Do Teachers Affect Learning in Developing Countries?: Evidence from Matched Student-Teacher Data from China*

Albert Park
Fairbank Center, Harvard University
Department of Economics, University of Michigan
alpark@umich.edu

Emily Hannum
Graduate School of Education, Harvard University
emily_hannum@gse.harvard.edu

Paper prepared for the conference
Rethinking Social Science Research on the Developing World in the 21st Century
Social Science Research Council, Park City Utah, June 7-11, 2001

Comments Welcome: Please Do Not Cite Without Permission

*We thank the Spencer Foundation Small Grants and Major Grants programs for supporting the collection and analysis of data for the Gansu Survey of Children and Families. Albert Park acknowledges an An Wang Post-doctoral Fellowship from the Fairbank Center, Harvard University. We thank Phil Brown, Dong Ruijun, Liu Xiaodong, Tang Tiemin, Wang Jiayi, Yang Hong, and Zhang Yanhong for their help in data collection and data management.

I. Introduction

Studies of the effect of teacher quality and educational inputs on academic achievement have produced decidedly mixed results for both developed and developing countries, leading to considerable controversy (Burtless, 1996; Hanushek, 1995; and Kremer, 1995).[1] Hanushek has argued that two decades of research in the U.S. has found no systematic evidence that teacher education, experience, salaries, or other measures such as teacher-pupil ratios or spending per pupil affect student performance (Hanushek 1986; 1989, 1996). However, other recent studies have found stronger evidence of positive school and teacher effects on learning and labor market outcomes (Hanushek, Kain, and Rivkin, 1998; Card and Krueger, 1996). In developing countries, a number of studies have found that teacher experience, as well as basic material resources, including textbooks and libraries do affect achievement, but many others have presented a mixed verdict on teacher and school effects (Heyneman and Jamison, 1980; Heyneman and Loxley, 1983; Lockheed et al., 1986, Behrman and Birdsall, 1983; Hanushek, 1995; Kremer, 1995).

Many of these conflicting research results can be attributed to a set of identification problems that have beset previous studies of teacher effects on student achievement. These problems can be grouped into three categories: omitted variables, selection bias, and measurement problems. No previous studies have adequately dealt with all of these concerns, especially the studies of student achievement in developing countries. In particular, because nearly all of the studies use average teacher characteristics for schools or

---

1 Much of this research traces its roots to the Coleman Report (Coleman et al., 1966) in the United States, which concluded that school resources mattered less than family background as a determinant of children's educational achievement. In a widely-cited study, Heyneman and Loxley (1983) found that the portion of explained variance in achievement attributable to family background was generally much smaller, and that attributable to school quality generally much larger, in developing than industrialized countries. The methodology of both studies were subsequently widely criticized. See Buchmann and Hannum, 2001 for a discussion of the impact of the Coleman Report on research in developing countries and a review of empirical studies of school effects in developing countries (see also Fuller and Clarke, 1994).

regions, they cannot control effectively for bias from unobserved school or community factors or deal with selection problems caused by school choice and student mobility. In this paper, we adopt an identification strategy that for the first time in a developing country setting exploits within-school variation in teacher characteristics and student performance. We exploit matched student-teacher data from a survey of primary school students, their families, teachers, and schools that we conducted in rural China and take advantage of unique features of China's educational system to estimate teacher effects on student test scores. One such feature is that teachers undergo rigorous, multifaceted evaluations of their performance each year, providing us with an excellent measure of school quality unavailable in most other contexts. A second feature is that teachers often stay with a cohort of students for more than one year, advancing with them from one grade to the next, allowing us to examine whether time together is a substitute or complement to other teacher attributes. Our goal is to address two questions: do teachers vary in their effectiveness, and if so, what factors explain the variation?

The paper is organized as follows. First, in section II, we describe the identification challenges of measuring teacher effects in greater detail. Then in sections III and IV we introduce the Gansu Survey of Children and Families and China's rural educational system, highlighting the ways in which they help us address the problems described in II. We then proceed with the empirical analysis in two stages. In section V, we first examine the extent to which teachers explain variation in math and language achievement. Then, in section VI, we identify the factors that explain the effects of teacher attributes on student test scores after first explaining how our identification strategy deals with all of the issues raised in II. A final section concludes.

II. Identifying teacher effects

In this section, we describe in greater detail the identification challenges related to omitted variables, selection bias, and mis-measurement. There are three possible types of omitted variables—school and community variables, teacher characteristics, and student background. In estimating teacher effects on learning, nearly all studies use data on average teacher characteristics for the school or region. However, these school-level measures are likely to be correlated with other unobserved school and community characteristics that affect learning, easily leading to mistaken inferences about the relationship between student performance and specific school attributes. Second, most studies use data sets with limited information about teachers—typically only a few commonly used variables such as education, experience, and wages. This leads to two potential problems. The true importance of teachers could be underestimated if the data do not capture important unobserved dimensions of teacher quality. At the same time, over-estimation of the importance of specific characteristics could occur if measured attributes pick up the effects of correlated, unmeasured teacher characteristics. A third and final type of omitted variable bias is lack of child background information (Burtless, 1996). If children's unobserved socio-economic status is positively correlated with teacher quality, failure to fully measure children's background would inflate estimates of the effects of teacher quality indicators.

Potential selection biases affect both between-school and within-school comparisons of student performance. School choice and residential (geographic) mobility can lead to different types of schools having students with different levels of unobserved motivation, ability, or family support. In the US, non-trivial proportions of families of students relocate based on perceived school quality (xxx). In many developing countries, children often are not restricted to attending a certain school, so that more motivated parents are likely to send their children to higher quality schools (Glewwe and Jacoby, 1994). Within schools, additional selection problems emerge. In poor settings characterized by delayed school entry, dropping out, and grade repetition, the usual approach of drawing school-based samples is problematic because children selected from a certain grade (whose test scores are comparable) will be unrepresentative of children in one age cohort (Glewwe and Jacoby, 1994). This could lead to nonrandom sorting of students across grade levels, which could lead to bias if teachers are also assigned non-randomly to different grades. Within grades, comparisons may be problematic if students are tracked into “gifted” or “accelerated” classes or remedial classes, or if motivated parents can affect which teacher is assigned to their children (more below).

The final set of concerns relate to measurement issues. One potential problem is the common use of school quality measures that are highly aggregated, such as the district level in South Africa (Case and Deaton, 1996); or states in the US (Card and Krueger, 1992). Empirical studies that use highly aggregated data to proxy for school-level measures tend to yield very different estimates of effects on student outcomes, usually more significant, than studies using data at the school level (see Moffitt, 1996; and Hanushek, Rivkin and Taylor 1996).[2] A second problem is the appropriateness of treating current school attributes as a proxy for what is a cumulative process. School resources could have changed only recently, or with student mobility some students may have spent fewer years in the school benefiting from its resources (Burtless, 1996). Finally, error in measurement of teacher characteristics would lead to estimated effects biased toward zero.

---

2 Aggregation bias can result from omitted school or regional variables, greater selection bias (e.g., sorting through migration, family background differences) across larger regional units, or non-linearities in true effects (Betts; 1996; Heckman, et al., 1996). Alternatively, aggregation can smooth out measurement error or sorting bias (if sorting is more local).

III. Survey Data

The Gansu Survey of Children and Families (GSCF), conducted by the authors in the summer of 2000, is a survey of 2000 children aged 9-12 and their families in rural areas of 20 counties in Gansu Province in northwest China. The data draw from extensive, separate questionnaires measuring attributes of children, parents, teachers, schools, and communities, and thus avoid many of the omitted variable and measurement problems of previous studies. Our multi-stage sampling scheme draws children from lists of all school-aged children in selected villages, enabling us to avoid concerns about selection bias that afflict school-based samples. Achievement tests in math and Chinese were administered to all children in the sample (described below in more detail). A teacher questionnaire was also administered to all teachers in schools attended by sample children (including teachers who did not teach any sample children); providing us with a sample of over 1000 primary school teachers.

Gansu, the study site in northwest China (see Map 1), is one of the nation's poorest provinces. Gansu encompasses 390,000 square kilometers of flat Loess Plateau, Gobi desert, mountainous and hilly areas, and vast grasslands. The province has a population of about 23 million. Gansu's socioeconomic and educational profiles resemble those of other interior provinces: relative to the nation as a whole, Gansu exhibits low per capita income, high rates of agricultural labor force participation, high rates of illiteracy, and low per-child educational expenditures.

IV. Teachers and Learning in Rural China

Schooling in rural China is similar to the situation found in many other developing country settings. First, rural schools in China face human resource constraints. Under-qualified teachers are prevalent in rural areas, and particularly in poor rural areas (Lo 1984; World Bank 1992).[3] Lack of preparation is exacerbated by the challenging conditions faced by rural teachers on the job. Rural teachers are poorly paid, they have little decision-making authority and face a very heavy workload (Lin 1993; World Bank 1992). Second, also following a global pattern, China has decentralized educational finance, a trend that has increased the private costs of schooling and tightened the links between local area revenues and the provision of schooling.[4] These common features of rural education found in China highlight the policy importance of learning more about the consequences that teacher quality has for learning outcomes.

A few unique aspects of the Chinese case enable us to more effectively address some of the methodological problems outlined above. First, Chinese schools are exceptional in conducting rigorous and systematic annual evaluations of teacher performance. Nationally mandated, these reviews include principal evaluations, student evaluations, publication counts, student test scores, the teacher's attendance record, and other aspects of teacher service. Principals evaluate the teacher's enthusiasm, use of innovative teaching methods, ability to maintain order and manage student concerns and problems. The different aspects of the review are scored using a point system. Based on these scores and the teacher's

---

3 Even at a national level, under-qualified teachers are a serious problem. A 1994 report issued by the Ministry of Education's Department of Planning and Construction indicates that 15.3% of primary teachers and 40.5% of junior secondary teachers did not have the required training or formal qualifications to be teachers (Ashmore and Cao 1997). Further, those with formal qualifications may lack the skills and motivation to be effective teachers: teacher education programs reportedly admit students of questionable capabilities and, more strikingly, little interest in teaching (Chang and Paine 1992, Wu and Chang 1990).

4 The quality of schools, the qualifications of teachers, and student expenditures vary not only across the urban-rural divide, but also with level of development (Cheng, 1996, pp. 24-29; Tsang, 1994; Lewin and Wang, 1994; Lin, 1993; Lo, 1984; World Bank, 1992).

educational background and years of teaching, the teacher may be considered for permanent ranking adjustments. For primary schools in Gansu, there are four quality rankings: intern (*jianxiqi*), second level (*xiaojiao erji*), first level (*xiaojiao yiji*), and highest level (*xiaojiao gaoji*).[5] Because the official quality rankings integrate information on so many aspects of teacher performance, they provide a uniquely informative measure of teacher quality that is unavailable in other developing and developed countries. Our empirical analysis enables us to say something about whether the Chinese rank system rewards dimensions of quality that actually do affect student performance as measured by test scores.

Second, in China it is quite common for the same teacher to teach a cohort of students for more than one year, teaching a higher grade level each year until the cohort graduates or the cohort is passed on to another teacher at higher grade levels. This practice is based on the premise that students learn better when they feel more comfortable with a teacher, both because they are less self-conscious and because they become accustomed to how the teacher teaches. As seen in Table 1, in our sample we see considerable variation in the number of years that teachers have taught students, even among teachers teaching the same grade level. In every grade, 40-50 percent of students have been with their current teacher for only one year, but the majority of students have been with their current teacher for two or more years.

The tendency of teachers to follow students to higher grades presents a potential opportunity to better identify teacher effects on learning, and to test whether systems that

[5] Rules govern the years of service required before one can apply for the next rank. All teachers begin as interns in their first year, regardless of educational background, and can apply immediately for second level beginning the second year. Those graduating from a secondary teacher training school can apply for first level after 7 years, and highest level after 15 years. Those graduating from a normal college can apply for first level after 3 years and highest level after 7 years. Those graduating from universities can apply for first class after one year and for highest level after 5 years. These three levels correspond to the third class, second class, and first class quality rankings described by Ding and Lehrer (2001). Only middle school teachers can qualify to be "superior" teachers, the highest possible rank.

encourage multiyear assignment of teachers to student cohorts improves student performance. One question of interest is whether the number of years taught acts as a complement or substitute to other aspects of teacher quality. Does longer teacher-student interaction accentuate the effect of teacher differences or allow less effective teachers to compensate for other limitations and do better in a relative sense? Do teachers teach as capably when they change grade curriculums each year, leading to interaction effects between years taught and teacher quality? We can test these questions by interacting the variable years taught (YRS) with teacher variables. An obvious concern is that who teaches students longer within the same school is not a random outcome, but reflects unobserved characteristics of the teacher that may also affect test scores. This could lead to erroneous inference about the effect of years taught on test scores.

Finally, various aspects of Chinese families, communities, and schools in remote rural areas lead to many fewer selection problems than occurs in many other contexts. First, enrollment rates in the study area tend to drop with the transition to junior high school rather than at the primary stage (Hannum, 1999a, 2000; Brown and Park 2001). Second, geographic mobility is very limited administratively and nearly all children attend the nearest primary school, which is usually the sole primary school in the village. Third, within schools, sorting of students into classes of different abilities is rare because most schools have only one class per grade.

V. Analysis of Variance of Test Scores

To quantify the amount of variation in test scores explained by teacher differences versus other factors, we decompose the variance of student math and language test scores into the shares of variance explained by school differences, within-school teacher differences, and remaining unexplained factors. Because of our inability to control for covariances between school, teacher, and individual factors; multiple possible interpretations of school and teacher differences; and the likely prevalence of measurement error in test scores, our ability to make inferences is limited. Nonetheless, we argue that the analysis of variance suggests that a significant amount of test score variation is likely explained by teacher differences. We also examine the effects of longer teacher-student relationships on test score variation by comparing the decomposition results for students who have been taught by their current teacher for one year and for those who have been taught for two or more years.

Test scores are from examinations administered to all 2000 children in the Gansu survey sample. Divided on a random basis, half of the children in each village were given mathematics exams and half were given language exams. To ensure that the tests were calibrated to test an appropriate range of knowledge, separate exams were given for children in grades 3 and below and those in grades 4 and above. The tests were designed by educational experts at the Gansu Educational Commission to cover the range of official primary school curriculum, which is standardized nationally. The tests were scored from zero to 100, and were converted to standard deviations from mean score by grade level.

School differences explain 39 percent of the variation in math scores and 38 percent of the variation in language scores; within-school teacher differences explain 24 and 23 percent; and unexplained variation accounts for 37 and 39 percent (Table 2). The relative magnitudes thus appear highly consistent for math and language scores.

How much of the actual variation in child learning is accounted for by teacher differences? A number of factors should be kept in mind in interpreting the above numbers. First, test-scores are differenced from grade means, so the decompositions exclude test score differences that are systematic by grade. Grade differences are likely to include teacher differences if the quality of teachers differs systematically by grade level, which we find in subsequent analysis below. They are less likely to include differences in other learning inputs since within schools, the quality of classrooms and other facilities is likely to be quite similar for different grades.

Second, between-school differences are likely to have a sizable component that is due to teacher differences (due to the covariance of teacher and school quality), since average teacher quality varies across schools. However, it is very difficult to quantify this share because of unobserved school and community factors.

Third, most of the measured within-school between-teacher variation in test scores is likely due to teacher differences. Components of this variation that may be unrelated to teacher characteristics include differences in the allocation of other inputs that vary among teachers, differences in class size, and variation in average student or family characteristics. All of these biases must not be systematic by grade, but they can include within-school grade differences that are unique to the school.

Our measure of within-school teacher variation might be measuring other factors that vary among classes within schools (but not systematically by grade), such as allocation of other learning inputs or class size. As noted above, we expect allocation of other learning inputs to be equitable within schools, especially those that are not systematically grade-related. Most class size differences within schools reflect enrollment differences since most grades have only one class. Enrollment variation should tend to vary systematically by grade if demographic changes are similar within the province. In any case class size is not found to significantly affect test scores (below).

Finally, average characteristics of students can vary if students are grouped by ability when there are multiple classes per grade, or if there is variation in the differences in average ability of grade cohorts, due perhaps to different propensities in different schools for children to enroll at different ages or to be promoted, or to drop out. To exclude the effects of the former source of bias, which should be limited since only one third of students are in multiple-class grades, we define within-school teacher variation as within-school grade variation, banishing the within-grade teacher differences to the unexplained variance. We find that the share of variance falls to 0.19 and 0.18 for math and language scores, respectively, which includes a systematic underestimation of within-school teacher variation. To test the effects of the bias from average student differences unrelated to systematic grade differences, we redefine the test scores not as the difference from grade-level mean scores but as the difference from other students in the same grade enrolled at the same age and held back the same number of years. This will reduce shares of both the within-school between-teacher variation and the unexplained share of variance. We can get a sense of the importance of differences in average student characteristics by examining how the ratio of within-school teacher variation to between-school variation changes. We find that this ratio is almost identical to the results in Table 2, suggesting that the effects are small. Thus we conclude that the within-school among-teacher variation is primarily due to true teacher differences.

Fourth, the unexplained share of variance is likely includes significant measure, which will lead to an understatement of the relative shares of between-school and within-school between-teacher variation.

Overall, we conclude that the share of between-school within-teacher variation (24 and 23 percent) is likely to substantially underestimate true teacher effects, highlighting the importance of teachers in learning.

Analysis of Variance by Years Taught

If longer exposure to teachers increases the impact of teacher quality on learning, we might expect teachers to explain a greater share of variation in the test scores of students who are with their teachers for longer periods of time. By this theory, a system in which children stay with the same teacher throughout primary school will produce much larger teacher effects than a system which passes students from one teacher to another. On the other hand, time with students could be a substitute rather than a complement to other teacher characteristics. Familiarity with students may improve child performance in ways that help teachers overcome limitations in other measured quality dimensions.

To test the relative validity of these two competing views, we divide the samples into two groups that turn out to be of roughly equal size—students who have been with their current teacher for one year and students who have been with their current teacher for more than one year (see Table 2). We find that the share of variance in math scores explained by within-school between-teacher variation is significantly lower for children who have been with their teachers longer (0.18 versus 0.27 for those who have been with their teachers for one year). However, for language scores, there is not much difference; in fact the share of within-school between-teacher variation is slightly higher for the children who have been with their current teachers for more than one year (0.20 versus 0.17). Thus, years taught seems to be a substitute for other dimensions of teacher quality for math learning but not to matter or be a slight complement for language learning. The fact that the results are different for math and language also provides evidence against the possibility that other correlates that might differ with years taught, such as differences in student cohort differences, could explain the result since such factors should affect math and language samples in the same way.

VI. Teacher Effects on Student Test Scores

Identification Strategy

We briefly summarize the strengths of our identification strategy with reference to the numerous issues raised section II. Turning first to omitted variable bias, we avoid bias from unobserved school and community factors by identifying teacher effects based on *within*-school variation in teacher characteristics and student performance. To reduce the possibility of bias from omitted teacher and student variables, our survey collected extensive information on teachers, students, and parents. In addition, in China, quality rankings of teachers based on systematic teacher evaluations undertaken by all schools provide us with a unique teacher variable likely to capture quality attributes that normally are unobserved.

Next we consider selection bias. In rural communities in northwest China, nearly 100 percent of children attend the nearest primary school and residential movement is prohibited by a strict residence permit system, so that selection problems due to school choice or residential mobility do not exist. We control for differences in age of enrollment and grade repetition among students in the same grade by controlling directly for these differences with a set of flexible interaction dummies (described below). Selection problems due to dropouts are not a significant concern because very few children in our residence-based sample of 9-12 year-olds drop out. Finally, non-random teacher-student matching within grades is not an issue for most children because most primary schools have only one class per grade. In the cases where there are two or more classes per grade, we take grade averages, so that identification comes exclusively from between-grade comparisons.

With regard to measurement problems, our detailed information on the teachers of each student avoids bias from using aggregate teacher quality measures. Our focus on within-school variation and the lack of student mobility means that we are typically making comparisons among students who have been in the same school for the same length of time, eliminating bias from use of variables the measure *current* school attributes. We do not expect our teacher characteristic variables to be subject to significant measurement error given the unambiguous nature of the variable definitions.

Teacher and Student Variables

We focus on eight teacher characteristics, which are taken from survey questionnaires completed by each sampled child's primary teacher (*ban zhuren*) in the past school year, and by all teachers in primary schools attended by sample children. We briefly explain how teacher variables are defined and how they should be interpreted in the context of rural Chinese schools. The sample means for the teacher variables (for the full teacher sample) are presented in Table 3, and a correlation matrix for the different measures is presented in Table 4. Teacher quality rankings are most highly correlated with experience, wages, and government employment status, with a somewhat weaker but positive correlation with the education variables.

Quality Rank

We construct two indicator (1/0) variables, TQUAL1 and TQUAL2, for whether the teacher has a first level quality ranking (46 percent of all teachers) or a highest level quality ranking (16 percent). We do not distinguish between intern and second class because promotion from intern to second level is almost automatic after the first year of service.

Education

We describe education by two indicator variables, TEDH and TEDC, for whether the teacher has completed high school-level education or college-level education. High school level education includes those who graduate from regular high school or from a specialized teacher training school attended after middle school (*zhongzhuan*). College-level education includes those that graduate from regular universities (very few) or from a normal college (*dazhuan*) following the completion of high school. There are some teachers who take correspondence courses to receive accreditation for *dazhuan* without ever having completed high school. Because of the design of the survey instrument, we have difficulty distinguishing these cases from teachers who receive *zhongzhuan* degrees, and so they are categorized as having high school-level education. Thus TEDC is equal to one only when the teacher completes both high school *and* college-level schooling (14 percent of all teachers), and TEDH captures all other educational outcomes except those teachers who hold only a middle school degree (63 percent).

Experience

We have two experience variables. TEXPT is the number of total years of teaching experience, and TEXPS is the number years the teacher has taught in the current school. Mean values for TEXPT and TEXPS are 14.8 and 7.3 years, respectively.

Wage

Wages are the monthly salary and bonuses received by teachers. Interviews throughout the province found that because of budgetary shortfalls, bonuses were extremely rare. Wage scales are set by each county, based on the teacher's quality rank and years of teaching (Ding and Lehrer, 2001). The average monthly wage was 516 yuan, or $62 (8.3 yuan/$). TWAGE is the monthly wage and TLWAGE is the log of the monthly wage in yuan.

Sex

TSEX is an indicator variable for whether the teacher is male. Sixty-two percent of all teachers are male.

Other Work

Many teachers also are farmers with their own plots of land. Others engage in secondary income-earning activities such as petty trade. Teachers who are native to the village are more likely to engage in non-teaching activities. China has a long history of farmer-teachers (*minban laoshi*), who received partial salaries. Under recent reforms, however, most *minban* teachers have been re-certified as official teachers or let go. However, many continue to farm their land. TOWORK is the number of hours spent on work unrelated to teaching. Fifty-two percent of teachers report positive hours of other work, and those that work spend an average of 10.8 hours per week in the non-teaching activity.

Government Employee

In China, most teachers are government employees whose wages are paid out of county government budgets and who are assigned to schools by county education bureaus (80 percent of all teachers). In addition, some schools also hire "unofficial" teachers using locally raised funds, especially when there is a shortage of available qualified teachers. These teachers often are local villagers who have a relatively high level of educational attainment (e.g., high school) and who are paid a negotiated wage. The variable TGOV is an indicator for whether the teacher is an official government employee.

Hometown

Thirty-six percent of teachers are native villagers, 33 percent were born in other villages in the same township, 26 percent in other townships in the same county, four percent in other counties in Gansu, and one percent in other provinces. An indicator variable, TVILL, is equal to one if the teacher was born and raised in the village in which the school is located. In China's centralized job allocation system, managed by education bureaus at different administrative levels, teachers are often assigned as close to their homes as possible.

Child and Household Characteristics

In addition to the information on age of enrollment, years held back, and grade incorporated into the interactive dummies that control for cohort differences, we also include four additional child and household characteristics as regressors: CSEX, the sex of the sample child; FEDUC, father's years of education; MEDUC, mother's years of education, and LEXPPC, log of expenditures per capita.

Empirical Specification

A straightforward specification for estimating teacher effects using matched student-teacher data is the following:

$$T_{its} = \alpha C_i + \beta X_t + \gamma D_s \tag{1}$$

In equation (1), T_{its} is the test score of child i with teacher t and school s, C_i is a vector of child and family characteristics, X_t is a vector of teacher attributes, and D_s is a vector of school dummy variables.

If teachers were randomly matched to students within schools, the β estimated in (1) would be unbiased. However, the possibility of non-random matching within schools leads to concern that teachers with unobserved quality traits will tend to be assigned to students with positive or negative unobserved background factors, creating an artificial statistical correlation between teacher characteristics and test scores. In fact, the lack of these potential sources of bias is an advantage of using school average measures of teacher quality.

One difficulty faced by studies in the U.S. that have attempted to identify teacher and class size effects using within-school variation is the possibility of nonrandom matching of students and teachers when there are multiple classes in each grade level, for example if schools organize accelerated or remedial classes (Boozer and Rouse, 1995; Hanushek, Kain, and Rivkin, 1998). Intentional nonrandom matching of this type is much more likely than *intentional* sorting among grade levels because schools have much less discretion to move students among grade levels than between classes, so that grade cohorts tend to be quite similar from year to year (although not necessarily from grade to grade in any given year—more below). In rural schools in China, identification is facilitated by the fact that most schools are small village schools that have only one class per grade (73.2 percent of students in the sample were in grades with only one class). For those children in grades with multiple classes, we can avoid within-grade selection bias by taking the average teacher characteristics of all teachers in the grade (almost always 2 teachers, and 3 at the most), and identify teacher effects solely from between-grade teacher differences. Of course, this is a possible strategy in any study, but the lower prevalence of multiple classes per grade in China should reduce aggregation bias and make the teacher effects more pronounced.

Nonetheless, there still remain many sources of non-random matching across grade levels within schools, including the following:

1. Differential student attrition. More able teachers may have lower dropout rates, leading to larger class size and different student characteristics.

2. Promotion decisions. Effective teachers may have higher promotion rates and so retain lower ability students on average. Further, effective teachers may also inherit those not promoted from higher years. These students may have lower ability, but would also be older and would have seen the curricular material already.

3. Nonrandom teacher assignment by schools. More able teachers may be assigned to upper grade levels, to larger classes, or on the basis of other characteristics of grade cohorts.

4. Nonrandom assignment of complementary inputs. Teachers with specific characteristics (age, quality, etc.) may be given better classrooms or other types of support.

5. Strategic behavior by parents. Parents may influence teacher-student matches by lobbying for teacher assignment, strategically timing their children's enrollment, arranging for children to skip or be held back, etc.

If non-random sorting corresponds to observable factors for which we have data, such as grade level, we can control effectively for such differences. In this case, having detailed data on child, household, and teacher characteristics can help reduce potential bias. Even if non-random sorting is based on unobservable characteristics, it may still not be a problem if the unobservables do not affect test scores.

Tables 3 and 4 show that both average teacher characteristics and average student characteristics differ by grade. Teachers in higher grades tend to be better educated, higher quality, more experienced, male, from outside the village, earn higher wages, and be formal employees. Higher grades tend to find more boys than girls and children from richer families with parents who are better educated. Neither of these patterns is surprising. Because subject matter is more difficult in higher grades and preparation is important to determine middle school placement, schools tend to put their best teachers in upper grades. Because the sample consists of children in a specific age range (9-12) and there is considerable variation in the range of ages of children in each grade, the children in the lower grades will be over-represented by children who enrolled later or were held back more often than the children in higher grades.

If cohort and teacher differences across grades exhibited a common tendency across all schools, the inclusion of grade dummy variables would effectively control for possible bias associated with unobservables that affect sorting among grades. Then the identification of teacher effects would be based on how students performed relative to kids in other grades in their school compared to average grade differences common to all schools. However, if the sorting of students and teachers across grades differs across schools, then grade dummies will not sweep out potential sources of bias. For example, it might be the case that in poorer villages, there are sharper differences across grades in both teacher characteristics (because rich villages have a greater proportion of effective teachers) and in student characteristics (because socio-economic status differences have more pronounced effects on age of enrollment and whether kids are held back). In this case, even after controlling for average grade differences, there would be a positive association between teacher quality and student quality in poorer areas. If these correlations are associated only with observable characteristics, then there is still no bias as long as we include the observable characteristics as controls, but there is no reason to be confident that unobservables are not similarly correlated.

To the extent that non-random sorting of students across grades ultimately is manifest in differences in the age of enrollment, years held back, years skipped, or years out of school (including dropouts and temporary suspensions of schooling), one way to more convincingly control for student sorting is to include direct controls for these specific decisions. In our sample, skipped years, dropouts, and temporary suspensions of schooling are extremely rare, so we disregard them as serious concerns.[6] To be as unrestricting as possible in how we assume non-random student sorting affects grade cohort comparisons, we include not just grade dummies, but a full set of interactive dummies for grade level (D_g), age of enrollment (D_a), and years held back (D_h).[7] The specification can now be written as follows:

$$T_{its} = \alpha C_i + \beta X_t + \lambda D_g \times D_a \times D_h + \eta CS_t + \gamma D_s \qquad (2)$$

In essence, all comparisons are now made among individuals not just in the same grade, but among those who enrolled at the same age, were held back the same number of years, and are in the same grade. The grade controls also control for systematic teacher differences across grades. We have no reason to believe that remaining differences in how unobservables correlate with these different student outcomes should correlate systematically with differences in how unobservable teacher characteristics vary across grades.[8]

---

6 The percentage of children who had skipped a grade or ever withdrawn from school temporarily was about one percent each, and only 4 students enrolled in primary school and had already dropped out before graduating.

7 One concern about using past promotion decisions as controls is that the past held back decision may have been made by the current teacher and reflect his or her characteristics. This is true for a maximum of four percent of the held back decisions in the sample, calculated based on information on the grades held back and the years the teacher has taught the child. Because we believe the sorting bias is likely to be a higher order problem, we include the held back decisions. If there is bias, it should bias toward zero the effects of teacher characteristics.

8 Note that it is possible to add another level of interactions with the child's sex if we think that sorting is substantially different by sex. We tested this and found that the results changed little.

In (2), we also add class size (CSIZE) as an additional control. Assuming demographic changes do not differ widely across the province, this should help control for sorting due to enrollment or promotion decisions that respond to or are correlated with teacher quality, and also controlling for actual effects of class size on student learning.

Results

Table 6A and 6B summarize the estimation results for (2) for math and language scores, respectively. We report robust standard errors adjusted for clustering by teacher. On average, there are 3.2 students per teacher in the sample. Separate regressions are run for math and language test scores, respectively. In addition to the preferred specification 1, we also report the results when years taught and its interactions are dropped (which reveal average effects not conditioning on years taught) and when child and family characteristics are dropped (but we maintain the grade, age enrolled, and years held back interactive dummies).

Before turning to a detailed discussion of specific teacher characteristics, we offer several general characterizations of the results for math and language test scores. First, the effects of many teacher characteristics are much more pronounced for math scores than for language scores, indicating either that these characteristics matter less for language learning than math scores, or that the important teacher characteristics are different for math and language and we do not have adequate measures for the characteristics that affect language learning. Second, the interactions between years taught and teacher characteristics are very significant for math scores but unimportant for language scores, and the conclusions about teacher effects are very sensitive to whether the specification accounts for the effect of years taught and its interactions with teacher characteristics. Third, once we have controlled for grade, age of enrollment, and years held back differences, teacher effects are not sensitive to the inclusion of additional child and family variables. This result makes us optimistic that unobserved student characteristics also are unlikely to bias our estimates. Finally, and most importantly, overall the results suggest that teacher characteristics have a significant effect on student learning, leading us to respond affirmatively to the paper's motivating question.

Quality rankings have large effects on test scores for both math and language. For math, the effects weaken as years taught increase. In the first year, relative to a second level ranking, a first level ranking increases math scores by 0.270 standard deviations and a highest ranking increases scores by 0.481 standard deviations. But the effects disappear by the second year and are negative by the third year. On average the effects of quality rankings are modestly positive (mean years taught is 1.79 years), about 0.10 standard deviations for both first level and highest level rankings. For language scores, the average effects of a first level ranking is small and not statistically significant but a highest ranking increases test scores by a very large 0.459 standard deviations. Interactions with year taught are not statistically significant, but if anything suggest an increasing importance of quality as years taught increase.

The negative interaction between years taught and quality rankings for math scores but not language scores suggests that in mathematics, the relative benefits to learning of staying with students longer is stronger for teachers with lower quality rankings than those with high quality rankings. This suggests that time spent with children may be a substitute for teacher quality. It could be that less effective teachers do relatively better as their contact with students increase. Perhaps effective teachers connect with children right away, but less qualified teachers take more time. Another possibility is that teacher attributes matter less when teachers are shifting the curriculum they teach from year to year. In other words, teacher quality matters more when teachers specialize in one year of curriculum. Either explanation can offer a rationale for why interaction effects are more important for teaching mathematical concepts rather than language. Math learning, which involves more abstract conceptualization, could depend more on student comfort levels or teacher skills, or could benefit more from specialization in grade curriculum. Another possibility is an endogeneity story in which teachers with negative unobservables are likely to teach kids longer, if for instance, better teachers are used exclusively to teach 5th or 6th grade (and so change students yearly). We will return to this selection argument later, but for now simply note that it cannot explain the lack of similar negative interactions for language scores.

Conditioning on teacher quality ranks, we find that for math scores, the coefficients on college education (TEDC) and its interaction with YRS are of the opposite sign of those for TQUAL2. Assuming that TQUAL is positively associated with teacher effectiveness, this result suggests that the criteria for highest level quality rank (TQUAL2) may overstate the importance of college education. The coefficients for high school education are not statistically significant in either regression, nor is that for TEDC in the language regression.

There is some evidence that math test scores are higher when the gender of the teacher and the student is the same (about 0.35 standard deviations higher). The coefficient of the gender interaction term is also positive for language scores, but much smaller and not statistically significant in any specification. The average effect of gender of teachers or students is not statistically significant, but the signs are consistent in all specifications and suggest (weakly) that male teachers and male students have lower test scores (all effects in the 0.10 to 0.20 standard deviations range). Interactions of gender variables and years taught all have statistically insignificant coefficients.

We find that teachers' wages act similarly to teacher quality (positive initial effect and a negative interaction with years taught) but do not affect language scores. Working in non-teaching activities is never statistically significant with or without interactions.

For math scores, both being a government employee and being from the local village have positive and statistically significant interactions with years taught, suggesting that these characteristics are complementary to building trust with students and their families. "Informal" teachers may be less committed to the teaching profession and so devote less time to cultivating relationships, and native villagers can build upon existing social ties. "Informal" employees could, however start with an initial greater familiarity with the community, explaining the initial negative effect of government employment on math scores. For language, being a native villager has a negative interaction with years taught (recall that quality ranks also had opposite signs for language and math). This again is consistent with the story that the duration of teacher-student contact matters less for language learning but familiarity with students may provide an initial advantage. As years taught increase, being a native villager matters less but being an effective teacher matters more. But for math learning, longer periods teaching the same children apparently has much bigger payoffs.

Finally we note that evaluated at sample means, the average effects of years taught is very small, just 0.03 and 0.02 standard deviations per year for math and language scores, not statistically different from zero. Thus, it is hard to say anything definitive about the benefits of China's policy of having teachers stay with students as they progress through school. The policy certainly affects which teacher characteristics matter. The benefits of the policy appear greater for teachers with lower quality ranks, lower wages, lack of government employment status, and village outsiders. Thus, the policy of keeping teachers and students together for longer periods of time could have the effect of reducing inequities in learning associated with teacher qualifications, and may be desirable for schools with less effective teachers. Overall, the policy may hold fewer advantages in schools with better teachers, and so perhaps will hold less attraction as overall teacher quality improves over time.

Among the household variables, father's education has a significant positive effect on test scores, with an additional year of education increasing test scores by about 0.02 standard deviations. No other effects are significant, and as noted earlier, excluding these variables does not significantly affect the estimated teacher effects.

Is the number of years taught endogenous?

We examine which teacher characteristics are associated with the number of years that teachers accompany the same grade cohort by estimating an ordered probit model in which years taught is explained by our set of teacher characteristics, grade dummies, and school dummies. We exclude first grade teachers because years taught is equal to one by definition. Specification 2 also excludes all teachers who have taught in the school for less than five years, to try to isolate the differences among those whose years taught is dictated by recent arrival.

The results presented in Table 7 reveal that quality rankings are not related to years taught. This suggests that principals do not strategically use better teachers in different rotation schemes than worse teachers, and corresponds with our understanding from field interviews that within schools, teachers generally received equal treatment with regard to whether they followed students to higher grades. The results do find other teacher characteristics to be important, however, including experience, gender, and whether the employee is a government worker. Overall, characteristics associated with less mobility appear to predict more years teaching the same students (more years of experience, women, those with official jobs). While it is certainly possible that such characteristics are correlated with unobserved teacher ability, the lack of correlation with quality ranks suggests that selection is not occurring directly on teacher effectiveness. Also, even if years taught is picking up unobserved teacher quality, it is not obvious that these unobservables should interact with observable teacher characteristics and bias coefficients on the interaction terms. Finally, as mentioned earlier, strong endogeneity bias is not very consistent with the very different effects of the interactions for math and language scores.

Do quality ranks accurately measure quality?

China's unique system of systematic teacher quality evaluations and its direct link to pay offers a potentially valuable model for introducing merit-pay systems to other developing or developed countries. We have shown that quality ranks strongly predict differences in test scores. For this reason, the Chinese system and available data on ranks has enabled us to more convincingly demonstrate quality effects than would be possible using data from other parts of the world.

Assuming that test scores are a good measure of teachers' performance goals, if quality ranks perfectly measure all relevant aspects of quality, the inclusion of teacher attribute variables in addition to quality ranks in the test score regressions should yield insignificant coefficients. The sign of the coefficients on other attributes indicates whether the quality ranks undervalue or overvalue the attribute. To see which factors most determine quality ranks, we estimate regressions that directly assess the effect of teacher attributes on quality ranks (Table 8). Then, to accurately measure the true importance of specific attributes, which includes the effects captured by quality ranks and the effects not captured by quality ranks, we estimate a reduced form regression that excludes quality ranks (specification 2 in Tables 6A and 6B).

The quality rank (TQUAL1, TQUAL2) regressions reported in Table 8, separately test which factors affect whether teachers have second level versus first level quality rank (excluding those with highest level quality rank), and whether teachers have highest level versus first level quality rank (excluding those with second level). This specification may introduce some selection bias but is most in accord with the sequential nature of rank assignments (one moves from second level to first level to highest level). It also allows us to estimate conditional logits that control for school effects. Not surprisingly, we find highly significant positive effects for education, teacher experience, and government employment across all specifications in which the variables are included, with similar magnitudes whether or not we control for school effects (suggesting a standardized ranking system). We also find that men are more likely than women to attain 1st level rank only when controlling for school effects, and that being a native villager reduces the likelihood of attaining a higher rank only when not controlling for school effects (likely because teachers with higher quality ranks are more likely to be moved to other schools).

Next we compare the results in specification 1 (with quality ranks) and 2 (without quality ranks) in Tables 6A and 6B. Because few of the estimates for language scores is statistically significant, we focus on the math score results. First, we observe that interaction terms for a number of variables have significant coefficients when quality ranks are included, suggesting that they are not adequately reflected in quality ranks. If we assume quality ranks are in fact associated with more effective teaching, the signs of the interaction term coefficients suggest that the quality ranks overvalue education, undervalue experience, and undervalue government employment and being a native villager. The latter is not surprising since being a native villager is not taken into account in quality ranks. We also find that the magnitude of the effects of teacher characteristics change relatively little even when TQUAL1 and TQUAL2 are removed from the regressions, so that the reduced form effects are very similar to the effects that are independent of TQUAL1 and TQUAL2. This suggests that the quality rank variables contain a significant amount of information about teaching quality which is independent of education and experience. Thus, the failure of these variables to show up as significant in empirical studies should not imply that teacher quality is not important.

VII. Conclusion

Do teachers affect learning in developing countries? In this paper, we conduct an analysis of variance of math and language test scores and utilize rarely available matched student-teacher data from China and unique features of China's educational system to estimate teacher effects on student test scores, employing an identification strategy which exploits within-school variation in teacher characteristics and student performance for the first time in a developing country setting.

We find that teacher quality matters. We cautiously infer that much of the variation in test scores (at least about one fourth) is likely due to teacher differences. For math scores, higher teacher quality rankings substantially increases test scores, but if the teachers continue to teach students for multiple years, the increased teacher-student contact appears to substitute for teacher quality ranking, so that the quality effects diminish over time. For language scores, the effects do not appear to interact with years taught, but the average effects of the highest teacher rank is very substantial (nearly half of a standard deviation). Other teacher characteristics matter, too, in particular for math scores (including education, experience, wages, whether the teacher is a government employee, and whether the teacher is a native villager). In general, in teaching math, less effective teachers tend to benefit relatively more from longer interaction with the same student cohort, suggesting that China's system of having teachers teach the same student cohort over multiple years may promote equity in education quality and be more appropriate when teachers are less qualified.

The quality ranks appear to contain a substantial amount of information about teacher quality not contained in conventional measures such as teacher education and experience. This suggests that China may be a good setting for more in-depth studies of the effects of teacher quality, and that China's system of teacher evaluations can serve as an effective model for other countries interested in monitoring and rewarding teacher quality.

References

Aitkin M, Longford N. 1986. Statistical modeling issues in school effectiveness studies (with discussion). *J. R. Stat. Soc., Ser. B* 149:1-43.

Baker DP, Goesling B, LeTendre GK. 1999. Social class, school quality, and national economic development: a cross-national analysis of the "Heyneman-Loxley" effect. Presented at Annu. Meet. Am. Sociol. Assoc., 94th, Chicago.

Behrman JR, and N. Birdsall. 1983. The quality of schooling: quantity alone is misleading. *Am. Econ. Rev.* 73:928-46

Betts, Julian. 1995. "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth," *The Review of Economics and Statistics* 77(2): 231-250.

Bidwell C.E., Kasarda J.D. 1980. Conceptualizing and measuring the effects of school and schooling. *Am. J. Educ.* 88:401-30

Boozer, Michael, and Cecilia Rouse. 1995. "Intraschool Variation in Class Size: Patterns and Implications," NBER Working Paper No. 5144.

Brown, Phil, and Albert Park. 2001 (forthcoming). "Education and Poverty in Rural China." *Economics of Education Review*.

Buchmann, Claudia, and Emily Hannum. 2001. "Education and Stratification in Developing Countries: A Review of Theories and Research," *Annual Review of Sociology* 27.

Burtless, Gary. 1996. "Introduction and Summary," in Gary Burtless, ed., *Does Money Matter?: The Effect of School Resources on Student Achievement and Adult Success* (Washington, D.C.: Brookings Institution Press, pp. 1-42).

Card, David and Alan Krueger. 1996. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy* 100: 1-40.

Coleman, JS, Campbell E, Hobson C, McPartland J, Mood A et al. 1966. *Equality of Educational Opportunity.* Washington, DC: Department of Health, Education, and Welfare.

Ding, Weili, and Steven Lehrer. 2001. *The Optimal Policy to Reward the Value Added by Educators: Theory and Evidence from China*, University of Pittsburgh, mimeo.

Ferguson, Ronald F. 1991. "Paying for Public Education: New Evidence on How and Why Money Matters." *Harvard Journal on Legislation* 28 (Summer): 465-498.

Fuller B, Clarke P. 1994. Raising school effects while ignoring culture? *Rev. Educ. Res.* 64:119-57

Fuller B. 1987. What school factors raise achievement in the developing world? *Rev. Educ. Res.* 57:255-92.

Glewwe, P., & Jacoby, H. (1994). Student achievement and schooling choice in low income countries: Evidence from Ghana. *Journal of Human Resources* 29(3), 841-64.

Goldstein H. 1987. Multilevel models in educational and social research. New York: Oxford

Hannum, E. 1999a. "Political Change and the Urban-Rural Gap in Education in China, 1949-1990." *Comparative Education Review*. 43(2): 193-211.

Hannum, E. 1999b. "Poverty and Basic-Level Schooling in the People's Republic of China: Equity Issues in the 1990s." *Prospects: the Quarterly Journal of Comparative Education*. 29(4): 561-577.

Hanushek, E. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24: 1141-77.

Hanushek, E. 1989. "The Impact of Differential Expenditures on School Performance." *Educational Researcher* 1: 45-51.

Hanushek, E. 1995. Interpreting recent research on schooling in developing countries. *World Bank Res. Observ*. 10:247-254

Hanushek, E. 1996. "School Resources and Student Performance," in Gary Burtless, ed., *Does Money Matter?: The Effect of School Resources on Student Achievement and Adult Success* (Washington, D.C.: Brookings Institution Press, pp. 43-73).

Hanushek, Eric, John F. Kain, and Steven G. Rivkin. 1998. "Teachers, Schools, and Academic Achievement," NBER Working Paper No. 6691.

Heckman, J., A. Layne-Farrar, and P. Todd. 1996. "Does Measured School Quality Really Matter? An Eamination of the Earnings-Quality Relationship," in Gary Burtless, ed., *Does Money Matter?: The Effect of School Resources on Student Achievement and Adult Success* (Washington, D.C.: Brookings Institution Press, pp. 192-289).

Hedge, Larry, and Rob Greewald. 1996. "Have Times Changed?: The Relation between School Resources and Student Performance," in Gary Burtless, ed., *Does Money Matter?: The Effect of School Resources on Student Achievement and Adult Success* (Washington, D.C.: Brookings Institution Press, pp. 74-92).

Heyneman SP, Jamison DT. 1980. Student learning in Uganda: textbook availability and other factors. *Comp. Educ. Rev.* 24:206-20

Heyneman SP. 1976. Influences on academic achievement: a comparison of results from Uganda and more industrialized societies. *Sociol. Educ.* 49:200-11

Heyneman, Stephen, and William Loxley. 1983. "The Effect of Primary-School Quality on Academic Achievement Across Twenty-nine High-and Low-Income Countries," *American Journal of Sociology* 88(6): 1162-1194.

Kremer M. 1995. Research on schooling: what we know and what we don't (a comment on Hanushek). *World Bank Res. Observ.* 10:247-54

Lloyd CB, Mensch BS, Clark WH. 2000. The effects of primary school quality on school dropout among Kenyan girls and boys. *Comp. Educ. Rev.* 44:113-47

Lockheed ME, Longford NT. 1991. School effects on mathematics achievement gain in Thailand. In *Schools, Classrooms and Pupils: International Studies of Schooling From a Multi-level Perspective*, ed. SW Raudenbush, JD Willms, pp. 131-48. San Diego: Academic

Moffitt, Robert. 1996. "Symposium on School Quality and Educational Outcomes: Introduction." *Review of Economics and Statistics* 128 (4): 559-561.

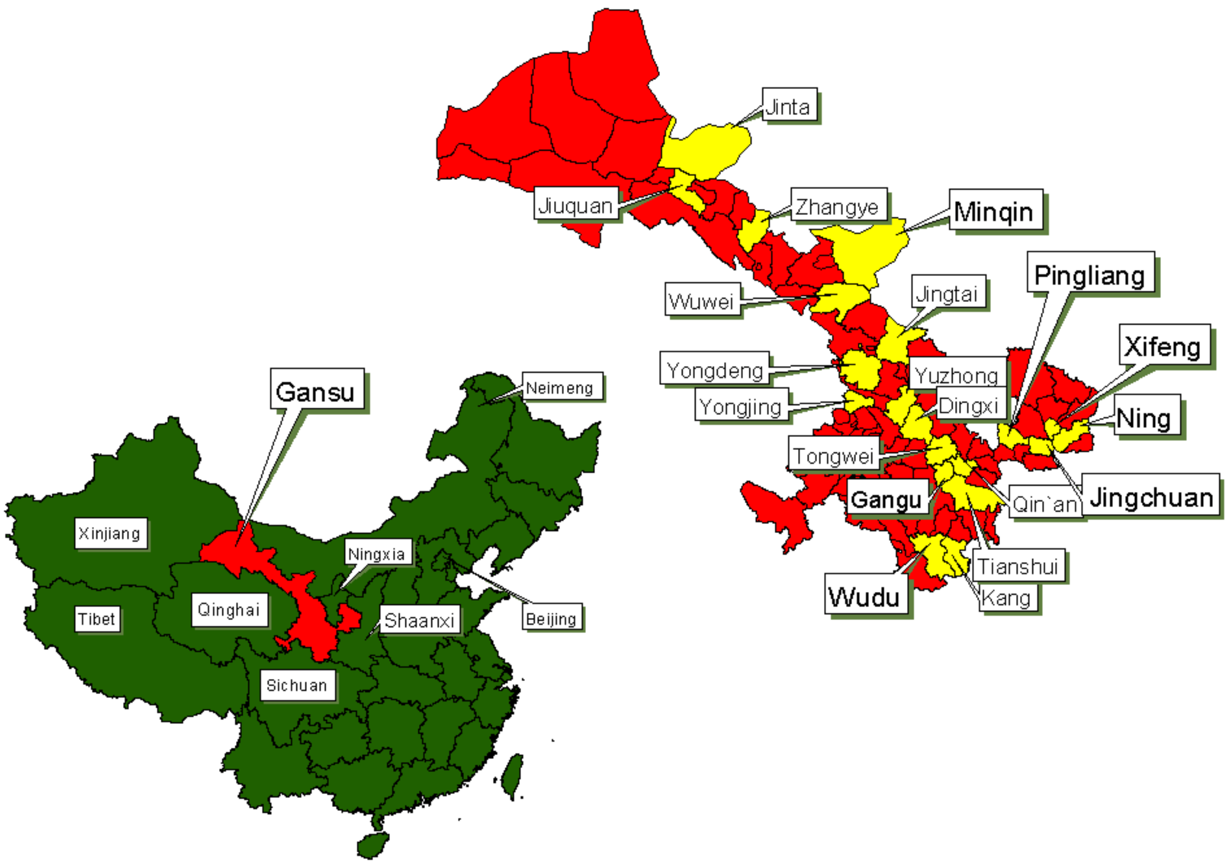
Murnane, Richard.

Peaker G. 1971. *The Plowden children four years later*. London: National Foundation for Educational Research in England and Wales

Saha LJ. 1983. Social structure and teacher effects on academic achievement: a comparative analysis. *Comp. Educ. Rev.* 27:69-88

Appendix
Sampling Method for the Gansu Survey of Children and Families

The sample design for the GSCF consisted of a primary sample of 2000 children in 20 rural counties aged 9-12 in July 2000; five linkable secondary samples of children's mothers, household heads, home-room teachers, school principals, and village leaders; and a linkable census of primary school teachers and school principals in sampled villages. The sample was drawn using a multi-stage, cluster design with random selection procedures employed at each stage. First, a systematic random sample of 20 counties was selected from the total of 86 counties in Gansu, ordered according to per capita income level in each county (see Map 1). Tibetan counties from which foreign access is restricted were excluded from the initial frame. The number of households selected from each county was determined according to the proportion of the rural population in each selected county. A random-start, systematic sample of two townships was then selected from the list of all townships for each county, and a random-start, systematic sample of five villages was selected from each sampled township (townships and villages were listed in "natural" or geographic order). Finally, a random sample of 20 children was selected from a listing of all 9-12 year old children in each selected village.



Map 1. Gansu Province, GSCF Counties Marked

Table 1
Number of Years Student has been with Current Teacher by Grade

	Years with Current Teacher (% students)					
Grade	1	2	3	4	5	N
1	100	0	0	0	0	156
2	39	62	0	0	0	361
3	47	21	32	0	0	567
4	43	32	13	12	0	474
5	45	29	10	4	13	284
6	47	29	14	4	6	51
All grades	48	31	15	4	2	1893

Table 2
Test Score Analysis of Variance

	N	Between school share of variance (S)	Within school, between teacher share of variance (T)	Unexplained share of variance (U)	Teacher share of explained variance (T/T+S)
Math					
Full sample	918	0.39	0.24	0.37	0.38
Yrs=1	461	0.42	0.27	0.31	0.39
Yrs>1	457	0.50	0.18	0.32	0.26
Language					
Full sample	974	0.38	0.23	0.39	0.38
Yrs=1	456	0.49	0.17	0.34	0.26
Yrs>1	518	0.46	0.20	0.34	0.30

Table 3
Teacher Characteristics By Grade

			Grade					
		All	1	2	3	4	5	6
High school education	TEDH	0.63	0.57	0.60	0.61	0.61	0.68	0.67
College education	TEDC	0.14	0.09	0.13	0.12	0.15	0.17	0.23
Quality grade 1	TQUAL1	0.46	0.36	0.40	0.42	0.47	0.52	0.65
Quality grade 2	TQUAL2	0.16	0.11	0.12	0.15	0.20	0.16	0.15
Experience (years)	TEXP	14.8	13.4	14.2	14.9	15.4	15.3	15.6
Male	TSEX	0.62	0.52	0.50	0.64	0.74	0.74	0.75
Native villager	TVILL	0.36	0.46	0.35	0.43	0.36	0.32	0.29
Monthly wage (yuan)	TWAGE	516	430	462	481	546	584	558
Govt. employee	TGOV	0.80	0.67	0.73	0.77	0.85	0.89	0.85
Other work (hours)	TWORK	5.61	6.89	5.94	4.79	6.36	5.23	6.53
N		1012	195	218	246	247	226	79

Table 4
Correlation Matrix: Teacher Characteristics

	TEDHC	TEDC	TQUAL1	TQUAL2	TSEX	TEXP	TVILL	TLWAGE	TGOV	TWORK
TEDHC	1.00									
TEDC	0.22	1.00								
TQUAL1	0.18	0.09	1.00							
TQUAL2	0.11	-0.02	0.34	1.00						
TSEX	0.05	-0.05	0.28	0.20	1.00					
TEXP	0.05	-0.11	0.60	0.50	0.42	1.00				
TVILL	-0.11	-0.14	-0.07	0.00	0.20	0.17	1.00			
TLWAGE	0.19	0.09	0.55	0.27	0.23	0.39	-0.21	1.00		
TGOV	0.11	0.04	0.53	0.21	0.23	0.37	-0.20	0.83	1.00	
TWORK	-0.02	-0.05	-0.07	0.01	0.19	0.05	0.23	-0.25	-0.25	1.00

Table 5
Child and Family Characteristics by Grade

	Variable		Grade					
Description	Name	All	1	2	3	4	5	6
Child sex (1=M, 0=F)	CSEX	0.54	0.48	0.50	0.53	0.58	0.57	0.55
Child age	CAGE	11.0	10.0	10.4	10.7	11.4	12.0	12.5
Father education (years)	FEDUC	6.99	5.35	6.18	6.83	7.69	7.88	8.25
Mother's education (years)	MEDUC	4.19	2.53	3.07	4.25	4.57	5.46	5.86
Expenditures p.c. (yuan)	EXPPC	2662	2207	2174	2721	2865	2933	3453
Ever held back (1/0)	HBACK	0.38	0.44	0.54	0.41	0.35	0.19	0.04
Age of enrollment	ENROL	7.39	8.37	7.76	7.34	7.14	7.00	6.68
	N	1893	156	361	567	474	284	51

Table 6A
Teacher Effects on Math Scores

	1		2		3		4	
	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
TQUAL1	**0.524	0.237			0.113	0.132	**0.513	0.238
TQUAL2	***1.069	0.365			0.138	0.222	***1.140	0.363
TEDH	-0.018	0.252	0.087	0.271	-0.161	0.124	-0.027	0.253
TEDC	**-0.768	0.344	*-0.644	0.359	-0.068	0.188	**-0.718	0.338
TEXPT	-0.0204	0.0127	-0.0001	0.0113	-0.0033	0.0068	-0.0204	0.0128
TLWAGE	**0.505	0.243	***0.638	0.243	0.050	0.148	**0.531	0.243
TSEX	-0.171	0.209	-0.158	0.213	-0.204	0.139	-0.129	0.194
TSEX*CSEX	*0.346	0.203	*0.355	0.205	0.189	0.155	0.229	0.162
TWORK	0.0158	0.0102	0.0161	0.0110	0.0019	0.0077	0.0141	0.0104
TGOV	-0.464	0.372	-0.394	0.396	0.198	0.247	-0.572	0.373
TVILL	-0.089	0.205	-0.086	0.210	***0.302	0.121	-0.064	0.206
YRS	*1.101	0.583	***1.554	0.583			**1.169	0.595
TQUAL1*YRS	**-0.224	0.109					**-0.229	0.110
TQUAL2*YRS	***-0.554	0.152					***-0.592	0.151
TEDH*YRS	-0.047	0.116	*-0.096	0.126			-0.036	0.118
TEDC*YRS	***0.544	0.177	**0.474	0.193			***0.511	0.174
TEXPT*YRS	*0.0107	0.0061	0.0003	0.0054			*0.0116	0.0061
TLWAGE*YRS	*-0.221	0.117	**-0.291	0.119			**-0.240	0.120
TSEX*YRS	-0.047	0.091	-0.034	0.092			-0.045	0.089
TSEX*CSEX*YRS	-0.082	0.081	-0.093	0.080			-0.080	0.081
TWORK*YRS	-0.0075	0.0051	-0.0070	0.0054			-0.0068	0.0052
TGOV*YRS	*0.367	0.199	*0.341	0.202			**0.407	0.202
TVILL*YRS	***0.226	0.090	**0.222	0.093			**0.208	0.091
CLSIZE	0.0016	0.0061	0.0022	0.0061	0.0003	0.0060	0.0023	0.0059
CSEX	-0.108	0.129	-0.093	0.130	-0.091	0.129		
FEDUC	*0.0210	0.0110	**0.0233	0.0110	**0.0225	0.0107		
MEDUC	0.0099	0.0104	0.0086	0.0105	0.0078	0.0104		
LEXPPC	-0.128	0.095	-0.146	0.093	-0.129	0.093		
N	884		884		884		884	
Adjusted R2	0.339		0.331		0.316		0.333	

Notes: 1) Dependent variable is standard deviations from grade-level mean test score; 2) All specifications include school dummies and interactive dummies for grade level, age of enrollment, and number of grade repetitions; 3) Standard errors adjusted for clustering

Table 6B
Teacher Effects on Language Scores

	1		2		3		4	
	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
TQUAL1	-0.284	0.249			0.007	0.121	-0.310	0.247
TQUAL2	0.201	0.362			***0.443	0.179	0.162	0.359
TEDH	0.004	0.281	0.136	0.281	0.061	0.138	-0.004	0.277
TEDC	0.040	0.388	0.071	0.411	0.033	0.163	0.021	0.386
TEXPT	-0.0214	0.0141	-0.0188	0.0120	-0.0059	0.0068	-0.0211	0.0139
TLWAGE	0.152	0.254	0.130	0.260	-0.120	0.138	0.159	0.253
TSEX	-0.194	0.274	-0.117	0.277	-0.087	0.145	-0.134	0.244
TSEX*CSEX	0.244	0.209	0.237	0.208	0.093	0.151	0.151	0.162
TWORK	0.0046	0.0126	0.0082	0.0132	-0.0010	0.0069	0.0048	0.0126
TGOV	-0.236	0.476	-0.370	0.480	0.083	0.248	-0.229	0.472
TVILL	*0.400	0.216	0.324	0.218	0.052	0.108	**0.427	0.218
YRS	0.632	0.639	0.471	0.634			0.649	0.635
TQUAL1*YRS	0.166	0.126					0.177	0.126
TQUAL2*YRS	0.132	0.164					0.150	0.164
TEDH*YRS	0.037	0.137	-0.002	0.143			0.032	0.136
TEDC*YRS	0.029	0.226	0.014	0.237			0.041	0.226
TEXPT*YRS	0.0080	0.0062	*0.0100	0.0055			0.0078	0.0061
TLWAGE*YRS	-0.149	0.136	-0.113	0.138			-0.151	0.135
TSEX*YRS	0.086	0.129	0.040	0.127			0.083	0.128
TSEX*CSEX*YRS	-0.090	0.083	-0.088	0.082			-0.086	0.082
TWORK*YRS	-0.0032	0.0060	-0.0033	0.0062			-0.0031	0.0059
TGOV*YRS	0.149	0.248	0.211	0.238			0.146	0.245
TVILL*YRS	*-0.183	0.101	-0.144	0.099			*-0.193	0.101
CLSIZE	0.007	0.006	0.005	0.006	0.005	0.006	0.007	0.006
CSEX	-0.074	0.129	-0.082	0.128	-0.099	0.126		
FEDUC	0.0069	0.0102	0.0050	0.0103	0.0091	0.0100		
MEDUC	-0.0016	0.0108	-0.0006	0.0110	-0.0014	0.0107		
LEXPPC	-0.073	0.066	-0.085	0.067	-0.072	0.064		
N	937		937		937		937	
Adjusted R2	0.292		0.284		0.291		0.294	

Notes: 1) Dependent variable is standard deviations from grade-level mean test score; 2) All specifications include school dummies and interactive dummies for grade level, age of enrollment, and number of grade repetitions; 3) Standard errors adjusted for clustering

Table 7
Ordered Probit Model for Years Taught

	1		2	
Variable	Coefficient	Standard Error	Coefficient	Standard Error
TQUAL1	0.026	0.171	-0.282	0.366
TQUAL2	-0.009	0.273	0.209	0.489
TLWAGE	0.199	0.211	0.537	0.476
TEXPT	*0.017	0.010	-0.007	0.024
TEXPS	**0.020	0.011	**0.050	0.024
TVILL	**0.308	0.157	0.021	0.326
TSEX	***-0.581	0.169	***-1.077	0.340
TGOV	**-1.070	0.362	*-1.566	0.940
TWORK	-0.010	0.009	*-0.031	0.018
GRADE3	***0.691	0.170	***1.182	0.336
GRADE4	***0.857	0.175	***1.116	0.349
GRADE5	***0.911	0.197	***1.940	0.396
GRADE6	**0.690	0.309	0.185	0.648
N	497		219	
Pseudo R2	0.2259		0.3703	

Notes: 1) Specification 1 includes all teachers in grades 2-6, Specification 2 includes all teachers in grades 2-6 who have taught in the school for at least 5 years; 2) Both models include school dummy variables; 3) *=significant at 10 percent level, **=5 percent level, and ***=1 percent level.

Table 8
Determinants of Quality Rank
Conditional Logit Model

	TQUAL1 (first level versus second level)				TQUAL2 (highest level versus first level)			
	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
TEDH	***0.99	0.32	***0.70	0.24	***1.75	0.54	***1.62	0.37
TEDC	***1.65	0.44	***1.69	0.33	***2.04	0.75	***2.07	0.51
TEXPT	***0.179	0.022	***0.187	0.018	***0.298	0.038	***0.226	0.023
TEXPS	-0.007	0.024	*-0.039	0.021	-0.016	0.023	0.017	0.014
TWORK	0.010	0.017	0.002	0.013	*0.040	0.023	0.012	0.016
TVILL	-0.42	0.30	*-0.43	0.25	-0.60	0.42	***-0.87	0.28
TSEX	***0.71	0.27	0.10	0.20	0.70	0.45	0.08	0.32
TGOV	***3.11	0.47	***2.43	0.29				
School fixed effects	yes		no		yes		no	
N	761		834		443		610	
Pseudo-R2	0.515		0.407		0.480		0.280	

Notes: 1) TQUAL models exclude teachers with highest quality rank, TQUAL2 models exclude teachers with quality rank lower than first level; 2) TGOV is excluded from TQUAL2 model because all highest level teachers are government employees; 3) *=significant at 10 percent level, **=5 percent level, and ***=1 percent level.