Discussion Papers In Economics And Business

The Effects of Student–Teacher Gender Matching on Students' Performance in Junior High Schools in Japan

Hisanobu Kakizawa

Discussion Paper 17-29

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

The Effects of Student–Teacher Gender Matching on Students' Performance in Junior High Schools in Japan

Hisanobu Kakizawa

Discussion Paper 17-29

September 2017

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

The Effects of Student–Teacher Gender Matching on Students' Performance in Junior High Schools in Japan*

Hisanobu Kakizawa†

September 28, 2017

ABSTRACT

This study analyzes the student–teacher gender matching effect on students' academic performance and questioning behavior. The results indicate as follows: 1. Positive effects of same gender teachers on students' performance are observed, especially for girls. 2. The gender-matching effect appears to be most significant in the study of English, followed by math and science. 3. Gender matching has an effect on students' questioning behavior. 4. Changes in questioning behavior may partly explain the improvement in performance. 5. Even when the effects of questioning behavior are controlled for, female teachers still have a positive effect on girls' performance.

JEL classification: I20, I21

Keywords: Academic performance; Gender-matching effect; Questioning behavior

* This research forms part of the results of the "Investigative Research Regarding the Effects of Education", a research project of the National Institute for Educational Policy Research (NIER). The author has received many excellent comments regarding the drafts of this paper from members of the project. The author would like to thank for these contributions here. Any remaining errors are the author's own.

† Center for Education in Liberal Arts and Sciences, Osaka University, 1-16 Machikaneyama, Toyonaka, Osaka 560-0043, Japan; e-mail: hkakizawa@celas.osaka-u.ac.jp

I. Introduction

Much research has been conducted on the effects of gender matching between students and teachers on the academic behavior and performance of students as well as their life choices. However, the results of previous research vary depending on subject and grade. Some studies have found that same-gender teachers have a positive effect on student academic performance, while other studies indicate a negative effect or no effect at all.

A prominent hypothesis that explains the mechanism by which gender has an effect on performance is the role model effect. According to this effect, the mere presence of a teacher serves as a role model for students, and functions to stimulate academic achievement. If this hypothesis is correct, then, as same-gender teachers are typically easier for students to identify with as role models, students taught by same-gender teachers should have increased performance in comparison to students taught by teachers of a different gender. Another oft-cited hypothesis is stereotype threat. Stereotype threat considers the effects of various social stereotypes that exist concerning gender and academic performance. Typically, this hypothesis deals with common ideas, like the notion that girls are good at the liberal arts and boys are good at science, technology, engineering, and math (STEM) subjects. Stereotypes pose a threat of being internalized by students and of leading to behavior of teachers that could affect students. Either way, if this hypothesis is correct, the behavior of students and teachers should tend to occur in a way that conforms to such stereotypes.

On the other hand, in the study of education, interactions between students and teachers in the classroom have been examined in detail. In particular, the questioning behavior of students is regarded as an important factor that influences grades and other measures of academic performance. Several previous works have already pointed out that gender matching between students and teachers has an effect on questioning behavior. Furthermore, it is possible that gender matching affects academic performance by causing changes in questioning behavior. This study attempts to analyze these points using micro data from a large survey conducted by the National Institute for Educational Policy Research (NIER). The subjects of this analysis are the three grades of Japanese middle school.

The structure of the rest of this paper is as follows. Previous research is detailed in Section II. In Section III, a summary of the data used in this study is presented. In Section IV, the effects of gender matching on test performance, which is a major component of academic performance, is estimated. In Section V, an analysis that considers questioning behavior is conducted. Section VI presents the conclusion.

II. Literature Review

A. Relationship between Gender Matching and Performance

The effects of gender matching between students and teachers on students' academic performance have been analyzed at many stages of education. In particular, regarding higher education at university level and beyond, there is a comparatively large body of research indicating that a role model effect exists between female students and female teachers. Nixon

and Robinson (1989) discovered a higher rate of majoring in STEM fields in university for girls who attended high schools with a large number of female STEM teachers, indicating the possibility of a role model effect. Rothstein (1995) showed that the percentage of female university students going on to graduate school increased with the ratio of female teachers within the university. However, this had no effect on wages earned upon entering the labor market later. Neumark and Gardecki (1998) analyzed the relationship between the number of female teachers and performance of female students in doctoral courses in economics, and showed that, although the presence of female teachers had no special effect on the rate at which students successfully found employment in research positions, it did have the effect of somewhat shortening the amount of time it took female students to finish graduate school. Hoffman and Orepoulos (2009) discovered that female teachers had a small effect on boosting female students' performance in university-level education. Bettinger and Long (2005) discovered a positive effect on the number of credits acquired by female students at university level in some subjects, including math, and a negative effect for other subjects. Carrell et al. (2010) discovered a matching effect between female students with a high level of mathematical proficiency at the time of enrollment and female teachers, but observed no such effect for male students. Conversely, Canes and Rosen (1995), after analyzing the relationship between the ratio of female students choosing STEM subjects at a number of universities and the ratio of female teachers for those subjects, concluded that no special relationship existed between the two. Price (2010) showed that the number of female teachers had a negative effect on the tendency of female university students to major in STEM subjects. Overall, although the relationship is not entirely clear, there is much research that indicates that matching between female students and female teachers at the level of university higher education and beyond has a positive effect on academic performance and life choices.

By comparison, the results vary for research on primary and secondary education. Dee (2007), who analyzed US middle schoolers, found that female teachers had a positive effect on the performance of girls in English and history. However, female teachers had a negative effect on both male and female students on performance in math. Muralidharan and Sheth (2016) performed an analysis of Indian elementary school students using large-scale panel data, and reported that female teachers had a positive effect only on the performance of girls. Parades (2014), in a study of middle school students in Chile, found that female teachers had the same positive effect on girls' performance, and concluded upon further analysis that this was not due to stereotype threat but rather the role model effect. Conversely, Steele (1997) and Spencer et al. (1999) conducted an analysis that connected girls' lower performance in math to stereotype threat. Lavy (2008) analyzed the effects of stereotype threat using data for Israeli high school students and, contrary to initial expectations, the results indicated that male students faced negative bias from teachers.

However, there is also a great deal of research that argues that female teachers have a negative effect or no effect at all on performance. Ehrenberg et al. (1995) indicated that,

although white, female teachers have no improvement effect on the performance of white, female students in STEM fields, such teachers tend to evaluate such students relatively positively when performing subjective evaluations. Beilcock et al. (2010) argued that, when women with math anxiety became primary school teachers, they had a negative effect on girls' performance in math. Furthermore, Antecol et al. (2015) discovered that, although female teachers had a negative effect on average on girls' test scores in math at the primary school level, female teachers who majored in math as students actually had a positive effect. This indicates the possibility that a lack of female teachers majoring in STEM fields produces an average negative effect. Holmlund and Sund (2008), who use data on Swedish high schoolers, and Cho (2012), who uses the Trends in International Mathematics and Science Study data from 15 countries, reported a negative result for the existence of a gender-matching effect.

B. Questioning Behavior

In this subsection, we shift our focus to previous research on the questioning behavior of students. Much research has already been undertaken on this topic, primarily in the field of pedagogy, which regards students' questioning behavior as an important factor that affects academic performance. Zoller (1987) argued that questioning behavior is an essential skill for solving problems. Rosenshine et al. (1996) performed a meta-analysis of reading comprehension using previous research, and calculated an average of 0.36 for the effect of questions on performance in research using standardized tests and of 0.86 in research using independent tests. Concerning the effect of questioning behavior in STEM fields, an increase in text comprehension in physics (Koch and Eckstein 1991) and the promotion of deeper autonomous thinking (King 1992) have been indicated. King and Rosenshine (1993) reported that students who had received training to produce thought-provoking questions performed better on average than students who did not receive training. Harper et al. (2003) indicated that the content of questions, more than their quantity, contributed to increased understanding of concepts in physics. Chin and Osborne (2010a, 2010b) argued that questioning behavior might raise the quality of group discussions in science classes. Chin and Osborne (2008) provided a survey of previous research regarding the effect of questioning behavior on learning in STEM fields.

In addition, much previous research indicates that the gender of the student and teacher has an effect on the questioning behavior of students, or on the mutual exchange between students and teachers, which includes questioning behavior. However, the various conclusions from these studies do not necessarily coincide. Sternglanz and Lyberger-Ficek (1977), in a survey targeting university students, found that male students were more assertive in mutual exchanges with instructors that occurred through asking questions. Brooks (1982) indicated that male university students were more assertive in asking questions when their professors were female. Bowers (1986), in a survey conducted at Iowa University, determined that students were more comfortable when their teachers were female. Pearson and West

(1991) discovered that, in universities, male students were more likely to ask questions in the classroom, while having a female teacher led to a small increase in the rate at which female students asked questions. Canada and Pringle (1995) analyzed cases in which women's universities transitioned to co-educational institutions, and discovered that not only did the gender of students and professors have an effect on the relationship between students and teachers, but so did the ratio of men and women within the class. Conversely, Good et al. (1987) reported a tendency for boys to ask questions assertively in childhood, but this difference narrowed gradually as students grew older. Crawford and MacLeod (1990), although observing a tendency for male university students to participate in class in an assertive manner, determined that the gender of the teacher had no significant effect on this tendency. Keeling et al. (2009), in a survey performed for high-level molecular biology courses in universities, concluded there was no difference between the two genders in the quality and number of questions asked. Blonder et al. (2015), concerning the number and content of questions asked in an inquiry chemistry laboratory, reported that either there was no difference between the two genders or female students tended to ask more questions of a higher quality than male students did. In other related research, Rocca (2010) offered a survey of the literature.

III. Data

The NIER is a research institute established under the Ministry of Education, Sport, Science, Education and Technology for the purpose of conducting investigative research related to education policy. This study used micro data from the "Survey on the Course of Study Implementation at Elementary Schools" conducted by the NIER. The subjects of this survey were students between elementary grade 5 and middle school grade 3 (aged 11-15 years) as well as the teachers that supervised these students. This national survey was conducted between January and February 2004.

The sample for this survey was chosen as follows. First, all elementary and middle schools in Japan were divided into four groups: Public schools in the 23 wards of Tokyo or in government designated cities (public metropolitan); public schools in cities (public urban); public schools in towns and villages (public rural); and national and private schools (national/private). Next, subject schools were randomly selected from each group. Furthermore, for each selected school, a subject class was randomly selected for each grade. The sample for this survey comprised all students in this class, as well as the teachers responsible for instructing this class in each subject. This study used data for the three middle school grades from this survey. Approximately 240,000 students selected from over 2,500 schools comprise the total sample, which includes every subject for these three grades.

The classes in the sample participated in a test and a questionnaire on the learning attitudes for five subjects: English, Japanese, math, science, and social studies. However, not all classes were surveyed for all subjects. Instead, three of five possible subjects were

assigned at random. In addition, for each subject, three types of problem booklets were prepared. Each class was given one of these problem booklets, also assigned at random, as an examination. These three types of problem booklets were produced so as to be of identical scope and difficulty. A questionnaire was distributed to the teachers responsible for these students in each subject, asking about fundamental personal attributes and teaching methods.

IV. Estimation of Gender-Matching Effects

First, the effect on students' performance of having female teachers was estimated for both girls and boys. A potential problem here is that correlation might exist between the unobservable individual effects for each student and the distribution of female teachers.

In compulsory education in Japan, students themselves do not determine their course subjects and teachers. Classes are determined by schools. Normally, a class set up in April does not change for at least a year. The teacher responsible for each subject visits the classroom to present a lesson. In addition, the assignment of these teachers is determined by the school. As a result, each teacher is assigned in a manner unrelated to the wishes of the students. This is similar to the case studies performed by Carrell and West (2010) and Carrel et al. (2010) concerning the United States Air Force Academy (USAFA). Under such a system, for matching between students and teachers, it is considered unlikely that self-selection by students themselves occurs.

However, there is no uniform system for how to consider student ability and individuality when determining classes, with such decisions being largely dependent on each school's policy. If students were non-randomly sorted in the determination of classes, and if the assignments of teachers to each class were biased based on gender, it would be possible for correlation to occur. In addition, as many schools are included in the data used in this study, it is possible that correlation exists between trends in student attributes and the distribution of female teachers at the school level. For instance, there are cases in which areas with strong average academic performance have a greater number of teachers of a particular gender. Accordingly, in this study, it is necessary to consider ways of dealing with such correlations when proceeding with the analysis.

Antecol et al. (2015) avoided this problem by using data obtained from a randomized experiment. Carrell and West (2010) used data on USAFA students from before they enrolled in the academy to confirm by means of a permutation test that matching between students and teachers was random. Neither of these methods is possible for the cross-sectional survey data used in this study. A further method to consider is first-difference estimation, used by Dee (2007) and Cho (2012). However, in this method, modeling that considers the endogeneity of questioning behavior becomes difficult. Therefore, in this study, we attempt to deal with this problem by using a correlated random-effects model.[1]

[1] See Mundlak (1978) and Wooldridge (2010, p.332).

A. Empirical Strategy

The five subjects are represented by s_j, with $j = 1, \cdots, 5$. Each student takes tests in three of these subjects. The three tests that student i takes are s_{ik} with $k = 1,2,3$. Here, we focus on one subject, s_j^*. To estimate the gender-matching effects in s_j^*, we consider the following model.

$$(1)\quad y_{ik} = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{x}_{Tik}'\boldsymbol{\beta}_T + \gamma_1 d_{s_j^*} + \gamma_2 d_{Gi} + \gamma_3 d_{Fik}$$
$$+\gamma_4 d_{s_j^*} d_{Gi} + \gamma_5 d_{Gi} d_{Fik} + \gamma_6 d_{s_j^*} d_{Fik} + \gamma_7 d_{s_j^*} d_{Gi} d_{Fik} + \nu_i + \varepsilon_{ik}$$
$$= \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{x}_{Tik}'\boldsymbol{\beta}_T + \boldsymbol{d}_{ik}'\boldsymbol{\gamma} + \nu_i + \varepsilon_{ik}$$

y_{ik} is the test score of student i in subject s_{ik}. $\boldsymbol{x}_i$ is the shared curriculum exogeneity variable vector, that is, variables that express the individual attributes of student i that are not the student's gender, as well as the attributes of the class and the school to which the student belongs. $\boldsymbol{x}_{Tik}$ denotes the attributes other than gender of the teacher who teaches subject s_{ik} to student i. $d_{s_j^*}$ is a dummy variable that expresses subject s_j^* and takes a value of 1 if $s_{ik} = s_j^*$, and otherwise 0. d_{Gi} is a female student dummy variable. d_{Fik} is a female teacher dummy that takes a value of 1 if the teacher teaching subject s_{ik} to student i is female, and 0 if the teacher is male. $\boldsymbol{d}_{ik}$ is a vector composed of these three variables and their cross-terms. ν_i is a random variable with mean 0 and variance σ_ν^2. It expresses the unobservable individual effect for student i on performance regardless of subject. ε_{ik} is an error term that satisfies the standard assumptions.

Furthermore, we consider the possibility that correlation exists between the female teacher dummy d_{Fik}, the variables for other teacher attributes $\boldsymbol{x}_{Tik}$, and the student individual effects variable ν_i. To express this correlation, the following linear relationship is assumed.

$$(2)\quad \nu_i = \bar{\boldsymbol{x}}_{Ti}'\boldsymbol{\delta}_T + \delta_F \bar{d}_{Fi} + \mu_i$$
$$= \boldsymbol{m}_i'\boldsymbol{\delta} + \mu_i$$

$\boldsymbol{m}_i = (\bar{\boldsymbol{x}}_{Ti}' \quad \bar{d}_{Fi})'$ is a vector made up of the averages of teacher attributes for the three subjects in which student i took tests. Substituting equation 2 into 1, the following model is obtained, which incorporates the correlation between these variables.[2]

[2] To perform this estimation, it is necessary for teacher attributes $\boldsymbol{x}_{Tik}$ and d_{Fik} to have variance for each i. In other words, each subject must have a different teacher. However, this condition is not satisfied, as in primary school, a single teacher typically teaches all subjects. As a result, the subjects of the analysis in this study are limited to middle schools.

(3) $y_{ik} = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{x}_{\boldsymbol{T}ik}'\boldsymbol{\beta}_T + \boldsymbol{d}_{ik}'\boldsymbol{\gamma} + \boldsymbol{m}_i'\boldsymbol{\delta} + \mu_i + \varepsilon_{ik}$

The versions of equation 3 for each subject for which student i took tests are as follows:

(4) $\boldsymbol{y_i} = \mathbf{x}_i\boldsymbol{\beta} + \mathbf{x}_{\boldsymbol{T}i}\boldsymbol{\beta}_T + \mathbf{d}_i\boldsymbol{\gamma} + \mathbf{m}_i\boldsymbol{\delta} + \mathbf{u}_i$
$\mathrm{E}(\mathbf{u}_i) = \mathbf{0}$
$\mathrm{E}(\mathbf{u}_i\mathbf{u}_i') = \sigma_\mu^2\boldsymbol{\iota}_3\boldsymbol{\iota}_3' + \sigma_\varepsilon^2\boldsymbol{I}_3$

$\boldsymbol{y_i} = (y_{i1} \quad y_{i2} \quad y_{i3})'$, $\mathbf{x}_i = \boldsymbol{\iota}_3\boldsymbol{x}_i'$, $\mathbf{x}_{\boldsymbol{T}i} = (\boldsymbol{x}_{\boldsymbol{T}i1} \quad \boldsymbol{x}_{\boldsymbol{T}i2} \quad \boldsymbol{x}_{\boldsymbol{T}i3})'$, $\mathbf{d}_i = (\boldsymbol{d}_{i1} \quad \boldsymbol{d}_{i2} \quad \boldsymbol{d}_{i3})'$, $\mathbf{m}_i = \boldsymbol{\iota}_3\boldsymbol{m}_i'$, and $\mathbf{u}_i = (\mu_i + \varepsilon_{i1} \quad \mu_i + \varepsilon_{i2} \quad \mu_i + \varepsilon_{i3})'$. $\boldsymbol{\iota}_3$ is a three-element unit vector, and $\boldsymbol{I}_3$ is a rank three identity matrix. Equation 4 is estimated for each grade by changing the subject of focus, s_j^*, in a sequential manner.

If male student i takes a test in subject s_j^* under the instruction of a female teacher, the values of the three dummy variables become $d_{s_j^*} = 1$, $d_{Gi} = 0$, and $d_{Fik} = 1$. In this case, the test score is expressed as $y_{ik(1,0,1)}$. Similarly, if this student is under the instruction of a male teacher, the test score is expressed as $y_{ik(1,0,0)}$. With the independent variable from equation 4 as the condition, if conditional independence is established between $(y_{ik(1,0,1)}, y_{ik(1,0,0)})$ and the female teacher dummy d_{Fik}, then the conditional average treatment effect (CATE) produced by the assignment to female teachers can be written in the following way using the estimates from equation 4.

$\tau_B := \hat{y}_{ik(1,0,1)} - \hat{y}_{ik(1,0,0)} = \hat{\gamma}_3 + \hat{\gamma}_6$

In other words, τ_B is a value obtained by adding the inherent female teacher effect ($\hat{\gamma}_6$) for subject s_j^* to the average effect ($\hat{\gamma}_3$) on boys for female teachers in the four subjects other than s_j^*.

Similarly, the CATE of female teachers on girls who took a test in subject s_j^* is as follows.

$\tau_G := \hat{y}_{ik(1,1,1)} - \hat{y}_{ik(1,1,0)} = \hat{\gamma}_3 + \hat{\gamma}_5 + \hat{\gamma}_6 + \hat{\gamma}_7$

Furthermore, the differences between the CATEs produced by boy and girl students are as follows.

$\tau_D := \tau_G - \tau_B = \hat{\gamma}_5 + \hat{\gamma}_7$

I obtain all these values.

B. Variables

The dependent variable in this estimation is the test scores. As described in section III, for the tests conducted in this study, three types of problem booklets of equal difficulty were used for each subject. The problem booklets were randomly distributed to each middle school. Therefore, the standardized scores for each test booklet is used as the dependent variable in this study.

The sample size and means by gender for standardized scores in each subject are shown in Table 1. Asterisks denote the results of t-tests performed on the differences in means. This table shows that girls outperform boys in English and Japanese for all three grades. The difference is particularly large in Japanese. However, boys perform better in science in all three years. Girls perform better in math in both grades 2 and 3. In social studies, boys do better in grade 1 and girls do better in grade 3, and thus, no consistent trend is observed.

Table 1
Sample Sizes and Means of Standardized Scores

	Grade 1			Grade 2			Grade 3		
	Girls	Boys		Girls	Boys		Girls	Boys	
English	23990	25759		23843	25628		21731	22998	
	0.143	-0.133	***	0.147	-0.137	***	0.122	-0.115	***
Japanese	24150	25981		23888	25462		21704	23208	
	0.183	-0.170	***	0.192	-0.180	***	0.226	-0.211	***
Math	24129	25641		23936	25445		21722	23141	
	-0.008	0.007	*	0.018	-0.017	***	0.012	-0.010	**
Science	24243	25692		24067	25413		17526	18332	
	-0.010	0.009	**	-0.025	0.024	***	-0.017	0.017	***
Social studies	24385	25782		24120	25626		21802	23425	
	-0.010	0.010	**	-0.007	0.007		0.088	-0.082	***

Note: The upper rows show the sample sizes. The lower rows show the mean standardized scores by gender. The asterisks indicate the results of t-tests of differences in means. * $p < .10$, ** $p < .05$, *** $p < .01$.

However, the independent variable that is the focus of this analysis is the teacher's gender. A dummy variable is used for this; it takes a value of 1 for female teachers and 0 for male teachers. The ratios of female teachers at every grade and in every subject are shown in Table 2. While female teachers comprise the majority in English and Japanese for all 3 years, male teachers make up around 80% of teachers in math and science, and the ratio of female teachers in social studies is even lower than in math and science.

Table 2
Ratios of Female Teachers (in percentages)

	English	Japanese	Math	Science	Social studies
Grade 1	59.83	61.88	27.21	23.14	15.36
Grade 2	61.94	59.09	27.58	21.09	15.96
Grade 3	56.18	59.67	20.77	19.30	16.48

Other independent variables are as follows. First, years of teaching experience is used as a non-gender teacher attribute. Next, the number of students and the ratio of female students are used as class attributes, and school classification is used as a school attribute. The number of students and the ratio of female students are both continuous variables. School classification indicates the division into the groups used when separating subject classes into "public metropolitan," "public urban," "public rural," and "national/private." Here, dummy variables created using "public metropolitan" as the reference category are used. Finally, a girl dummy is created as a variable expressing individual student attributes. In addition, four variables are used to express attitude towards life and study outside of school. These variables represent hours of daily sleep, whether a student eats breakfast daily, whether items the student brings to school are checked, and whether a cram school or home tutor is used in any subject. The descriptive statistics for these are shown in Table A1 in appendix.

C. Results

The CATEs obtained from the estimated values of equation 4 are compiled in Table 3 below. Asterisks in this table indicate the results of testing linear constraints with the null hypothesis that each value from equation 4 is zero (for all estimation results, see appendix Tables A2, A3, and A4).

First, to outline results by student gender, regarding girls, for three subjects in grade 1, all subjects in grade 2, and four subjects in grade 3, the estimated values of τ_G are significantly positive. In other words, regarding performance in these subjects, female teachers have a relatively positive effect in comparison to male teachers. Non-significant estimated values for other subjects are also all positive. However, regarding boys, for English in grade 1 and English and Japanese in grade 3, the estimated values of τ_B are significantly negative, meaning that male teachers had a relatively positive effect. The only case in which teachers of the opposite gender were more effective than those of the same gender was for boys in grade 3 science.

Next, we consider the results by subject. The effect of same-gender teachers was strongest through all three grades in English. The effect of female teachers on girls, τ_G, is significantly positive for all three grades. The average taken for all three grades is 0.0534, the largest of all five subjects. Conversely, the effect on boys τ_B was negative for all three

grades, and significant for grades 1 and 3. In other words, regarding boys' performance in English, male teachers tend to have a relatively positive effect. The average for all three grades for boys is -0.0202. These results for English are similar to those of Dee (2007). The estimates from that study showed an effect of 0.045 for girls and -0.047 for boys. By comparison, the effect from the results in this study is rather large for girls, and the absolute value of the effect for boys is rather small. For all three grades, the difference between the effects for male and female students, τ_D, is significant.

Table 3

Estimated CATEs of Female Teachers

		English		Japanese		Math		Science		Social studies	
Grade 1											
	τ_B	-0.0253	***	-0.0101		-0.0103		0.0067		-0.0191	
	τ_G	0.0559	***	0.0075		0.0359	***	0.0257	**	0.0170	
	τ_D	0.0812	***	0.0176		0.0462	***	0.0190		0.0361	**
Grade 2											
	τ_B	-0.0141		0.0011		0.0074		0.0170		0.0053	
	τ_G	0.0453	***	0.0273	***	0.0712	***	0.0359	***	0.0449	***
	τ_D	0.0594	***	0.0262	**	0.0639	***	0.0190		0.0397	**
Grade 3											
	τ_B	-0.0214	**	-0.0237	**	-0.0050		0.0633	***	-0.0039	
	τ_G	0.0589	***	0.0001		0.0457	***	0.0482	***	0.0317	***
	τ_D	0.0803	***	0.0238	*	0.0508	***	-0.0151		0.0356	**
Average											
	τ_B	-0.0202		-0.0109		-0.0026		0.0290		-0.0059	
	τ_G	0.0534		0.0116		0.0510		0.0366		0.0312	
	τ_D	0.0736		0.0225		0.0536		0.0076		0.0371	

Note: The asterisks indicate the results of chi-square tests of linear constraints in equation 4. * $p < .10$, ** $p < .05$, *** $p < .01$.

In math, the effect on girls of having a female teacher was, as with English, significantly positive for all three grades. The absolute value of the effect for grade 2 was 0.0712, the highest value for any grade or subject. The average over all three grades was 0.0510, slightly below the average for English. However, the effect on boys was not significant for any of the three grades. In other words, unlike for girls, for boys, gender matching between students and teachers did not have any particular effect on performance of math. The difference in effect between male and female students was significant.

In science, as well as math and English, the effect on girls of having a female teacher

was significantly positive for all three grades. However, the estimated values were comparatively small, with an average over all three grades of 0.0366. As stated earlier in this subsection, in grade 3 science class, female teachers had a significantly positive effect on boys. In grades 1 and 2 too, the effect was positive, although not at a significant level. As a result, the difference between male and female students was judged to not be significant for all three grades.

In social studies, for grades 2 and 3, the effect on girls of having a female teacher was significantly positive. The estimated value averaged over all three grades was 0.0312. However, as with math, there was no significant effect on boys for all three grades.

Although the difference between test scores for boys and girls was the largest in Japanese (Table 1), the effect of gender matching between students and teachers was not especially clear. The effect of female teachers on girls was significantly positive for only grade 2, and the effect on boys was significantly negative only for grade 3.

From the abovementioned effects, it is possible to observe that gender matching between students and teachers has some effect on students' performance in Japanese middle schools. As the results for boys and girls differ, it is unlikely that the problem is a difference in average ability between male and female teachers. For both girls and boys, there are subjects in which the instruction of a teacher of the same gender had a stronger effect on performance than the instruction of a teacher of the opposite gender. This tendency was particularly remarkable for girls. Concerning education in STEM subjects, which has often been the focus of previous research, this study confirms that, for all three grades in middle school, female teachers had some positive effect on girls' performance in both math and science. This result is shared by Paredes (2014), Muralidharan and Sheth (2016), and Carrell et al. (2010) with regard to university students, among others. Furthermore, it is clear that, in Japan, the gender-matching effect appears to be larger for English than for STEM subjects. These results for English are similar to those from Dee (2007). However, for Japanese, which is the native language of the sample in this study, although girls perform better than boys in all grades, having female teachers does not appear to have any clear effect.

V. Effects of Questioning Behavior

A. Empirical Strategy

Based on the analysis results presented in Section IV, we continue our analysis in this section by considering students' questioning behavior. The questionnaire distributed during the survey asked students what behavior they engaged in for each subject when they did not understand something during class. The possible answers to this question were: (i) Ask the teacher on the spot, (ii) Ask the teacher after class, (iii) Ask a friend, (iv) Ask a family member, (v) Ask a cram school teacher or home tutor, (vi) Investigate independently, and (vii) Do nothing. Students were allowed to choose multiple answers. In this section, we make estimates focusing in particular on answers (i) and (ii), which represent questions posed to

teachers responsible for instruction at the school.

As in Section IV, with regard to the test score of student i in subject s_{ik}, the following model is considered.

$$(6) \qquad q_{ik}^{*} = \boldsymbol{x}_i'\boldsymbol{\beta}_q + \boldsymbol{x}_{Tik}'\boldsymbol{\beta}_{Tq} + \boldsymbol{d}_{ik}'\boldsymbol{\gamma}_q + \boldsymbol{m}_i'\boldsymbol{\delta}_q + \boldsymbol{z}_i'\boldsymbol{\zeta}_q + \mu_{qi} + \varepsilon_{qik}$$

$$(7) \qquad y_{ik} = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{x}_{Tik}'\boldsymbol{\beta}_T + \boldsymbol{d}_{ik}'\boldsymbol{\gamma} + \boldsymbol{m}_i'\boldsymbol{\delta} + q_{ik}\boldsymbol{d}_{ik}'\boldsymbol{\zeta} + \mu_i + \varepsilon_{ik}$$

$$q_{ik} = \begin{cases} 1 & if \;\; q_{ik}^{*} > 0 \\ 0 & if \;\; q_{ik}^{*} \leq 0 \end{cases}$$

Here, q_{ik} is a dummy variable that takes a value of 1 if, of the choices listed above, (i), (ii), or both (i) and (ii) were selected, and 0 if neither of these was selected. In other words, this variable indicates whether students directly queried the teacher responsible for their instruction about subject matters they did not understand, either during or after class. Henceforth, this is referred to as the question dummy. q_{ik}^{*} is the latent variable for q_{ik}. $\boldsymbol{x}_i$, $\boldsymbol{x}_{Tik}$, $\boldsymbol{d}_{ik}$, and μ_i are defined as in equation 3.

μ_{qi} in equation 6 is the individual effect for student i regarding questioning behavior. In this equation, as in equation 3, a correlation between students' individual effects and teacher attributes is assumed with regard to questioning behavior. In other words, the tendency for a certain student to ask questions of a teacher in a certain subject should be shared and observed across subjects, as it is considered to be related to the student's personality or other individual characteristics. Furthermore, it is possible that a correlation exists between those individual characteristics and the distribution of teachers. For these reasons, vector $\boldsymbol{m}_i$, which is composed of the averages of teacher attributes, is included in equation 6.

ε_{ik} and ε_{qik} are error terms that satisfy the typical assumptions. $q_{ik}\boldsymbol{d}_{ik}$ in equation 7 is a cross-term vector between the question dummy and the other dummy variables $\boldsymbol{d}_{ik}$. $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \cdots, \zeta_7)'$ is a coefficient vector that corresponds to each element of $q_{ik}\boldsymbol{d}_{ik}$. $\boldsymbol{z}_i$ from equation 6 should correlate with q_{ik}, but not with ε_{ik}. This variable is discussed later in this subsection.

This study deals with the effect of gender matching between students and teachers on questioning behavior. Therefore, the endogenous variable q_{ik} is used as a binary variable, and estimates are performed using the following procedure. First, equation 6 is estimated as a probit model, and the effect of the teacher's gender on the probability that a student asks a direct question (the questioning probability) is checked. Next, from these estimation results, the following predicted probability is obtained.

$$\widehat{\Phi}_{ik} = \Phi\left(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_q + \boldsymbol{x}_{Tik}'\widehat{\boldsymbol{\beta}}_{Tq} + \boldsymbol{d}_{ik}'\hat{\boldsymbol{\gamma}}_q + \boldsymbol{m}_i'\widehat{\boldsymbol{\delta}}_q + \boldsymbol{z}_i'\hat{\boldsymbol{\zeta}}_q + \hat{\mu}_{qi}\right)$$

Here, Φ is the standard normal distribution function and $\hat{\mu}_{qi}$ is the best linear

unbiased estimator of individual effects. Finally, the instrumental variable estimation of equation 7 is performed using $\widehat{\Phi}_{ik}$ and $\widehat{\Phi}_{ik}\boldsymbol{d}_{ik}$ as instrumental variables.[3]

Although this estimation procedure itself is possible without variable $\boldsymbol{z}_i$, to avoid multiple collinearity between $\widehat{\Phi}_i$ and other variables, a variable that satisfies the conditions mentioned earlier should be appended. For this purpose, $\boldsymbol{z}_i$ is used; it is the answer to the question from the questionnaire targeting students that asks, "When you study, do your parents praise you?" Students had to respond on a four-point scale, ranging from "I do not think so" to "I think so." Using the response "I do not think so" as a reference, three dummy variables are created and added to the estimate in equation 6. In addition, the results of a test of over-identifying restrictions that considers q_{ik} as a continuous variable are shown in Tables A8–A10 in the appendix. J-statistics are sufficiently small in all grades and subjects, and thus, exogeneity is not rejected. Furthermore, F-statistics regarding weak instrumental variables show sufficient size for all five subjects in grade 3. However, F-statistics in grades 1 and 2 are rather small overall. In the data set for this study, no other variables with conditions better than these exist.

B. Gender-matching Effect on Questioning Behavior

First, the partial effects of the female teacher dummy on the questioning probability are calculated using the probit estimation of equation 6.

For simplicity of description, the linear combination of independent variables from equation 6 is again expressed as $\hat{\theta} = \boldsymbol{d}_{ik}'\hat{\boldsymbol{\gamma}}_q + \overline{\boldsymbol{X}}_q\widehat{\boldsymbol{B}}_q$. $\overline{\boldsymbol{X}}_q$ is a row vector that contains the averages of all independent variables except $\boldsymbol{d}_{ik}$, and $\widehat{\boldsymbol{B}}_q$ is a column vector made up of corresponding coefficient estimates for each independent variable. Using this notation, the partial effect due to female teachers, evaluated using the averages of independent variables other than $\boldsymbol{d}_{ik}$, is written as the following equations 8 and 9. First, for a boy taking a test in some subject s_j^*,

$$(8)\qquad \tau_{PB} := \left.\frac{\Delta\Phi(\hat{\theta})}{\Delta d_{Fik}}\right|_{d_{s_j^*}=1, d_{Gi}=0} = \Phi\left(\hat{\gamma}_{q1} + \hat{\gamma}_{q3} + \hat{\gamma}_{q6} + \overline{\boldsymbol{X}}_q\widehat{\boldsymbol{B}}_q\right) - \Phi\left(\hat{\gamma}_{q1} + \overline{\boldsymbol{X}}_q\widehat{\boldsymbol{B}}_q\right)$$
$$= \hat{P}_{BF} - \hat{P}_{BM}$$

Here, $\hat{P}_{BF} = \Phi\left(\hat{\gamma}_{q1} + \hat{\gamma}_{q3} + \hat{\gamma}_{q6} + \overline{\boldsymbol{X}}_q\widehat{\boldsymbol{B}}_q\right)$ denotes the probability that the boy will ask a female teacher a direct question, and $\hat{P}_{BM} = \Phi\left(\hat{\gamma}_{q1} + \overline{\boldsymbol{X}}_q\widehat{\boldsymbol{B}}_q\right)$ denotes the probability that the boy will ask a male teacher a direct question. Similarly, for a girl taking a test in some subject s_j^*,

$$(9)\qquad \tau_{PG} := \left.\frac{\Delta\Phi(\hat{\theta})}{\Delta d_{Fik}}\right|_{d_{s_j^*}=1, d_{Gi}=1} = \Phi\left(\sum_{k=1}^{7}\hat{\gamma}_{qk} + \overline{\boldsymbol{X}}_q\widehat{\boldsymbol{B}}_q\right) - \Phi\left(\hat{\gamma}_{q1} + \hat{\gamma}_{q2} + \hat{\gamma}_{q4} + \overline{\boldsymbol{X}}_q\widehat{\boldsymbol{B}}_q\right)$$

[3] See Wooldridge (2010, p.939).

$$= \hat{P}_{GF} - \hat{P}_{GM}$$

Here, $\hat{P}_{GF} = \Phi\left(\sum_{k=1}^{7} \hat{\gamma}_{qk} + \overline{\boldsymbol{X}}_q \widehat{\boldsymbol{B}}_q\right)$ denotes the probability that the girl will directly ask a female teacher a question, and $\hat{P}_{GM} = \Phi\left(\hat{\gamma}_{q1} + \hat{\gamma}_{q2} + \hat{\gamma}_{q4} + \overline{\boldsymbol{X}}_q \widehat{\boldsymbol{B}}_q\right)$ denotes the probability that the girl will directly ask a male teacher a question. The difference between equations 8 and 9 is

$$\tau_{PD} := \tau_{PG} - \tau_{PB}$$

The calculation results of these probabilities are compiled in Table 4. Asterisks in this table indicate the results of a z test performed using standard errors determined by the delta method (see appendix Tables A5, A6, and A7 for all estimation results from equation 6.

First, considering the effect that female teachers have on questioning probability ($\hat{P}_{BF}$ and $\hat{P}_{GF}$), these values are all significantly positive regardless of the student's gender. Furthermore, the partial effects of female teachers (τ_{PB} and τ_{PG}) are all positive and significant, except boys in grade 1 English. Considering the average effect for all three grades, the largest value is for girls in math (0.1295), followed by girls in science (0.0984), Japanese (0.0889), and English (0.0855). The next largest average effect is for boys in math, but at a grade-3 average of 0.0722, this value is almost the same as the smallest value for girls, in social studies (0.0715). In addition, values of τ_{PD}, which denotes the difference in effect between boys and girls, are all positive and significant, except grade 3 social studies and math.

In summary, in every subject, if the teacher responsible for instruction were female, students were relatively more likely to ask questions. This result is not inconsistent with indications from previous research on university students (Brooks 1982; Bowers 1986; Pearson and West 1991). In addition, this trend is more pronounced for girls than for boys. However, comparing $\hat{P}_{BF}$ and $\hat{P}_{GF}$ with $\hat{P}_{BM}$ and $\hat{P}_{GM}$, the difference between the former tends to be smaller than the difference between the latter. In the data used in this study, the effect appears to be primarily due to the fact that girls experience a large decline in questioning probability if their teacher is male, rather than because having a female teacher encourages students to ask questions.

C. CATE with Questioning Behavior Taken into Consideration

Next, the CATE for female teachers that considers questioning behavior is calculated by conducting an independent variable estimation of equation 7 using the procedure described in SubsectionV.A.

A male student i who took a test in subject s_j^* under the instruction of a female teacher and who also asks direct questions of his teacher will have values for four dummy variables as follows: $q_{ik} = 1,\ d_{s_j^*} = 1,\ d_{Gi} = 0$, and $d_{Fik} = 1$. In this case, test scores are

Table 4

Estimated Questioning Probabilities and Female Teachers' Partial Effects

	English		Japanese		Math		Science		Social studies	
Grade 1										
$\hat{P}_{BF}$	0.2046	***	0.1829	***	0.3057	***	0.2214	***	0.1851	***
$\hat{P}_{BM}$	0.2025	***	0.1384	***	0.2385	***	0.1890	***	0.1445	***
τ_{PB}	0.0021		0.0444	***	0.0672	***	0.0324	***	0.0406	***
$\hat{P}_{GF}$	0.1841	***	0.1374	***	0.2998	***	0.1517	***	0.1069	***
$\hat{P}_{GM}$	0.1197	***	0.0747	***	0.1820	***	0.0799	***	0.0440	***
τ_{PG}	0.0644	***	0.0627	***	0.1177	***	0.0718	***	0.0629	***
τ_{PD}	0.0623	***	0.0183	**	0.0505	***	0.0394	***	0.0223	*
Grade 2										
$\hat{P}_{BF}$	0.2034	***	0.1926	***	0.3028	***	0.2541	***	0.1902	***
$\hat{P}_{BM}$	0.1860	***	0.1291	***	0.2617	***	0.2195	***	0.1339	***
τ_{PB}	0.0174	**	0.0635	***	0.0411	***	0.0346	***	0.0563	***
$\hat{P}_{GF}$	0.2064	***	0.1803	***	0.3572	***	0.2221	***	0.1211	***
$\hat{P}_{GM}$	0.1216	***	0.0926	***	0.2120	***	0.1241	***	0.0459	***
τ_{PG}	0.0847	***	0.0877	***	0.1452	***	0.0980	***	0.0752	***
τ_{PD}	0.0674	***	0.0241	***	0.1041	***	0.0634	***	0.0189	
Grade 3										
$\hat{P}_{BF}$	0.2278	***	0.1970	***	0.3954	***	0.2892	***	0.1869	***
$\hat{P}_{BM}$	0.1929	***	0.1289	***	0.2871	***	0.2532	***	0.1261	***
τ_{PB}	0.0349	***	0.0681	***	0.1084	***	0.0360	***	0.0608	***
$\hat{P}_{GF}$	0.2862	***	0.2449	***	0.4671	***	0.3340	***	0.1514	***
$\hat{P}_{GM}$	0.1788	***	0.1286	***	0.3414	***	0.2086	***	0.0750	***
τ_{PG}	0.1073	***	0.1163	***	0.1257	***	0.1254	***	0.0765	***
τ_{PD}	0.0725	***	0.0482	***	0.0173		0.0894	***	0.0157	
Average										
$\hat{P}_{BF}$	0.2119		0.1908		0.3347		0.2549		0.1874	
$\hat{P}_{BM}$	0.1938		0.1321		0.2624		0.2205		0.1348	
τ_{PB}	0.0181		0.0587		0.0722		0.0343		0.0525	
$\hat{P}_{GF}$	0.2255		0.1875		0.3747		0.2359		0.1265	
$\hat{P}_{GM}$	0.1401		0.0986		0.2451		0.1375		0.0549	
τ_{PG}	0.0855		0.0889		0.1295		0.0984		0.0715	
τ_{PD}	0.0674		0.0302		0.0573		0.0641		0.0190	

Note: The asterisks indicate the results of z-tests based on delta-method standard errors.
* p < .10, ** p < .05, *** p < .01.

expressed as $y_{ik(1,1,0,1)}$. Similarly, when direct questions are not asked, test scores are written as $y_{ik(0,1,0,1)}$. Then, the expected value of the performance of the boy receiving instruction from a female teacher, as the probability-weighted sum of these scores based on questioning probability, is expressed as follows.

$$\begin{aligned}\hat{P}_{BF}\hat{y}_{ik(1,1,0,1)} + \left(1-\hat{P}_{BF}\right)\hat{y}_{ik(0,1,0,1)} &= \hat{y}_{ik(0,1,0,1)} + \hat{P}_{BF}\left(\hat{y}_{ik(1,1,0,1)} - \hat{y}_{ik(0,1,0,1)}\right)\\ &= \hat{y}_{ik(0,1,0,1)} + \hat{P}_{BF}\left(\hat{\zeta}_1 + \hat{\zeta}_3 + \hat{\zeta}_6\right)\\ &= \hat{y}_{ik(0,1,0,1)} + \hat{P}_{BF}\widehat{M}_{BF}\end{aligned}$$

Here $\widehat{M}_{BF} = \hat{\zeta}_1 + \hat{\zeta}_3 + \hat{\zeta}_6$ is the marginal effect of the question dummy. In other words, it is the increment in marginal test score that the male student obtains by asking questions of a female teacher.

Similarly, the performance of this student if he receives instruction from a male teacher is expressed as $y_{ik(1,1,0,0)}$ and $y_{ik(0,1,0,0)}$. Then, the expected value is

$$\begin{aligned}\hat{P}_{BM}\hat{y}_{ik(1,1,0,0)} + \left(1-\hat{P}_{BM}\right)\hat{y}_{ik(0,1,0,0)} &= \hat{y}_{ik(0,1,0,0)} + \hat{P}_{BM}\left(\hat{y}_{ik(1,1,0,0)} - \hat{y}_{ik(0,1,0,0)}\right)\\ &= \hat{y}_{ik(0,1,0,0)} + \hat{P}_{BM}\hat{\zeta}_1\\ &= \hat{y}_{ik(0,1,0,0)} + \hat{P}_{BM}\widehat{M}_{BM}\end{aligned}$$

$M_{BM} = \hat{\zeta}_1$ is the increment in marginal test score obtained by asking direct questions of a male teacher. The CATE of having a female teacher for boys is expressed as the difference between these:

$$\begin{aligned}(10)\quad \tau_{QB} &:= \hat{y}_{ik(0,1,0,1)} + \hat{P}_{BF}M_{BF} - \left[\hat{y}_{ik(0,1,0,0)} + \hat{P}_{BM}\widehat{M}_{BM}\right]\\ &= \left[\hat{y}_{ik(0,1,0,1)} - \hat{y}_{ik(0,1,0,0)}\right] + \left[\hat{P}_{BF}\widehat{M}_{BF} - \hat{P}_{BM}\widehat{M}_{BM}\right]\\ &= \left(\hat{\gamma}_3 + \hat{\gamma}_6\right) + \left[\hat{P}_{BF}\widehat{M}_{BF} - \hat{P}_{BM}\widehat{M}_{BM}\right]\\ &= \tau_{Q_0B} + \tau_{Q_1B}\end{aligned}$$

Here $\tau_{Q_0B} = \hat{\gamma}_3 + \hat{\gamma}_6$ expresses the difference in effects between male and female teachers if questions are not asked. Likewise, $\tau_{Q_1B} = \hat{P}_{BF}\widehat{M}_{BF} - \hat{P}_{BM}\widehat{M}_{BM}$ expresses the difference in effects between male and female teachers if questions are asked. These are totaled to obtain the CATE for female teachers, τ_{QB}.

The CATE of female teachers for girls is denoted in a similar fashion:

$$\begin{aligned}(11)\quad \tau_{QG} &:= \hat{y}_{ik(0,1,1,1)} + \hat{P}_{GF}\sum_{k=1}^{7}\hat{\zeta}_k - \left[\hat{y}_{ik(0,1,1,0)} + \hat{P}_{GM}\left(\hat{\zeta}_1 + \hat{\zeta}_2 + \hat{\zeta}_4\right)\right]\\ &= \left[\hat{y}_{ik(0,1,1,1)} - \hat{y}_{ik(0,1,1,0)}\right] + \left[\hat{P}_{GF}\sum_{k=1}^{7}\hat{\zeta}_k - \hat{P}_{GM}\left(\hat{\zeta}_1 + \hat{\zeta}_2 + \hat{\zeta}_4\right)\right]\\ &= \left(\hat{\gamma}_3 + \hat{\gamma}_5 + \hat{\gamma}_6 + \hat{\gamma}_7\right) + \left(\hat{P}_{GF}\widehat{M}_{GF} - \hat{P}_{GM}\widehat{M}_{GM}\right)\end{aligned}$$

$$= \tau_{Q_0G} + \tau_{Q_1G}$$

$\tau_{Q_0G} = \hat{\gamma}_3 + \hat{\gamma}_5 + \hat{\gamma}_6 + \hat{\gamma}_7$ denotes the difference in effects between male and female teachers if questions are not asked. $\widehat{M}_{GF} = \sum_{k=1}^{7} \hat{\zeta}_k$ and $\widehat{M}_{GM} = \hat{\zeta}_1 + \hat{\zeta}_2 + \hat{\zeta}_4$ are the increments in marginal test score obtained by asking direct questions of female and male teachers, respectively. $\tau_{Q_1G} = \hat{P}_{GF}\widehat{M}_{GF} - \hat{P}_{GM}\widehat{M}_{GM}$ is the difference between the expected value of effects when questions are asked of male and female teachers.

D. Results

Of the elements in equations 10 and 11, questioning probability is examined in the previous subsection. Next, we review the marginal increment in test scores obtained by asking questions. In Table 5 below, each estimated value and difference between male and female teachers, that is, $\widehat{M}_{BD} = \widehat{M}_{BF} - \widehat{M}_{BM}$ and $\widehat{M}_{GD} = \widehat{M}_{GF} - \widehat{M}_{GM}$, is shown. Asterisks in this table indicate the results of testing linear constraints with the null hypothesis that each value from equation 7 is zero (see appendix Tables A8, A9, and A10 for a complete set of estimation results from equation 7).

First, comparing male and female students, in general, increments in girls' test scores tend to be greater than increments in boys' test scores. The question dummy used here expresses whether a direct question is asked of the teacher when a student does not understand something. The specific frequency and content of the question or questions are not controlled. Several previous studies have pointed out the importance of frequency and content (King and Rosenshine, 1993; Harper et al., 2003). There might be differences in tendencies between boys and girls when questioning behavior is considered from such a standpoint.

It is particularly important at this point to confirm the presence or absence of difference due to the gender of the teacher. First, in reviewing $\widehat{M}_{BD}$, which concerns boys, although values for grade 1 social studies and grade 3 English are significantly negative at the 10% level, no other values are significant. In reviewing $\widehat{M}_{GD}$, which concerns girls, significant values at the 5% level are seen in grade 1 social studies, grade 2 English, and grade 3 Japanese. All these values are negative, indicating that increments are larger when girls have a male teacher. However, the other 12 values are not significant. Generally, concerning the effect of questions on performance itself, it can be said that there is no clear difference due to teacher gender.

Based on this, we confirm the CATE for female teachers taking questioning behavior into consideration. Calculation results for equations 8 and 9 are compiled in Table 6 below. Asterisks indicate the significance of each value. For τ_{Q_0B} and τ_{Q_0G}, the results are from testing linear constraints with the null hypothesis that each value from equation 7 is zero. For all other values, the results are from a z test using a standard error obtained through the delta method.

First, we review the results for male and female students. The total effect on boys of

Table 5
Estimated Marginal Increments of Scores by Questioning Behavior

	English		Japanese		Math		Science		Social studies	
Grade 1										
$\widehat{M}_{BF}$	0.1065	***	0.0365	**	0.0928	***	0.0926	***	0.0379	
$\widehat{M}_{BM}$	0.0743	***	0.0390	**	0.0868	***	0.0965	***	0.0939	***
$\widehat{M}_{BD}$	0.0322		-0.0025		0.0060		-0.0039		-0.0560	*
$\widehat{M}_{GF}$	0.1835	***	0.1486	***	0.1931	***	0.1671	***	0.1347	***
$\widehat{M}_{GM}$	0.1927	***	0.1184	***	0.1734	***	0.1957	***	0.2139	***
$\widehat{M}_{GD}$	-0.0092		0.0302		0.0197		-0.0286		-0.0792	**
Grade 2										
$\widehat{M}_{BF}$	0.1366	***	0.0381	**	0.1160	***	0.1045	***	0.0618	*
$\widehat{M}_{BM}$	0.1482	***	0.0614	***	0.1128	***	0.0996	***	0.1043	***
$\widehat{M}_{BD}$	-0.0116		-0.0233		0.0031		0.0049		-0.0425	
$\widehat{M}_{GF}$	0.2558	***	0.1813	***	0.2373	***	0.2208	***	0.2558	***
$\widehat{M}_{GM}$	0.3267	***	0.1914	***	0.2335	***	0.2552	***	0.2786	***
$\widehat{M}_{GD}$	-0.0709	***	-0.0102		0.0038		-0.0344		-0.0228	
Grade 3										
$\widehat{M}_{BF}$	0.1172	***	0.0565	***	0.1312	***	0.0543	*	0.1431	***
$\widehat{M}_{BM}$	0.1617	***	0.0280		0.1517	***	0.0838	***	0.1206	***
$\widehat{M}_{BD}$	-0.0444	*	0.0285		-0.0204		-0.0295		0.0224	
$\widehat{M}_{GF}$	0.3564	***	0.1963	***	0.3011	***	0.2272	***	0.3280	***
$\widehat{M}_{GM}$	0.3699	***	0.2550	***	0.3355	***	0.2867	***	0.3031	***
$\widehat{M}_{GD}$	-0.0135		-0.0587	**	-0.0343		-0.0595	*	0.0249	
Average										
$\widehat{M}_{BF}$	0.1201		0.0437		0.1133		0.0838		0.0809	
$\widehat{M}_{BM}$	0.1280		0.0428		0.1171		0.0933		0.1063	
$\widehat{M}_{BD}$	-0.0079		0.0009		-0.0038		-0.0095		-0.0253	
$\widehat{M}_{GF}$	0.2653		0.1754		0.2438		0.2050		0.2395	
$\widehat{M}_{GM}$	0.2964		0.1883		0.2474		0.2459		0.2652	
$\widehat{M}_{GD}$	-0.0312		-0.0129		-0.0036		-0.0408		-0.0257	

Note: The asterisks indicate the results of chi-square tests of linear constraints in equation (7).
* p < .10, ** p < .05, *** p < .01.

having a female teacher, τ_{QB}, is significantly negative for grades 1 and 2 English and grade 3 Japanese. In other words, in these grades and subjects, having a male teacher has a relatively positive effect. Most other values are not significant, with only grade 3 science being

Table 6
Estimated CATEs of Female Teachers Considering Questioning Behavior

	English		Japanese		Math		Science		Social studies	
Grade 1										
τ_{Q_0B}	-0.0380	***	-0.0035		-0.0170		0.0051		0.0100	
τ_{Q_1B}	0.0067		0.0013		0.0077		0.0023		-0.0066	
τ_{QB}	-0.0312	***	-0.0022		-0.0093		0.0074		0.0034	
τ_{Q_0G}	0.0491	***	-0.0038		0.0113		0.0196		0.0198	
τ_{Q_1G}	0.0107	***	0.0116	***	0.0263	***	0.0097	**	0.0050	
τ_{QG}	0.0598	***	0.0077		0.0376	***	0.0293	**	0.0248	*
Grade 2										
τ_{Q_0B}	-0.0170		0.0024		0.0018		0.0139		0.0069	
τ_{Q_1B}	0.0002		-0.0006		0.0056		0.0047		-0.0022	
τ_{QB}	-0.0168	*	0.0018		0.0074		0.0186		0.0047	
τ_{Q_0G}	0.0459	***	0.0118		0.0450	***	0.0275	**	0.0276	*
τ_{Q_1G}	0.0130	***	0.0149	***	0.0353	***	0.0174	***	0.0182	***
τ_{QG}	0.0590	***	0.0268	***	0.0803	***	0.0449	***	0.0458	***
Grade 3										
τ_{Q_0B}	-0.0112		-0.0342	**	-0.0012		0.0712	***	-0.0170	
τ_{Q_1B}	-0.0045		0.0075		0.0084		-0.0055		0.0115	*
τ_{QB}	-0.0157		-0.0266	**	0.0071		0.0657	***	-0.0055	
τ_{Q_0G}	0.0428	***	0.0004		0.0374	**	0.0561	***	-0.0041	
τ_{Q_1G}	0.0358	***	0.0153	***	0.0261	**	0.0161		0.0270	***
τ_{QG}	0.0786	***	0.0157		0.0636	***	0.0722	***	0.0229	*
Average										
τ_{Q_0B}	-0.0221		-0.0118		-0.0055		0.0301		-0.0001	
τ_{Q_1B}	0.0008		0.0027		0.0072		0.0005		0.0009	
τ_{QB}	-0.0212		-0.0090		0.0017		0.0306		0.0009	
τ_{Q_0G}	0.0459		0.0028		0.0313		0.0344		0.0144	
τ_{Q_1G}	0.0199		0.0139		0.0292		0.0144		0.0167	
τ_{QG}	0.0658		0.0167		0.0605		0.0488		0.0311	

Note: The asterisks for τ_{Q_0B} and τ_{Q_0G} indicate the results of chi-square tests of linear constraints in equation (7). The asterisks for other values indicate the results of z-tests based on delta-method standard errors. * p < .10, ** p < .05, *** p < .01.

significantly positive. This is largely similar to the results from Table 3. However, the total effect of female teachers on girls, τ_{QG}, is significantly positive for four subjects in grade 1, all subjects in grade 2, and four subjects in grade 3. This, too, is similar to the results from Table

3.

Next, we review the effects of questions. Here the difference between male and female students is striking. First, for boys, τ_{Q_1B} is significant at the 10% level for only grade 3 social studies. All other values are not significant. In comparison, for girls, τ_{Q_1G} is significantly positive for four subjects in grade 1, all subjects in grade 2, and four subjects in grade 3. In other words, while there is almost no difference in the size of the effect of asking questions of the teacher arising from teacher gender for boys, there is a clear difference for girls. As confirmed earlier in this subsection, in addition to having a comparatively large marginal increment for girls' questions, the questioning probability for girls shows a sharp decline for male teachers. This is connected to the overall effect.

Next, we review effects when questions are not asked. τ_{Q_0B}, which deals with boys, has a significantly negative value for grade 1 English and grade 3 Japanese. On the contrary, the value for grade 3 science is significantly positive. All other values are not significant. In comparison, with regard to τ_{Q_1G}, which deals with girls, significantly positive values are present for grade 1 English and grades 2 and 3 English, math, and science. Furthermore, the value for grade 2 social studies is significantly positive at the 10% level. In other words, the effect of having same-gender teachers in these grades and subjects is due in part to causes that cannot be attributed to the presence or absence of questioning behavior.

Considering the effect of female teachers on girls by subject, the most clear effect appears in English. Through all three grades, the effects of female teachers on girls τ_{QG}, τ_{Q_0G}, and τ_{Q_1G}, are all significantly positive. In other words, an effect produced by asking questions and a female teacher effect that cannot be attributed to questioning behavior are both observed. Calculating the portion of the total effect made up by the question effect, τ_{Q_1G}/τ_{QG}, yields 17.9% for grade 1, 22.1% for grade 2, and 45.6% for grade 3.

In addition, in STEM subjects, a comparatively clear effect is obtained. For math, τ_{QG} and τ_{Q_1G} are significantly positive for all grades, and τ_{Q_0G} is significantly positive in grades 2 and 3. Calculating τ_{Q_1G}/τ_{QG} for grades 2 and 3 yields proportions of 43.9% and 41.1%, respectively. For science, τ_{QG} is significantly positive for all grades, while τ_{Q_1G} is significantly positive for grades 1 and 2, and τ_{Q_0G} is significantly positive for grades 2 and 3. Calculating τ_{Q_1G}/τ_{QG} for grade 2, when all three values are significant, this yields a figure of 38.8%.

Concerning Japanese and social studies, the question effect τ_{Q_1G} is significantly positive for all cases, except grade 1 social studies. Although τ_{Q_0G} is significantly positive at the 10% level for grade 2 social studies, it is not significant for all other cases. In other words, the CATE for female teachers in these two subjects is determined by the difference in ease of asking questions between male and female teachers.

VI. Conclusion

In this research, for all grades in Japanese middle schools, the effect of gender

matching between students and teachers on academic performance was empirically analyzed using individual data from the "Survey on the Course of Study Implementation at Elementary Schools."

The main conclusions are as follows. First, gender matching between students and teachers has an effect on student performance. In subjects for both male and female students, when taught by a same-gender teacher, some improvement in performance occurs. This is particularly striking for girls. For boys, of all five subjects over all three grades, a positive effect was observed for three subjects due to a same-gender teacher. By contrast, the same was for girls for 12 subjects.

Second, the effect of gender matching between students and teachers differs depending on subject. The clearest trend was observed in English, in which matching between female teachers for girls and male teachers for boys improved the respective performance of each gender. In addition, concerning STEM subjects, on which previous research has frequently focused, although matching between girls and female teachers improved performance in math and science through all grades, no effect was observed consistently for matching between boys and male teachers.

Third, gender matching between students and teachers had an effect on students' questioning behavior. The probability that students would ask questions was lower when their teacher was male in comparison to when their teacher was female. This trend was observed regardless of student gender. However, through all grades and subjects, the decrease in the probability of asking a question under a male teacher was larger for girls than for boys. However, because a trend of lower questioning probability was also observed for boys, this result should be considered an indication of a reaction to the teacher's gender on the part of the students, rather than an indication of the existence of bias in the behavior of the teachers.

Fourth, a part of the positive effect that female teachers have on girls' performance might be caused by a difference in questioning behavior. This kind of question effect was hardly seen for boys. This effect is due to the fact that the decrease in the probability of asking a question for girls under the instruction of male teachers is comparatively large, and the marginal increment in girls' performance from asking questions was greater than for boys. This result indicates the importance of questioning behavior in learning for female students. In particular, it is possible that male teachers improve the performance of female students by encouraging them to ask further questions.

Fifth, even controlling for the effect of questions, a positive effect of same-gender teachers that cannot be attributed to questioning behavior remains, particularly for girls in English, math, and science. In other words, this effect is observed for both liberal arts and STEM subjects. The percentage of the total effect made up by this non-questioning effect is around 60% in math, science, and grade 3 English and around 80% in grades 1 and 2 English. Conversely, in Japanese, in which girls outperform boys by the greatest margin, this effect was not observed. In addition, this positive effect was hardly seen for boys, confirmed in only

grade 1 English and grade 3 Japanese. It is difficult to think of a stereotype that fits all these results. However, as there was a positive effect for same-gender teachers for both male and female students in almost all remaining subjects, excluding grade 3 science for boys, the results obtained here are not inconsistent with the role model hypothesis. However, naturally, the possibility of some separate, unconsidered effect cannot be excluded.

Finally, I touch on issues that the analysis presented herein did not address. The analysis in this study pooled cross-sectional data from five subjects and performed estimates that controlled individual effects shared across subjects. However, it was not possible to control individual effects inherent in each subject. For example, if these effects had been controlled through the use of panel data, it is possible that different results would have been obtained. Furthermore, in this study, questioning behavior was treated as an endogenous variable, and individual effects common to subjects were addressed. However, as the question dummy was included unilaterally in the estimates of test scores, it was not possible to deal with the possibility that performance and questioning behavior were determined simultaneously. Therefore, there is room for improvement on this point.

References

Antecol, Heather, Ozkan Eren, and Serkan Osbeklik 2015. "The effect of teacher gender on student achievement in primary school: evidence from a randomized experiment." *Journal of Labor Economics* 33(1): 63-89.

Beilcock, Sian. L., Elizabeth A. Gunderson, Gerardo Ramirez, and Suzan C. Levine 2010. "Female teacher's math anxiety affects girls' math achievement." *Proceedings of the National Academy of Sciences* 107(5): 1060-1063.

Bettinger, Eric P. and Bridget Terry Long 2005. "Do faculty serve as role models? The impact of instructor gender on female students." *American Economic Review* 95: 152-157.

Blonder, Ron, Shelley Rap, Rachel Mamlok-Naaman, and Avi Hofstein 2015. "Questioning behavior of students in the inquiry chemistry: differences between sectors and genders in the Israeli context." *International Journal of Science and Mathematics Education* 13: 705-732.

Bowers, John W. 1986. "Classroom communication apprehension: a survey." *Communication Education* 35(4): 372-378.

Brooks, Virginia R. 1982. "Sex differences in student dominance behavior in female and male professors' classrooms." *Sex Roles* 8(7): 683-690.

Canada, Katherine and Richard Pringle 1995. "The role of gender in college classroom interactions: a social context approach." *Sociology of Education* 68: 161-186.

Canes, Brandice J. and Harvey S. Rosen 1995. "Following in her footsteps? Faculty gender composition and women's choice of college majors." *Industrial and Labor Relations Review* 48: 486-504.

Carrell, Scott E., Marianne E. Page, and James E. West 2010. "Sex and science: how professor gender perpetuates the gender gap." *Quarterly Journal of Economics* 125(3): 1101-1114.

Carrell, Scott E. and James E. West 2010. "Does professor quality matter? Evidence from random assignment of students to professors." *Journal of Political Economy* 118: 439-464.

Chin, Christine and Jonathan Osborne 2008. "Students' questions: a potential for teaching and learning science." *Studies in Science Education* 44(1): 1-39.

Chin, Christine and Jonathan Osborne 2010a. "Students' questions and discursive interaction: their impact on argumentation during collaborative group discussions in science." *Journal of Research in Science Teaching* 47(7): 883-908.

Chin, Christine and Jonathan Osborne 2010b. "Supporting argumentation through students' questions: case studies in science classrooms." *Journal of the Learning Sciences* 19: 230-284.

Cho, Insook. 2012. "The effect of teacher–student gender matching: evidence from OECD countries." *Economics of Education Review* 31: 54-67.

Crawford, Mary and Margo MacLeod 1990. "Gender in the college classroom: an assessment of the 'chilly climate'' for women." *Sex Roles* 23(3): 101-122.

Dee, Thomas S. 2007. "Teachers and the gender gaps in student achievement." *The Journal of Human Resources* 42(3): 158-165.

Ehrenberg, Ronald G., Daniel D. Goldhaber, and Dominic J. Brewer 1995. "Do teachers' race, gender, and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988." *Industrial and Labor Relations Review* 48: 547-561.

Good, Thomas L., Ricky L. Slavings, Kathleen Hobson Harel, and Hugh Emerson 1987. "Student passivity: a study of question asking in K-12 classrooms." *Sociology of*

Education 60: 181-199.
Harper, Kathleen A., Eugenia Etkina, and Yuhfen Lin 2003. "Encouraging and analyzing student questions in a large physics course: meaningful patterns for instructors." *Journal of Research in Science Teaching* 40:776-791.
Hoffman, Florian and Philip Orepoulos 2009. "A professor like me: the influence of instructor gender on college attainment." *The Journal of Human Resources* 44(2): 479-494.
Holmlund, Helena and Krister Sund 2008. "Is the gender gap in school performance affected by the sex of the teacher?" *Labor Economics* 15: 37-53.
Keeling, Elena L., Kelly M. Polacek, and Ella L. Ingram 2009. "A statistical analysis of student questions in a cell biology laboratory." *Life Science Education* 8: 131-139.
King, A. 1992. "Facilitating elaborative learning through guided student-generated questioning." *Educational Psychologist* 27(1): 111-126.
King, Alison and Barak Rosenshine 1993. "Effects of guided cooperative-questioning on children's knowledge construction." *Journal of Experimental Education* 6: 127-148.
Koch, Adina and Shulamith G. Eckstein 1991. "Improvements of reading comprehension of physics texts by students' question formulation." *International Journal of Science Education* 13: 473-485.
Lavy, Victor 2008. "Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment." *Journal of Public Economics* 92: 2083-2105.
Mundlak, Yair 1978. "On the pooling of time series and cross section data." *Econometrica* 46(1): 69-85.
Muralidharan, Karthik and Ketki Sheth 2016. "Bridging education gender gaps in developing countries: the role of female teachers." *Journal of Human Resources* 51(2): 269-297.
Neumark, David and Rosella Gardecki 1998. "Women helping women? Role model and mentoring effects on female Ph.D. students in economics." *The Journal of Human Resources* 33(1): 220-246.
Nixon, Lucia A. and Michael D. Robinson 1989. "The educational attainment of young women: role model effects of high school faculty." *Demography* 36(2): 185-194.
Paredes, Valentina 2014. "A teacher like me or student like me? Role model versus teacher bias effect." *Economics of Education Review* 39: 38-49.
Pearson, Judy C. and Richard West 1991. "An initial investigation of the effects of gender on student questions in the classroom: developing a descriptive base." *Communication Education* 40: 22-32.
Price, Joshua 2010. "The effect of instructor race and gender on student persistence in STEM fields." *Economics of Education Review* 29: 901-910.
Rocca, Kelly A. 2010. "Student participation in the college classroom: an extended multidisciplinary literature review." *Communication Education* 59(2): 185-213.
Rosenshine, Barak, Carla Meister, and Saul Chapman 1996. "Teaching students to generate questions: a review of the intervention studies." *Review of Educational Research* 66(2): 181-221.
Rothstein, Donna S. 1995. "Do female faculty influence female students educational and labor market attainments?" *Industrial and Labor Relations Review* 48: 515-530.
Spencer, Steven J., Claude M. Steele, and Diane M. Quinn 1999. "Stereotypes threat and women's math performance." *Journal of Experimental Social Psychology* 35: 4-28.
Steele, Claude M. 1997. "A threat in the air: how stereotypes shape intellectual identity and performance." *American Psychologist* 52(6): 613-629.
Sternglanz, Sarah H. and Shirley Lyberger-Ficek 1977. "Sex differences in student–teacher

interactions in the college classroom." *Sex Roles* 3(4): 345-352.
Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data (Second Edition)*, The MIT Press, Cambridge, Massachusetts.
Zoller, Uri 1987. "The fostering of question-asking capability: a meaningful aspect of problem-solving in chemistry." *Journal of Chemical Education* 64: 510-512.

Appendix Table A1
Descriptive Statistics of Independent Variables

		Grade 1			Grade 2			Grade 3		
		N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.
Teacher attributes										
Female teacher	=1 if the teacher is a female, 0 otherwise	2517	0.331	0.471	2491	0.348	0.477	2497	0.323	0.468
Experience	Years of teaching experience	2493	16.390	9.074	2473	16.323	8.725	2487	17.295	8.148
Class attributes										
Class size	Number of students in the class	2463	34.199	5.878	2411	34.704	5.585	2466	34.534	5.568
Girl ratio	Ratio of girls in the class	2463	0.486	0.098	2411	0.486	0.097	2466	0.485	0.094
School attributes										
Public urban	=1 if the school is a public school in an urban area, 0 otherwise	2557	0.539	0.499	2535	0.540	0.499	2530	0.543	0.498
Public town	=1 if the school is a public school in a rural area, 0 otherwise	2557	0.248	0.432	2535	0.249	0.432	2530	0.249	0.433
Private/national	=1 if the school is a national school or a private school, 0	2557	0.057	0.233	2535	0.057	0.232	2530	0.049	0.217
Student attributes										
Girl	=1 if the student is a girl, 0 otherwise	83477	0.484	0.500	82853	0.485	0.500	75404	0.485	0.500
Outside education	=1 if a cram school or home tutor is used, 0 otherwise (English)	49749	0.518	0.500	49471	0.575	0.494	44789	0.686	0.464
	=1 if a cram school or home tutor is used, 0 otherwise (Japanese)	50131	0.314	0.464	49350	0.350	0.477	44986	0.499	0.500
	=1 if a cram school or home tutor is used, 0 otherwise (Math)	49770	0.484	0.500	49381	0.550	0.498	44924	0.679	0.467
	=1 if a cram school or home tutor is used, 0 otherwise (Science)	49935	0.284	0.451	49480	0.344	0.475	35926	0.528	0.499
	=1 if a cram school or home tutor is used, 0 otherwise (Social	50167	0.270	0.444	49746	0.320	0.466	45284	0.501	0.500
Amount of sleep	Sleeping hours per day. 1: less than 6 hours, 2: 6–7 hours, 3: 7–8 hours, 4: 8–9 hours, 5: 9–10 hours, 6: 10 or more hours	82677	3.010	1.169	82035	2.746	1.130	74387	2.340	1.086
Breakfast	Frequency of eating breakfast. 1: Never or almost never, 2: Often do not eat breakfast, 3: Generally eat breakfast, 4: Always eat	82617	3.615	0.748	81968	3.580	0.790	74318	3.522	0.848
Belongings checked	Frequency with which items brought to school are checked the day before or in the morning. 1: Almost never checked, 2: Often not checked, 3: Generally checked, 4: Always checked	82819	3.160	0.925	82166	3.098	0.958	74475	3.108	0.946

Appendix Table A2
Estimated Results for Equation (4) (Grade 1)

	English			Japanese			Math			Science			Social studies		
	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E
Subject (γ_1)	-0.1406	***	0.0070	-0.1625	***	0.0075	0.0064		0.0052	0.0735	***	0.0052	0.0803	***	0.0050
Girl (γ_2)	-0.0051		0.0064	-0.0177	***	0.0064	0.0420	***	0.0066	0.0453	***	0.0065	0.0504	***	0.0066
Female teacher (γ_3)	-0.0850	***	0.0048	-0.0802	***	0.0047	-0.1089	***	0.0046	-0.0976	***	0.0045	-0.0819	***	0.0045
Subject×Girl (γ_4)	0.2133	***	0.0098	0.3441	***	0.0105	-0.0912	***	0.0073	-0.0983	***	0.0072	-0.1043	***	0.0070
Girl×Female teacher (γ_5)	0.1782	***	0.0065	0.1551	***	0.0064	0.2060	***	0.0062	0.1999	***	0.0061	0.1796	***	0.0061
Subject×Female teacher (γ_6)	0.0597	***	0.0100	0.0702	***	0.0105	0.0986	***	0.0105	0.1042	***	0.0111	0.0628	***	0.0130
Subject×Girl×Female teacher (γ_7)	-0.0970	***	0.0139	-0.1376	***	0.0145	-0.1598	***	0.0145	-0.1809	***	0.0156	-0.1435	***	0.0180
Experience	0.0051	***	0.0005	0.0051	***	0.0005	0.0051	***	0.0005	0.0054	***	0.0005	0.0052	***	0.0005
Experience^2	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000
Class size	-0.0005		0.0012	-0.0006		0.0012	-0.0006		0.0012	-0.0007		0.0012	-0.0004		0.0012
Class size^2	0.0000	**	0.0000	0.0000	**	0.0000	0.0000	**	0.0000	0.0000	**	0.0000	0.0000	**	0.0000
Girl ratio	-0.3933	***	0.0240	-0.3996	***	0.0240	-0.3875	***	0.0240	-0.3874	***	0.0240	-0.3960	***	0.0240
Public urban	0.0294	***	0.0083	0.0296	***	0.0083	0.0292	***	0.0083	0.0294	***	0.0083	0.0294	***	0.0083
Public town	0.0043		0.0096	0.0045		0.0096	0.0044		0.0096	0.0045		0.0096	0.0043		0.0096
Private/national	0.7686	***	0.0120	0.7695	***	0.0120	0.7667	***	0.0120	0.7666	***	0.0120	0.7665	***	0.0120
Outside education	0.1822	***	0.0041	0.1774	***	0.0041	0.1831	***	0.0041	0.1798	***	0.0041	0.1801	***	0.0041
Amount of sleep: 6–7h	0.0747	***	0.0112	0.0747	***	0.0112	0.0747	***	0.0112	0.0746	***	0.0112	0.0744	***	0.0112
Amount of sleep: 7–8h	0.1141	***	0.0110	0.1138	***	0.0110	0.1144	***	0.0110	0.1141	***	0.0110	0.1137	***	0.0110
Amount of sleep: 8–9h	-0.0239	**	0.0117	-0.0245	**	0.0117	-0.0236	**	0.0117	-0.0242	**	0.0117	-0.0245	**	0.0117
Amount of sleep: 9–10h	-0.2323	***	0.0148	-0.2323	***	0.0148	-0.2318	***	0.0148	-0.2321	***	0.0148	-0.2326	***	0.0148
Amount of sleep: >10h	-0.5523	***	0.0252	-0.5544	***	0.0252	-0.5524	***	0.0252	-0.5530	***	0.0252	-0.5528	***	0.0252
Breakfast: Often do not eat	0.0763	***	0.0205	0.0766	***	0.0205	0.0763	***	0.0205	0.0765	***	0.0205	0.0765	***	0.0205
Breakfast: Generally eat	0.2460	***	0.0178	0.2467	***	0.0178	0.2455	***	0.0178	0.2460	***	0.0178	0.2461	***	0.0178
Breakfast: Always eat	0.5755	***	0.0168	0.5757	***	0.0168	0.5750	***	0.0168	0.5754	***	0.0168	0.5755	***	0.0168
Belongings checked: Often not checked	0.1674	***	0.0140	0.1666	***	0.0140	0.1667	***	0.0140	0.1669	***	0.0140	0.1672	***	0.0140
Belongings checked: Generally checked	0.3352	***	0.0121	0.3349	***	0.0121	0.3347	***	0.0121	0.3349	***	0.0121	0.3353	***	0.0121
Belongings checked: Always checked	0.4876	***	0.0120	0.4870	***	0.0120	0.4868	***	0.0120	0.4873	***	0.0120	0.4871	***	0.0120
Female teacher (Average)	-0.0043		0.0113	0.0008		0.0113	0.0008		0.0113	-0.0028		0.0112	-0.0052		0.0113
Experience (Average)	0.0073	***	0.0017	0.0072	***	0.0017	0.0072	***	0.0017	0.0072	***	0.0017	0.0072	***	0.0017
Experience^2 (Average)	-0.0002	***	0.0000	-0.0002	***	0.0000	-0.0002	***	0.0000	-0.0002	***	0.0000	-0.0002	***	0.0000
Constant term	-0.8849	***	0.0367	-0.8824	***	0.0367	-0.9075	***	0.0367	-0.9229	***	0.0367	-0.9249	***	0.0367
sigma_u	0.7377			0.7378			0.7377			0.7375			0.7376		
sigma_e	0.5540			0.5522			0.5540			0.5540			0.5544		
rho	0.6394			0.6410			0.6393			0.6392			0.6390		
No. of students	81944			81944			81944			81944			81944		
No. of obs.	234575			234575			234575			234575			234575		
R-sq: within	0.0244			0.0306			0.0241			0.0242			0.0230		
R-sq: between	0.1652			0.1660			0.1650			0.1654			0.1652		
R-sq: overall	0.1361			0.1378			0.1358			0.1361			0.1357		
Chi-sq(3) : H0 : m=0	18.5473			17.5400			17.3103			17.6888			18.2654		

Note: * p < .10, ** p < .05, *** p < .01.

Appendix Table A3
Estimated Results for Equation (4) (Grade 2)

	English			Japanese			Math			Science			Social studies		
	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E
Subject (γ_1)	-0.1385	***	0.0072	-0.1675	***	0.0074	-0.0105	**	0.0053	0.0971	***	0.0052	0.0736	***	0.0053
Girl (γ_2)	-0.0078		0.0065	-0.0246	***	0.0065	0.0290	***	0.0066	0.0543	***	0.0066	0.0460	***	0.0066
Female teacher (γ_3)	-0.0708	***	0.0049	-0.0717	***	0.0048	-0.1011	***	0.0047	-0.0792	***	0.0046	-0.0763	***	0.0046
Subject×Girl (γ_4)	0.2173	***	0.0101	0.3492	***	0.0099	-0.0506	***	0.0074	-0.1420	***	0.0072	-0.1084	***	0.0072
Girl×Female teacher (γ_5)	0.1827	***	0.0066	0.1616	***	0.0064	0.2081	***	0.0062	0.1879	***	0.0062	0.1829	***	0.0061
Subject×Female teacher (γ_6)	0.0567	***	0.0101	0.0728	***	0.0105	0.1085	***	0.0106	0.0962	***	0.0115	0.0816	***	0.0134
Subject×Girl×Female teacher (γ_7)	-0.1233	***	0.0140	-0.1354	***	0.0141	-0.1442	***	0.0145	-0.1689	***	0.0157	-0.1432	***	0.0180
Experience	0.0037	***	0.0005	0.0038	***	0.0005	0.0036	***	0.0005	0.0037	***	0.0005	0.0038	***	0.0005
Experience^2	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000
Class size	-0.0037	***	0.0010	-0.0030	***	0.0010	-0.0028	***	0.0010	-0.0031	***	0.0010	-0.0031	***	0.0010
Class size^2	0.0001	***	0.0000	0.0001	***	0.0000	0.0001	***	0.0000	0.0001	***	0.0000	0.0001	***	0.0000
Girl ratio	-0.3824	***	0.0250	-0.3857	***	0.0251	-0.3770	***	0.0251	-0.3758	***	0.0251	-0.3817	***	0.0250
Public urban	0.0098		0.0084	0.0100		0.0084	0.0101		0.0084	0.0098		0.0084	0.0098		0.0084
Public rural	-0.0061		0.0096	-0.0065		0.0096	-0.0058		0.0096	-0.0071		0.0096	-0.0066		0.0096
Private/national	0.6721	***	0.0125	0.6748	***	0.0125	0.6715	***	0.0125	0.6738	***	0.0125	0.6704	***	0.0125
Outside education	0.1673	***	0.0041	0.1621	***	0.0040	0.1661	***	0.0041	0.1630	***	0.0040	0.1628	***	0.0040
Amount of sleep: 6–7h	0.0732	***	0.0097	0.0730	***	0.0097	0.0731	***	0.0097	0.0730	***	0.0097	0.0731	***	0.0097
Amount of sleep: 7–8h	0.0752	***	0.0098	0.0749	***	0.0097	0.0753	***	0.0098	0.0748	***	0.0097	0.0749	***	0.0098
Amount of sleep: 8–9h	-0.0853	***	0.0111	-0.0861	***	0.0111	-0.0854	***	0.0111	-0.0862	***	0.0111	-0.0858	***	0.0111
Amount of sleep: 9–10h	-0.3213	***	0.0169	-0.3221	***	0.0169	-0.3218	***	0.0169	-0.3215	***	0.0169	-0.3223	***	0.0169
Amount of sleep: >10h	-0.5768	***	0.0288	-0.5770	***	0.0288	-0.5757	***	0.0288	-0.5762	***	0.0288	-0.5774	***	0.0288
Breakfast: Often do not eat	0.0410	**	0.0193	0.0406	**	0.0192	0.0413	**	0.0193	0.0417	**	0.0193	0.0411	**	0.0193
Breakfast: Generally eat	0.2307	***	0.0164	0.2302	***	0.0164	0.2306	***	0.0164	0.2317	***	0.0164	0.2306	***	0.0164
Breakfast: Always eat	0.5547	***	0.0153	0.5544	***	0.0153	0.5548	***	0.0153	0.5558	***	0.0153	0.5548	***	0.0153
Belongings checked: Often not checked	0.2163	***	0.0134	0.2160	***	0.0134	0.2161	***	0.0134	0.2164	***	0.0134	0.2157	***	0.0134
Belongings checked: Generally checked	0.3718	***	0.0115	0.3711	***	0.0115	0.3714	***	0.0115	0.3716	***	0.0115	0.3714	***	0.0115
Belongings checked: Always checked	0.5288	***	0.0115	0.5286	***	0.0115	0.5287	***	0.0115	0.5289	***	0.0115	0.5286	***	0.0115
Female teacher (Average)	0.0395	***	0.0113	0.0442	***	0.0113	0.0444	***	0.0113	0.0414	***	0.0113	0.0409	***	0.0113
Experience (Average)	0.0017		0.0018	0.0016		0.0018	0.0020		0.0018	0.0020		0.0018	0.0018		0.0018
Experience^2 (Average)	0.0000		0.0001	0.0000		0.0001	0.0000		0.0001	0.0000		0.0001	0.0000		0.0001
Constant term	-0.8000	***	0.0338	-0.8019	***	0.0337	-0.8303	***	0.0338	-0.8516	***	0.0337	-0.8440	***	0.0338
sigma_u	0.7417			0.7419			0.7416			0.7417			0.7416		
sigma_e	0.5517			0.5496			0.5523			0.5512			0.5520		
rho	0.6438			0.6456			0.6433			0.6442			0.6435		
No. of students	81416			81416			81416			81416			81416		
No. of obs.	231821			231821			231821			231821			231821		
R-sq: within	0.0198			0.0273			0.0178			0.0218			0.0188		
R-sq: between	0.1645			0.1651			0.1642			0.1646			0.1644		
R-sq: overall	0.1342			0.1362			0.1336			0.1347			0.1340		
Chi-sq(3) : H0 : m=0	29.4056			31.2405			31.0766			28.3025			28.4358		

Note: * p < .10, ** p < .05, *** p < .01.

Appendix Table A4
Estimated Results for Equation (4) (Grade 3)

	English			Japanese			Math			Science			Social studies		
	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E
Subject (γ_1)	-0.0760	***	0.0073	-0.1406	***	0.0079	0.0398	***	0.0054	0.1069	***	0.0059	-0.0090	*	0.0053
Girl (γ_2)	0.0333	***	0.0068	0.0058		0.0068	0.0855	***	0.0070	0.0876	***	0.0069	0.0277	***	0.0070
Female teacher (γ_3)	-0.0733	***	0.0053	-0.0415	***	0.0053	-0.0740	***	0.0050	-0.0704	***	0.0049	-0.0806	***	0.0050
Subject×Girl (γ_4)	0.0977	***	0.0103	0.3538	***	0.0106	-0.1471	***	0.0075	-0.1995	***	0.0081	0.0680	***	0.0073
Girl×Female teacher (γ_5)	0.1863	***	0.0072	0.1081	***	0.0072	0.1619	***	0.0067	0.1608	***	0.0066	0.1997	***	0.0067
Subject×Female teacher (γ_6)	0.0519	***	0.0107	0.0178		0.0111	0.0690	***	0.0122	0.1338	***	0.0133	0.0767	***	0.0134
Subject×Girl×Female teacher (γ_7)	-0.1060	***	0.0150	-0.0843	***	0.0150	-0.1111	***	0.0168	-0.1760	***	0.0184	-0.1641	***	0.0180
Experience	0.0029	***	0.0007	0.0032	***	0.0007	0.0030	***	0.0007	0.0032	***	0.0007	0.0030	***	0.0007
Experience^2	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000
Class size	0.0011		0.0011	0.0009		0.0011	0.0011		0.0011	0.0013		0.0011	0.0013		0.0011
Class size^2	0.0000		0.0000	0.0000		0.0000	0.0000		0.0000	0.0000		0.0000	0.0000		0.0000
Girl ratio	-0.2718	***	0.0279	-0.2757	***	0.0280	-0.2668	***	0.0279	-0.2657	***	0.0279	-0.2646	***	0.0279
Public urban	-0.0166	*	0.0091	-0.0166	*	0.0091	-0.0163	*	0.0091	-0.0167	*	0.0091	-0.0167	*	0.0091
Public rural	-0.0709	***	0.0103	-0.0705	***	0.0103	-0.0704	***	0.0103	-0.0714	***	0.0103	-0.0709	***	0.0103
Private/national	0.4103	***	0.0139	0.4134	***	0.0139	0.4119	***	0.0139	0.4113	***	0.0139	0.4090	***	0.0139
Outside education	0.1854	***	0.0046	0.1860	***	0.0046	0.1880	***	0.0047	0.1803	***	0.0046	0.1838	***	0.0046
Amount of sleep: 6–7h	0.0193	**	0.0079	0.0193	**	0.0079	0.0192	**	0.0079	0.0190	**	0.0079	0.0189	**	0.0079
Amount of sleep: 7–8h	-0.0313	***	0.0088	-0.0311	***	0.0088	-0.0312	***	0.0088	-0.0319	***	0.0088	-0.0319	***	0.0088
Amount of sleep: 8–9h	-0.2470	***	0.0124	-0.2466	***	0.0124	-0.2469	***	0.0124	-0.2482	***	0.0124	-0.2475	***	0.0124
Amount of sleep: 9–10h	-0.4661	***	0.0232	-0.4653	***	0.0232	-0.4652	***	0.0232	-0.4676	***	0.0232	-0.4669	***	0.0232
Amount of sleep: >10h	-0.6712	***	0.0339	-0.6704	***	0.0338	-0.6706	***	0.0338	-0.6723	***	0.0338	-0.6717	***	0.0339
Breakfast: Often do not eat	0.0877	***	0.0185	0.0875	***	0.0185	0.0876	***	0.0185	0.0870	***	0.0185	0.0880	***	0.0185
Breakfast: Generally eat	0.2506	***	0.0160	0.2506	***	0.0159	0.2504	***	0.0159	0.2502	***	0.0160	0.2508	***	0.0160
Breakfast: Always eat	0.5437	***	0.0147	0.5436	***	0.0147	0.5436	***	0.0147	0.5433	***	0.0147	0.5437	***	0.0147
Belongings checked: Often not checked	0.1510	***	0.0147	0.1513	***	0.0147	0.1508	***	0.0147	0.1512	***	0.0147	0.1509	***	0.0147
Belongings checked: Generally checked	0.2792	***	0.0128	0.2793	***	0.0128	0.2792	***	0.0128	0.2795	***	0.0128	0.2791	***	0.0128
Belongings checked: Always checked	0.4241	***	0.0128	0.4240	***	0.0128	0.4241	***	0.0128	0.4244	***	0.0128	0.4241	***	0.0128
Female teacher (Average)	0.0159		0.0118	0.0231	*	0.0118	0.0200	*	0.0118	0.0187		0.0118	0.0155		0.0118
Experience (Average)	0.0122	***	0.0022	0.0119	***	0.0022	0.0121	***	0.0022	0.0122	***	0.0022	0.0121	***	0.0022
Experience^2 (Average)	-0.0003	***	0.0001	-0.0003	***	0.0001	-0.0003	***	0.0001	-0.0003	***	0.0001	-0.0003	***	0.0001
Constant term	-0.8750	***	0.0370	-0.8698	***	0.0369	-0.9034	***	0.0370	-0.9130	***	0.0369	-0.8907	***	0.0370
sigma_u	0.7490			0.7495			0.7492			0.7494			0.7489		
sigma_e	0.5530			0.5492			0.5518			0.5509			0.5530		
rho	0.6472			0.6507			0.6483			0.6492			0.6471		
No. of students	73905			73905			73905			73905			73905		
No. of obs.	205339			205339			205339			205339			205339		
R-sq: within	0.0070			0.0201			0.0110			0.0142			0.0069		
R-sq: between	0.1446			0.1458			0.1451			0.1446			0.1446		
R-sq: overall	0.1156			0.1193			0.1169			0.1171			0.1156		
Chi-sq(3) : H0 : m=0	41.0368			40.9579			39.6665			40.6286			39.5446		

Note: * p < .10, ** p < .05, *** p < .01.

Appendix Table A5
Estimated Results for Equation (6) (Grade 1)

	English			Japanese			Math			Science			Social studies		
	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E
Subject (γ_{q1})	0.0789	***	0.0223	-0.2044	***	0.0227	0.2708	***	0.0175	0.0270		0.0170	-0.2118	***	0.0168
Girl (γ_{q2})	-0.4326	***	0.0176	-0.4326	***	0.0175	-0.5208	***	0.0186	-0.3870	***	0.0181	-0.3558	***	0.0183
Female teacher (γ_{q3})	0.1164	***	0.0148	0.1349	***	0.0151	0.1211	***	0.0145	0.0988	***	0.0142	0.0432	***	0.0141
Subject×Girl (γ_{q4})	0.0888	***	0.0332	0.0782	**	0.0348	0.3242	***	0.0260	-0.1370	***	0.0256	-0.2903	***	0.0257
Girl×Female teacher (γ_{q5})	0.2635	***	0.0214	0.3182	***	0.0216	0.3536	***	0.0209	0.2662	***	0.0203	0.2277	***	0.0203
Subject×Female teacher (γ_{q6})	-0.1090	***	0.0312	0.0480		0.0315	0.0819	**	0.0336	0.0154		0.0349	0.1211	***	0.0406
Subject×Girl×Female teacher (γ_{q7})	0.0058		0.0460	-0.1513	***	0.0471	-0.1740	***	0.0493	-0.0038		0.0511	0.0713		0.0601
Experience	-0.0243	***	0.0018	-0.0226	***	0.0018	-0.0226	***	0.0018	-0.0247	***	0.0018	-0.0232	***	0.0018
Experience^2	0.0004	***	0.0001	0.0004	***	0.0001	0.0004	***	0.0001	0.0004	***	0.0001	0.0004	***	0.0001
Class size	-0.0191	***	0.0035	-0.0185	***	0.0035	-0.0189	***	0.0035	-0.0186	***	0.0035	-0.0206	***	0.0035
Class size^2	0.0001	*	0.0000	0.0001		0.0000	0.0001	*	0.0000	0.0001	*	0.0000	0.0001	**	0.0000
Girl ratio	0.2950	***	0.0716	0.2732	***	0.0718	0.2919	***	0.0726	0.2912	***	0.0716	0.3085	***	0.0720
Public urban	-0.0813	***	0.0212	-0.0803	***	0.0212	-0.0840	***	0.0214	-0.0810	***	0.0212	-0.0824	***	0.0213
Public rural	-0.0222		0.0243	-0.0233		0.0243	-0.0279		0.0246	-0.0219		0.0243	-0.0236		0.0245
Private/national	0.4085	***	0.0349	0.4118	***	0.0350	0.4116	***	0.0354	0.4121	***	0.0349	0.4063	***	0.0351
Outside education	0.1286	***	0.0125	0.1217	***	0.0123	0.0628	***	0.0126	0.1394	***	0.0123	0.0960	***	0.0124
Amount of sleep: 6–7h	0.0018		0.0280	0.0018		0.0281	0.0020		0.0284	0.0025		0.0280	0.0023		0.0282
Amount of sleep: 7–8h	0.0437		0.0276	0.0439		0.0277	0.0407		0.0280	0.0446		0.0276	0.0418		0.0278
Amount of sleep: 8–9h	0.0589	**	0.0290	0.0579	**	0.0291	0.0525	*	0.0294	0.0599	**	0.0290	0.0562	*	0.0292
Amount of sleep: 9–10h	0.1649	***	0.0364	0.1642	***	0.0365	0.1615	***	0.0369	0.1663	***	0.0364	0.1620	***	0.0367
Amount of sleep: >10h	0.2010	***	0.0532	0.1998	***	0.0533	0.1990	***	0.0539	0.2014	***	0.0532	0.2011	***	0.0536
Breakfast: Often do not eat	0.1427	***	0.0506	0.1434	***	0.0507	0.1442	***	0.0513	0.1430	***	0.0506	0.1449	***	0.0510
Breakfast: Generally eat	0.1769	***	0.0439	0.1775	***	0.0440	0.1807	***	0.0445	0.1773	***	0.0439	0.1802	***	0.0442
Breakfast: Always eat	0.2774	***	0.0415	0.2788	***	0.0415	0.2842	***	0.0420	0.2772	***	0.0415	0.2821	***	0.0418
Belongings checked: Often not checked	0.1407	***	0.0337	0.1414	***	0.0338	0.1437	***	0.0341	0.1411	***	0.0337	0.1413	***	0.0339
Belongings checked: Generally checked	0.2635	***	0.0288	0.2645	***	0.0288	0.2685	***	0.0291	0.2637	***	0.0288	0.2651	***	0.0290
Belongings checked: Always checked	0.4102	***	0.0286	0.4120	***	0.0287	0.4176	***	0.0290	0.4105	***	0.0286	0.4126	***	0.0288
Female teacher (Average)	0.0357		0.0296	-0.0005		0.0297	0.0024		0.0300	0.0286		0.0296	0.0692	**	0.0299
Experience (Average)	-0.0069		0.0047	-0.0074		0.0047	-0.0068		0.0047	-0.0066		0.0047	-0.0068		0.0047
Experience^2 (Average)	0.0001		0.0001	0.0001		0.0001	0.0001		0.0001	0.0001		0.0001	0.0001		0.0001
Praised by parents: Not praised more than praised	-0.0311		0.0221	-0.0308		0.0221	-0.0318		0.0224	-0.0317		0.0221	-0.0317		0.0222
Praised by parents: Praised more than not praised	0.0026		0.0201	0.0035		0.0201	0.0050		0.0203	0.0021		0.0201	0.0036		0.0202
Praised by parents: Praised	0.0858	***	0.0214	0.0875	***	0.0215	0.0931	***	0.0217	0.0847	***	0.0214	0.0897	***	0.0216
Constant term	-0.8141	***	0.1006	-0.7773	***	0.1007	-0.8647	***	0.1017	-0.8156	***	0.1007	-0.7590	***	0.1010
var_u	2.5886	***	0.0413	2.6050	***	0.0416	2.7041	***	0.0437	2.5886	***	0.0413	2.6432	***	0.0424
No. of students	72505			72505			72505			72505			72505		
No. of obs.	207623			207623			207623			207623			207623		
Log likelihood	-106603.22			-106498.57			-105759.14			-106613.00			-106165.08		
Chi-sq(3) : H_0 : **m=0**	8.4813			7.7294			5.8115			7.3439			12.6119		

Note: * p < .10, ** p < .05, *** p < .01.

Appendix Table A6
Estimated Results for Equation (6) (Grade 2)

	English			Japanese			Math			Science			Social studies		
	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E
Subject (γ_{q1})	-0.0151		0.0224	-0.2809	***	0.0226	0.3374	***	0.0170	0.1445	***	0.0165	-0.3001	***	0.0170
Girl (γ_{q2})	-0.3339	***	0.0170	-0.3467	***	0.0170	-0.4056	***	0.0181	-0.3072	***	0.0177	-0.2644	***	0.0178
Female teacher (γ_{q3})	0.1205	***	0.0149	0.1150	***	0.0150	0.1477	***	0.0144	0.1215	***	0.0141	0.0252	*	0.0141
Subject×Girl (γ_{q4})	0.0599	*	0.0336	0.1527	***	0.0334	0.2440	***	0.0253	-0.0735	***	0.0245	-0.3135	***	0.0257
Girl×Female teacher (γ_{q5})	0.3053	***	0.0213	0.3406	***	0.0211	0.3509	***	0.0206	0.3041	***	0.0201	0.2648	***	0.0199
Subject×Female teacher (γ_{q6})	-0.0574	*	0.0311	0.1474	***	0.0315	-0.0259		0.0334	-0.0092		0.0357	0.2057	***	0.0403
Subject×Girl×Female teacher (γ_{q7})	-0.0207		0.0458	-0.1925	***	0.0461	-0.0390		0.0483	-0.0266		0.0511	0.0209		0.0588
Experience	-0.0321	***	0.0017	-0.0302	***	0.0017	-0.0273	***	0.0018	-0.0319	***	0.0017	-0.0301	***	0.0018
Experience^2	0.0006	***	0.0000	0.0006	***	0.0000	0.0005	***	0.0000	0.0006	***	0.0000	0.0006	***	0.0000
Class size	-0.0022		0.0035	-0.0024		0.0035	-0.0041		0.0035	-0.0029		0.0035	-0.0034		0.0035
Class size^2	0.0000		0.0001	0.0000		0.0001	0.0000		0.0001	0.0000		0.0001	0.0000		0.0001
Girl ratio	0.4260	***	0.0720	0.4176	***	0.0722	0.4301	***	0.0732	0.4215	***	0.0720	0.4519	***	0.0727
Public urban	-0.0215		0.0207	-0.0215		0.0208	-0.0261		0.0211	-0.0199		0.0208	-0.0249		0.0210
Public rural	0.0186		0.0236	0.0166		0.0237	0.0105		0.0240	0.0199		0.0237	0.0119		0.0239
Private/national	0.3552	***	0.0347	0.3571	***	0.0348	0.3410	***	0.0352	0.3613	***	0.0347	0.3439	***	0.0350
Outside education	0.1495	***	0.0119	0.1216	***	0.0118	0.0530	***	0.0120	0.1573	***	0.0118	0.0831	***	0.0119
Amount of sleep: 6–7h	0.0006		0.0236	0.0004		0.0237	-0.0005		0.0240	0.0012		0.0236	0.0012		0.0239
Amount of sleep: 7–8h	0.0291		0.0239	0.0272		0.0240	0.0232		0.0243	0.0300		0.0239	0.0276		0.0242
Amount of sleep: 8–9h	0.0211		0.0268	0.0179		0.0269	0.0112		0.0272	0.0226		0.0268	0.0169		0.0271
Amount of sleep: 9–10h	0.0342		0.0389	0.0308		0.0390	0.0198		0.0395	0.0363		0.0390	0.0272		0.0394
Amount of sleep: >10h	0.1753	***	0.0596	0.1733	***	0.0597	0.1700	***	0.0605	0.1772	***	0.0596	0.1775	***	0.0602
Breakfast: Often do not eat	0.1949	***	0.0457	0.1972	***	0.0458	0.2048	***	0.0464	0.1956	***	0.0457	0.1992	***	0.0462
Breakfast: Generally eat	0.1573	***	0.0392	0.1594	***	0.0393	0.1696	***	0.0397	0.1580	***	0.0392	0.1631	***	0.0396
Breakfast: Always eat	0.2944	***	0.0366	0.2977	***	0.0367	0.3100	***	0.0371	0.2950	***	0.0366	0.3017	***	0.0370
Belongings checked: Often not checked	0.1354	***	0.0311	0.1366	***	0.0312	0.1380	***	0.0315	0.1357	***	0.0311	0.1380	***	0.0314
Belongings checked: Generally checked	0.2560	***	0.0266	0.2565	***	0.0267	0.2584	***	0.0270	0.2564	***	0.0267	0.2591	***	0.0269
Belongings checked: Always checked	0.3982	***	0.0267	0.3996	***	0.0267	0.4035	***	0.0270	0.3991	***	0.0267	0.4034	***	0.0270
Female teacher (Average)	-0.0440		0.0288	-0.0657	**	0.0289	-0.0725	**	0.0292	-0.0514	*	0.0288	-0.0084		0.0291
Experience (Average)	-0.0173	***	0.0048	-0.0185	***	0.0048	-0.0223	***	0.0048	-0.0173	***	0.0048	-0.0189	***	0.0048
Experience^2 (Average)	0.0003	**	0.0001	0.0003	**	0.0001	0.0004	***	0.0001	0.0003	**	0.0001	0.0003	**	0.0001
Praised by parents: Not praised more than praised	-0.0007		0.0211	0.0001		0.0212	0.0022		0.0214	-0.0007		0.0211	0.0016		0.0214
Praised by parents: Praised more than not praised	0.0260		0.0190	0.0274		0.0191	0.0304		0.0193	0.0260		0.0190	0.0291		0.0192
Praised by parents: Praised	0.0936	***	0.0210	0.0965	***	0.0210	0.1054	***	0.0213	0.0931	***	0.0210	0.1017	***	0.0212
Constant term	-1.0946	***	0.0928	-1.0466	***	0.0929	-1.1208	***	0.0937	-1.1232	***	0.0929	-1.0123	***	0.0933
var_u	2.3529	***	0.0362	2.3693	***	0.0365	2.4585	***	0.0383	2.3572	***	0.0363	2.4287	***	0.0377
No. of students	71998			71998			71998			71998			71998		
No. of obs.	205071			205071			205071			205071			205071		
Log likelihood	-107618.75			-107481.48			-106679.01			-107572.25			-106865.81		
Chi-sq(3) : H_0 : **m=0**	36.6560			46.0436			51.8036			37.4840			38.6027		

Note: * p < .10, ** p < .05, *** p < .01.

Appendix Table A7
Estimated Results for Equation (6) (Grade 3)

	English		Japanese		Math		Science		Social studies	
	Coef.	R.S.E	Coef.	R.S.E	Coef.	R.S.E	Coef.	R.S.E	Coef.	R.S.E
Subject (γ_{q1})	-0.0342	0.0221	-0.3259 ***	0.0234	0.4092 ***	0.0174	0.2224 ***	0.0180	-0.4043 ***	0.0174
Girl (γ_{q2})	-0.0623 ***	0.0175	-0.0706 ***	0.0175	-0.1569 ***	0.0189	-0.0381 **	0.0179	0.0050	0.0184
Female teacher (γ_{q3})	0.1226 ***	0.0161	0.1618 ***	0.0163	0.1746 ***	0.0155	0.1475 ***	0.0149	0.0164	0.0152
Subject×Girl (γ_{q4})	0.0098	0.0321	0.0693 **	0.0341	0.3103 ***	0.0253	-0.1089 ***	0.0264	-0.2997 ***	0.0258
Girl×Female teacher (γ_{q5})	0.1880 ***	0.0225	0.2039 ***	0.0227	0.2946 ***	0.0217	0.1853 ***	0.0208	0.1636 ***	0.0211
Subject×Female teacher (γ_{q6})	-0.0016	0.0323	0.1175 ***	0.0330	0.1221 ***	0.0381	-0.0388	0.0410	0.2392 ***	0.0415
Subject×Girl×Female teacher (γ_{q7})	0.0461	0.0463	-0.0407	0.0476	-0.2653 ***	0.0545	0.0885	0.0587	-0.0097	0.0599
Experience	-0.0226 ***	0.0022	-0.0211 ***	0.0022	-0.0187 ***	0.0022	-0.0226 ***	0.0022	-0.0180 ***	0.0022
Experience^2	0.0004 ***	0.0001	0.0004 ***	0.0001	0.0003 ***	0.0001	0.0004 ***	0.0001	0.0003 ***	0.0001
Class size	-0.0227 ***	0.0038	-0.0203 ***	0.0039	-0.0221 ***	0.0039	-0.0223 ***	0.0039	-0.0232 ***	0.0039
Class size^2	0.0001 **	0.0001	0.0001 *	0.0001	0.0001 **	0.0001	0.0001 **	0.0001	0.0001 **	0.0001
Girl ratio	0.2809 ***	0.0749	0.2692 ***	0.0755	0.2967 ***	0.0769	0.2699 ***	0.0752	0.2860 ***	0.0762
Public urban	0.0486 **	0.0217	0.0457 **	0.0219	0.0404 *	0.0223	0.0490 **	0.0218	0.0410 *	0.0221
Public rural	0.1133 ***	0.0247	0.1098 ***	0.0248	0.1024 ***	0.0253	0.1138 ***	0.0247	0.0994 ***	0.0251
Private/national	0.4148 ***	0.0370	0.4077 ***	0.0373	0.3830 ***	0.0381	0.4197 ***	0.0371	0.3807 ***	0.0377
Outside education	0.2224 ***	0.0131	0.1861 ***	0.0130	0.1159 ***	0.0133	0.2332 ***	0.0130	0.1506 ***	0.0132
Amount of sleep: 6–7h	-0.0248	0.0195	-0.0260	0.0196	-0.0291	0.0200	-0.0241	0.0196	-0.0281	0.0199
Amount of sleep: 7–8h	-0.0535 **	0.0215	-0.0577 ***	0.0216	-0.0671 ***	0.0221	-0.0518 **	0.0216	-0.0610 ***	0.0219
Amount of sleep: 8–9h	-0.0509 *	0.0289	-0.0580 **	0.0291	-0.0734 **	0.0297	-0.0491 *	0.0290	-0.0656 **	0.0294
Amount of sleep: 9–10h	-0.0199	0.0503	-0.0277	0.0506	-0.0401	0.0516	-0.0173	0.0504	-0.0310	0.0512
Amount of sleep: >10h	0.1090	0.0687	0.1053	0.0691	0.1022	0.0707	0.1100	0.0689	0.1067	0.0700
Breakfast: Often do not eat	0.0473	0.0424	0.0484	0.0426	0.0510	0.0435	0.0467	0.0424	0.0480	0.0431
Breakfast: Generally eat	0.0991 ***	0.0370	0.1020 ***	0.0372	0.1083 ***	0.0380	0.0991 ***	0.0370	0.1043 ***	0.0376
Breakfast: Always eat	0.2427 ***	0.0339	0.2476 ***	0.0341	0.2590 ***	0.0348	0.2431 ***	0.0339	0.2526 ***	0.0344
Belongings checked: Often not checked	0.2501 ***	0.0332	0.2507 ***	0.0334	0.2559 ***	0.0340	0.2512 ***	0.0333	0.2554 ***	0.0337
Belongings checked: Generally checked	0.3547 ***	0.0285	0.3557 ***	0.0286	0.3619 ***	0.0291	0.3569 ***	0.0285	0.3607 ***	0.0289
Belongings checked: Always checked	0.4547 ***	0.0285	0.4563 ***	0.0287	0.4635 ***	0.0292	0.4574 ***	0.0286	0.4619 ***	0.0290
Female teacher (Average)	0.0356	0.0294	0.0013	0.0296	0.0005	0.0302	0.0278	0.0295	0.0918 ***	0.0300
Experience (Average)	-0.0251 ***	0.0056	-0.0259 ***	0.0056	-0.0258 ***	0.0058	-0.0260 ***	0.0056	-0.0290 ***	0.0057
Experience^2 (Average)	0.0006 ***	0.0002	0.0006 ***	0.0002	0.0006 ***	0.0002	0.0006 ***	0.0002	0.0007 ***	0.0002
Praised by parents: Not praised more than praised	0.0266	0.0212	0.0276	0.0213	0.0288	0.0217	0.0260	0.0212	0.0291	0.0215
Praised by parents: Praised more than not praised	0.0634 ***	0.0192	0.0657 ***	0.0194	0.0676 ***	0.0198	0.0630 ***	0.0193	0.0661 ***	0.0196
Praised by parents: Praised	0.1553 ***	0.0220	0.1596 ***	0.0221	0.1674 ***	0.0226	0.1553 ***	0.0220	0.1645 ***	0.0224
Constant term	-0.6750 ***	0.0996	-0.6685 ***	0.1004	-0.8144 ***	0.1024	-0.7178 ***	0.0999	-0.5647 ***	0.1009
var_u	2.2943 ***	0.0361	2.3334 ***	0.0368	2.4746 ***	0.0396	2.3096 ***	0.0364	2.4108 ***	0.0383
No. of students	65569		65569		65569		65569		65569	
No. of obs.	182246		182246		182246		182246		182246	
Log likelihood	-99581.94		-99353.19		-98189.71		-99474.74		-98493.40	
Chi-sq(3) : H_0 : **m=0**	28.8929		30.0134		25.9763		29.5093		44.3146	

Note: * p < .10, ** p < .05, *** p < .01.

Appendix Table A8
Estimated Results for Equation (7) (Grade 1)

	English			Japanese			Math			Science			Social studies		
	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E
Subject (γ_1)	-0.1663	***	0.0097	-0.1770	***	0.0100	-0.0195	***	0.0072	0.0435	***	0.0070	0.0511	***	0.0065
Girl (γ_2)	-0.0475	***	0.0075	-0.0647	***	0.0075	0.0044		0.0075	0.0007		0.0076	0.0059		0.0077
Female teacher (γ_3)	-0.1035	***	0.0065	-0.1096	***	0.0065	-0.1330	***	0.0061	-0.1244	***	0.0061	-0.1129	***	0.0060
Subject×Girl (γ_4)	0.2290	***	0.0134	0.3760	***	0.0138	-0.0772	***	0.0102	-0.0653	***	0.0096	-0.0686	***	0.0090
Girl×Female teacher (γ_5)	0.1943	***	0.0089	0.1708	***	0.0089	0.2246	***	0.0083	0.2213	***	0.0084	0.2052	***	0.0084
Subject×Female teacher (γ_6)	0.0655	***	0.0140	0.1061	***	0.0143	0.1160	***	0.0151	0.1296	***	0.0152	0.1229	***	0.0176
Subject×Girl×Female teacher (γ_7)	-0.1072	***	0.0194	-0.1712	***	0.0195	-0.1964	***	0.0211	-0.2068	***	0.0211	-0.1954	***	0.0241
Question×Subject (ζ_1)	0.0743	***	0.0179	0.0390	**	0.0198	0.0868	***	0.0127	0.0965	***	0.0135	0.0939	***	0.0135
Question×Girl (ζ_2)	0.1879	***	0.0134	0.2027	***	0.0133	0.1844	***	0.0137	0.1895	***	0.0133	0.1748	***	0.0128
Question×Female teacher (ζ_3)	0.0669	***	0.0117	0.0864	***	0.0114	0.0830	***	0.0109	0.0885	***	0.0107	0.0942	***	0.0106
Question×Subject×Girl (ζ_4)	-0.0694	**	0.0272	-0.1233	***	0.0302	-0.0978	***	0.0199	-0.0903	***	0.0209	-0.0548	***	0.0212
Question×Girl×Female teacher (ζ_5)	-0.1039	***	0.0177	-0.1033	***	0.0172	-0.1161	***	0.0170	-0.1140	***	0.0164	-0.1080	***	0.0161
Question×Subject×Female teacher (ζ_6)	-0.0347		0.0258	-0.0889	***	0.0275	-0.0770	***	0.0262	-0.0924	***	0.0289	-0.1503	***	0.0343
Question×Subject×Girl×Female teacher (ζ_7)	0.0625	*	0.0377	0.1359	***	0.0406	0.1298	***	0.0378	0.0893	**	0.0428	0.0848	*	0.0512
Experience	0.0057	***	0.0005	0.0057	***	0.0005	0.0057	***	0.0005	0.0060	***	0.0005	0.0058	***	0.0005
Experience^2	-0.0002	***	0.0000	-0.0002	***	0.0000	-0.0002	***	0.0000	-0.0002	***	0.0000	-0.0002	***	0.0000
Class size	0.0002		0.0013	0.0002		0.0013	0.0001		0.0013	0.0001		0.0013	0.0003		0.0013
Class size^2	0.0000	*	0.0000	0.0000	*	0.0000	0.0000	*	0.0000	0.0000	*	0.0000	0.0000	*	0.0000
Girl ratio	-0.4010	***	0.0254	-0.4088	***	0.0254	-0.3946	***	0.0255	-0.3960	***	0.0254	-0.4034	***	0.0254
Public urban	0.0282	***	0.0088	0.0285	***	0.0088	0.0278	***	0.0088	0.0285	***	0.0088	0.0283	***	0.0088
Public town	-0.0036		0.0101	-0.0032		0.0101	-0.0037		0.0101	-0.0030		0.0101	-0.0035		0.0101
Private/national	0.7490	***	0.0127	0.7502	***	0.0127	0.7476	***	0.0127	0.7468	***	0.0127	0.7471	***	0.0127
Outside education	0.1765	***	0.0044	0.1708	***	0.0043	0.1777	***	0.0044	0.1729	***	0.0043	0.1744	***	0.0043
Amount of sleep: 6–7h	0.0772	***	0.0119	0.0773	***	0.0119	0.0774	***	0.0119	0.0772	***	0.0119	0.0771	***	0.0119
Amount of sleep: 7–8h	0.1136	***	0.0117	0.1133	***	0.0117	0.1140	***	0.0117	0.1133	***	0.0117	0.1133	***	0.0117
Amount of sleep: 8–9h	-0.0229	*	0.0124	-0.0236	*	0.0124	-0.0226	*	0.0124	-0.0237	*	0.0124	-0.0235	*	0.0124
Amount of sleep: 9–10h	-0.2404	***	0.0158	-0.2407	***	0.0158	-0.2401	***	0.0158	-0.2407	***	0.0158	-0.2408	***	0.0158
Amount of sleep: >10h	-0.5430	***	0.0269	-0.5450	***	0.0268	-0.5432	***	0.0268	-0.5442	***	0.0268	-0.5439	***	0.0268
Breakfast: Often do not eat	0.0673	***	0.0220	0.0678	***	0.0220	0.0673	***	0.0220	0.0674	***	0.0219	0.0677	***	0.0220
Experience2 (Average)	0.2411	***	0.0190	0.2418	***	0.0190	0.2406	***	0.0190	0.2409	***	0.0190	0.2413	***	0.0190
Breakfast: Always eat	0.5610	***	0.0180	0.5612	***	0.0180	0.5602	***	0.0180	0.5605	***	0.0180	0.5607	***	0.0180
Belongings checked: Often not checked	0.1515	***	0.0150	0.1512	***	0.0150	0.1507	***	0.0150	0.1509	***	0.0150	0.1513	***	0.0150
Belongings checked: Generally checked	0.3137	***	0.0129	0.3139	***	0.0129	0.3132	***	0.0129	0.3131	***	0.0129	0.3136	***	0.0129
Belongings checked: Always checked	0.4618	***	0.0128	0.4618	***	0.0128	0.4610	***	0.0128	0.4610	***	0.0128	0.4610	***	0.0128
Female teacher (Average)	-0.0080		0.0120	-0.0017		0.0120	-0.0021		0.0119	-0.0063		0.0119	-0.0097		0.0119
Experience (Average)	0.0067	***	0.0019	0.0066	***	0.0019	0.0065	***	0.0019	0.0065	***	0.0019	0.0065	***	0.0019
Experience2 (Average)	-0.0002	***	0.0001	-0.0002	***	0.0001	-0.0002	***	0.0001	-0.0002	***	0.0001	-0.0002	***	0.0001
Constant term	-0.8575	***	0.0390	-0.8559	***	0.0390	-0.8825	***	0.0390	-0.8966	***	0.0390	-0.8962	***	0.0390
sigma_u	0.7337			0.7342			0.7340			0.7333			0.7337		
sigma_e	0.5526			0.5513			0.5527			0.5529			0.5531		
rho	0.6381			0.6394			0.6382			0.6376			0.6377		
No. of students	72505			72505			72505			72505			72505		
No. of obs.	207623			207623			207623			207623			207623		
R-sq: within	0.0224			0.0285			0.0230			0.0220			0.0209		
R-sq: between	0.1348			0.1363			0.1349			0.1353			0.1348		
R-sq: overall	0.1642			0.1647			0.1642			0.1650			0.1646		
Chi-sq(3) : H_0 : **m=0**	14.2952			13.1774			13.0272			13.5001			14.2760		
Hansen J statistic (P-value)	3.5260		(0.1715)	3.4720		(0.1762)	3.4890		(0.1748)	3.7100		(0.1565)	3.5520		(0.1693)
Kleibergen-Paap rk Wald F statistic	8.1410			8.4010			9.6470			7.8500			8.9030		

Note: * p < .10, ** p < .05, *** p < .01.

Appendix Table A9
Estimated Results for Equation (7) (Grade 2)

	English			Japanese			Math			Science			Social studies		
	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E
Subject (γ_1)	-0.1823	***	0.0098	-0.1839	***	0.0098	-0.0512	***	0.0074	0.0612	***	0.0072	0.0458	***	0.0068
Girl (γ_2)	-0.0722	***	0.0076	-0.0962	***	0.0076	-0.0330	***	0.0076	-0.0131	*	0.0077	-0.0259	***	0.0079
Female teacher (γ_3)	-0.1009	***	0.0068	-0.1136	***	0.0066	-0.1382	***	0.0062	-0.1183	***	0.0062	-0.1161	***	0.0062
Subject×Girl (γ_4)	0.2373	***	0.0137	0.3892	***	0.0132	-0.0290	***	0.0105	-0.1100	***	0.0098	-0.0585	***	0.0094
Girl×Female teacher (γ_5)	0.2059	***	0.0092	0.1844	***	0.0090	0.2336	***	0.0084	0.2154	***	0.0085	0.2186	***	0.0085
Subject×Female teacher (γ_6)	0.0839	***	0.0139	0.1160	***	0.0142	0.1399	***	0.0156	0.1323	***	0.0163	0.1230	***	0.0179
Subject×Girl×Female teacher (γ_7)	-0.1429	***	0.0193	-0.1750	***	0.0193	-0.1903	***	0.0218	-0.2019	***	0.0222	-0.1979	***	0.0240
Question×Subject (ζ_1)	0.1482	***	0.0188	0.0614	***	0.0201	0.1128	***	0.0129	0.0996	***	0.0132	0.1043	***	0.0142
Question×Girl (ζ_2)	0.2467	***	0.0127	0.2712	***	0.0127	0.2640	***	0.0129	0.2582	***	0.0129	0.2447	***	0.0122
Question×Female teacher (ζ_3)	0.0816	***	0.0121	0.1154	***	0.0116	0.1049	***	0.0111	0.1077	***	0.0109	0.1119	***	0.0107
Question×Subject×Girl (ζ_4)	-0.0683	**	0.0285	-0.1412	***	0.0283	-0.1433	***	0.0197	-0.1026	***	0.0202	-0.0704	***	0.0220
Question×Girl×Female teacher (ζ_5)	-0.1315	***	0.0178	-0.1346	***	0.0171	-0.1520	***	0.0167	-0.1448	***	0.0164	-0.1462	***	0.0159
Question×Subject×Female teacher (ζ_6)	-0.0932	***	0.0265	-0.1387	***	0.0281	-0.1017	***	0.0269	-0.1028	***	0.0297	-0.1543	***	0.0366
Question×Subject×Girl×Female teacher (ζ_7)	0.0722	*	0.0386	0.1478	***	0.0391	0.1526	***	0.0378	0.1055	**	0.0419	0.1658	***	0.0522
Experience	0.0045	***	0.0006	0.0047	***	0.0005	0.0043	***	0.0006	0.0046	***	0.0005	0.0046	***	0.0006
Experience^2	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000
Class size	-0.0041	***	0.0011	-0.0035	***	0.0011	-0.0032	***	0.0011	-0.0036	***	0.0011	-0.0036	***	0.0011
Class size^2	0.0001	***	0.0000	0.0001	***	0.0000	0.0001	***	0.0000	0.0001	***	0.0000	0.0001	***	0.0000
Girl ratio	-0.4029	***	0.0263	-0.4104	***	0.0264	-0.3987	***	0.0264	-0.3983	***	0.0263	-0.4044	***	0.0263
Public urban	0.0078		0.0089	0.0078		0.0089	0.0082		0.0089	0.0077		0.0089	0.0078		0.0089
Public town	-0.0152		0.0101	-0.0157		0.0101	-0.0148		0.0101	-0.0163		0.0101	-0.0155		0.0101
Private/national	0.6444	***	0.0132	0.6482	***	0.0132	0.6451	***	0.0132	0.6463	***	0.0132	0.6439	***	0.0132
Outside education	0.1579	***	0.0043	0.1535	***	0.0043	0.1607	***	0.0043	0.1534	***	0.0043	0.1563	***	0.0043
Amount of sleep: 6–7h	0.0643	***	0.0102	0.0642	***	0.0102	0.0642	***	0.0102	0.0643	***	0.0102	0.0642	***	0.0102
Amount of sleep: 7–8h	0.0627	***	0.0103	0.0625	***	0.0103	0.0631	***	0.0103	0.0625	***	0.0103	0.0625	***	0.0103
Amount of sleep: 8–9h	-0.0996	***	0.0117	-0.1003	***	0.0117	-0.0994	***	0.0117	-0.1005	***	0.0117	-0.0999	***	0.0117
Amount of sleep: 9–10h	-0.3447	***	0.0178	-0.3451	***	0.0178	-0.3448	***	0.0178	-0.3446	***	0.0178	-0.3457	***	0.0178
Amount of sleep: >10h	-0.5805	***	0.0312	-0.5811	***	0.0311	-0.5800	***	0.0311	-0.5808	***	0.0312	-0.5827	***	0.0311
Breakfast: Often do not eat	0.0235		0.0206	0.0240		0.0206	0.0240		0.0206	0.0245		0.0206	0.0240		0.0206
Experience2 (Average)	0.2098	***	0.0174	0.2099	***	0.0174	0.2096	***	0.0174	0.2110	***	0.0174	0.2100	***	0.0174
Breakfast: Always eat	0.5302	***	0.0163	0.5304	***	0.0163	0.5300	***	0.0163	0.5312	***	0.0163	0.5304	***	0.0163
Belongings checked: Often not checked	0.1907	***	0.0143	0.1905	***	0.0143	0.1906	***	0.0143	0.1907	***	0.0143	0.1905	***	0.0143
Belongings checked: Generally checked	0.3483	***	0.0122	0.3478	***	0.0122	0.3477	***	0.0122	0.3479	***	0.0122	0.3480	***	0.0122
Belongings checked: Always checked	0.4984	***	0.0122	0.4984	***	0.0122	0.4980	***	0.0122	0.4981	***	0.0122	0.4982	***	0.0122
Female teacher (Average)	0.0389	***	0.0119	0.0443	***	0.0119	0.0447	***	0.0119	0.0412	***	0.0119	0.0392	***	0.0119
Experience (Average)	0.0020		0.0019	0.0018		0.0019	0.0023		0.0019	0.0022		0.0019	0.0021		0.0019
Experience2 (Average)	0.0000		0.0001	0.0000		0.0001	0.0000		0.0001	0.0000		0.0001	0.0000		0.0001
Constant term	-0.7236	***	0.0355	-0.7214	***	0.0354	-0.7551	***	0.0355	-0.7709	***	0.0355	-0.7657	***	0.0355
sigma_u	0.7321			0.7324			0.7319			0.7313			0.7318		
sigma_e	0.5500			0.5480			0.5504			0.5495			0.5504		
rho	0.6393			0.6411			0.6388			0.6391			0.6386		
No. of students	71998			71998			71998			71998			71998		
No. of obs.	205071			205071			205071			205071			205071		
R-sq: within	0.0179			0.0249			0.0167			0.0194			0.0163		
R-sq: between	0.1354			0.1372			0.1354			0.1362			0.1354		
R-sq: overall	0.1671			0.1674			0.1673			0.1677			0.1675		
Chi-sq(3) : H_0 : **m=0**	28.6958			30.0520			30.6693			27.2443			26.6762		
Hansen J statistic (P-value)	0.8490		(0.6541)	0.8210		(0.6632)	0.7320		(0.6935)	0.8650		(0.6488)	0.7620		(0.6832)
Kleibergen-Paap rk Wald F statistic	6.8800			7.4790			9.2510			6.6470			8.3580		

Note: * p < .10, ** p < .05, *** p < .01.

Appendix Table A10
Estimated Results for Equation (7) (Grade 3)

	English			Japanese			Math			Science			Social studies		
	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E	Coef.		R.S.E
Subject (γ_1)	-0.1266	***	0.0098	-0.1496	***	0.0105	-0.0183	**	0.0077	0.0796	***	0.0083	-0.0424	***	0.0069
Girl (γ_2)	-0.0623	***	0.0083	-0.0982	***	0.0083	0.0024		0.0082	-0.0108		0.0083	-0.0828	***	0.0086
Female teacher (γ_3)	-0.1074	***	0.0074	-0.0788	***	0.0073	-0.1134	***	0.0067	-0.1082	***	0.0066	-0.1202	***	0.0068
Subject×Girl (γ_4)	0.1261	***	0.0140	0.3945	***	0.0144	-0.1478	***	0.0111	-0.1676	***	0.0114	0.1478	***	0.0097
Girl×Female teacher (γ_5)	0.2301	***	0.0102	0.1220	***	0.0103	0.1900	***	0.0093	0.1952	***	0.0093	0.2528	***	0.0096
Subject×Female teacher (γ_6)	0.0962	***	0.0149	0.0446	***	0.0152	0.1122	***	0.0187	0.1794	***	0.0191	0.1031	***	0.0178
Subject×Girl×Female teacher (γ_7)	-0.1762	***	0.0212	-0.0875	***	0.0210	-0.1513	***	0.0265	-0.2102	***	0.0271	-0.2398	***	0.0241
Question×Subject (ζ_1)	0.1617	***	0.0196	0.0280		0.0211	0.1517	***	0.0129	0.0838	***	0.0146	0.1206	***	0.0143
Question×Girl (ζ_2)	0.3065	***	0.0123	0.3269	***	0.0124	0.3011	***	0.0125	0.3197	***	0.0125	0.3110	***	0.0120
Question×Female teacher (ζ_3)	0.0979	***	0.0132	0.1042	***	0.0128	0.1147	***	0.0120	0.1112	***	0.0117	0.1120	***	0.0117
Question×Subject×Girl (ζ_4)	-0.0982	***	0.0284	-0.0999	***	0.0292	-0.1173	***	0.0191	-0.1168	***	0.0213	-0.1285	***	0.0211
Question×Girl×Female teacher (ζ_5)	-0.1805	***	0.0186	-0.1046	***	0.0186	-0.1560	***	0.0175	-0.1593	***	0.0170	-0.1700	***	0.0170
Question×Subject×Female teacher (ζ_6)	-0.1423	***	0.0287	-0.0756	**	0.0296	-0.1351	***	0.0303	-0.1406	***	0.0337	-0.0896	**	0.0356
Question×Subject×Girl×Female teacher (ζ_7)	0.2115	***	0.0406	0.0174		0.0408	0.1421	***	0.0422	0.1292	***	0.0476	0.1725	***	0.0498
Experience	0.0040	***	0.0007	0.0043	***	0.0007	0.0039	***	0.0007	0.0043	***	0.0007	0.0039	***	0.0007
Experience^2	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000	-0.0001	***	0.0000
Class size	0.0028	**	0.0012	0.0025	**	0.0012	0.0027	**	0.0012	0.0028	**	0.0012	0.0030	**	0.0012
Class size^2	0.0000		0.0000	0.0000		0.0000	0.0000		0.0000	0.0000		0.0000	0.0000		0.0000
Girl ratio	-0.2912	***	0.0295	-0.2979	***	0.0296	-0.2878	***	0.0295	-0.2864	***	0.0295	-0.2835	***	0.0294
Public urban	-0.0217	**	0.0096	-0.0214	**	0.0096	-0.0210	**	0.0095	-0.0219	**	0.0096	-0.0212	**	0.0096
Public town	-0.0774	***	0.0108	-0.0767	***	0.0108	-0.0767	***	0.0108	-0.0781	***	0.0108	-0.0767	***	0.0108
Private/national	0.3794	***	0.0146	0.3829	***	0.0146	0.3840	***	0.0146	0.3802	***	0.0146	0.3804	***	0.0146
Outside education	0.1765	***	0.0049	0.1770	***	0.0049	0.1842	***	0.0050	0.1704	***	0.0049	0.1773	***	0.0049
Amount of sleep: 6–7h	0.0208	**	0.0083	0.0208	**	0.0083	0.0206	**	0.0083	0.0204	**	0.0083	0.0205	**	0.0083
Amount of sleep: 7–8h	-0.0273	***	0.0093	-0.0272	***	0.0093	-0.0269	***	0.0093	-0.0281	***	0.0093	-0.0279	***	0.0093
Amount of sleep: 8–9h	-0.2469	***	0.0131	-0.2466	***	0.0131	-0.2461	***	0.0131	-0.2481	***	0.0131	-0.2470	***	0.0131
Amount of sleep: 9–10h	-0.4606	***	0.0246	-0.4600	***	0.0246	-0.4595	***	0.0246	-0.4628	***	0.0246	-0.4616	***	0.0246
Amount of sleep: >10h	-0.6688	***	0.0357	-0.6676	***	0.0356	-0.6681	***	0.0356	-0.6705	***	0.0356	-0.6714	***	0.0357
Breakfast: Often do not eat	0.0778	***	0.0197	0.0776	***	0.0197	0.0771	***	0.0197	0.0774	***	0.0197	0.0780	***	0.0197
Experience2 (Average)	0.2476	***	0.0170	0.2475	***	0.0170	0.2464	***	0.0170	0.2474	***	0.0170	0.2472	***	0.0170
Breakfast: Always eat	0.5347	***	0.0157	0.5348	***	0.0157	0.5332	***	0.0157	0.5346	***	0.0157	0.5340	***	0.0157
Belongings checked: Often not checked	0.1410	***	0.0157	0.1419	***	0.0157	0.1402	***	0.0157	0.1410	***	0.0157	0.1407	***	0.0157
Belongings checked: Generally checked	0.2611	***	0.0136	0.2624	***	0.0136	0.2606	***	0.0136	0.2613	***	0.0136	0.2608	***	0.0136
Belongings checked: Always checked	0.3996	***	0.0136	0.4007	***	0.0136	0.3992	***	0.0136	0.3999	***	0.0136	0.3996	***	0.0136
Female teacher (Average)	0.0206	*	0.0124	0.0277	**	0.0124	0.0254	**	0.0124	0.0229	*	0.0124	0.0174		0.0124
Experience (Average)	0.0132	***	0.0023	0.0128	***	0.0023	0.0129	***	0.0023	0.0130	***	0.0023	0.0131	***	0.0023
Experience2 (Average)	-0.0003	***	0.0001	-0.0003	***	0.0001	-0.0003	***	0.0001	-0.0003	***	0.0001	-0.0003	***	0.0001
Constant term	-0.8766	***	0.0394	-0.8665	***	0.0393	-0.8995	***	0.0394	-0.9096	***	0.0393	-0.8897	***	0.0394
sigma_u	0.7371			0.7388			0.7383			0.7374			0.7374		
sigma_e	0.5537			0.5482			0.5504			0.5519			0.5518		
rho	0.6393			0.6450			0.6427			0.6409			0.6410		
No. of students	65569			65569			65569			65569			65569		
No. of obs.	182246			182246			182246			182246			182246		
R-sq: within	0.0061			0.0164			0.0105			0.0111			0.0062		
R-sq: between	0.1208			0.1235			0.1234			0.1219			0.1216		
R-sq: overall	0.1539			0.1538			0.1553			0.1535			0.1546		
Chi-sq(3) : H_0 : **m=0**	42.5284			42.1232			41.1315			41.5016			41.1377		
Hansen J statistic (P-value)	0.2560		(0.8798)	0.3260		(0.8497)	0.3210		(0.8517)	0.2720		(0.8728)	0.2240		(0.894)
Kleibergen-Paap rk Wald F statistic	15.5630			16.8630			18.6950			15.3880			17.7480		

Note: * p < .10, ** p < .05, *** p < .01.