothed version of the non-parametrically estimated production function for experience gains described in equation (13). Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Standard errors (in brackets) are calculated by applying the delta method to the cluster-robust standard errors for the experience-cell fixed effects, which were clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Appendix E for methodological details.

Table H.19: Average Accumulated Marginal Effects Derived from Non-Parametric Experience Production Function (Restricted Specification with Year-Based Definition of Experience)

Years Experience	General (1)	Subject (2)	Level (3)	Subj.-Level (4)
1 yr	0.017***	0.029***	0.020***	0.013***
	[0.004]	[0.007]	[0.006]	[0.004]
2 yrs	0.062***	0.035***	0.029***	0.017***
	[0.016]	[0.010]	[0.009]	[0.005]
3 yrs	0.052**	0.058***	0.025**	0.015**
	[0.027]	[0.013]	[0.011]	[0.008]
4 yrs	0.041*	0.067***	0.020**	0.014*
	[0.027]	[0.016]	[0.012]	[0.010]
5-6 yrs	0.096***	0.065***	0.025**	0.015*
	[0.027]	[0.020]	[0.014]	[0.011]
7-10 yrs	0.092***	0.065***	0.011	0.011
	[0.031]	[0.025]	[0.016]	[0.014]
11-14 yrs	0.116***	0.069**	0.059***	0.005
	[0.036]	[0.031]	[0.020]	[0.027]

Notes: Refer to the notes below Table 6 for a full description of the control variables. Experience profiles are generated by integrating partial derivatives of extra experience in each experience dimension, evaluated at each level of experience, over all the levels of experience. These partial derivatives are derived from a smoothed version of the non-parametrically estimated production function for experience gains described in equation (13), but where the school-teacher-subject-level fixed effects μ_{srjl} are restricted to be common across subject-levels within a school-teacher combination: $\mu_{srjl} = \bar{\mu}_{st}\ \forall\ (j,l)$ and (s,r). Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Standard errors (in brackets) are calculated by applying the delta method to the cluster-robust standard errors for the experience-cell fixed effects, which were clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Appendix E for methodological details.

Table H.20: Counterfactual Simulations: Achievement Gains from Optimal Allocation Relative to Actual and Random Allocations Separately by Field (Year-Based Measure of Experience, Excluding Teachers Without Full Histories)

		Math				Science				Social Studies			
Eligible Teach.		**Static**		**Dynamic**		**Static**		**Dynamic**		**Static**		**Dynamic**	
		Actual	Random	Actual	Random	Actual	Random	Actual	Random	Actual	Random	Actual	Random
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
2	**Total**	**.015**	**.018**	**.016**	**.021**	**.017**	**.026**	**.017**	**.028**	**.018**	**.027**	**.019**	**.028**
	Talent	.015	.015	.015	.017	.019	.021	.020	.022	.020	.022	.020	.022
	Exper.	-.000	.003	.000	.004	-.002	.005	-.003	.005	-.002	.004	-.001	.005
3	**Total**	**.029**	**.035**	**.030**	**.041**	**.024**	**.040**	**.026**	**.041**	**.025**	**.034**	**.027**	**.038**
	Talent	.029	.031	.030	.034	.026	.033	.027	.033	.026	.029	.027	.029
	Exper.	-.000	.004	.000	.007	-.002	.006	-.001	.008	-.001	.005	.000	.008
4	**Total**	**.033**	**.043**	**.035**	**.045**	**.030**	**.043**	**.032**	**.044**	**.031**	**.038**	**.033**	**.044**
	Talent	.034	.039	.034	.036	.032	.034	.032	.033	.032	.032	.032	.034
	Exper.	-.001	.004	.001	.009	-.002	.009	-.000	.011	-.001	.006	.001	.010
5-6	**Total**	**.041**	**.049**	**.043**	**.051**	**.034**	**.050**	**.037**	**.052**	**.038**	**.046**	**.040**	**.053**
	Talent	.041	.044	.041	.042	.036	.039	.037	.039	.038	.039	.038	.042
	Exper.	-.000	.005	.002	.009	-.002	.011	.000	.013	-.000	.007	.002	.011
7-10	**Total**	**.039**	**.049**	**.042**	**.054**	**.041**	**.051**	**.044**	**.057**	**.037**	**.046**	**.038**	**.050**
	Talent	.039	.042	.039	.043	.043	.040	.044	.042	.036	.038	.036	.038
	Exper.	.000	.007	.003	.012	-.002	.011	.000	.015	.001	.009	.002	.012
11+	**Total**	**.042**	**.053**	**.045**	**.064**	**—**	**—**	**—**	**—**	**.039**	**.052**	**.038**	**.045**
	Talent	.039	.044	.039	.048	—	—	—	—	.037	.039	.037	.029
	Exper.	.003	.009	.005	.016	—	—	—	—	.003	.013	.001	.017

Notes: Each cell presents simulated achievement gains from the optimal allocation of teachers to classrooms relative to either the observed allocation (in columns labeled "Actual") or a randomly selected feasible allocation (columns labeled "Random") among all school-year-field combinations with the number of eligible teachers specified by the row label in the field specified by the column label. The top entry in each cell displays the total achievement gains, while the middle and bottom entries display the components of the gains attributable to task-specific experience and task-specific talent, respectively. *Static* refers to simulations in which teacher experience stocks are held fixed as they were in the actual sample through year $t-1$ prior to simulated reassignment in year t. *Dynamic* refers to simulations in which teacher experience stocks used as the basis for simulated reassignment in year t are based on simulated assignments from 1995 through year $t-1$. See Section 7.1 and Appendix Section Appendix G for further detail about simulation methodology. A teacher is eligible for reassignment if their full teaching history is observed in the data. Estimates of gains from task-specific experience and of teachers' task-specific talent are derived from the Full Specification (equation (13)). The principal incorporates information from empirical Bayes posterior beliefs about each teacher's task-specific talent based on our school-teacher-subject-level fixed effect estimates for any school-teacher-subject-level combination that is observed in our sample. We assign task-specific productivities of 0 for any school-teacher-subject-level combination that we do not observe.

Table H.21: Counterfactual Simulations: Fraction of Classrooms Reallocated (Year-Based Measure of Experience, Excluding Teachers Without Full Histories)

	Math		**Science**		**Social Studies**	
Eligible Teachers	**Static** (1)	**Dynamic** (2)	**Static** (3)	**Dynamic** (4)	**Static** (5)	**Dynamic** (6)
2	0.253	0.303	0.230	0.329	0.267	0.317
3	0.331	0.391	0.337	0.412	0.370	0.444
4	0.404	0.449	0.410	0.482	0.422	0.479
5-6	0.428	0.469	0.434	0.505	0.463	0.518
7-10	0.427	0.469	0.484	0.574	0.488	0.542
11+	0.456	0.519	0.000	0.000	0.500	0.540

Notes: Each cell presents the fraction of classroom assignments in which a reallocation takes place (i.e. the simulated teacher assignment does not match the actual teacher assignment) among all school-year-field combinations with the number of eligible teachers specified by the row label in the field specified by the column label. *Static* refers to simulations in which teacher experience stocks are held fixed as they were in the actual sample through year $t-1$ prior to simulated reassignment in year t. *Dynamic* refers simulations in which teacher experience stocks used as the basis for simulated reassignment in year t are based on simulated assignments from 1995 through year $t-1$. See Section 7.1 and Appendix Section Appendix G for further detail about simulation methodology. A teacher is eligible for reassignment if their full teaching history is observed in the data.

Table H.22: Counterfactual Simulations: Achievement Gains from Optimal Allocation Relative to Actual and Random Allocations (Year-Based Measure of Experience, Including Teachers Without Full Histories)

Eligible Teach.		**Static**		**Dynamic**	
		Actual (1)	Random (2)	Actual (3)	Random (4)
2	**Total**	**.005**	**.010**	**.005**	**.011**
	Talent	.004	.005	.004	.005
	Exper.	.001	.005	.000	.006
3	**Total**	**.011**	**.018**	**.011**	**.021**
	Talent	.010	.010	.011	.011
	Exper.	.001	.008	.001	.010
4	**Total**	**.014**	**.023**	**.014**	**.027**
	Talent	.012	.014	.012	.014
	Exper.	.001	.010	.001	.013
5-6	**Total**	**.018**	**.030**	**.019**	**.034**
	Talent	.017	.018	.017	.018
	Exper.	.001	.011	.002	.016
7-10	**Total**	**.022**	**.034**	**.023**	**.040**
	Talent	.020	.021	.020	.021
	Exper.	.002	.013	.003	.019
11+	**Total**	**.023**	**.035**	**.025**	**.042**
	Talent	.020	.021	.020	.021
	Exper.	.003	.014	.004	.021

Notes: Each cell presents simulated achievement gains from the optimal allocation of teachers to classrooms relative to either the observed allocation (in columns labeled "Actual") or a randomly selected feasible allocation (columns labeled "Random") among all school-year-field combinations with the number of eligible teachers specified by the row label. Classroom-level gains are pooled across the three fields (math, science, and social studies). The top entry in each cell displays the total achievement gains, while the middle and bottom entries display the contribution to this total of gains from task-specific experience and task-specific talent, respectively. *Static* refers to simulations in which teacher experience stocks are held fixed as they were in the actual sample through year $t-1$ prior to simulated reassignment in year t. *Dynamic* refers to simulations in which teacher experience stocks used as the basis for simulated reassignment in year t are based on simulated assignments from 1995 through year $t-1$. See Section 7.1 and Appendix Section Appendix G for further detail about simulation methodology. Eligible teachers consist of teachers who taught a test subject in the chosen school-year-field. Teachers who begin teaching prior to 1995 for whom full teaching histories were not observed are assigned imputed teaching histories as of 1995. See Section Appendix G for a description of the imputation procedure. Estimates of gains from task-specific experience and of teachers' task-specific talent are derived from the Full Specification (presented in equation (13)). The principal incorporates information from empirical Bayes posterior beliefs about each teacher's task-specific talent based on our school-teacher-subject-level fixed effect estimates for any school-teacher-subject-level combination that is observed in our sample. We assign task-specific productivities of 0 for any school-teacher-subject-level combination that we do not observe.

Table H.23: Identification Example: Experience Stocks for Hypothetical Teachers in Each Year

Panel A: Identifying Variation for Experience Profiles, Example 1

	Teacher 1: New Subj/Lvl					Teacher 2: New Subj Only				
Year	Crs.	Gen.	Subj.	Lvl.	Subj.-Lvl.	Crs.	Gen.	Subj.	Lvl.	Subj.-Lvl.
1	BP	0	0	0	0	HP	0	0	0	0
2	HC	1	0	0	0	HC	1	0	1	0
3	HC	2	1	1	1	HC	2	1	2	1
4	HC	3	2	2	2	HC	3	2	3	2
5	HC	4	3	3	3	HC	4	3	4	3

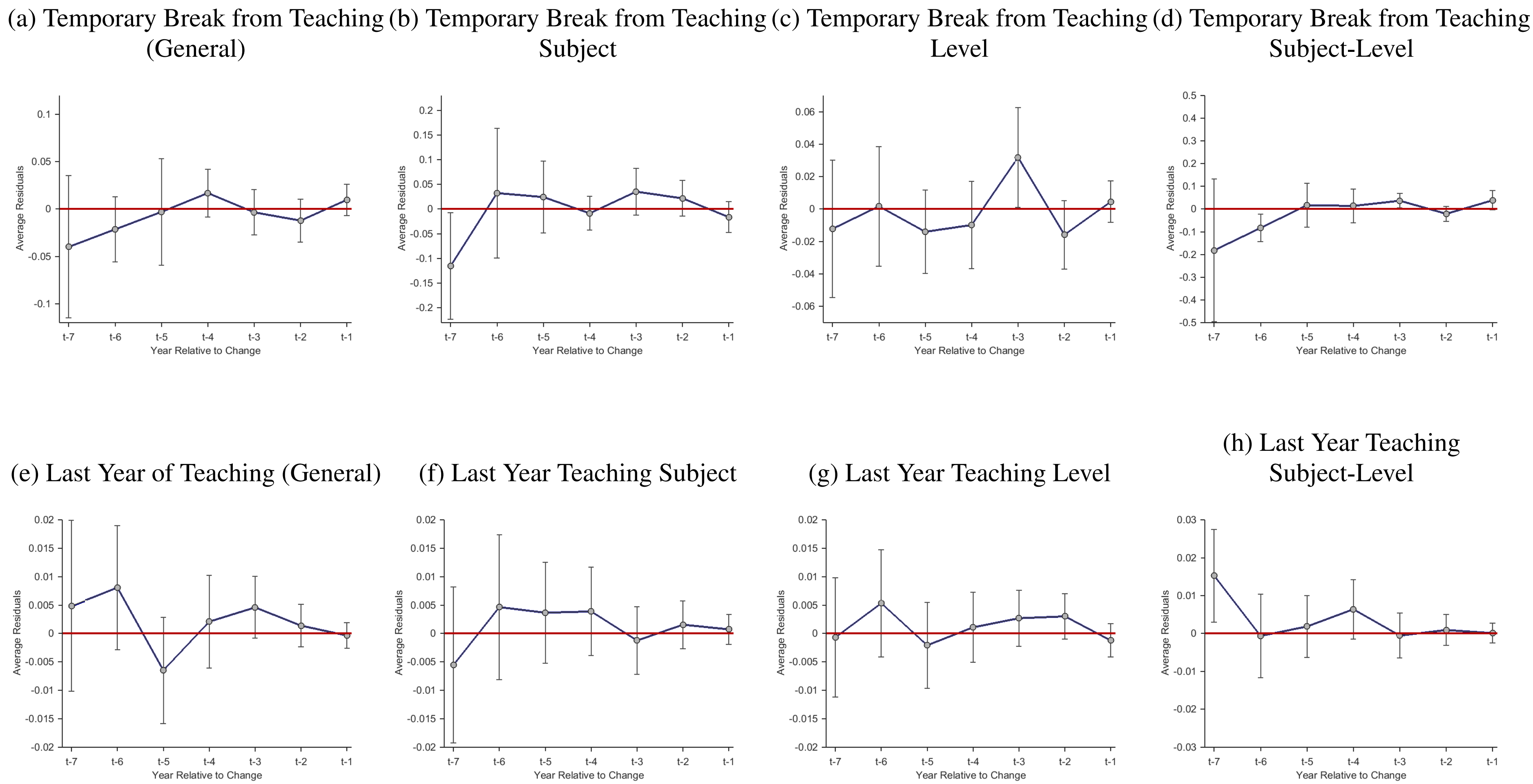
	Teacher 3: New Lvl Only					Teacher 4: Same Subj/Lvl				
Year	Crs.	Gen.	Subj.	Lvl.	Subj.-Lvl.	Crs.	Gen.	Subj.	Lvl.	Subj.-Lvl.
1	BC	0	0	0	0	HC	0	0	0	0
2	HC	1	1	0	0	HC	1	1	1	1
3	HC	2	2	1	1	HC	2	2	2	2
4	HC	3	3	2	2	HC	3	3	3	3
5	HC	4	4	3	3	HC	4	4	4	4

Panel B: Identifying Variation for Experience Profiles, Example 2

	Teacher 1					Teacher 2				
Year	Crs.	Gen.	Subj.	Lvl.	Subj.-Lvl.	Crs.	Gen.	Subj.	Lvl.	Subj.-Lvl.
1	BC	0	0	0	0	BP	0	0	0	0
2	HC	1	1	0	0	HC	1	0	0	0
3	BP	2	0	1	0	BC	2	1	1	0
4	HC	3	2	1	1	HC	3	2	1	1

Notes: This table provides the path of experience stocks for each teacher in each of the two examples illustrating experience profile identification that occur in Appendix A. Each entry provides the level of general or task-specific experience in the dimension indicated by the column heading at the beginning of the year associated with the row. "B"-Basic, "H"-Honors, "P"-Physics, "C"-Chemistry.

Figure H.1: Tests for Dynamic Course Assignment Responses to Unobserved Time-Varying Endogenous Inputs



Notes: Figures display average school-teacher-year residuals (Figures H.1a and H.1e), school-teacher-subject-year residuals (Figures H.1b and H.1f), school-teacher-level-year residuals (Figures H.1c and H.1g), and school-teacher-subject-level-year residuals (Figures H.1d and H.1h), respectively, in the years leading up to a change in classroom assignment (using residuals from the *Restricted Specification*). *Restricted Specification* refers to a specification in which the school-teacher-subject-level fixed effects μ_{srjl} from Equation (2) are restricted to be common across subject-levels within a school-teacher combination: $\mu_{srjl} = \bar{\mu}_{st} \; \forall \, (j, l)$ and (s, r). A permanent change in general course assignment (H.1e) is defined as a teacher-year combination in which the teacher is not observed teaching any course in a subsequent sample year. A permanent change in subject assignment (H.1f) is defined as a teacher-subject-year combination in which the teacher teaches the chosen subject, but is not observed teaching the chosen subject again in subsequent sample years. Permanent changes in level (H.1g) and subject-level (H.1h) assignments are defined analogously to permanent subject changes. Temporary changes in assignment are defined in a similar manner as permanent course assignment changes, except the teacher is observed returning to teach (H.1a) or teach in the chosen subject (H.1b), level (H.1c), or subject-level (H.1d) in a subsequent sample year. Bootstrap standard errors (in brackets) are computed using 1,000 iterations.