

Does Money Matter? The Effect of Private Tutoring
Expenditures on Academic Performance: Evidence from a
Nonparametric Bounding Method

Changhui Kang

Department of Economics
National University of Singapore

Singapore

and

Department of Economics

Chung-Ang University

Seoul, South Korea

Abstract

The causal relationship between educational investments and student outcomes continues to attract attention. The majority of studies have examined the effectiveness of public school expenditures or private school attendance on student outcomes. This paper contributes to the literature by examining the effectiveness of an unexplored dimension of educational inputs—private tutoring expenditures of South Korean parents. In the face of difficulties in causal estimation, the paper employs a nonparametric bounding method that is recently gaining popularity. Using the method we show that the true effect of private tutoring remains at most modest. The tightest bounds suggest that a 10 percent increase in expenditure raises a student's test score by 0.764 percent at the largest. Such a modest effect remains similar across male and female students, and across students of different ability levels.

JEL Classification : I20, C30

Keywords : Private Tutoring, Test Scores, Nonparametric Bounds, South Korea

* Address: Department of Economics, National University of Singapore, AS2 Level 6, 1 Arts Link, Singapore 117570, Singapore; E-mail: eckch@nus.edu.sg, Phone: +65-6516-6830, Fax: +65-6775-2646.

Does Money Matter? The Effect of Private Tutoring Expenditures on Academic Performance: Evidence from a Nonparametric Bounding Method

Abstract

The causal relationship between educational investments and student outcomes continues to attract attention. The majority of studies have examined the effectiveness of public school expenditures or private school attendance on student outcomes. This paper contributes to the literature by examining the effectiveness of an unexplored dimension of educational inputs—private tutoring expenditures of South Korean parents. In the face of difficulties in causal estimation, the paper employs a nonparametric bounding method that is recently gaining popularity. Using the method we show that the true effect of private tutoring remains at most modest. The tightest bounds suggest that a 10 percent increase in expenditure raises a student’s test score by 0.764 percent at the largest. Such a modest effect remains similar across male and female students, and across students of different ability levels.

JEL Classification : I20, C30

Keywords : Private Tutoring, Test Scores, Nonparametric Bounds, South Korea

1 Introduction

The causal relationship between educational investments and student outcomes continues to attract the attention of many. Despite decades of intensive study, there is no general consensus regarding the effectiveness of monetary educational inputs for student outcomes.

In particular, papers that summarize the debate on the effects of public school expenditures often advocate conflicting views. For example, Card and Krueger (1996), Guryan (2003), Greenwald et al. (1996) and Krueger (1999, 2003) present evidence for the effectiveness of public school expenditures. In contrast, Betts (1996), Hanushek (1986, 1997, 2003) and Leuven et al. (2007) cast doubt on the conclusion of these researchers and suggest several factors that can explain discrepancies in conclusion (e.g., Betts, 1996).

Studies focusing on private schools (e.g., Catholic schools) do not shed much light on the impacts of educational investments as well. While Evans and Schwab (1995) and Neal (1997) show educational benefits of attending Catholic high school, Altonji et al. (2005), Figlio and Stone (1999) and Goldhaber (1996) find no significant gaps in test scores between public and

private schools. Importantly, many studies on private and Catholic schooling seem to suffer from lack of reliable exogenous variation for identifying the causal effect. Altonji et al. (2005b) argue that two frequently used instrumental variables—religious affiliation of the parents and geographical proximity of Catholic schools—are not a useful source of identification of Catholic school effects.

Given such lack of consensus about the effectiveness of educational investments for outputs, this paper aims to contribute to the literature by examining the effect of an unexplored dimension of educational inputs—expenditures on private tutoring that is widely employed by South Korean parents in order to supplement public school education.

In South Korea, secondary school students have little freedom in the choice of their middle and high schools in a school district. Since 1969 student allocation to public and private schools has largely been under the strict control of the government, especially in urban regions. Under this system (labeled ‘Leveling Policy’), students are basically assigned—not admitted upon application—to secondary schools within their residential school district by either a pure lottery or an application-accompanied-by-lottery system under the supervision of the local Ministry of Education office. Moreover, within school ability grouping is rarely implemented due to the government’s egalitarian policy on secondary education and parents’ objections. Curricula are also controlled by the Ministry of Education nationwide for the most part.

In response to such a rigid public education system, parents in South Korea spend a great deal of money on private tutoring for their children. According to a statistic, South Korean parents spend on private tutoring 85 percent as much as they spend on public schooling (Korean Educational Development Institute, *Survey on Educational Expenditures*, 1998). Given the large amount of expenditures on private tutoring, many including parents as well as educational policy makers are concerned about its effectiveness for student academic performance. From a broader perspective, an examination into the effect of private tutoring serves to illuminate the debates on the impacts of educational investments on student outcomes.

It is well known that educational expenditures on a student are not exogenously and randomly determined; there is little doubt that private tutoring expenditures are endogenous and correlated with a student’s personal, family and academic characteristics. In the absence of a randomized experiment on private tutoring, a causal estimation in principle calls for an instrumental variable (IV) that is strongly related with the expenditures but exogenous to the

determinants of academic performance. However, it seems fairly difficult if not impossible to find such an exogenous variable.

In the face of such difficulties in estimation, the current paper examines causal effects of private tutoring by employing an alternative strategy that is recently gaining popularity in empirical analysis—a nonparametric bounding method. Nonparametric bounds analysis was first introduced in economics by Manski (1990) and further developed in Manski (1997), Manski and Pepper (2000), and Hotz et al. (1997). Some recent examples of this method include Blundell et al. (2007), Gerfin and Schellhorn (2006), Gonzalez (2005), Kreider and Pepper (2007), Lechner (1999), Manski and Nagin (1998), and Pepper (2000) among others.

The basic idea of the nonparametric bounding method is that instead of obtaining point estimates that often rely on questionable assumptions, one may calculate lower and upper bounds of the treatment effect given a few weaker assumptions. A unique advantage of this approach in an estimation of the effect of private tutoring is that a potential IV does not have to be fully exogenous. To the extent that the IV is monotonically (either positively or negatively) related with the outcome variable, the method, in combination of some other assumptions, is able to draw fairly tight bounds of the causal effect (Manski and Pepper 2000). Provided that the bounds are sufficiently narrow and informative to locate the causal effect, we interpret that the magnitude of the true effect is somewhere between the estimated lower and upper bounds.

We apply such a nonparametric bounding method to a data set from South Korea that has an information on private tutoring expenditures and college-entrance examination scores of high school graduates. Using birth order, the first-born status to be specific, of a student in the family as a monotone IV (MIV) of Manski and Pepper (2000), the current paper shows that the true effect of private tutoring remains at most modest. The tightest—hence most informative—bounds suggest that a 10 percent increase in private tutoring expenditure raises a student’s test score by 0 percent at the smallest and by 0.764 percent at the largest. Such a modest effect of private tutoring remains similar across male and female students, on the one hand, and across students of different ability levels, on the other.

The rest of the paper is organized as follows. Section 2 outlines the empirical strategy of the paper. Section 3 gives a description on private tutoring in South Korea. Data are discussed in section 4; empirical results are shown in section 5. Potential explanations for the small effect of private tutoring are reviewed in section 6. Section 7 concludes.

2 Empirical Framework

Let us first define y_i as a normalized—with mean zero and variance one—test score of student i . Let T_i be a treatment indicator that is equal to zero if the average monthly spending for i 's tutoring is equal to zero; one if it is greater than zero but less than or equal to H_1 ; and two if it is greater than H_1 . In empirical analysis below, we set H_1 equal to KRW 300,000—approximately US\$251.¹ Finally, let an indicator F_i be equal to one if i is first-born in the family and zero otherwise.

Each student receives treatment $t \in T = \{0, 1, 2\}$. The response function $y_i(\cdot) : T \rightarrow Y$ maps treatments into outcomes. The realized outcome $y_i \equiv y_i(z)$ is the level of y for student i who actually receives treatment z . The latent outcome $y_i(t)$ ($t \neq z$) describes what level of performance student i would have achieved had he or she received treatment t . Of primary interest are the causal effects of increasing tutoring expenditures on i 's test score: $E[y_i(1) - y_i(0)]$, $E[y_i(2) - y_i(1)]$, and $E[y_i(2) - y_i(0)]$.

Following Gonzalez (2005), Manski (1990) and Manski and Pepper (2000), in order to set up bounds for the treatment effects, we first decompose $E[y(t)]$ by

$$E[y(t)] = E[y|z = t]Pr(z = t) + E[y(t)|z \neq t]Pr(z \neq t) \quad (1)$$

To make bounds analysis feasible, let us suppose that y is bounded by $[K_0, K_1]$. Since the unobservable counterfactual $E[y(t)|z \neq t]$ is also bounded by $[K_0, K_1]$, we have the worst-case (WC) bounds of $E[y(t)]$ given by

$$\begin{aligned} & E[y|z = t]Pr(z = t) + K_0Pr(z \neq t) \\ & \leq E[y(t)] \leq \\ & E[y|z = t]Pr(z = t) + K_1Pr(z \neq t) \end{aligned} \quad (2)$$

In order to further tighten the bounds of $E[y(t)]$, a few assumptions can be invoked below individually as well as jointly. The first assumption to be employed is monotone treatment response (MTR), which is specified as follows:

$$t_n < t_m \longrightarrow y(t_n) \leq y(t_m) \quad (3)$$

This assumption is drawn from a theory that there will be non-negative impacts of increased educational spending on a student’s academic performance. A majority of empirical studies support the validity of such an assumption. Although the exact magnitude of a positive effect of educational spending on a student’s performance is hardly agreed, it is rare that studies find strong *negative* impacts of monetary educational investments (see Hanushek (1997, 2003); an exception is a study of Leuven et al. (2007)). Under MTR, the following relationships hold:

$$\begin{aligned} z > t &\longrightarrow K_0 \leq y(t) \leq y(z) \\ z = t &\longrightarrow y(t) = y(z) \\ z < t &\longrightarrow y(z) \leq y(t) \leq K_1 \end{aligned}$$

Hence, the MTR bounds of $E[y(t)]$ can be expressed by

$$\begin{aligned} &E[y|z \leq t]Pr(z \leq t) + K_0Pr(z > t) \\ &\leq E[y(t)] \leq \\ &E[y|z \geq t]Pr(z \geq t) + K_1Pr(z < t) \end{aligned} \tag{4}$$

The second assumption is monotone treatment selection (MTS), which is specified by:

$$t_n < t_m \longrightarrow E[y(t)|z = t_n] \leq E[y(t)|z = t_m] \tag{5}$$

This assumption supposes that sorting into treatment is not exogenous but monotone in the sense that the average latent outcome $y(t)$ is greater for those students whose parents spend a large amount of money on private tutoring ($z = t_m$) than for those whose parents spend a small amount ($z = t_n, t_n < t_m$). For instance, high income parents are more likely to spend a large amount of money on private tutoring for their child than low-income parents, while children of high income parents tend to be more academically able and smarter than those of low-income parents (see, e.g., Haveman and Wolfe (1995)). While it specifies a source of endogeneity in a conventional OLS method of examining the impacts of educational investments, the MTS assumption can make an important contribution to tightening the bounds of the true effect in a bounds analysis.

Under MTS, the following relationships hold:

$$\begin{aligned}
 u > t &\longrightarrow E[y|z = t] \leq E[y(t)|z = u] \leq K_1 \\
 u = t &\longrightarrow E[y(t)|z = u] = E[y|z = t] \\
 u < t &\longrightarrow K_0 \leq E[y(t)|z = u] \leq E[y|z = t]
 \end{aligned}$$

Hence, the MTS assumption yields the bounds of $E[y(t)]$ given by

$$\begin{aligned}
 &E[y|z = t]Pr(z \geq t) + K_0Pr(z < t) \\
 &\leq E[y(t)] \leq \\
 &E[y|z = t]Pr(z \leq t) + K_1Pr(z > t)
 \end{aligned} \tag{6}$$

The third assumption is the presence of a monotone instrumental variable (MIV) that is proposed by Manski and Pepper (2000). While an instrumental variable (IV) v should satisfy mean-independence (i.e., $E[y(t)|v = u_1] = E[y(t)|v = u_2]$, $u_1 \neq u_2$), an MIV is assumed to satisfy the following mean-monotonicity:

$$u_1 < u_2 \longrightarrow E[y(t)|v = u_1] \leq E[y(t)|v = u_2] \tag{7}$$

Following an idea of Black et al. (2005, p.695), we use the first-born indicator (F_i) as an MIV for our bounds analysis, where F_i is a dummy variable that takes 1 if i is a first-born child in the family and 0 otherwise. The MIV assumption supposes that for a given level of tutoring expenditure first-born students ($F_i = 1$) on average performs as well or better than students who are not first-born (i.e., later-born) in the family ($F_i = 0$). There are at least two reasons why such an assumption is valid in our current context. First, although there are studies showing little impact of birth order on a child's intelligence (Retherford and Sewell 1991, Rodgers et al. 2000), papers that report strong, if any, birth order effects usually show negative rather than positive effects of birth order on intelligence (Bjerkedal et al. 2007, Black et al. 2007, Zajonc 1976, Zajonc and Mullally 1997). Namely, intelligence of older siblings is either as high or higher than that of younger siblings. It is a well-established empirical regularity that a child's high intelligence leads to high academic performance in school (Herrnstein and Murray 1994). Second, previous empirical studies show that parents favor the first-born relative to the later-born with respect

to educational investments in general (Behrman and Taubman 1986, Black et al. 2005). Our results, which will be presented in Table 2, also show that parents spend significantly more on private tutoring for the first-born than for the later-born. To the extent that parents favor the first-born with respect to monetary educational investments, they will tend to support the same child more over other educational dimensions as well, say, by providing better emotional and non-financial supports for the first-born.² Such supports can lead to better performance of the first-born independently of monetary investments.

To the extent that the MIV assumption is valid in our framework, it yields the bounds of $E[y(t)]$ that are expressed by

$$\begin{aligned} & \sum_{u \in F} Pr(F = u) \{ \sup_{u_1 \leq u} [E(y|F = u_1, z = t)Pr(z = t|F = u_1) + K_0Pr(z \neq t|F = u_1)] \} \\ & \leq E[y(t)] \leq \\ & \sum_{u \in F} Pr(F = u) \{ \inf_{u_2 \geq u} [E(y|F = u_2, z = t)Pr(z = t|F = u_2) + K_1Pr(z \neq t|F = u_2)] \} \end{aligned} \quad (8)$$

Given each of the MTR, MTS and MIV assumptions, we can impose multiple assumptions jointly in order to further tighten the bounds of $E[y(t)]$ and those of treatment effects $E[y(t_m)] - E[y(t_n)]$ ($t_n < t_m$). We below experiment with all possible combinations of the individual assumptions: MIV+MTR, MIV+MTS, MTR+MTS, and MIV+MTR+MTS.

Under MIV+MTR, the bounds of $E[y(t)]$ are given by

$$\begin{aligned} & \sum_{u \in F} Pr(F = u) \{ \sup_{u_1 \leq u} [E(y|F = u_1, z \leq t)Pr(z \leq t|F = u_1) + K_0Pr(z > t|F = u_1)] \} \\ & \leq E[y(t)] \leq \\ & \sum_{u \in F} Pr(F = u) \{ \inf_{u_2 \geq u} [E(y|F = u_2, z \geq t)Pr(z \geq t|F = u_2) + K_1Pr(z < t|F = u_2)] \} \end{aligned} \quad (9)$$

Under MIV+MTS, the bounds of $E[y(t)]$ are given by

$$\begin{aligned} & \sum_{u \in F} Pr(F = u) \{ \sup_{u_1 \leq u} [E(y|F = u_1, z = t)Pr(z \geq t|F = u_1) + K_0Pr(z < t|F = u_1)] \} \\ & \leq E[y(t)] \leq \\ & \sum_{u \in F} Pr(F = u) \{ \inf_{u_2 \geq u} [E(y|F = u_2, z = t)Pr(z \leq t|F = u_2) + K_1Pr(z > t|F = u_2)] \} \end{aligned} \quad (10)$$

Under MTR+MTS, the bounds of $E[y(t)]$ are given by

$$\begin{aligned}
& \sum_{h < t} E(y|z = h)Pr(z = h) + E(y|z = t)Pr(z \geq t) \\
& \leq E[y(t)] \leq \\
& \sum_{h > t} E(y|z = h)Pr(z = h) + E(y|z = t)Pr(z \leq t)
\end{aligned} \tag{11}$$

Finally, under MIV+MTR+MTS, the bounds of $E[y(t)]$ are given by

$$\begin{aligned}
& \sum_{u \in F} Pr(F = u) \cdot \\
& \left\{ \sup_{u_1 \leq u} [\sum_{h < t} E(y|F = u_1, z = h)Pr(z = h|F = u_1) + E(y|F = u_1, z = t)Pr(z \geq t|F = u_1)] \right\} \\
& \leq E[y(t)] \leq \\
& \sum_{u \in F} Pr(F = u) \cdot \\
& \left\{ \inf_{u_2 \geq u} [\sum_{h > t} E(y|F = u_2, z = h)Pr(z = h|F = u_2) + E(y|F = u_2, z = t)Pr(z \leq t|F = u_2)] \right\}
\end{aligned} \tag{12}$$

Given the bounds of $E[y(t)]$ under varying assumptions, the lower bound (LB) of average treatment effects (ATE), $E[y(t_m) - y(t_n)]$ ($t_m > t_n$), is calculated by the difference between the lower bound of $E[y(t_m)]$ and the upper bound of $E[y(t_n)]$; the upper bound (UB) of ATE is obtained by the difference between the upper bound of $E[y(t_m)]$ and the lower bound of $E[y(t_n)]$. Along with the bounds of $E[y(t)]$ and ATEs are calculated bootstrap 5th and 95th percentiles of the lower and upper bound, respectively. The interval between these percentiles shows a conservative 90% confidence interval for the estimated bounds. The number of the bootstrap samples is 200.

3 Private Tutoring in South Korea

There are potentially many forms of private tutoring that a student may receive for various reasons. It varies from a swimming lesson for an exercise to math tutoring for a slow-learning child. Here we focus our attention on a private supplementary instruction of academic subjects that involves financial transactions outside the formal school system. Such private tutoring is generally observed in many countries where public education system is poorly equipped or the existing system fails to satisfy highly motivated parents. While it is apparently most prominent in East Asian societies such as Japan, Hong Kong, Taiwan and South Korea (Time Asia 2006), studies report the presence of private tutoring in a wide range of countries from Egypt, Kenya to India, Romania, Canada and UK (Baker et al. 2001, Bray 1999).

In South Korea there exist widespread and large-scale markets for private tutoring outside the public education system. For example, Ministry of Education (1999, 2000) reports that about 58.2 percent of all pre-college students experience various kinds of private tutoring in 2000. It also documents that such a proportion is highest (70.7) among elementary school students, medium (59.5) among middle school students, and relatively low (46.8) among college-bound students of general high school. As a consequence of widespread popularity, private tutoring expenditures by parents are also quite large in Korea. Ministry of Education (1999, 2000) shows that private tutoring expenses are about 9 percent of incomes of the households that have school-aged children for all income groups. At the national level, total household expenditures on private tutoring of year 2003 amount to 2.3 percent of the national GDP and 55 percent of the national annual budget for public education (Korean Educational Development Institute, *Media Briefing*, November 19, 2003). A major reason for such widespread private tutoring in the country is that there are virtually no private secondary schools that are independent of the government's control: in Korea private middle and high schools are little different from public schools with respect to school administration, curriculum and student placement, because they are heavily subsidized and controlled by the government.³

Among potentially many channels of private instruction, only two broad types of private tutoring are permitted by the government and practiced in the market in Korea. One is a relatively formal instruction offered by *hakwons*—private for-profit school-like learning institutions. The other is an informal private instruction by individual university students. All other forms of private tutoring including private instruction by full-time school teachers outside the school, that by *hakwon* instructors outside the *hakwon*, and private tutoring through mails, phones and TVs are prohibited by the government. Of these two legal forms of private tutoring, the government maintains a strong control over *hakwons*, while it has little control over individual tutors: the government imposes some requirements for establishing a *hakwon*, and exerts administrative controls with respect to pricing, academic qualification of tutors, physical facilities, etc. (Kim and Lee 2001)

According to our data from the Korean Education and Employment Panel (discussed shortly), the proportion of grade 11 students who receive private tutoring for any subject is 79.9 percent. And their overall average monthly spending on private tutoring is about KRW 297,400—approximately USD\$249, which amounts to 9 percent of monthly family income (see Table 1).

Among academic subjects, private tutoring is most frequently practiced for mathematics (55.7 percent). Of those students who receive private tutoring for math, 44.8 percent uses *hakwons* and 48.0 percent employs one-to-one or one-to-many tutoring offered by individual tutors.

4 Data

4.1 Description of the Main Sample

For empirical analysis this study employs the Korean Education and Employment Panel (KEEP). KEEP is an annual longitudinal survey that is conducted from year 2004 by the Korea Research Institute for Vocational Education and Training (KRIVET)—a government-funded research institute. The basic structure of KEEP follows the National Educational Longitudinal Studies (NELS) of the U.S. The beginning cohorts of KEEP consist of 6,000 students from three different populations: 2,000 students each from middle school (grade 9), general high school and vocational high school (both grade 12, the final year of secondary education). Students of each group are sampled by a stratification method to reflect the national population of the group. More specifically, for each group 100 schools are selected in consideration of the regional distribution of schools and students. For each school 4 different classes are randomly chosen, and for each class 5 students are sampled at random. The sampled students are administered a variety of personal, family and school-related questionnaires. In addition, students' homeroom teachers, school principals and parents are separately surveyed to collect a range of background information on the sampled student.

An important feature of the KEEP data is that the survey collects detailed information on a student's private tutoring experience and expenditures, and the sibling composition from the parent questionnaire. It enables us to construct main explanatory variables and the instrumental variable of this study. Also unique in the KEEP data is the availability of the College Scholastic Ability Test (CSAT) scores for high school graduates. CSAT is the national college-entrance examination of Korea that is annually administered under the supervision of the Ministry of Education and whose scores are used as an important factor by colleges and universities to determine the admission of the applicants. Using the resident registration number of the student, the KEEP data are linked to the administrative data base of the 2004 CSAT scores for the test writers. As a measure of a student's academic performance, we employ the raw CSAT scores

of the following three subjects: the Korean language, mathematics, and English.⁴ The score of each individual subject ranges from lowest 36 to highest 150; mean and standard deviation of each subject score are around 100 and 17, respectively. Given such a raw score of each subject, a simple average of the three individual scores—or two individual scores if a student chooses to take only two tests—is calculated. This average score is further normalized to have mean zero and variance one for subsequent analysis.

Although vocational high school graduates are in principle eligible for CSAT, the majority of the CAST writers are general high school graduates; they are also the majority of students taking private tutoring. Therefore, we restrict our analysis to the general high-school sample of 2,000 students.⁵ For main analysis, the original general high-school sample has been further narrowed down. First, those students whose guardian is not one of the parents are excluded from our main sample. Patterns of educational investments and academic performance among these students may be far from typical due to the absence of both parents. A total of 85 students are removed from the original sample as their guardian is either a grandparent, sibling, or close relatives. Students from single-parent families, however, are retained.

Second, students are excluded from the main sample if they either attend a special high school for music, fine arts and athletics, take private tutoring to major in these subjects for higher education in universities, or both. Tutoring costs among them are generally much greater than costs of a normal tutoring of academic subjects. And these students are likely to be poor performers in such a general subject test as CSAT.⁶ The number of such students are 168 in total. The preceding two restrictions leave a total of 1,752 students for further analysis.

Third, we further exclude from the analysis those whose average CSAT scores are unknown due to the failure to take the test⁷ and those who have missing values for the key explanatory variables. Such conditions further remove a total of 285 students. Thus the final sample size is 1,467. Descriptive statistics of this main sample and their differences between first-born and later-born students are documented in Table 1.

INSERT TABLE 1 HERE.

4.2 Descriptive Statistics

Among CSAT test-writers, the mean score of the Korean language is 100.9; the mean scores of math and English are 98.6 and 99.7, respectively. While mean math score is close between the two groups, mean scores of Korean and English among the first-born significantly exceed those of the later-born students. And the mean Z-score of the raw scores averaged over the three subjects (or two subjects if a student has taken only two) is also significantly greater for first-born than for later-born students. Yet it is not clear whether these differences between the two groups are causally created by variations in educational investments.

As for the amount of spending on private tutoring, first-born students receive larger educational investments from their parents than later-born counterparts do.⁸ While the overall average monthly spending on private tutoring is about $W297,400$ —approximately \$249, the average spending for first-born students ($W336,800$) is 34 percent greater than that for the later-born ($W251,100$). This amount of gap is significantly different from zero. The proportion of those who have ever received private tutoring—those with positive monthly spending—is also far higher among first-born students (85.4 percent) than among the later-born (73.5 percent).

When tutoring experiences are focused on the three main subjects, math is the subject for which students most likely take private tutoring (55.7 percent); it is followed by English (42.9 percent) and Korean (31.3 percent) in frequency. Weekly tutoring hours are also longest for math (2.6 hours), followed by English (1.8 hours) and Korean (1.3 hours) in terms of duration. Whether in terms of frequency or duration, the preceding statistics suggest that first-born students receive significantly greater educational investments than the later-born do.

Using the questionnaire for grade-12 homeroom teachers, we create a measure of a student's pre-tutoring quality. This measure is important to our analysis, since it reveals pre-existing differences in student quality before the treatment (i.e., private tutoring). The teachers are asked to report a student's approximate rank from 0 (lowest quality) to 100 (highest quality) within a school or a classroom during the second semester of grade 11.⁹ According to this measure, students are on average rated to have a medium quality (44.5 percentile). The later-born have a quality slightly higher than the first-born. The difference, however, is indistinguishable from zero. Weekly hours of self-study excluding private tutoring hours are also similar between first-born (12.4 hours) and later-born students (11.6 hours). The preceding statistics indicate that

there seem to exist no substantial differences in student quality and self-investment in study between first-born and later-born students. Parents, on the other hand, invest more intensively in first-born than in later-born children.

A student’s age, sex, presence of both parents and school characteristics do not vary substantially between first-born and later-born students. The average age of parents of the first-born is lower than that of the later-born, since a couple gives birth to the first child earlier in life. The parents’ average education level and family monthly income is higher among the first-born than among the later-born. This reflects a tendency that better-educated (hence higher-income) people get married, give birth to a first child later in life and have fewer offsprings (Rosenzweig 1986). In our main sample, 54.0 percent is first-born students that include only children.¹⁰ 30.7 percent is first-born boys; 23.2 percent is first-born girls. In the sample, only children account for 7.3 percent; the mean number of children in a family is 2.2 and the maximum is 6.

5 Estimation Results

5.1 OLS and 2SLS Results

An an intermediate step, we estimate the following conventional model for academic effects of educational expenditures by OLS and 2SLS:

$$y_i = \beta_0 + S_i\beta_1 + X_i\beta_2 + u_i \tag{13}$$

where y_i is the Z-score of student i that is normalized from the averaged CSAT score (Y_i) to have mean zero and variance one in the entire general high school sample; S_i is an amount of a monthly spending on tutoring for i ; X_i is the vector of i ’s personal and family backgrounds as well as school characteristics; and u_i is the random error term. For 2SLS estimation, F_i is employed as an IV for S_i .

Table 2 presents the OLS and 2SLS estimation results. In the square bracket under each of OLS and 2SLS estimates is reported a percentage change in test score due to a 10 percent increase in tutoring expenditure, which is evaluated at the mean values of the average test score (99.64) and expenditure (297.4).

INSERT TABLE 2 HERE.

The OLS estimate for the effect of private tutoring expenditures in column (1) suggests that the association between the tutoring expenditure and normalized test score is positive but quite small and significantly indistinguishable from zero. A W10,000 greater monthly expenditure on private tutoring is related to no more than a 0.001 SD higher test score. Such magnitude implies that a 10 percent greater expenditure is related to only a 0.06 percent higher test score. Such an association, however, may not be consistent and causal. Depending on the correlation between S_i and u_i , the OLS estimate may be biased either upward or downward.

The first-stage results of the 2SLS regression of tutoring expenditures (in level) on IV and explanatory variables are presented in column (2). Being first-born significantly and positively affects private tutoring expenditures for a student. First-born students receive on average W47,620 greater expenditures on tutoring than later-born students. Such an amount is significantly different from zero. Thus to the extent that F_i is fully exogenous, it can serve as an IV for S_i in a 2SLS estimation.¹¹

The 2SLS estimates of the effect of tutoring expenditures are shown in columns (3) and (4). The estimate in column (3) is based on a specification that uses a level of a tutoring expenditure for S_i ; that in column (4) is based on a natural log of a tutoring expenditure.¹² The estimate in column (3) suggests that a W10,000 increase in expenditure enhances a student’s performance by 0.02 SD. Evaluated at the mean values of the test score and expenditure, the estimate implies that a 10 percent increase in expenditure on private tutoring raises test score by 0.76 percent. The estimate in column (4) implies a similar magnitude of the effect of private tutoring; a 10 percent greater expenditure leads to a 0.41 percent increase in test score. Both estimated effects, however, are statistically indistinguishable from zero. Furthermore, to the extent that there exists a potentially positive rather than negative correlation between F_i and u_i as is implied by our MIV assumption in (7), our 2SLS estimates are more likely to even overstate the true effect of private tutoring rather than understate it. As explained in the introduction, a potentially positive correlation between F_i and u_i can be better exploited by nonparametric bounding methods than by conventional 2SLS methods.

5.2 Results of Bounds Analysis

The estimated bounds of $E[y(t)]$ and average treatment effects (ATE) for the whole sample are presented in Table 3. We also present a graphical illustration of the bounds of ATE under

different combinations of the assumptions in Figure 1.

INSERT TABLE 3 AND FIGURE 1 HERE.

In our specification y is a normalized test score that may in principle vary from negative infinity to positive infinity. In the data, however, y varies from -2.845 to 2.571. In order to make bounds analysis feasible, we arbitrarily impose lower and upper bounds of y , setting K_0 equal to -3 and K_1 equal to 3. Alternative values of K_0 and K_1 , however, do not affect the results of the paper substantially.¹³

Let us first examine the bounding results for an average treatment effect $E[y(1) - y(0)]$, which we call ‘*ATE10*.’ We will discuss the results for other ATEs, $E[y(2) - y(1)]$ (‘*ATE21*’) and $E[y(2) - y(0)]$ (‘*ATE20*’), subsequently. The worst-case (WC) bounds of *ATE10*, which are generated under no particular assumptions on statistical properties of $E[y(t)]$, suggest that *ATE10* could be anywhere between -3.840 SD and 3.993 SD. Such bounds imply that a 10 percent increase in expenditure may change the test score of the student, who has $t = 1$ and mean values of expenditure (183.5) and test score (99.3), by negative 589 percent to positive 613 percent. Although such bounds are not very informative to infer the true effect of tutoring expenditure, they are much narrower than the bounds that range between -6 SD and 6 SD under no assumption whatsoever. The MTR assumption, which supposes non-negative effects of private tutoring, cuts the size of the WC bounds by more than a half, ruling out negative *ATE10* by design. The MTS assumption tightens the WC bounds mainly by lowering the upper bound of *ATE10* to 1.177 SD. However, the MIV assumption by itself fails to narrow the WC bounds substantially.

If we put the individual assumptions together to further tighten the bounds of *ATE10*, the joint MIV+MTR assumption narrows the MIV bounds of *ATE10* primarily by raising the lower bound (LB) to 0; the joint MIV+MTS assumption narrows the MIV bounds by lowering the upper bound (UB) to 1.162 SD. In addition, the joint MTR+MTS assumption considerably tightens the bounds of *ATE10* to a range between 0 SD and 0.36 SD mainly by ruling out negative *ATE10* under MTS by design. Finally, the joint MIV+MTR+MTS assumption yields the tightest bounds of *ATE10* between 0 SD and 0.345 SD. If one accepts the three assumptions of MIV, MTR, and MTS jointly to estimate the true effect of private tutoring, she can infer that the true magnitude of the effect of private tutoring is somewhere between 0 SD and 0.345

SD, while the expenditure (S_i) on average increases from 0 to 183.5.

In order to gain some perspective of our bounding results, in Table 6 we report percentage increases in test score due to a 10 percent increase in expenditure for ATE_{10} , ATE_{21} and ATE_{20} and for different samples, which are calculated at the mean values of the test score and expenditure of each sample. To facilitate interpretations and compare results of the bounding method with the 2SLS estimate in Table 2, it is necessary to note that the mean value of the tutoring expenditure (S_i) is 0 for $t = 0$, 183.5 for $t = 1$, and 677.2 for $t = 2$. (Recall that the cutoff expenditure (H_1) of 1 and 2 of t is set at 300.) Therefore, $S^{10} \equiv \widehat{E}(S_i|t = 1) - \widehat{E}(S_i|t = 0) = 183.5$; $S^{21} \equiv \widehat{E}(S_i|t = 2) - \widehat{E}(S_i|t = 1) = 493.8$; and $S^{20} \equiv \widehat{E}(S_i|t = 2) - \widehat{E}(S_i|t = 0) = 677.2$; the mean value of the averaged raw test score (Y_i) is $\widehat{E}(Y_i|t = 0) = 94.9$, $\widehat{E}(Y_i|t = 1) = 99.3$, and $\widehat{E}(Y_i|t = 2) = 103.3$. Using these figures and estimated bounds of ATEs, one can calculate a percentage change in test score due to a 10 percent increase in expenditure by

$$10 \times \frac{\widehat{\gamma}_{mn} \times 15.2 \times \widehat{E}(S_i|t = m)}{S^{mn} \times \widehat{E}(Y_i|t = m)}$$

where $15.2 = 1$ SD; $\widehat{\gamma}_{mn}$ is an estimated bound of ATE, $E[y(m) - y(n)]$ ($m > n$; $m, n = 0, 1, 2$); and Y_i is the averaged raw CSAT score before normalization.

INSERT TABLE 6 HERE.

Evaluated at the mean values of the test score (99.3) and expenditure (183.5), the preceding MIV+MTR+MTS bounds of ATE_{10} suggest that a 10 percent increase in expenditure on average raises the test score by at least 0 and at most 0.53 percent. The bootstrap confidence interval for the MIV+MTR+MTS bounds implies that the LB of ATE_{10} fails to be strictly greater than zero at the 10 percent significance level. In addition, the UB of ATE_{10} rules out substantial effects of private tutoring. Even the 95th percentile of the UB of ATE_{10} suggests no more than a 0.69 percent increase in test score due to a 10 percent rise in expenditure. Therefore, it seems doubtful that there exist strong positive causal impacts of private tutoring on student academic performance. Even the estimated UB of the MIV+MTR+MTS bounds of ATE_{10} is below the size of the effect of private tutoring estimated by 2SLS. However, to the extent that the 2SLS estimates in Table 2 are likely to overstate the true effect of private tutoring, our bounding results are not incompatible with the 2SLS result.

Next, if we examine an ATE $E[y(2) - y(1)]$ (*'ATE21'*), the primary results and patterns of changes in bounds with varying assumptions are largely similar to those of *ATE10*. The WC bounds of *ATE21* are not very informative in tightening the bounds of ATE, suggesting that *ATE21* could be anywhere between -3.537 SD and 3.669 SD. MTR cuts the size of the WC bounds by more than a half, ruling out negative *ATE21*. MTS tightens the WC bounds by lowering the UB of *ATE21* to 0.874 SD. MIV does not lead to tightening the WC bounds. The joint MIV+MTR assumption tightens the MIV bounds by raising the LB of *ATE21* to 0; the joint MIV+MTS assumption tightens the MIV bounds by lowering the UB to 0.864 SD. The joint MTR+MTS assumption considerably tightens the bounds of *ATE21*. Finally, the joint MIV+MTR+MTS assumption yields the tightest bounds of *ATE21* between 0 SD and 0.304 SD. From Table 6, evaluated at the mean values of the test score (103.3) and expenditure (677.2) for $t = 2$, the upper MIV+MTR+MTS bounds of *ATE21* suggest that a 10 percent increase in expenditure on average raises the test score by 0.615 percent at the largest.

The bootstrap confidence interval for the MIV+MTR+MTS bounds implies that the LB of *ATE21* also fails to be strictly greater than zero at the 10 percent significance level; the UB estimate and its 95 percentile of *ATE21* rule out strong effects of private tutoring. Therefore, it is also the case for *ATE21* that there do not exist strong positive causal impacts of private tutoring on student academic performance.

Finally, as for an ATE $E[y(2) - y(0)]$ (*'ATE20'*), the primary results and patterns of changes in bounds with varying assumptions are largely similar to the previous two cases. The joint MIV+MTR+MTS bounds of *ATE20* remain between 0 SD and 0.518 SD. From Table 6, they suggest that a 10 percent increase in expenditure raises the test score by at most 0.764 percent. The bootstrap confidence interval implies that the LB of *ATE20* also fails to be strictly greater than zero; the UB again rules out substantial effects of private tutoring. Namely, there do not seem to exist strong positive causal impacts of private tutoring on student academic performance. Such results largely reiterate the findings of 2SLS methods.

Given such a modest effect of private tutoring expenditures, it would be instructive to compare our 2SLS and bounding estimates with corresponding estimates of previous studies, precision of the estimates set apart, in order to gain some perspective of our results. Unfortunately, however, existing literature on the effect of private educational expenditures on test scores is quite scarce. Thus, we rely on estimates of the effect of *public* school expenditures on

student outcomes for comparison.

In the analysis of a randomization experiment on class size (Project STAR), Krueger (1999, Table VII) presents that a one student decrease in class size in grades K to 3 leads to a 0.67-0.88 percentile point increase in test score. Evaluated at the mean values of 21 students per class and 51 percentile test score (Appendix Table), these estimates imply a 2.8-3.6 percent improvement in test score due to a 10 percent decrease in class size and the accompanying 10 percent increase in per-pupil expenditure.¹⁴ Guryan (2003) shows that a \$1,000 increase in per-pupil spending leads to a third to a half of a standard-deviation increase in average test scores of 4th to 8th students in Massachusetts. Provided that pre-treatment levels of average spending and average test scores were roughly \$4,169 and 1,340 (Table 1) and that one standard deviation of the test score was approximately 74 points, his estimates suggest that a 10 percent increase in spending raises test scores by 0.77 to 1.15 percent. In stark contrast, Leuven et al. (2007) shows that roughly a 10 percent increase in school budget *reduces* language and arithmetic scores of eight graders by 0.1 SD, i.e., approximately 1.5 percent in our metric.

In terms of earnings in the labor market, Card and Krueger (1996, p.37) summarize that a 10 percent increase in public school spending leads to about a 1-2 percent increase in subsequent earnings. For example, they report that a reduced-form re-analysis of Card and Krueger (1992) presents a 1.1 percent increase in weekly earnings associated with a 10 percent reduction in the average pupil-teacher ratio.¹⁵ Other researchers find slightly weaker effects on earnings. Betts (1995) suggests that a 10 percent reduction in the average teacher-pupil ratio leads to a 0.4 percent increase in earnings. Grogger (1996) shows that a 10 percent increase in mean spending per student leads to a 0.7 percent increase in wages. Our 2SLS and bounds estimates of the effect of private tutoring expenditures are fairly comparable to the estimated effect of public school expenditures, while some estimates of the previous studies are slightly above our estimated bounds.

5.3 Disaggregated Results of the Bounds Analysis

Until now, our bounding analysis does not address heterogeneity in ATE, focusing exclusively on the whole sample and assuming that the effect of private tutoring is homogeneous for the entire population. In contrast, previous research suggests that the effects of educational resources may vary according to sex, race and the ability level of the student (Bedard 2003, Krueger 1999,

Rivkin et al. 2005). Moreover, to the extent that one takes our 2SLS estimates seriously, the estimated effect of private tutoring does not necessarily represent the ATE among students who are randomly chosen in the population; it is valid only for those students whose private tutoring expenditures are likely to be affected by birth order (Angrist et al. 1996). Given potential implications of heterogeneous effects for the current findings, there is a need to explore whether the causal effects of private tutoring vary by the characteristics of the student.

In order to address heterogeneity in ATE, we disaggregate the whole sample by the following two characteristics of the student: a student’s ability that is measured by the level of pre-tutoring quality, and sex. The bounding results of the subsamples are reported in Table 4 for different-ability students, and in Table 5 for male and female students.

5.3.1 Disaggregation by Student Ability

INSERT TABLE 4 HERE.

Let us first examine the results for low-ability students whose pre-tutoring quality is below the 50th percentile. The joint assumption bounds of ATEs are shown in Panel A of Table 4 for this group of students. Patterns of changes in bounds with varying assumptions are similar to those of previous cases. The tightest MIV+MTR+MTS bounds of ATE_{10} show that the true magnitude of the effect of private tutoring is somewhere between 0 SD and 0.322 SD for low-ability students, while the expenditure (S_i) on average increases from 0 to 189.4. From Table 6, the UB of ATE_{10} suggests that a 10 percent increase in expenditure on average raises the test score by at most 0.532 percent. The bootstrap confidence interval for the MIV+MTR+MTS bounds of ATE_{10} implies that the LB fails to be strictly greater than zero; the UB does not reveal substantial effects of private tutoring. Thus, there do not seem to exist strong positive causal impacts of private tutoring on academic performance of low-ability students.

The primary results for the bounds of ATE_{21} and ATE_{20} are largely similar to those of ATE_{10} , although the UB of ATE_{21} is slightly smaller and that of ATE_{20} is slightly greater than the UB of ATE_{10} . From Table 6, the MIV+MTR+MTS upper bound of ATE_{21} suggests that a 10 percent increase in expenditure raises the test score by at most 0.451 percent; that of ATE_{20} implies that a 10 percent increase in expenditure increases the test score by 0.672 percent at the largest. The LBs of both bounds fail to be strictly greater than zero; the UBs

do not present strong effects of the expenditure. Again, there exist negligible causal impacts of private tutoring on low-ability students.

Examining the bounding results for high-ability students whose pre-tutoring quality is above the 50th percentile in Panel B of Table 4, we find little evidence for heterogeneity of the effect of private tutoring across different-ability students. The main results for the bounds of ATE_{10} , ATE_{21} and ATE_{20} are largely similar to the counterpart of each ATE for low-ability students, although the UBs of ATEs are slightly greater for high-ability students. For high-ability students, the MIV+MTR+MTS upper bound of ATE_{10} suggests at most a 0.596 percent increase in test score due to a 10 percent increase in expenditure. The MIV+MTR+MTS upper bound of ATE_{21} reveals at most a 0.580 percent increase in test score, and that of ATE_{20} shows at most a 0.813 percent increase in test score due to a 10 percent increase in expenditure.

The bootstrap confidence intervals for the MIV+MTR+MTS bounds imply that each LB of the bounds for ATE_{10} , ATE_{21} and ATE_{20} fails to be strictly greater than zero; each UB does not show substantial effects of the expenditure. Namely, there are negligible causal impacts of private tutoring on academic performance of high-ability students as is the case of low-ability students.

5.3.2 Disaggregation by Sex

INSERT TABLE 5 HERE.

In addition to disaggregation by student ability, we divide the whole sample by sex of the student. As for male students, the tightest MIV+MTR+MTS bounds of ATE_{10} in Panel A of Table 5 show that the true magnitude of the effect of private tutoring is somewhere between 0 SD and 0.308 SD, while the expenditure (S_i) on average increases from 0 to 188.7. From Table 6, the UB of ATE_{10} suggests at most a 0.476 percent increase in test score due to a 10 percent increase in expenditure. The bootstrap confidence interval implies that the LB fails to be strictly greater than zero; the UB does not reveal strong effects of private tutoring.

The primary results for the bounds of ATE_{21} and ATE_{20} are largely similar to those of ATE_{10} . From Table 6, the MIV+MTR+MTS upper bound of ATE_{21} (ATE_{20}) suggests that a 10 percent increase in expenditure raises the test score by at most 0.582 (0.695) percent.

As for female students, we find little evidence for heterogeneity of the effect of private tutor-

ing across male and female students. The main results for the bounds of ATE_{10} , ATE_{21} and ATE_{20} are largely similar to the counterpart of each ATE for male students. For female students, the MIV+MTR+MTS upper bound of ATE_{10} suggests at most a 0.445 percent increase in test score due to a 10 percent increase in expenditure. The MIV+MTR+MTS upper bound of ATE_{21} (ATE_{20}) reveals a 0.661 (0.711) percent increase in test score due to a 10 percent increase in expenditure. There do not exist strong positive causal impacts of private tutoring on female students and male students alike.

6 Potential Explanations for the Modest Effect

Our main conclusion of the preceding analysis is that private tutoring expenditures have modest, if any, causal effects on student test performance. Such a finding may be contrary to expectations of many. Here we review a few potential explanations for the finding.

First, monetary educational investments may not always matter in children's educational outcomes. Although controversial, some researchers summarize that public school resources are not an important determinant of student outcomes. Our main results lend further support to the claim in light of private tutoring expenditures.

As potential explanations about small effects of public educational investments, researchers suggest poor management of educational resources due to large bureaucracies in the public sector (Anderson et al. 1991) and unionization of public school teachers (Hoxby 1996). Our study offers an answer to such explanations: private use of educational funds also has no more positive impacts on student outcomes than public management does. In Korea most of private educational institutions such as *hakwons* are small-sized and non-unionized. Thus public mismanagement due to bureaucracies and unionization is unlikely to be responsible for small effects of public educational expenditures.

Second, in response to the lack of significant effects of public school resources in the U.S., researchers pose a possibility that if added resources have diminishing effects on student achievement, the current level of public school expenditures may be on the flat portion of the production function. The expenditures may be more effective at lower levels than at higher levels (Betts 1996, Hanushek 1997). International educational statistics show that South Korea is placed below median among OECD countries with respect to educational expenditures: OECD

(2004) reports that the educational expenditures per student (private and public in total) of Korea is 20th largest among 26 OECD countries that have valid statistics. In 2001 Korea spent \$5,159 per student for secondary education, and \$8,873 for both primary and secondary education; in contrast, the U.S. spent \$8,779 for secondary, and \$16,339 for both primary and secondary education. Educational resources do not seem to be more effective at lower levels than at higher levels.

Third, as Hanushek (2003) implies, overall quality of teachers in the private sector may be responsible for small effects of private expenditures. In Korea, full-time public school teachers are tenured up to 62 years of age and enjoy the same employment benefits as government officials. In addition, school teachers enjoy public respect. In contrast, contracts of *hakwon* instructors are usually short-term in nature and fairly unstable as in other private small firms. This will cause teachers' quality in the private sector to be worse than that in the public sector. To the extent that teachers' quality can make differences in student outcomes, private investment of educational funds may have small effects.

Fourth, peer pressure among parents may explain the lack of the effect. According to the recent theory of group socialization, there are cultural factors in parenting practices, and their attitudes and beliefs about children; they are shared by parents' peer groups (Harris, 1995; 1998). When private tutoring is a norm in parents' peer groups, the decision to invest in children's tutoring may not be based on an objective cost-benefit analysis; they may be based on a subjective/cultural belief about the effectiveness of private tutoring, or the concern about their being viewed by the peers as neglectful of children's education. If the decision about tutoring is based on peer pressure, small effects of private tutoring will not be a big surprise.

7 Concluding Remarks

In order to shed light on the effectiveness of educational investments for student outcomes, this paper examines an unexplored dimension of educational inputs—private tutoring expenditures of South Korean parents. In the face of difficulties in causal estimation, the paper employs a nonparametric bounding method. With the method we show that the true effect of private tutoring remains at most modest. A 10 percent increase in expenditure on private tutoring leads to at most a 0.764 percent improvement in test score. Such a modest effect, however, is fairly

comparable to the effects of public school expenditures on test scores and earnings estimated by previous studies.

Disaggregating the entire sample by a student's sex and ability level reveals no considerable differences in the effect of tutoring expenditures. There exists only a modest effect of private tutoring for male and female students alike, and for students of different ability levels. As potential explanations for modest effects of private tutoring, we propose lower teacher quality in the private education sector of Korea and peer pressure of parents. Yet, whether monetary educational investments raise student educational outcomes in different contexts remains to be further examined.

Notes

¹Such a threshold is arbitrary. As robustness checks, we construct two alternative T_i 's by employing KRW 200,000 and 400,000 as a new threshold between 1 and 2 of T_i . The results based on such thresholds are qualitatively similar to those reported in the current paper. The alternative results are available upon request.

²In our data set discussed shortly, there are not many variables available by which we can examine such unobservable supports of parents for a certain-parity child. Nevertheless, we attempt an indirect method to address how differently parents treat a first-born child relative to other siblings by focusing on early childhood. Using the information on how often parents read books to a child in pre-school period (1 if very often and 0 otherwise) and how often they help the child's homework and examination in elementary school period (1 if very often or often and 0 otherwise), we run linear probability models in which the student's personal and family backgrounds are controlled for together with the first-born indicator. The results show that parents more often read to a first-born child in pre-school period (coefficient 0.125; SE 0.027), and help the first-born child's homework and examination in elementary school period (coefficient 0.073; SE 0.029). In addition, parents have greater educational aspiration for a first-born child (a post-graduate degree as opposed to a university degree or below) than for other siblings (coefficient 0.052; SE 0.026).

³For an overview of secondary education and private tutoring in South Korea, see Kim and Lee (2001) and OECD (1998). For an alternative explanation about widespread private tutoring that focuses on the ability-mixing school system of Korea, see Hur and Kang (2007).

⁴Guidelines of CSAT stipulate that students are free to choose individual subjects for their examination. Nevertheless, the majority of students choose either Korean, mathematics, English or all, because they are required by many universities for application. Out of a national total of 574,218 CSAT writers in 2004, 98.9 percent select Korean, 87.8 percent mathematics, 99.3 percent English, 34 percent science and 59.1 percent social studies (Korea Institute of Curriculum and Evaluation (KICE), Media Briefing in December, 2004). In the KEEP data set, 98.2 percent of a total of 1,733 CSAT writers in 2004 choose Korean. The corresponding figures are 89.6 and 98.0 percent for math and English, respectively.

⁵For 2004 general high school accounts for 70.5 percent of 1.75 million high school students in the nation; vocational high school accounts for the remainder (*Yearbook of Educational Statistics 2004*, National Statistical Office). As a robustness check, we have estimated the model same as that of Table 2, including a total of 356 vocational graduates who have valid CSAT scores. The 2SLS (weighted by sampling weights) estimates are similar to those reported in the paper, suggesting that a 10 percent increase in expenditure enhances a student’s performance by a 0.015 SD of test score—by 0.23 percent. Such results are available upon request.

⁶According to the KEEP data, students who attend special high schools for music, fine arts and athletics or take private tutoring of these fields expend a monthly average of $W513,700$ —approximately \$430. Other students, on the other hand, spend nearly half as much on private tutoring—a monthly average of $W283,200$. In addition, the average CSAT scores are substantially different between these two groups of students. The average CSAT score of the arts and athletic students is 88.5; that of the non-specialist group is 99.5. The difference is significantly different from zero.

⁷We consider a student to have no intention of writing the 2004 CSAT, thus treat him or her as a non-writer, if the student either does not apply for the 2004 CSAT, or is absent (upon application) for at least one test of the three main subjects (Korean, math, and English). The total number of such non-writers is 262. Although not reported in the paper, we attempt to take those non-writers of CSAT into consideration by substituting hypothetical test scores for missing scores. The results, which are suppressed but available upon request, are qualitatively

similar.

⁸The KEEP survey asks the monthly average amount of overall expenditures on private tutoring during the last six months before grade 12—roughly nine to fourteen months prior to the CSAT test.

⁹Because of the traditional leveling policy of Korean government in secondary education, ability mixing is widely applied in Korean high schools. As a result, a student’s rank will not vary substantially, whether an entire school or a single classroom is employed as a reference group for ranking students. We do not attempt to convert the ranks reported by grade-12 homeroom teachers across the school and class levels, since there is no sufficient information to support objectively this conversion in the KEEP data set.

As the KEEP survey collects information on private tutoring during the second semester of grade 11 (as mentioned in footnote 8), an assumption required in our analysis is that it takes some time for private tutoring to take effect in a student’s performance. Otherwise, our measure of a student’s pre-treatment quality is error-ridden, because it contains part of short-term effects of private tutoring.

¹⁰Given cross-sectional data, the current study makes inter-household comparisons of children in the analysis rather than intra-household comparisons. As a result, it will not be inappropriate to have only children in the analysis sample and compare them as first-born children with later-born children in other households. (If we rely on intra-household comparisons, however, we should exclude only children because they have no siblings to compare with.) In our regression analysis we include only children and control for their status with a dummy variable, since they may have unique advantages and disadvantages in education. As a robustness check, we exclude only children from the analysis sample but the primary results are not substantially affected.

¹¹Although the coefficient for F_i is significantly different from zero in the regression setting, F_i fails to be a strong IV for tutoring expenditures for the student (F-statistic is 7.26), according to a criterion of Stock et al. (2002). Hence we apply testing methods that are robust to the presence of weak IVs (Stock et al. 2002). At the bottom of columns (3) and (4) of Table 2, we report the 95 percent confidence intervals for β_1 that are based on test statistics of the Anderson-Rubin (AR) test (Anderson and Rubin 1949), the Lagrangian Multiplier (LM) test (Kleibergen 2002), and the Conditional Likelihood Ratio (CLR) test (Moreira 2003).

¹²To deal with a zero spending in the log transformation, a value of 10 is added to every

student's raw value of tutoring expenditure. The value of 10 is used since it is the smallest accounting unit reported in the survey ($W10,000$) and it is about 3.5 percent of the mean expenditure on private tutoring. Whether a smaller value (e.g., 1) is added to every expenditure or the level of raw values is employed rather than the log, the results are qualitatively similar.

¹³This sensitivity analysis is unnecessary for the MTR+MTS and MIV+MTR+MTS bounds, because they are not a function of K_0 and K_1 . The results based on alternative values of K_0 and K_1 are available upon request.

¹⁴Krueger (2003, F55-F56) infers that a one percent decrease in class size will be approximately converted into a one percent increase in annual per pupil cost.

¹⁵Our estimated effect of tutoring expenditures is dwarfed by a meta-analysis of Hedges et al. (1994) that yields an estimate that a 10 percent increase in public school expenditure produces an improvement in student performance of approximately 0.7 standard deviations. This amount is equivalent to a 10.7 percent improvement in test score in our metric. The summary of Hedges et al. (1994) are, however, criticized by Hanushek (1997) for being biased in favor of large positive effects of school expenditures.

References

- Altonji, J.G., Elder, T.E., Taber, C.R., 2005. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools, *Journal of Political Economy* 113 (1), 151-184.
- Altonji, J.G., Elder, T.E., Taber, C.R., 2005b. An evaluation of instrumental variable strategies for estimating the effects of Catholic schooling, *Journal of Human Resources* 40 (4), 791-821.
- Anderson, G.M., Shughart, W.F., Tollison, R., 1991. Educational achievement and the cost of bureaucracy, *Journal of Economic Behavior & Organization* 15 (1), 29-45.
- Anderson, T.W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations, *Annals of Mathematical Statistics* 20 (1), 46-63.
- Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables, *Journal of the American Statistical Association* 91 (434), 444-455.
- Baker, D.P., Akiba, M., LeTendre, G.K., Wiseman, A.W., 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.
- Bedard, K., 2003. School quality and the distribution of male earnings in Canada, *Economics of Education Review* 22, 395-407.
- Behrman, J.R., Taubman, P., 1986. Birth order, schooling, and earnings, *Journal of Labor Economics* 4 (3), S121-S145.
- Betts, J.R., 1995. Does school quality matter? Evidence from the national longitudinal survey of youth, *Review of Economics and Statistics* 77 (2), 231-250.
- Betts, J.R., 1996. Is there a link between school inputs and earnings? Fresh scrutiny of an old literature, In: Burtless, G. (Eds.), *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Brookings, Washington, DC, pp. 141-191.
- Bjerkedal, T., Kristensen, P., Skjeret, G.A., Brevik, J.I., 2007. Intelligence test scores and birth order among young Norwegian men (conscripts) analyzed within and between families *Intelligence* 35 (5), 503-514.
- Black, S.E., Devereux, P.J., Salvanes, K.G., 2005. The more the merrier? The effect of family size and birth order on children's education, *Quarterly Journal of Economics* 120 (2), 669-700.

- Black, S.E., Devereux, P.J., Salvanes, K.G., 2007. Older and Wiser? Birth Order and IQ of Young Men, NBER Working Paper 13237.
- Blundell, R., Gosling, A., Ichimura, H., Meghir, C., 2007. Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds, *Econometrica* 75 (2), 323-363.
- Bray, M., 1999. The shadow education system: Private tutoring and its implications for planners, UNESCO International Institute for Educational Planning, Paris.
- Card, D., Krueger, A.B., 1992. Does school quality matter? returns to education and the characteristics of public schools in the united states, *Journal of Political Economy* 100 (1), 1-40.
- Card, D., Krueger, A. B., 1996. School resources and student outcomes: An overview of the literature and new evidence from North and South Carolina, *Journal of Economic Perspectives* 10 (4), 31-50.
- Evans, W.N., Schwab, R.M., 1995. Finishing high school and starting college: Do Catholic schools make a difference?, *Quarterly Journal of Economics* 110 (4), 941-974.
- Figlio, D.N., Stone, J.A., 1999. Are private schools really better?, In: Polachek, S. (Eds.), *Research in Labor Economics*, Vol.18, JAI Press, Stamford, Connecticut, pp. 115-140.
- Gerfin, M., Schellhorn, M., 2006. Nonparametric bounds on the effect of deductibles in health care insurance on doctor visits - Swiss evidence, *Health Economics* 15 (9), 1011-1020.
- Goldhaber, D.D., 1996. Public and private high schools: Is school choice an answer to the productivity problem?, *Economics of Education Review* 15 (2), 93-109.
- Gonzalez, L., 2005. Nonparametric Bounds on the Returns to Language Skills, *Journal of Applied Econometrics* 20 (6), 771-795.
- Greenwald, R., Hedges, L.V., Laine, R.D., 1996. The effect of school resources on student achievement, *Review of Educational Research* 66 (3), 361-396.
- Grogger, J., 1996. School expenditures and post-schooling earnings: Evidence from High School and Beyond, *Review of Economics and Statistics* 78 (4), 628-637.
- Guryan, J., 2003. Does Money Matter? Estimates from Education Fianance Reform in Massachusetts, Mimeo.
- Hanushek, E.A., 1986. The economics of schooling: Production and efficiency in public schools, *Journal of Economic Literature* 24 (3), 1141-1177.

- Hanushek, E.A., 1997. Assessing the effects of school resources on student performance: An update, *Educational Evaluation and Policy Analysis* 19 (2), 141-164.
- Hanushek, E.A., 2003. The failure of input-based schooling policies, *Economic Journal* 113 (485), F64-F98.
- Harris, J.R., 1995. Where is the child's environment? A group socialization theory of development, *Psychological Review* 102 (3), 458-489.
- Harris, J.R., 1998. *The Nurture assumption: Why children turn out the way they do*, New York, NY:Touchstone.
- Haveman, R., Wolfe, B., 1995. The Determinants of Children's Attainments: A Review of Methods and Findings, *Journal of Economic Literature* 33 (4), 1829-1878.
- Hedges, L.V., Laine, R.D., Greenwald, R., 1994. Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes, *Educational Researcher* 23 (3), 5-14.
- Herrnstein, R., Murray, C., 1994. *The Bell Curve*, The Free Press, New York.
- Hotz, V.J., Mullin, C.H., Sanders, S.G., 1997. Bounding Causal Effects Using Data from a Contaminated Natural Experiment: Analysing the Effects of Teenage Childbearing, *Review of Economic Studies* 64 (4), 575-603.
- Hoxby, C.M., 1996. How teachers' unions affect education production, *Quarterly Journal of Economics* 111 (3), 671-718.
- Hur, J., Kang, C., 2007. *School Systems and Efficiency and Equity of Education*, Working Paper, Department of Economics, National University of Singapore.
- Kim, S., Lee, J.-H., 2001. Demand for education and developmental state: Private tutoring in South Korea, *Social Science Research Network Electronic Paper Collection*: <http://ssrn.com/abstract=268284>.
- Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental variables regression, *Econometrica* 70 (5), 1781-1803.
- Kreider, B., Pepper, J.V., 2007. Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors, *Journal of the American Statistical Association* 102 (478), 432-441.
- Krueger, A.B., 1999. Experimental estimates of education production functions, *Quarterly Journal of Economics* 114 (2), 497-532.

- Krueger, A.B., 2003. Economic considerations and class size, *Economic Journal* 113 (485), F34-F62.
- Lechner, J., 1999. Nonparametric bounds on employment and income effects of continuous vocational training in East Germany, *Econometrics Journal* 2 (1), 1-28.
- Leuven, E., Lindahl, M., Oosterbeek, H., Webbink, D., 2007. The Effect of Extra Funding for Disadvantaged Pupils on Achievement, *Review of Economics and Statistics* 89 (4), 721-736.
- Manski, C.F., 1990. Nonparametric Bounds on Treatment Effects, *American Economic Review* 80 (2), 319-323.
- Manski, C.F., 1997. Monotone Treatment Response, *Econometrica* 65(6), 1311-1334.
- Manski, C.F., Nagin, D.S., 1998. Bounding Disagreements About Treatment Effects: A Case Study of Sentencing and Recidivism, *Sociological Methodology* 28, 99-137
- Manski, C.F., Pepper, J.V., 2000. Monotone Instrumental Variables, with an Application to the Returns to Schooling, *Econometrica* 68 (4), 997-1012.
- Ministry of Education, 1999, 2000. Sakyoyuk Siltae Josa [Survey on Private Tutoring in 1999 (2000)], Seoul, South Korea (in Korean).
- Moreira, M.J., 2003. A conditional likelihood ratio test for structural models, *Econometrica* 71 (4), 1027-1048.
- Neal, D., 1997. The effects of Catholic secondary schooling on educational achievement, *Journal of Labor Economics* 15 (1), 98-123.
- OECD (Organization for Economic Co-Operation and Development), 1998. *Reviews of National Policies for Education: Korea*, OECD Publications, Paris.
- OECD (Organization for Economic Co-Operation and Development), 2004. *Education at a glance 2004*, OECD Publications, Paris.
- Pepper, J.V., 2000. The Intergenerational Transmission Of Welfare Receipt: A Nonparametric Bounds Analysis, *Review of Economics and Statistics* 82 (3), 472-488.
- Retherford, R.D., Sewell, W.H., 1991. Birth order and intelligence: Further tests of the confluence model, *American Sociological Review* 56 (2), 141-158.
- Rivkin, S.G., Hanushek, E.A., Kain, J.F., 2005. Teachers, schools, and academic achievement, *Econometrica* 73 (2), 417-458.

- Rodgers, J.L., Cleveland, H.H., Oord, E. V.D., Rowe, D.C., 2000. Resolving the debate over birth order, family size, and intelligence, *American Psychologist* 55 (6), 599-612.
- Rosenzweig, M.R., 1986. Birth spacing and sibling inequality: Asymmetric information within the family, *International Economic Review* 27 (1), 55-76.
- Stock, J.H., Wright, J.H., Yogo, M., 2002. A survey of weak instruments and weak identification in Generalized Method of Moments, *Journal of Business and Economic Statistics* 20 (4), 518-529.
- Time Asia, 2006. Asia's overscheduled kids, 167 (12), 48-55.
- Zajonc, R.B., 1976. Family configuration and intelligence, *Science* 192 (4236), 227-236.
- Zajonc, R.B., Mullally, P.R., 1997. Birth order reconciling conflicting effects, *American Psychologist* 52 (7), 685-699.

Figure 1: Bounds of the Average Treatment Effects

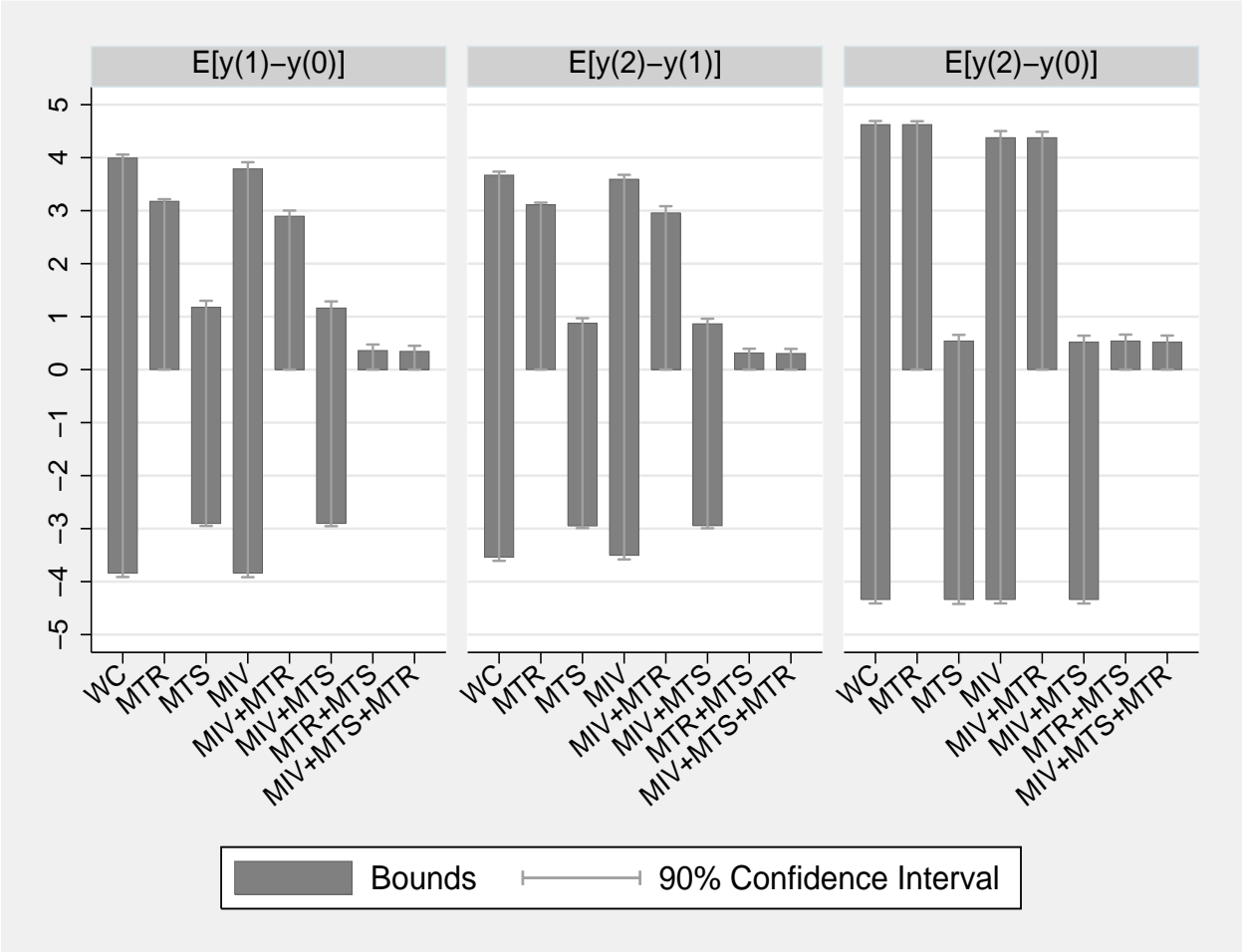


Table 1: Descriptive Statistics of the Main Sample

Variable	Total Sample			(1) First-born		(2) Later-born		Differences ((1)-(2))		
	N	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.E.	T-value
Normalized Z-score of three tests	1467	0.089	0.968	0.161	0.936	0.005	0.999	0.156	0.051	3.09
Average score of three tests	1467	99.64	15.23	100.77	14.72	98.31	15.71	2.459	0.795	3.09
Test score of Korean	1455	100.94	16.98	102.34	16.31	99.30	17.60	3.036	0.890	3.41
Test score of math	1395	98.60	17.99	98.70	17.95	98.47	18.06	0.229	0.968	0.24
Test score of English	1466	99.72	18.00	101.35	17.48	97.82	18.41	3.526	0.939	3.76
Tutoring Expenditure (W1, 000)	1467	297.4	342.5	336.8	377.8	251.1	289.3	85.63	17.81	4.81
Any tutoring (Yes=1)	1467	0.799	0.401	0.854	0.354	0.735	0.442	0.119	0.021	5.71
Tutoring hours for Korean	1467	1.327	2.509	1.492	2.611	1.133	2.371	0.359	0.131	2.74
Tutoring for Korean (Yes=1)	1467	0.313	0.464	0.352	0.478	0.267	0.443	0.086	0.024	3.54
Tutoring hours for math	1467	2.658	3.206	2.965	3.330	2.298	3.016	0.667	0.167	3.99
Tutoring for math (Yes=1)	1467	0.557	0.497	0.616	0.487	0.487	0.500	0.129	0.026	4.99
Tutoring hours for English	1467	1.793	2.575	2.057	2.544	1.484	2.578	0.572	0.134	4.27
Tutoring for English (Yes=1)	1467	0.429	0.495	0.496	0.500	0.350	0.477	0.147	0.026	5.71
Prior quality	1087	44.54	25.94	43.40	25.17	45.91	26.79	-2.504	1.578	-1.59
Hours of self-study	1467	12.01	10.21	12.36	10.34	11.58	10.06	0.780	0.535	1.46
Age	1467	17.73	0.517	17.72	0.528	17.75	0.503	-0.032	0.027	-1.20
Male (Yes=1)	1467	0.569	0.495	0.569	0.495	0.569	0.496	0.001	0.026	0.02
Only child (Yes=1)	1467	0.073	0.260	0.135	0.342	0.000	0.000	0.135	0.013	10.26
Sibship size	1467	2.185	0.647	2.001	0.536	2.400	0.698	-0.399	0.032	-12.37
Parents' average age	1467	46.14	3.194	44.97	2.848	47.51	3.033	-2.532	0.154	-16.47
Parents' average education	1467	12.20	2.470	12.54	2.410	11.79	2.479	0.750	0.128	5.86
Both parents present (Yes=1)	1467	0.928	0.258	0.928	0.259	0.929	0.257	-0.001	0.014	-0.06
Single father (Yes=1)	1467	0.026	0.159	0.029	0.168	0.022	0.148	0.007	0.008	0.82
Single mother (Yes=1)	1467	0.046	0.209	0.043	0.203	0.049	0.216	-0.006	0.011	-0.54
Books at home	1467	188.7	213.2	201.4	218.6	173.8	205.8	27.58	11.148	2.47
Family income (W1, 000)	1467	324.9	197.1	340.5	204.0	306.6	187.3	33.83	10.292	3.29
School characteristics										
Private school (Yes=1)	1467	0.552	0.497	0.576	0.495	0.524	0.500	0.051	0.026	1.97
Coed school (Yes=1)	1467	0.445	0.497	0.441	0.497	0.450	0.498	-0.010	0.026	-0.37
Boys-only school (Yes=1)	1467	0.309	0.462	0.321	0.467	0.296	0.457	0.024	0.024	1.01
Girls-only school (Yes=1)	1467	0.245	0.430	0.239	0.427	0.253	0.435	-0.015	0.023	-0.65
Rural area (Yes=1)	1467	0.155	0.362	0.140	0.347	0.173	0.379	-0.033	0.019	-1.75
Medium city (Yes=1)	1467	0.344	0.475	0.346	0.476	0.341	0.474	0.005	0.025	0.21
Metropolitan city (Yes=1)	1467	0.501	0.500	0.514	0.500	0.486	0.500	0.028	0.026	1.07
First-born child (Yes=1)	1467	0.540	0.499							
First-born boy (Yes=1)	1467	0.307	0.462							
First-born girl (Yes=1)	1467	0.232	0.423							

Table 2: OLS and 2SLS Estimates of the Effect of Tutoring Expenditures on Performance: Birth Order as an IV

Estimation method:	OLS		2SLS	
	Normalized test score	Tutoring Expenditure (level)	Normalized test score	
Dependent variable:	(1)	(2)	(3)	(4)
Tutoring Expenditure (in level)	0.0001 (0.0001) [0.060]		0.0017 (0.0011) [0.762]	
Tutoring Expenditure (in log)				0.268 (0.171) [0.410]
First-born child		47.62 (17.67)**		
Hours of self-study	0.015 (0.002)**	0.668 (0.772)	0.014 (0.002)**	0.015 (0.003)**
Prior quality (Q2)	0.619 (0.070)**	41.88 (29.67)	0.545 (0.099)**	0.541 (0.094)**
Prior quality (Q3)	0.925 (0.071)**	-0.270 (27.43)	0.917 (0.080)**	0.885 (0.087)**
Prior quality (Q4)	1.466 (0.075)**	4.814 (29.88)	1.454 (0.087)**	1.437 (0.096)**
Prior quality missing	0.824 (0.072)**	-59.49 (25.40)*	0.910 (0.103)**	0.890 (0.103)**
Single father	-0.276 (0.107)**	-24.83 (38.58)	-0.247 (0.131)	-0.243 (0.121)*
Single mother	0.050 (0.099)	-4.558 (23.79)	0.063 (0.098)	0.139 (0.110)
Books at home	0.000 (0.000)	0.070 (0.043)	0.000 (0.000)	0.000 (0.000)
Family income	0.000 (0.000)	0.646 (0.076)**	-0.001 (0.001)	0.000 (0.000)
Parents' avg edu	0.049 (0.010)**	14.17 (3.523)**	0.026 (0.020)	0.026 (0.019)
Parents' avg age	0.002 (0.007)	6.392 (2.502)*	-0.004 (0.008)	0.004 (0.006)
Age	-0.108 (0.039)**	5.369 (15.71)	-0.119 (0.047)*	-0.126 (0.044)**
Male	-0.004 (0.064)	-12.86 (21.64)	0.021 (0.074)	0.020 (0.081)
Only child	-0.117 (0.093)	2.398 (38.34)	-0.152 (0.108)	-0.113 (0.098)
Sibship size	-0.002 (0.039)	-33.67 (12.05)**	0.056 (0.058)	0.045 (0.049)
Intercept	-0.228 (0.753)	-507.9 (293.8)	0.362 (0.997)	-0.852 (0.860)
School characteristics	Yes	Yes	Yes	Yes
95% CI				
AR			(-0.0003, 0.0076)	(-0.048, 0.777)
LM			(-0.0003, 0.0076)	(-0.048, 0.777)
CLR			(-0.0003, 0.0076)	(-0.048, 0.777)
F (IVs excluded from the 2nd stage)		7.26		
R-square	0.373	0.270	0.155	0.277
Number of sample	1,467	1,467	1,467	1,467

Note: Standard errors are reported in parentheses. * and ** indicate that the estimate is significant at the 0.05 and 0.01 levels, respectively. The numbers in square brackets are percent changes in test score due to a 10 percent change in expenditure that are evaluated at mean values.

Table 3: Bounds of $E[y(t)]$ and Average Treatment Effects for the Whole Sample

Assumptions:	Worst-case (WC)				Monotone treatment response (MTR)			
	LB	UB	LB 5 pctile	UB 95 pctile	LB	UB	LB 5 pctile	UB 95 pctile
$E[y(0)]$	-2.440	2.354	-2.498	2.417	-2.440	0.089	-2.489	0.132
$E[y(1)]$	-1.486	1.553	-1.559	1.622	-0.926	0.736	-0.996	0.791
$E[y(2)]$	-1.985	2.183	-2.054	2.241	0.089	2.183	0.047	2.237
$E[y(1) - y(0)]$	-3.840	3.993	-3.913	4.059	0.000	3.176	0.000	3.217
$E[y(2) - y(1)]$	-3.537	3.669	-3.608	3.736	0.000	3.109	0.000	3.151
$E[y(2) - y(0)]$	-4.338	4.623	-4.412	4.692	0.000	4.623	0.000	4.688
Assumptions:	Monotone treatment selection (MTS)				Monotone IV (MIV)			
	LB	UB	LB 5 pctile	UB 95 pctile	LB	UB	LB 5 pctile	UB 95 pctile
$E[y(0)]$	-0.214	2.354	-0.305	2.415	-2.270	2.354	-2.362	2.414
$E[y(1)]$	-0.549	0.963	-0.621	1.039	-1.486	1.519	-1.558	1.604
$E[y(2)]$	-1.985	0.325	-2.059	0.395	-1.985	2.106	-2.057	2.191
$E[y(1) - y(0)]$	-2.903	1.177	-2.950	1.300	-3.840	3.788	-3.917	3.914
$E[y(2) - y(1)]$	-2.948	0.874	-2.990	0.971	-3.503	3.592	-3.581	3.676
$E[y(2) - y(0)]$	-4.338	0.539	-4.422	0.656	-4.338	4.376	-4.411	4.502
Assumptions:	MIV+MTR				MIV+MTS			
	LB	UB	LB 5 pctile	UB 95 pctile	LB	UB	LB 5 pctile	UB 95 pctile
$E[y(0)]$	-2.270	0.089	-2.349	0.130	-0.202	2.354	-0.310	2.414
$E[y(1)]$	-0.850	0.625	-0.943	0.707	-0.548	0.960	-0.619	1.030
$E[y(2)]$	0.089	2.106	0.049	2.193	-1.985	0.316	-2.053	0.389
$E[y(1) - y(0)]$	0.000	2.895	0.000	3.002	-2.902	1.162	-2.954	1.288
$E[y(2) - y(1)]$	0.000	2.957	0.000	3.084	-2.944	0.864	-2.994	0.961
$E[y(2) - y(0)]$	0.000	4.376	0.000	4.488	-4.338	0.518	-4.413	0.641
Assumptions:	MTR+MTS				MIV+MTR+MTS			
	LB	UB	LB 5 pctile	UB 95 pctile	LB	UB	LB 5 pctile	UB 95 pctile
$E[y(0)]$	-0.214	0.089	-0.318	0.130	-0.202	0.089	-0.300	0.131
$E[y(1)]$	0.011	0.146	-0.038	0.190	0.012	0.143	-0.036	0.188
$E[y(2)]$	0.089	0.325	0.049	0.391	0.089	0.316	0.048	0.391
$E[y(1) - y(0)]$	0.000	0.360	0.000	0.474	0.000	0.345	0.000	0.450
$E[y(2) - y(1)]$	0.000	0.314	0.000	0.395	0.000	0.304	0.000	0.392
$E[y(2) - y(0)]$	0.000	0.539	0.000	0.661	0.000	0.518	0.000	0.643

Table 4: Bounds of $E[y(t)]$ and Average Treatment Effects by Ability

Panel A: Low-ability students ($N = 454$)								
Assumptions:	MIV+MTR				MIV+MTS			
	LB	UB	LB 5 pctile	UB 95 pctile	LB	UB	LB 5 pctile	UB 95 pctile
$E[y(0)]$	-2.408	-0.402	-2.531	-0.336	-0.665	2.320	-0.858	2.441
$E[y(1)]$	-1.219	0.113	-1.353	0.253	-0.868	0.697	-0.976	0.827
$E[y(2)]$	-0.402	1.891	-0.467	2.053	-2.111	-0.248	-2.216	-0.123
$E[y(1) - y(0)]$	0.000	2.521	0.000	2.711	-3.188	1.362	-3.267	1.603
$E[y(2) - y(1)]$	0.000	3.109	0.000	3.307	-2.808	0.620	-2.876	0.777
$E[y(2) - y(0)]$	0.000	4.299	0.000	4.493	-4.431	0.417	-4.562	0.635
Assumptions:	MTR+MTS				MIV+MTR+MTS			
	LB	UB	LB 5 pctile	UB 95 pctile	LB	UB	LB 5 pctile	UB 95 pctile
$E[y(0)]$	-0.721	-0.402	-0.867	-0.336	-0.665	-0.402	-0.858	-0.336
$E[y(1)]$	-0.452	-0.341	-0.527	-0.274	-0.452	-0.343	-0.532	-0.273
$E[y(2)]$	-0.402	-0.235	-0.467	-0.105	-0.402	-0.248	-0.469	-0.131
$E[y(1) - y(0)]$	0.000	0.379	0.000	0.535	0.000	0.322	0.000	0.518
$E[y(2) - y(1)]$	0.000	0.217	0.000	0.369	0.000	0.204	0.000	0.337
$E[y(2) - y(0)]$	0.000	0.486	0.000	0.689	0.000	0.417	0.000	0.636
Panel B: High-ability students ($N = 633$)								
Assumptions:	MIV+MTR				MIV+MTS			
	LB	UB	LB 5 pctile	UB 95 pctile	LB	UB	LB 5 pctile	UB 95 pctile
$E[y(0)]$	-2.374	0.475	-2.502	0.532	0.118	2.531	-0.074	2.608
$E[y(1)]$	-0.589	0.916	-0.762	1.009	-0.103	1.284	-0.213	1.389
$E[y(2)]$	0.475	2.143	0.417	2.253	-1.783	0.702	-1.912	0.789
$E[y(1) - y(0)]$	0.000	3.290	0.000	3.443	-2.634	1.166	-2.707	1.389
$E[y(2) - y(1)]$	0.000	2.732	0.000	2.937	-3.067	0.805	-3.136	0.946
$E[y(2) - y(0)]$	0.000	4.517	0.000	4.686	-4.313	0.584	-4.436	0.800
Assumptions:	MTR+MTS				MIV+MTR+MTS			
	LB	UB	LB 5 pctile	UB 95 pctile	LB	UB	LB 5 pctile	UB 95 pctile
$E[y(0)]$	0.089	0.475	-0.076	0.532	0.118	0.475	-0.089	0.534
$E[y(1)]$	0.390	0.533	0.320	0.590	0.395	0.530	0.321	0.591
$E[y(2)]$	0.475	0.705	0.418	0.788	0.475	0.702	0.414	0.795
$E[y(1) - y(0)]$	0.000	0.444	0.000	0.618	0.000	0.412	0.000	0.624
$E[y(2) - y(1)]$	0.000	0.315	0.000	0.426	0.000	0.307	0.000	0.425
$E[y(2) - y(0)]$	0.000	0.617	0.000	0.792	0.000	0.584	0.000	0.801

Table 5: Bounds of $E[y(t)]$ and Average Treatment Effects by Sex

Panel A: Male students ($N = 835$)								
Assumptions:	MIV+MTR				MIV+MTS			
	LB	UB	LB 5 pctl	UB 95 pctl	LB	UB	LB 5 pctl	UB 95 pctl
$E[y(0)]$	-2.286	0.038	-2.393	0.097	-0.216	2.318	-0.369	2.416
$E[y(1)]$	-0.913	0.612	-1.035	0.727	-0.609	0.936	-0.723	1.036
$E[y(2)]$	0.038	2.072	-0.022	2.192	-1.998	0.250	-2.088	0.350
$E[y(1) - y(0)]$	0.000	2.898	0.000	3.052	-2.927	1.152	-2.999	1.326
$E[y(2) - y(1)]$	0.000	2.985	0.000	3.149	-2.934	0.859	-3.000	1.013
$E[y(2) - y(0)]$	0.000	4.358	0.000	4.517	-4.316	0.466	-4.424	0.635
Assumptions:	MTR+MTS				MIV+MTR+MTS			
	LB	UB	LB 5 pctl	UB 95 pctl	LB	UB	LB 5 pctl	UB 95 pctl
$E[y(0)]$	-0.253	0.038	-0.393	0.099	-0.216	0.038	-0.345	0.099
$E[y(1)]$	-0.034	0.095	-0.107	0.162	-0.033	0.092	-0.105	0.159
$E[y(2)]$	0.038	0.257	-0.024	0.362	0.038	0.250	-0.024	0.352
$E[y(1) - y(0)]$	0.000	0.348	0.000	0.501	0.000	0.308	0.000	0.451
$E[y(2) - y(1)]$	0.000	0.291	0.000	0.418	0.000	0.283	0.000	0.404
$E[y(2) - y(0)]$	0.000	0.510	0.000	0.688	0.000	0.466	0.000	0.627
Panel B: Female students ($N = 632$)								
Assumptions:	MIV+MTR				MIV+MTS			
	LB	UB	LB 5 pctl	UB 95 pctl	LB	UB	LB 5 pctl	UB 95 pctl
$E[y(0)]$	-2.248	0.158	-2.370	0.220	-0.084	2.401	-0.262	2.479
$E[y(1)]$	-0.767	0.642	-0.910	0.750	-0.469	0.990	-0.566	1.092
$E[y(2)]$	0.158	2.151	0.098	2.260	-1.968	0.405	-2.072	0.510
$E[y(1) - y(0)]$	0.000	2.890	0.000	3.060	-2.870	1.074	-2.933	1.272
$E[y(2) - y(1)]$	0.000	2.918	0.000	3.100	-2.957	0.874	-3.014	1.014
$E[y(2) - y(0)]$	0.000	4.400	0.000	4.570	-4.368	0.489	-4.477	0.694
Assumptions:	MTR+MTS				MIV+MTR+MTS			
	LB	UB	LB 5 pctl	UB 95 pctl	LB	UB	LB 5 pctl	UB 95 pctl
$E[y(0)]$	-0.157	0.158	-0.278	0.217	-0.084	0.158	-0.243	0.222
$E[y(1)]$	0.070	0.211	0.000	0.275	0.070	0.209	-0.003	0.278
$E[y(2)]$	0.158	0.416	0.100	0.518	0.158	0.405	0.094	0.512
$E[y(1) - y(0)]$	0.000	0.369	0.000	0.500	0.000	0.293	0.000	0.465
$E[y(2) - y(1)]$	0.000	0.346	0.000	0.468	0.000	0.335	0.000	0.454
$E[y(2) - y(0)]$	0.000	0.573	0.000	0.726	0.000	0.489	0.000	0.667

Table 6: Bounds of the Percentage Increase in Test Score due to a 10 Percent Increase in Expenditure

Panel A: Whole sample								
	LB	UB	LB 5 pctile	UB 95 pctile				
$E[y(1) - y(0)]$	0.000	0.529	0.000	0.690				
$E[y(2) - y(1)]$	0.000	0.615	0.000	0.792				
$E[y(2) - y(0)]$	0.000	0.764	0.000	0.948				
Panel B: Subsamples by ability								
	Low-ability				High-ability			
	LB	UB	LB 5 pctile	UB 95 pctile	LB	UB	LB 5 pctile	UB 95 pctile
$E[y(1) - y(0)]$	0.000	0.532	0.000	0.858	0.000	0.596	0.000	0.902
$E[y(2) - y(1)]$	0.000	0.451	0.000	0.746	0.000	0.580	0.000	0.802
$E[y(2) - y(0)]$	0.000	0.672	0.000	1.024	0.000	0.813	0.000	1.116
Panel C: Subsamples by sex								
	Male				Female			
	LB	UB	LB 5 pctile	UB 95 pctile	LB	UB	LB 5 pctile	UB 95 pctile
$E[y(1) - y(0)]$	0.000	0.476	0.000	0.697	0.000	0.445	0.000	0.707
$E[y(2) - y(1)]$	0.000	0.582	0.000	0.832	0.000	0.661	0.000	0.898
$E[y(2) - y(0)]$	0.000	0.695	0.000	0.934	0.000	0.711	0.000	0.969