After-School Tutoring and the Distribution of Student Performance

MIN-HSIUNG HUANG

As more primary and secondary students worldwide seek after-school tutoring in academic subjects, concerns are being raised about whether after-school tutoring can raise average test scores without widening the variability in student performance, and whether students of certain ability levels may benefit more than others from after-school tutoring. To address these questions, I compared the distributions of student performance across countries with differing levels of participation in after-school tutoring, while controlling for country-level unobserved heterogeneity using a fixed-effects model. Participating in either mathematics or science tutoring after school is found to raise national average performance without widening the dispersion in student performance. In science, low-performing students benefit more from tutoring than do high-performing students. In mathematics, high-performing students benefit more from tutoring than do low-performing students.

Background

Throughout Asia, particularly in Hong Kong, Japan, South Korea, and Taiwan, parents invest in after-school tutoring for their children in hopes of improving their academic performance (Zeng 1999; Bray and Kwok 2003). Some have described these out-of-school study sessions as a "shadow education" system (Stevenson and Baker 1992; Bray 1999, 2009).[1] The prevalence of after-school tutoring in these East Asian countries is often regarded as a distinctive educational phenomenon related to a Confucian cultural heritage

This research was supported by the National Science Council in Taiwan (grant no. 98-2410-H-001-049-MY2).

[1] Some researchers define "shadow" education as a type of out-of-school instruction that charges fees in the private sector (Stevenson and Baker 1992; Grodsky 2010). However, Claudia Buchmann, Dennis J. Condron, and Vincent J. Roscigno (2010) argue against the public/private conceptual distinction to define shadow education. They suggest that such a sharp public/private distinction may have been appropriate for the case of Japan in the late 1980s, but it does not exist in other countries, such as the United States, where the private mingles with the public dimensions of the education system. I do not include the public/private distinction as a defining characteristic of after-school tutoring, because the governments of some countries are supportive of such tutoring (Bray 2010). Throughout this article, I refer to "after-school tutoring" rather than "shadow" education, and I acknowledge that after-school tutoring is not limited to fee-paying tutoring.

Received September 19, 2011; revised March 22, 2012, June 29, 2012, and May 10, 2013; accepted May 10, 2013; electronically published June 25, 2013

Comparative Education Review, vol. 57, no. 4.

0010-4086/2013/5704-0001$10.00

or the tremendous competition for entrance into elite high schools and universities (Zeng 1999). However, tutoring is prevalent across countries and over time. Indeed, several studies make the point that the expansion of after-school tutoring is now a worldwide phenomenon (Baker and LeTendre 2005; Bray 2006, 2009, 2011; Dang and Rogers 2008). For example, the expansion of after-school tutoring has been observed in several formerly socialist countries, including China, Vietnam, Cambodia, and some eastern European countries, following their transitions to market-based systems (Bray 2006; Silova et al. 2006; Dang 2007). The scale of after-school tutoring has also expanded over time in Canada (Aurini and Davies 2004; Davies 2004), Kenya (Onsomu et al. 2005), Mauritius (Kulpoo and Soonarane 2005), Turkey (Tansel and Bircan 2006), South Korea (Kwak 2004), Singapore (Tan 2009), Japan (Ono 2007), Malawi, Namibia, Zambia, Zanzibar (Paviot et al. 2008), Germany (Mischo and Haag 2002), and most member states of the European Union (Bray 2011). David Baker and coauthors (2001) used data from the Trends in International Mathematics and Science Study, conducted in 1995 (TIMSS 1995), finding that, in 31 out of 41 TIMSS participating countries, more than 20 percent of eighth-grade students received weekly after-school tutorials.[2]

As after-school tutoring has spread worldwide, questions are arising over its impact on the distribution of student performance. Three key questions are, first, whether the use of after-school tutoring promotes student performance, second, whether students perform more unequally as after-school tutoring becomes more widely practiced, and, finally, whether students of certain ability levels benefit more from after-school tutoring than do others. This article addresses these three questions.

After-School Tutoring Participation and Average Performance

To assess the effects of after-school tutoring on academic performance, it is useful to investigate two distinct research areas. Most previous studies have addressed the question of whether or not individual students do better in school if they take extra lessons after school.[3] A second question is whether increasing the percentage of students receiving after-school tutoring would lead to a better national average performance. This article aims to address that second question.

The second question calls for a country-level analysis. It is important because individual students may benefit from after-school tutoring, but this benefit may come at the expense of those who do not receive such tutoring. As a consequence, the growth of after-school tutoring might not always lead

[2] This finding even led them to conclude that "as mass schooling as a world norm continues to intensify the importance of schooling, shadow education itself becomes an institutionalized component of mass education that grows and expands" (Baker et al. 2001, 14).

[3] Tansel and Bircan (2006); Dang (2007); Ono (2007); Gurun and Millimet (2008); Kang (2009); Kuan (2011).

to improvements in national average performance even though there is a positive effect on those who participate. Mark Bray (1999, 2003), for example, reported several ways in which after-school tutoring may have a positive learning effect on those who participate at the expense of those who do not. First, when some students in the classroom receive private tutoring after school but others do not, the result may be a more diverse range of student performance within the classroom. School teachers, when faced with a greater range of performance within the classroom, may gear their instruction to those at the upper range of the curve and allow achievement gaps between students to grow. As Bray wrote, "private tutoring can distort the curriculum in the mainstream system, upsetting the sequence of learning planned by mainstream teachers, and exacerbating diversity in classrooms" (1999, 18). Another reason there may be no national effect, despite apparent individual effects, is that those who participate in after-school tutoring may show less interest in school lessons because their tutors already covered the topics, or because their tutors do a better job of presenting the material. This lack of attention may create a relatively poorer learning atmosphere in the classroom for those who do not participate in after-school tutoring. Finally, when a system of after-school tutoring arises, the mainstream school system may lose talented teachers to private tutoring businesses, as is evident in Costa Rica, Lithuania, and Senegal (Bray 1999, 54). This loss of talent from the public sector would exacerbate the achievement gap within each country.

In estimating the effect of after-school tutoring participation on student performance, a common problem is that students are not randomly selected into after-school tutoring participation. Also, self-selection processes may work in opposite directions. On the one hand, a spurious positive correlation between student achievement and tutoring participation may be observed if students with greater motivation or more favorable family backgrounds are more likely to enroll in after-school tutoring, because these factors may have led to better academic performance even if they had not received after-school tutoring. On the other hand, a spurious negative correlation may be observed in countries where lower-performing students are more likely to participate in after-school tutoring. For example, parents may try to equalize academic achievement among children of different abilities by investing more in their lower performing children in a compensatory manner.

While the selection bias can be corrected with an instrumental variable methodology or propensity score matching, an individual-level analysis is likely to violate a crucial assumption—the stable unit treatment value assumption (SUTVA)—under which "a change in treatment status of any individual does not affect the potential outcomes of other individuals" (Winship and Morgan 1999, 663). If some students in the school take part in after-school tutoring, this may affect the atmosphere of school classrooms, thus the necessary assumption of SUTVA. Such a violation of SUTVA may lead to serious bias in research findings (Sobel 2006). By shifting the analysis to a more aggregate level and estimating macro treatment effects at that level, as the present study intends, SUTVA can be more reasonably maintained, according to Markus Gangl (2010, 40).

After-School Tutoring Participation and Performance Inequality

Many researchers are concerned that the rapid growth of after-school tutoring may magnify performance inequality and strengthen the correlation between family socioeconomic status and student achievement.[4] However, the correlation between the prevalence of after-school tutoring and the level of inequality in student performance has not been investigated directly. As after-school tutoring becomes more widely practiced, inequality in student performance may increase or decrease, depending on whether the use of tutoring is dominated by high-performing students seeking an enrichment strategy, or by low-performing students seeking remedial help. Assuming that tutoring is effective, inequality in student performance would increase when an enrichment strategy prevails, and inequality would decrease when its use is remedial. Previous studies using international data sets report that, worldwide, most students participating in after-school tutoring are those who are performing poorly but wish to reach an acceptable standard of achievement (Baker et al. 2001). This suggests that, as greater proportions of poorly performing students receive tutoring, their performance will become closer to the norms in each country

However, it is also likely that, within certain countries, both remedial and enrichment strategies coexist. For example, Claudia Buchmann and coauthors (2010) find that most American students use test preparation after school for remedial purposes, but those with favorable socioeconomic backgrounds are more likely to use it as an enrichment strategy. Soo-Yong Byun and Hyoonjun Park (2012) further suggest that East Asian American students, relative to students of other ethnic groups, are more likely to participate in after-school tutoring, such as taking a commercial SAT test preparation course, for enrichment purposes.

Who Benefits from Participating After-School Tutoring?

While after-school tutoring provides more learning opportunities for those who participate, it does not necessarily translate into higher level performances for everyone. According to a recent review by Mark Bray (2011, 47–51), previous research presents highly inconsistent findings with respect to the effectiveness of tutoring. Previously, Bray (1999, 50) suggested that whether or not after-school tutoring is effective "presumably depends on (a) the content and mode of delivery of the tutoring, (b) the motivation of the tutors and the tutees, (c) the intensity, duration and timing of tutoring; and

[4] Baker and LeTendre (2005); Buchmann et al. (2010); Byun and Kim (2010); Park et al. (2011).

(d) the types of pupils who receive tutoring." In assessing the effect of after-school tutoring on student performance, the present study considers whether students of certain ability levels benefit more than others from after-school tutoring.

In this article, I conduct a country-level analysis, using data from the TIMSS. Countries, however, differ in many respects. Simply comparing achievement levels in countries with higher and lower rates of tutoring participation is not an appropriate approach to evaluating the effects of tutoring participation. This is because country differences in student achievement are driven strongly by a number of unobserved social and educational factors other than the rate of tutoring participation after school. Unless a model is capable of controlling for these unobserved country factors, the effects of tutorial participation cannot be estimated correctly.[5]

To test the hypothesis that greater rates of after-school tutoring contribute to better student performance at the country level, I use data from fourth- and eighth-grade students on the TIMSS mathematics and science assessments for countries that participated at both grade levels in the 2003 survey year. To control for the effects of unmeasured country-level features, I compared the distributions of academic performance for students between grade 8 and grade 4 across countries with different levels of tutoring participation, using a country-level fixed-effects model. By modeling changes between grade levels within countries with a country-level fixed-effects strategy, I thus limit the effect of biases introduced by unobserved or uncontrolled-for differences between countries.

In addition to the application of the country-level fixed-effects method, this research improves on previous studies in two respects. First, this study conducts a subject-specific analysis, with separated effect estimated for each academic subject (mathematics and science). Many past studies were not subject specific. The effect of after-school tutoring on student performance can be more precisely estimated when the subject of after-school tutoring participated is the same as the subject of student performance. Even when past investigations conducted subject-specific analysis, it was often limited to one academic subject. A one-subject analysis, while subject specific, does not allow for a comparison between subjects in the effects of after-school tutoring on student performance. It is likely that the impact of after-school tutoring on student performance may differ by subject.

Second, student performance, in my research, has been measured by a

[5] For example, the analysis by Baker et al. (2001) lacked any controls for unobserved country factors. To assess the association between national use of shadow education and national achievement levels, Baker et al. used data from the TIMSS 1995 and conduct a country-level regression analysis. Baker et al. found the use of shadow education to be ineffective in raising national mathematics achievement, but this finding was based on a regression model that controlled only for one variable—the presence of high-stakes tests (see table 2 of Baker et al.). It would surprising if a more sophisticated regression model did not lead to different findings.

curriculum-based test, with a testing time of 72 minutes for fourth graders and 90 minutes for eighth graders (Martin et al. 2004). Student performance, when measured by a curriculum-based test, is sensitive to the use of after-school tutoring. After all, after-school tutoring aims to improve performance in school, not to raise general ability or problem-solving ability, which is usually measured by IQ or aptitude tests.

A Survey of Research

Due to selection bias, evaluating the consequences of after-school tutoring on student performance is not straightforward. To correct for selection bias, several researchers have employed an instrumental variable method to estimate the causal effect of after-school tutoring. An instrumental variable is useful only when it is highly correlated with participation in after-school tutoring, but not correlated with student performance. In Kang's (2009) study of South Korea, for example, a student's birth order was found to be highly correlated with expenditures on after-school tutoring, but there was no correlation between birth order and student test scores. A student's birth order, therefore, was used as an instrumental variable to address the possible endogeneity of household spending on after-school tutoring. With correction for selection bias, Kang found that a 10 percent increase in tutoring expenditure increases a student's test performance by 1.1 percentile. This was considered a modest effect, but the effect was more than five times greater than that estimated by the ordinary least square (OLS) model, which does not correct for selection bias. Kang suggested that the OLS estimates were seriously biased downward because parents spend their money in a compensating manner, with lower performing siblings receiving more private tutoring within a family. This interpretation contrasts with that of Baker et al. (2001), who suggested that private tutoring in South Korea is predominantly used by high-performing students as an enrichment strategy. One possible shortcoming in the Kang study is that neither the amount of tutoring expenditures nor student test performance is subject specific. In Kang's study, student test performance was measured as an average over the three main subjects (Korean, mathematics, and English).

Dang (2007) also applied an instrumental variable strategy and found a positive effect of tutoring expenditure on student performance in Vietnam. Dang's research has the advantage of using a nationally representative data set, but expenditures on private tutoring are not subject specific and may include spending on nonacademic subjects. Likewise, student performance was measured according to three ranks (poor, average, and good) and was not subject specific.

Gurun and Millimet (2008) used a large sample of 90,410 Turkish students taking the 2002 university entrance exam for the first time to discover which students received a university placement. Due to the lack of instrumental variables, Gurun and Millimet employed parametric and seminonparametric techniques from the program evaluation literature to address the possible endogeneity of tutoring participation. They distinguished between two treatment variables: (*a*) whether the students received any private tutoring prior to the exam, and (*b*) if they did receive tutoring prior to the exam, whether the student spent more than US$1,275 dollars on private tutoring. Once nonrandom selection was taken into account, they found that tutoring actually had a negative effect on university placement. However, for those who participated in tutoring, a high expenditure (spending more than US$1,275) on tutoring did pay off, resulting in a higher probability of entering university. This finding has an important implication for social inequality, because it means that private tutoring pays off only after a heavy investment. For the majority of parents who cannot spend so much on tutoring, a moderate investment would not be fruitful and could even be harmful.

A shortcoming common to the studies of Kang (2009), Dang (2007), as well as Gurun and Millimet (2008), is their insensitivity to academic subject. A subject-specific analysis permits a more rigorous research design because tutoring participation is usually specific to a subject, such as mathematics or science, and its impact on student performance is also limited to that subject. In a subject-specific analysis, the subject of tutoring participation would be consistent with the subject of student performance. One of the best recent examples of a subject-specific analysis is that of Ping-Yin Kuan (2011), who estimated the effects of participation in mathematics tutoring on mathematics performance. Using longitudinal data from the Taiwan Education Panel Study (TEPS) and adopting a technique of propensity score matching to tackle the problem of selection bias, Kuan found a positive average treatment effect of 2.8 in ninth-grade mathematics scores, which is equivalent to a test score gain of 0.13 standard deviations. While this effect is statistically significant, Kuan concludes: "The average causal effect of math cramming in Grade 9 is positive but fairly small" (2011, 363). However, one might question whether an effect of this size is "small," given the very short period of time (roughly 2 months) of exposure to after-school tutoring among those who participated. Kuan's "treatment" variable was a dummy variable: "one for participation in math cramming during the first semester of Grade 9 and zero for nonparticipation" (Kuan 2011, 353). Because student performance in mathematics was also assessed in the first semester of grade 9, this leaves a range of 1–3 months in exposure to after-school tutoring for those who use it. Kuan found a test-score gain of 0.13 standard deviations for tutorial participants in just 2 months. This could be regarded as a large effect that is substantively as well as statistically significant. Moreover, when students engage in after-school tutoring for more than 2 months, which is usually the case in Taiwan, the benefit may accumulate over time.[6]

[6] In a footnote, Kuan reported the cumulative effects of long-term participation in after-school

Methodologically, the four studies reviewed above go beyond previous studies by assessing the role of nonrandom selection into tutoring participation with an IV methodology or a propensity score matching approach. The methodology used by these four studies, however, assumes that students participating in after-school tutoring will not affect the educational outcomes of nonparticipants. If this assumption, formally called the SUTVA, were not to hold, findings of these studies could be biased.

Data

In the present study, I conduct country-level analyses and use data from fourth- and eighth-grade students on the TIMSS mathematics and science assessments for countries participating at both grade levels in the 2003 survey year.[7] Not all these countries provide sufficient data for an analysis of after-school tutoring, even though they meet the criterion of participating at both grade levels. For example, some countries are excluded from the analysis because their eighth graders did not report whether they participated in after-school tutoring. These include Armenia, Cyprus, Hungary, Latvia, Lithuania, Moldova, the Netherlands, Russian Federation, Slovenia, and the Flemish region of Belgium. In addition, England is excluded from the analysis because the eighth-grade sample from England does not satisfy guidelines for sample participation rates (Mullis et al. 2004). Morocco and Tunisia are also excluded from the analysis because more than one-fourth of the fourth graders in Morocco and nearly half of the fourth graders in Tunisia failed to respond to the question on after-school tutoring participation. An additional reason to exclude Morocco and Tunisia is that they have a much lower school enrollment ratio in secondary education relative to other TIMSS par-

tutoring on student performance, using the participation of after-school tutoring from grade 5 to grade 7 as the treatment variable and estimating its effects on the performance of grade 7 and grade 9 students in mathematics. Kuan found a positive, cumulative effect of two to three points in ninth-grade mathematics when students participated in after-school tutoring for 2 or 4 years. This is indeed a fairly small effect, given a very long period of exposure to after-school tutoring. However, this finding of Kuan should be taken with caution for several reasons. First, after-school tutoring experience reported by students is subject specific in grades 7 and 9, but not in grade 5. Therefore, Kuan's analyses estimating cumulative effects of tutoring starting from grade 5 are not subject specific. Second, the survey question on after-school tutoring experience is retrospective, asking students in grades 11 or 12 to report their after-school tutoring experience in grades 5, 7, and 9. Finally, the sample used to estimate the cumulative effects of after-school tutoring on student performance is limited to the core sample of TEPS, which is somewhat selective and derivative of a much smaller sample. Min-Hsiung Huang and Taissa S. Hauser (2010) provide a detailed introduction to the TEPS.

[7] In TIMSS 2003, the question on after-school tutorial participation in the student questionnaire is identical with those for grades 4 and 8. In other survey years, such as the TIMSS 1995, the wording of the question is different for these two grade levels. The TIMSS 1999 covers students in grade 8 only, and it cannot be compared with the grade 4 sample from TIMSS 1995 because the wording of the question is different. In TIMSS 2007, the student questionnaire no longer has a question on after-school tutorial participation in either grade level.

ticipating countries.[8] This may introduce a bias to a between-grade comparison across countries because eighth-grade students, relative to fourth-grade students, represent a more selective group in Tunisia and Morocco. After excluding countries with insufficient data, a total of 14 countries are available for this analysis.

TIMSS 2003 targeted those at the end of the fourth year of formal schooling and those at the end of the eighth year, who were essentially in the fourth and eighth grades in most countries. The TIMSS adopted a two-stage clustered sample design within each country, with schools sampled in the first stage and classrooms sampled in the second stage (Foy and Joncas 2000). In most countries, there was only one classroom randomly selected from a sampled school. When that classroom was sampled, all students in the classroom were surveyed.

The design of the TIMSS achievement test aimed to reflect as many curricular goals as possible in all participating countries. To ensure that test items reflected different curricula among countries, the construction of mathematics and science tests were based on the TIMSS curriculum frameworks, which were set by consensus and endorsed by all participating countries.

The TIMSS test included a series of multiple-choice test items and open-ended response questions requiring short or more elaborate explanations. Not all students were given the same test questions. To minimize the response burden on individual students, matrix-sampling techniques were used to divide the test item pool so that each sampled student responded to only a portion of the test questions, but a portion that nonetheless covered all subject areas (Garden 2000). Therefore, students answered different test items depending upon which test booklets they received. Based on item response theory, student responses were scaled to provide accurate estimates of achievement that could be compared across countries (Yamamoto 2000). In addition, a "plausible values" method was used to produce proficiency scores in mathematics and in science. The achievement scores, therefore, are available as a set of five plausible values for every individual student in each subject. Country-specific statistics used in this study, such as mean, standard deviation, interquartile range, and scores at various percentiles of the performance distribution, were estimated five times, each using a different plausible value. The final statistic used for analysis is the average of five estimates.

In the TIMSS base year, 1995, the international proficiency scores (mathematics or science) were scaled to have a mean of 500 and a standard deviation of 100, and the metric of the 1995 scale has been preserved, so test scores in later survey years can be compared with those in 1995. In the appendix,

[8] In secondary education, Morocco has a net school enrollment rate as low as 39 percent (Sarri 2002), and it is 65 percent in Tunisia (Dridi 2002). Similarly, Mullis et al. (2004, 29) report a net secondary school enrollment ratio of 31 percent in Morocco and 68 percent in Tunisia.

available online, table A1 presents the numbers of students, classrooms, and schools for grade 4 and grade 8 separately for each country.

Method

Measure of After-School Tutoring Participation

In the TIMSS 2003, the student questionnaire collects information on participation in after-school tutoring in mathematics by means of a question: During this school year, how often have you had extra lessons or tutoring in mathematics that were not a part of your regular class? The student was given four choices: (*a*) every, or almost every day, (*b*) once or twice a week, (*c*) sometimes, and (*d*) never or almost never. The same question is used for the subject of science. In the present study, participation in after-school tutoring is measured as a dichotomous variable with students who "never or almost never" receive any tutoring coded as 0. Students receiving tutoring sometimes or regularly are coded as 1.

The Dependent Variables

The dependent variable, student performance in mathematics or science, is measured in three aspects: (*a*) country mean test scores, (*b*) the variability of test scores within a country, and (*c*) student test score at the 5th, 25th, 50th, 75th, and 95th percentiles of the national distribution. The variability of test scores at a grade level within a country is indicated by multiple measures: the standard deviation, the interquartile range, and the test score range between the 5th and the 95th percentile of the test score distribution. Student test scores at the 5th, 25th, 50th, 75th, and 95th percentiles of the national distribution are also used as dependent variables because of concerns over whether the impact of tutoring differs in degree and direction for students at different points on the performance distribution. To make achievement test scores comparable between grade 4 and grade 8, test scores of each grade level for all students from all countries are standardized at the individual level to have a mean score of 500 and a standard deviation of 100. For each country, tables A2 and A3 (in the online appendix) report the mean score, the standard deviation, and the test scores at the 5th, 25th, 50th, 75th, and 95th percentiles by subject for grade 4 and grade 8, respectively.

Specification of the Fixed-Effects Regression Model

To assess the effects of tutoring participation on student performance, I adopt a fixed-effects model in order to control for unobserved country-level heterogeneity (Allison 1994). In the present study, data from two grade levels for each country are used to control for unobserved country heterogeneity.[9]

[9] While the fixed-effects model is often used with two- or multiple-period panel data, it can be applied to other data structures as well. For example, sibling data with a pair of observations within a family can be used to control for unobserved family and background characteristics (Hauser 1991; Geronimus and Korenman 1992).

This can be demonstrated from the subtraction of the following two grade-specific equations:

$$y_{i8} = (\beta_0 + \delta_0) + \beta_1 x_{i81} + \beta_2 x_{i82} + \beta_3 x_{i83} + \beta_4 x_{i84} + a_i + u_{i8} \qquad \text{(gr. 8)} \qquad (1)$$

$$y_{i4} = \beta_0 + \beta_1 x_{i41} + \beta_2 x_{i42} + \beta_3 x_{i43} + \beta_4 x_{i44} + a_i + u_{i4} \qquad \text{(gr. 4)} \qquad (2)$$

The subscript i in equations (1) and (2)denotes a country, a_i denotes an unobserved country effect that is constant between grade 4 and grade 8, and a_i may be correlated with the four explanatory variables in each equation. The first explanatory variable is an indicator of tutoring participation, which is x_{i81} in equation (1) for grade 8 and x_{i41} in equation (2) for grade 4. The second explanatory variable is net primary or secondary school enrollment ratio, which is x_{i82} in equation (1) for grade 8 and x_{i42} in equation (2) for grade 4. Note that x_{i82} indicates net secondary school enrollment ratio, while x_{i42} denotes net primary school enrollment ratio. The third explanatory variable is an indicator of classroom homogeneity in student performance, which is x_{i83} in equation (1) for grade 8 and x_{i43} in equation (2) for grade 4.[10] The fourth and final explanatory variable is subject specific, and it indicates the direction and the degree of selectivity in the participation of after-school tutoring, which is x_{i84} in equation (1) for grade 8 and x_{i44} in equation (2) for grade 4. By subtracting equation (2) from equation (1), we obtain:

$$(y_{i8} - y_{i4}) = \delta_0 + \beta_1(x_{i81} - x_{i41}) + \beta_2(x_{i82} - x_{i42}) + \beta_3(x_{i83} - x_{i43}) + \beta_4(x_{i84} - x_{i44}) + (u_{i8-}u_{i4}) \qquad (3)$$

or

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_{i1} + \beta_2 \Delta x_{i2} + \beta_3 \Delta x_{i3} + \beta_4 \Delta x_{i4} + \Delta u_i. \qquad (4)$$

Equation (3) is equivalent to equation (4), which is the model used in this study to estimate the effects of tutoring participation on student performance. Note that the unobserved country effect, a_i, does not appear in either equation (3) or (4) because it has been differenced away. Country-specific factors are differenced out because equation (4) models changes between grade 4 and grade 8 within countries.

The model presented in equation 4 controls for school enrollment ratio in order to avoid the bias when eighth-grade students, relative to fourth-grade students, represent a more selective group. The model also controls for classroom homogeneity in student performance because homogeneously grouped

[10] Huang's (2009) measure of classroom homogeneity is used for analysis. In that study, Huang decomposed the total mathematics score variance into within- and between-classroom components, and he suggested "the percentage of total mathematics score variation that occurs between classrooms reflects the intensity of sorting students by performance level into different classrooms or schools at a given grade within a country; the higher the percentage, the stronger the intensity of sorting. This percentage, therefore, is used to indicate classroom homogeneity. If a policy of tracking were introduced between Grade 4 and Grade 8 in a country, the percentage used to indicate classroom homogeneity should increase after tracking is introduced, and it should increase more for countries where tracking is more efficiently and extensively implemented" (2009, 784).

classrooms may be more prevalent at the grade 8 level than in grade 4, and between-grade changes in classroom homogeneity are likely to correlate with between-grade changes in the prevalence of after-school tutoring. For example, in an educational system with tracking, which usually occurs at the secondary level, students may be more actively engaged in after-school tutoring in order to attend better schools. Finally, the model controls for the direction and degree of selectivity in the participation of students in after-school tutoring. This control variable deals with the concern that participants in after-school tutoring may not be selected in the same way during grades 4 and 8. For example, it is possible that after-school tutoring is practiced more widely among low-performing students for remedial purposes in grade 4, but it is more likely to be used by high-performing students in grade 8 as an enrichment strategy.[11] The differential selectivity in the participation of after-school tutoring between fourth and eighth graders, if not taken into account, may bias the findings of the fixed-effects model.

Measuring the direction and the degree of selectivity into after-school tutoring is carried out in two steps. First, I examine the nature of the relationship between the use of after-school tutoring and academic performance at the student level for each academic subject at each grade level in each country. This is done by regressing a dichotomous indicator of tutoring participation on student test score, controlling for a set of variables previously shown to affect decisions to use after-school tutoring. Specifically, the logistic regression model is subject specific, and it includes the following independent variables: student mathematics or science test score; gender; number of books at home; whether or not the language spoken at home is the same as the language used in the TIMSS assessments; number of people living in the area where the school is located; whether or not students in the school were grouped by ability within mathematics (or science) classes; whether or not the school offers enrichment mathematics (or science); and whether or not the school offers remedial mathematics (or science).[12]

Second, the effect of test scores on tutoring participation, as indicated by the regression coefficient of the logistic regression model, is used to determine the extent to which the use of after-school tutoring in a grade level

[11] This is, in fact, the case for Taiwan.

[12] Note that mathematics test performance is likely the outcome of after-school tutoring in mathematics, since they were measured at the same time. Therefore, the problem of endogeneity occurs when mathematics test scores are used as a predictor of participation in mathematics tutorials. The same problem applies to the subject of science. To avoid the problem, I use mathematics test scores as a predictor of participation in science tutorials, and I use science test scores as a predictor of participation in mathematics tutorials. The rationale for adopting this strategy is that participating in after-school tutorials in mathematics is supposed to improve performance in mathematics, not science. Likewise, participating in after-school tutoring in science is supposed to raise performance in science, not mathematics. When either mathematics or science test scores are used as a predictor, they are treated as a proxy for the general level of student achievement. This strategy circumvents the problem of endogeneity because mathematics test performance is unlikely the outcome of after-school tutoring in science, and science test performance is unlikely the outcome of after-school tutoring in mathematics.

within a country is dominated by a remedial or enrichment purpose. A significant positive coefficient indicates that after-school tutoring is more likely to be used by high-performing students as an enrichment strategy. The higher the positive coefficient, the more that after-school tutoring is dominated by high-performing students. In contrast, a significant negative coefficient suggests that after-school tutoring is used mostly by low-performing students for remedial purposes. The stronger the negative effect, the more that after-school tutoring is dominated by low-performing students. An insignificant coefficient signifies an equal prevalence between enrichment and remedial strategies. I rescaled these coefficients to control for the direction and the degree of selectivity in the participation of after-school tutoring.

The measure of classroom homogeneity, the net school enrollment ratio, and the subject-specific measure of selectivity in the participation of after-school tutoring are presented in table A4, available online. As in the study by Eric Hanushek and Ludger Woessmann (2006, C67), I have used approximately contemporaneous measures of student performance at two different grade levels (grades 4 and 8, in 2003) in order to minimize any contamination due to variations in school policies.

Data Imputation

In most countries, no more than 4 percent of the sampled students failed to respond to the question on the participation of after-school tutoring in either subject. In Iran and Norway, however, the percentage of students not responding to the question is as high as 12 percent in grade 4 for mathematics as well as for science. For those who did not respond to the question on tutoring participation, data were imputed using the data of those students with nonmissing values and data on a set of explanatory variables available for all students. Take the subject of mathematics for example: this set of explanatory variables includes age, gender, number of books at home, whether or not the language spoken at home is the same as the language used in the TIMSS assessments, number of people living in the area where the school is located, whether or not students in the school were grouped by ability within mathematics classes, whether or not the school offers enrichment mathematics, and whether or not the school offers remedial mathematics. The small amount of missing data in these explanatory variables was imputed by taking the average value at the lowest level available, such as class average, school average, or country average. For each country in each grade level, this set of explanatory variables was then used to predict whether or not a student is likely to receive after-school tutoring in mathematics in the sample of students who have nonmissing values on tutoring participation in mathematics. The coefficients from these regressions were then used to impute the value of tutorial participation (either 1 or 0) for the students with missing data. For the subject of science, I followed the same procedure of

TABLE 1
PERCENTAGE OF STUDENTS PARTICIPATING IN AFTER-SCHOOL TUTORING BY SUBJECT AND GRADE, TIMSS 2003

	Mathematics			Science		
Country	Grade 4	Grade 8	Difference	Grade 4	Grade 8	Difference
New Zealand	63.6	24.8	−38.8	50.3	15.9	−34.4
Scotland	52.7	16.0	−36.7	41.0	12.3	−28.7
Australia	49.3	21.9	−27.4	35.5	11.7	−23.7
Japan	66.6	46.3	−20.3	52.4	33.0	−19.4
United States	53.9	34.4	−19.5	42.9	20.4	−22.5
Norway	29.2	11.7	−17.5	22.3	5.5	−16.8
Canada (Ontario)	45.7	29.7	−16.0	37.1	16.3	−20.8
Italy	34.2	20.6	−13.6	29.5	10.1	−19.4
Singapore	81.6	69.5	−12.1	64.4	51.1	−13.3
Canada (Quebec)	35.8	30.9	−4.9	26.9	14.2	−12.7
Philippines	78.4	75.7	−2.7	79.6	75.4	−4.1
China (Hong Kong)	56.3	55.3	−1.0	46.3	37.7	−8.6
Taiwan	60.3	59.5	−.8	40.4	40.5	.1
Iran	61.7	66.8	5.1	45.2	49.9	4.6

NOTE.—Subtracting the percentage in grade 4 from the percentage in grade 8 gives the values displayed in the "Difference" column. Countries are displayed in increasing order of the fourth column. Missing values are imputed.

data imputation. The differences in research findings introduced by the imputation of missing data, however, are found to be trivial. Results of the fixed-effects regression analysis are presented with the imputation of missing data.

Results

Prevalence of After-School Tutoring

Table 1 reports the percentage of students participating in after-school tutoring by subject and grade for each country.[13] After-school tutoring in mathematics is quite common among students in grade 4. The percentage of fourth graders participating in mathematics tutoring ranges from 29 percent in Norway to 82 percent in Singapore. The majority of fourth graders received some tutoring in mathematics after school in all countries except for students in Norway, Italy, and the two Canadian provinces (Quebec and Ontario). East and Southeast Asian countries generally show a higher percentage of students receiving mathematics tutoring after school in grade 4, but there are notable exceptions. For example, mathematics tutoring in grade 4 in New Zealand or Iran is more prevalent than in Taiwan or Hong Kong. To give another example, mathematics tutoring in grade 4 is almost equally prevalent in the United States and Hong Kong. Scotland and Australia have a level of prevalence close to 50 percent, which is not much lower than that in some East or Southeast Asian countries.

After-school tutoring in mathematics in grade 8 is not as widespread as

[13] Those who participate in after-school tutoring differ in the number of days per week that they are exposed to tutoring. Even those who have the same number of days of tutoring per week differ in the number of hours per day that they have tutorial classes. The present study does not distinguish between differing levels of participation within the set of those who participate in after-school tutoring.

in grade 4. Only in the Philippines, Singapore, Iran, Taiwan, and Hong Kong do the majority of eighth graders receive tutoring in mathematics. For most countries, there is less participation in mathematics tutoring in grade 8 than in grade 4. The difference is most dramatic in New Zealand, Scotland, and Australia, but it is also quite significant in Japan, the United States, Norway, Italy, Singapore, and Ontario, Canada. The difference is insignificant in Taiwan, Hong Kong, and the Philippines. In grade 8, the use of after-school tutoring in mathematics is dominated by East and Southeast Asian countries. Only Iran has a level of prevalence comparable with that in these Asian countries.

In science, as table 1 demonstrates, after-school tutoring is not as common as it is in mathematics. The Philippines, however, is exceptional, with an equally large share of students receiving tutoring in either subject. Compared with other countries, the Philippines stands out by having more than three-quarters of the students participating in tutoring after school, and this high rate of participation persists across different subjects and grades. As in the case of mathematics tutoring, there is a decrease in participation in science tutoring from grade 4 to grade 8 for most countries. Countries showing a decrease in participation in mathematics tutoring also demonstrate a comparable decrease in participation in science tutoring. In science, the dominance of East and Southeast Asian countries in using after-school tutoring is more apparent in grade 8 than in grade 4, as is the case in mathematics.

After-School Tutoring, Inequality, and Efficiency

As table 2 reveals, participation in mathematics tutoring after school does not widen the achievement gap in mathematics. The same applies to science tutoring. In science, as a matter of fact, the use of tutoring after school significantly reduces achievement inequality when inequality is indicated by interquartile range that is less sensitive to extreme values. At the same time, the use of after-school tutoring significantly raises national mean performance in both subjects. Consider mathematics tutoring effects in table 2. Table 2 shows a coefficient of about .8 for the effects of tutoring participation on national mean performance—that is, a 1 percent increase in tutoring participation is associated with a raise of .8 point in national mean scores. Suppose that after-school tutoring in mathematics were to become more prevalent worldwide with a 5 percent increase in tutoring participation. This increase would be expected to lead to an increase of four points in mathematics test scores globally. This is not a small effect, considering that over time, changes in average national mathematics test scores in most countries are not large—usually they fall into a range between an increase of 15 points and a decline of 15 points over 4 or 8 years (Mullis et al. 2004, 42–44).

In science, the positive effect of tutoring participation on test performance is quite substantial, nearly twice as large as that in mathematics, with

TABLE 2
FIRST-DIFFERENCED ESTIMATES OF THE EFFECTS OF AFTER-SCHOOL TUTORING PARTICIPATION ON THE LEVEL AND DISTRIBUTION OF STUDENT PERFORMANCE

	Mathematics		Science	
	B	SE	*B*	SE
Dependent variable: Standard deviation:				
Tutoring participation	.17	.20	−.27	.23
School enrollment ratio	.42+	.24	.15	.28
Classroom homogeneity	.17	.14	.34+	.16
Differential selectivity	.21*	.08	.21**	.06
Constant	3.05	6.18	−9.17	5.31
R^2	.73		.69	
Dependent variable: Interquartile range:				
Tutoring participation	.21	.28	−.66+	.33
School enrollment ratio	.63+	.30	.44	.49
Classroom homogeneity	.18	.20	.46	.29
Differential selectivity	.32*	.12	.37**	.10
Constant	7.03	9.57	−13.58	7.72
R^2	.74		.71	
Dependent variable: Score range between 5th and 95th percentiles:				
Tutoring participation	.48	.65	−.92	.76
School enrollment ratio	1.33	.78	.33	.88
Classroom homogeneity	.62	.47	1.15*	.52
Differential selectivity	.69*	.25	.65**	.20
Constant	7.42	20.08	−32.34+	17.87
R^2	.73		.68	
Dependent variable: Mean scores:				
Tutoring participation	.82*	.28	1.40**	.35
School enrollment ratio	.35	.33	.23	.31
Classroom homogeneity	.31	.27	.39	.31
Differential selectivity	−.15+	.07	−.13	.08
Constant	17.17*	5.92	25.79**	6.70
R^2	.48		.56	

NOTE.—Missing values are imputed for students who failed to respond to the question on participation in after-school tutoring in either subject. Huber-White robust standard errors are in parentheses.
+ Significant at 10%.
* Significant at 5%.
** Significant at 1%.

a 1 percent increase in tutoring participation associating with a 1.4-point difference in national mean scores. Supposing that after-school tutoring in science becomes more prevalent over time, with a 5 percent increase in tutoring participation, we might expect an increase of seven points in science test scores globally. This is a significant effect, considering the fact that the average national performance in science, like that in mathematics, does not change radically over time (Martin et al. 2004, 44–46).

The present study finds that after-school tutoring significantly raises national mean performance in mathematics and science. This finding is not consistent with others (Baker et al. 2001), who found after-school tutoring to be ineffective. What might explain the inconsistency between the findings of these two studies? Country differences in average performance are driven by many unobserved social and educational factors other than the rate of participation in after-school tutoring. To assess the effects of tutorial participation appropriately, these unobserved factors must be taken into account. The present study adopts a fixed-effects methodology to control for these unobserved country factors; the study by Baker et al. (2001) did not control for these unobserved factors.

To determine whether controlling for unobserved country-level heterogeneity would make a difference in the findings, I replicate the regression model conducted by Baker et al. (2001), except that I use data from the TIMSS 2003. Because the present study also uses data from the TIMSS 2003, this makes possible a comparison of the findings of the model with a control for unobserved country-level heterogeneity, and the model without such a control. Table A5 in the online appendix shows the findings when I replicate the model of Baker et al. (2001), which, as stated previously, does not control for unobserved country-level heterogeneity. As table A5 indicates, the use of after-school tutoring is ineffective in both subject and grade level (whether or not controlling for the presence of high-stakes tests)—a finding consistent with that of Baker et al. This suggests that Baker et al. did not find shadow education effective for student performance mainly because they did not control for unobserved country-level heterogeneity in their country-level regression analysis.

Does After-School Tutoring Benefit All?

Table 3 reports the effects of tutoring participation on test performance for students at the 5th, 25th, 50th, 75th, and 95th percentiles. Participating in mathematics tutoring after school has a significant and positive effect on mathematics performance for students performing at the 25th percentile, for average students, and for students at the upper tail of the performance distribution. For those performing at the 5th percentile, however, the effect is insignificant. High-performing students benefit more from mathematics tutoring than do low-performing students.

In the case of science, students from all ability levels benefit significantly from participating in after-school tutoring. However, the benefit to low-performing students is nearly twice as large as that for high-performing students. As a result, the use of science tutoring after school significantly reduces achievement inequality in science when inequality is measured by interquartile range. The positive effect of participating in science tutoring is stronger than that associated with participating in mathematics tutoring, especially for the students at lower performance levels. For students at the lowest performance level, participating in science tutoring after school seems much more effective and rewarding than participating in mathematics tutoring after school.

TABLE 3
First-Differenced Estimates of the Effects of After-School Tutoring Participation on Student Performance at the 5th, 25th, 50th, 75th, and 95th Percentiles of the Distribution

	Mathematics		Science	
	B	SE	*B*	SE
Dependent variable: Test scores at the 5th percentile:				
Tutoring participation	.49	.57	1.93*	.77
School enrollment ratio	−.10	.65	.33	.71
Classroom homogeneity	−.06	.45	−.28	.59
Differential selectivity	−.54**	.16	−.51**	.17
Constant	15.63	13.39	48.22**	14.47
R^2	.66		.55	
Dependent variable: Test scores at the 25th percentile:				
Tutoring participation	.73*	.32	1.70**	.44
School enrollment ratio	−.05	.36	−.03	.49
Classroom homogeneity	.23	.30	.19	.42
Differential selectivity	−.30**	.07	−.30*	.10
Constant	12.53$^+$	6.43	30.81**	8.78
R^2	.65		.59	
Dependent variable: Test scores at the 50th percentile:				
Tutoring participation	.90**	.26	1.38**	.36
School enrollment ratio	.24	.32	.00	.35
Classroom homogeneity	.33	.27	.41	.29
Differential selectivity	−.11	.08	−.11	.08
Constant	16.49*	6.80	22.45*	8.37
R^2	.47		.59	
Dependent variable: Test scores at the 75th percentile:				
Tutoring participation	.94**	.29	1.05**	.32
School enrollment ratio	.58	.35	.41	.28
Classroom homogeneity	.41	.28	.65*	.22
Differential selectivity	.02	.11	.06	.08
Constant	19.56*	8.87	17.22$^+$	8.49
R^2	.46		.63	
Dependent variable: Test scores at the 95th percentile:				
Tutoring participation	.97*	.35	1.01**	.34
School enrollment ratio	1.24**	.33	.66*	.30
Classroom homogeneity	.56$^+$	.28	.87**	.22
Differential selectivity	.15	.12	.14	.09
Constant	23.05*	9.78	15.88$^+$	7.60
R^2	.64		.74	

$^+$ Significant at 10%.
* Significant at 5%.
** Significant at 1%.

Conclusion

Does the use of after-school tutoring promote student performance? The research literature addressing this question is dominated by individual-level analyses, which may or may not take account of selection bias. A handful of individual-level analyses correct for selection bias with a propensity score matching approach or an IV (instrumental variable) methodology, but they run into a potential problem of violating the SUTVA assumption. The violation of the SUTVA assumption may bias their research findings. Therefore, this study presents results a country-level analysis that controls for country-level unobserved heterogeneity using a fixed-effects model. This research also improves on previous studies by conducting a subject-specific analysis, with separated effect estimated for each academic subject in which student performance is measured by a well-developed curriculum-based test.

The use of after-school tutoring is sometimes an enrichment strategy for high-performing students and sometimes a remedial plan for those who perform poorly. My findings suggest that the use of after-school tutoring in mathematics for remedial purposes may not be as effective as when it is used as an enrichment strategy. In contrast, the use of after-school tutoring in science for remedial purposes is more effective than when it is used as an enrichment strategy. The use of after-school tutoring in either subject raises national average performance without widening the dispersion of student performance. In fact, it may reduce achievement inequality in the subject of science.

The sequential and cumulative nature of mathematics knowledge may help explain why the use of after-school tutoring is not effective for those who perform poorly in that subject area. Learning mathematics requires a thorough understanding of a layer of knowledge in order to be able to understand the next layer. Thus, those who fall behind in the beginning are likely to be lost later on as the course proceeds. But science is relatively less sequential and cumulative. Therefore, the practice of after-school tutoring in science tends to be more effective, not only for those who perform above average, but also for those who are below average.

The survey question about after-school tutoring participation appears in the 1995, 1999, and 2003 TIMSS surveys. The 2007 TIMSS survey, however, dropped the question. To further understand the use of after-school tutoring and its consequences cross-nationally, future TIMSS surveys should not only re-pose but also improve the question by making possible a distinction between fee-paying and fee-free tutoring, between high-peak and low-peak seasons in the use of tutoring, between in-school and out-of-school tutoring, between one-to-one and one-to-many tutoring, between remedial and enrichment purposes, and between self-motivated students and those who attend tutorials because their parents ask them to. A finer classification in the use of after-school tutoring will better serve the various purposes of research.

Given the cross-sectional nature of the TIMSS data and the small number of countries analyzed, the present study is exploratory, and its findings are only suggestive. The effects of after-school tutoring on student performance are likely to differ across countries. To address fully this issue would require data from several national longitudinal surveys designed for cross-national comparisons. To inform policy, an in-depth investigation within each country is necessary, using data from a well-designed longitudinal survey. At both the national and international levels, further research on the effects of after-school tutoring is in order.

References

Allison, Paul D. 1994. "Using Panel Data to Estimate the Effects of Events." *Sociological Methods and Research* 23 (2): 174–99.

Aurini, Janice, and Scott Davies. 2004. "The Transformation of Private Tutoring: Education in a Franchise Form." *Canadian Journal of Sociology* 29 (3): 419–36.

Baker, David P., Motoko Akiba, Gerald K. LeTendre, and Alexander W. Wiseman. 2001. "Worldwide Shadow Education: Outside-School Learning, Institutional Quality of Schooling, and Cross-National Mathematics Achievement." *Educational Evaluation and Policy Analysis* 23 (1): 1–17.

Baker, David, and Gerald K. LeTendre. 2005. *National Differences, Global Similarities: World Culture and the Future of Schooling*. Stanford, CA: Stanford University Press.

Bray, Mark. 1999. *The Shadow Education System: Private Tutoring and Its Implications for Planners*. Paris: UNESCO International Institute for Educational Planning.

Bray, Mark. 2006. "Private Supplementary Tutoring: Comparative Perspectives on Patterns and Implications." *Compare: A Journal of Comparative and International Education* 36 (4): 515–30.

Bray, Mark. 2009. *Confronting the Shadow Education System: What Government Policies for What Private Tutoring?* Paris: UNESCO International Institute for Educational Planning.

Bray, Mark. 2010. "Researching Shadow Education: Challenges and Directions." *Asia Pacific Education Review* 11 (1): 3–13.

Bray, Mark. 2011. *The Challenge of Shadow Education: Private Tutoring and Its Implication for Policy Makers in the European Union*. Brussels: European Commission.

Bray, Mark, and Percy Kwok. 2003. "Demand for Private Supplementary Tutoring: Conceptual Considerations, and Socio-Economic Patterns in Hong Kong." *Economics of Education Review* 22 (6): 611–20.

Buchmann, Claudia, Dennis J. Condron, and Vincent J. Roscigno. 2010. "Shadow Education: Theory, Analysis and Future Directions: A Rejoinder." *Social Forces* 89 (2): 483–90.

Byun, Soo-Yong, and Kyung-Keun Kim. 2010. "Educational Inequality in South Korea: The Widening Socioeconomic Gap in Student Achievement." *Research in the Sociology of Education* 17:155–82.

Byun, Soo-Yong, and Hunjoon Park. 2012. "The Academic Success of East Asian American Youth: The Role of Shadow Education." *Sociology of Education* 85 (1): 40–60.

Dang, Hai-Anh. 2007. "The Determinants and Impacts of Private Tutoring Classes in Vietnam." *Economics of Education Review* 26 (6): 684–99.

Dang, Hai-Anh, and F. Halsey Rogers. 2008. "The Growing Phenomenon of Private Tutoring: Does It Deepen Human Capital, Widen Inequalities, or Waste Resources?" *World Bank Research Observer* 23 (2): 161–200.

Davies, Scott. 2004. "School Choice by Default? Understanding the Demand for Private Tutoring in Canada." *American Journal of Education* 110 (3): 233–55.

Dridi, Barbara L. 2002. "Tunisia." In *World Education Encyclopedia: A Survey of Ed-*

ucational Systems Worldwide, ed. Rebecca Marlow-Ferguson. 2nd ed. Detroit: Gale Group.

Foy, Pierre, and Marc Joncas. 2000. "TIMSS Sample Design." In *TIMSS 1999 Technical Report*, ed. Michael O. Martin, Kelvin D. Gregory, and Steven E. Stemler. Chestnut Hill, MA: Boston College.

Gangl, Markus. 2010. "Causal Inference in Sociological Research." *Annual Review of Sociology* 36:21–47.

Garden, Robert A. 2000. "TIMSS Test Development." In *TIMSS 1999 Technical Report*, ed. Michael O. Martin, Kelvin D. Gregory, and Steven E. Stemler. Chestnut Hill, MA: Boston College.

Geronimus, Arline T., and Sanders Korenman. 1992. "The Socioeconomic Consequences of Teen Childbearing Reconsidered." *Quarterly Journal of Economics* 107 (4): 1187–214.

Grodsky, Eric S. 2010. "Learning in the Shadows and in the Light of Day." *Social Forces* 89 (2): 475–81.

Gurun, Ayfer, and Daniel L. Millimet. 2008. *Does Private Tutoring Payoff?* Discussion Paper no. 3637, Institute for the Study of Labor (IZA), Schaumburg, IL.

Hanushek, Eric A., and Ludger Woessmann. 2006. "Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence across Countries." *Economic Journal* 116 (510): C63–C76.

Hauser, Robert M. 1991. "Simple Models of Sibling Resemblance: Reply to Gillespie." *American Journal of Sociology* 97 (1): 206–9.

Huang, Min-Hsiung. 2009. "Classroom Homogeneity and the Distribution of Student Math Performance: A Country-Level Fixed-Effects Analysis." *Social Science Research* 38 (4): 781–91.

Huang, Min-Hsiung, and Taissa S. Hauser. 2010. "Tracking Persons from High School through Adult Life: Lessons from the Wisconsin Longitudinal Study." *EurAmerica* 40 (2): 311–58.

Kang, Changhui. 2009. "Does Money Matter? The Effect of Private Education Expenditures on Academic Performance in the Republic of Korea." In *Emerging Evidence on Vouchers and Faith-Based Providers in Education: Case Studies from Africa, Latin America, and Asia*, ed. Felipe Barrera-Osorio, Harry A. Patrinos, and Quentin Wodon. Washington, DC: World Bank Publications.

Kuan, Ping-Yin. 2011. "Effects of Cram Schooling on Mathematics Performance: Evidence from Junior High Students in Taiwan." *Comparative Education Review* 55 (3): 342–68.

Kulpoo, Dhurumbeer, and Maya Soonarane. 2005. *The SACMEQ II Project in Mauritius: A Study of the Conditions of Schooling and the Quality of Education.* Harare, Zimbabwe: SACMEQ.

Kwak, Byong-Sun. 2004. "Struggle against Private Lessons in Korean Education Context." Paper presented at the 28th Annual Conference of the Pacific Circle Consortium in the Hong Kong Institute of Education, Hong Kong.

Martin, Michael O., Ina V. S. Mullis, and Steven J. Chrostowski. 2004. *TIMSS 2003 Technical Report.* Chestnut Hill, MA: Boston College, TIMSS, and PIRLS International Study Center.

Martin, Michael O., Ina V. S. Mullis, Eugenio J. Gonzalez, and Steven J. Chrostowski. 2004. *TIMSS 2003 International Science Report: Findings from IEA's Trends in Inter-*

national Mathematics and Science Study at the Fourth and Eighth Grades. Chestnut Hill, MA: Boston College, TIMSS, and PIRLS International Study Center.

Mischo, Christoph, and Ludwig Haag. 2002. "Expansion and Effectiveness of Private Tutoring." *European Journal of Psychology of Education* 17 (3): 263–73.

Mullis, Ina V. S., Michael O. Martin, Eugenio J. Gonzalez, and Steven J. Chrostowski. 2004. *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College, TIMSS, and PIRLS International Study Center.

Ono, Hiroshi. 2007. "Does Examination Hell Pay Off? A Cost-Benefit Analysis of 'Ronin' and College Education in Japan." *Economics of Education Review* 26 (3): 271–84.

Onsomu, Eldah, Juliana Nzomo, and Charles Obiero. 2005. *The SACMEQ II Project in Kenya: A Study of the Conditions of Schooling and the Quality of Education*. Harare, Zimbabwe: Southern and Eastern Africa Consortium for Monitoring Education Quality.

Park, Hyunjoon, Soo-Yong Byun, and Kyung-Keun Kim. 2011. "Parental Involvement and Students' Cognitive Outcomes in Korea: Focusing on Private Tutoring." *Sociology of Education* 84 (1): 3–22.

Paviot, Laura, Nina Heinsohn, and Julia Korkman. 2008. "Extra Tuition in Southern and Eastern Africa: Coverage, Growth, and Linkages with Pupil Achievement." *International Journal of Educational Development* 28 (2): 149–60.

Sarri, Samuel. 2002. "Morocco." In *World Education Encyclopedia: A Survey of Educational Systems Worldwide*, ed. Rebecca Marlow-Ferguson. 2nd ed. Detroit: Gale Group.

Silova, Iveta, Virginija Būdienė, and Mark Bray. 2006. *Education in a Hidden Marketplace: Monitoring of Private Tutoring*. New York: Open Society Institute.

Sobel, Michael E. 2006. "What Do Randomized Studies of Housing Mobility Demonstrate? Causal Inference in the Face of Interference." *Journal of American Statistical Association* 101 (476): 1398–407.

Stevenson, David L., and David P. Baker. 1992. "Shadow Education and Allocation in Formal Schooling: Transition to University in Japan." *American Journal of Sociology* 97 (6): 1639–57.

Tan, Jason. 2009. "Private Tutoring in Singapore: Bursting out of the Shadows." *Journal of Youth Studies* 12 (1): 93–103.

Tansel, Aysit, and Fatma Bircan. 2006. "Demand for Education in Turkey: A Tobit Analysis of Private Tutoring Expenditures." *Economics of Education Review* 25 (3): 303–13.

Winship, Christopher, and Stephen L. Morgan. 1999. "The Estimation of Causal Effects from Observational Data." *Annual Review of Sociology* 25:659–707.

Yamamoto, Kentaro. 2000. "Scaling Methodology and Procedures for The TIMSS Mathematics and Science Scales." In *TIMSS 1999 Technical Report*, ed. Michael O. Martin, Kelvin D. Gregory, and Steven E. Stemler. Chestnut Hill, MA: Boston College, TIMSS, and PIRLS International Study Center.

Zeng, Kangmin. 1999. *Dragon Gate: Competitive Examinations and Their Consequences*. New York: Cassell & Continuum.