Private Tutoring and Primary Students' Performance
Evidence From Gulf States

Imen Hentati

In their efforts to meet academic standards, students often rely on supplementary resources such as private tutoring. This study investigates the relationship between private tutoring and academic performance in mathematics and science among primary school students in the Gulf States. Using data from the 2019 Trends in International Mathematics and Science Study (TIMSS), we apply both Ordinary Least Squares (OLS) regression and Propensity Score Matching (PSM) techniques to account for student background characteristics and assess the differential effects of tutoring across performance levels. Results reveal a U-shaped pattern of effectiveness, with students in the lowest and highest achievement quartiles appearing to benefit the most from tutoring. These findings underscore the importance of targeted educational support and caution against interpreting the impact of private tutoring as uniformly beneficial.

Keywords: private tutoring, academic achievement, mathematics, science, Gulf States, propensity score technique, TIMSS 2019.

Imen Hentati is a full-time Assistant Professor at the Faculty of Economics and Management of Tunis, University of Tunis El Manar. She is also a researcher at the Laboratory for Research on Quantitative Development Economics (LAREQUAD). She holds a PhD in Economics from Cergy-Pontoise University (France), as well as an HDR in Quantitative Methods from the same faculty where she currently teaches. Her research interests include applied econometrics, family economics, and education outcomes in Arab countries. She is a member of the faculty's scientific and disciplinary councils and actively engages in academic collaborations. Email: imen.hintati@fsegt.utm.tn

Hentati, I., (2025). Private Tutoring and Primary Students' Performance. *Journal of Education in Muslim Societies, 7*(1), 76–100.
DOI 10.2979/jems.00042 •

INTRODUCTION

In the Gulf States, extensive research has explored various factors affecting student performance, including socioeconomic status, school resources, teacher performance, and community characteristics (Chapman & Miric, 2009; Bouhlila, 2017). Despite substantial government investment in education, student performance in international assessments such as the Trends in International Mathematics and Science Study (TIMSS) indicates that the quality of education in the Gulf States remains below international standards. Private tutoring, a widespread practice in the region, has been identified as a significant factor influencing student performance and warrants a comprehensive investigation. International research studies have highlighted the growing significance of private tutoring and its potential impact on the educational systems (Bray, 2003). The global proliferation of private tutoring highlights the necessity of examining its prevalence in the Gulf States and understanding its implications for both students and educational systems.

Globally, the prevalence and impact of private tutoring on national academic achievement have drawn attention, driven by comparative assessments such as TIMSS (Beaton et al., 1996) and PISA (OECD, 2001). The educational practices in high-performing countries such as Japan have sparked interest, particularly regarding the extent and effects of private tutoring. Despite this importance, private tutoring in the Gulf States has received limited scholarly attention. Accordingly, this paper aims to fill this gap by examining the impact of private tutoring on students' academic performance in the Gulf region. Across the Gulf States, students receive private tutoring in academic subjects outside regular school hours, effectively extending their learning time. This structured support, which aligns with the school curriculum, is commonly referred to as "shadow education." This practice can significantly enhance students' academic performance and represents a considerable financial investment by families in their children's education. However, empirical evidence remains inconclusive regarding whether private tutoring improves overall student achievement (Bray, 2007; Dang & Rogers, 2008; Mori & Baker, 2010). The inconsistent findings of these studies may be due to cultural and institutional differences across countries and methodological variations (Brehm & Silova, 2014; Guill & Bos, 2014; Ireson & Rushforth, 2014).

The Gulf States present a unique context due to the significant proportion of expatriate residents. Non-Arab expatriates often bring their cultural backgrounds and enroll their children in schools following the curricula

of their home countries, which are frequently supplemented by shadow education. Using the Gulf States as a case study, this research aims to examine the disparities in academic achievements between students who receive private tutoring and those who do not. Specifically, this study aims to provide empirical evidence on the relationship between private tutoring and academic performance, as measured by the TIMSS 2019 assessments.

This study contributes to the existing empirical literature on the relationship between academic performance and private tutoring. It has two main objectives. First, it extends the research on the relationship between early-age private tutoring and academic achievement. Second, it seeks to provide insights for developing effective policies for private tutoring in the Gulf States. By addressing these objectives, this research aims to enhance understanding of the dynamics surrounding private tutoring and its impact on academic outcomes in the region.

CONTEXT OF THE RESEARCH STUDY

The Gulf States have invested substantially in developing educational systems that prioritize access to education for all and emphasize the acquisition of knowledge and skills in a globalized economy. Concurrently, the private tutoring business has witnessed a remarkable expansion within these States. This growth is evident in diverse educational settings, where schools may adopt different curricula and offer varying educational programs. In this context, the inclusion or exclusion of supplementary tutoring as part of the overall educational package varies across institutions (Bray & Ventura, 2022).

Private tutoring is prevalent in the Gulf States, prompting numerous national and subnational studies to investigate its prevalence and impact. These studies aim to understand the landscape of private tutoring by examining factors such as participation rates, subjects covered, student demographics, and the reasons driving families to seek additional educational support (Bray & Hajar, 2022). By examining the prevalence of private tutoring, these studies make a significant contribution to ongoing discussions about educational practices in the Gulf States. They provide valuable insights into the educational decisions made by students and their families, highlighting the diverse factors influencing educational choices in the region.

In Bahrain, Albuhi and Alsadah (1994) surveyed 290 students across 28 schools and found that 69% were receiving private tutoring. Among

primary students, 84% were receiving tutoring. Moreover, students attending private schools were more inclined to seek supplementary tutoring compared to their counterparts in public schools. Similar findings were observed in Kuwait, where a 2009 survey of 785 students from various secondary schools in five cities found that 69% of the participants were receiving private tutoring (Al-Salhi et al., 2009). A 2013 study that randomly sampled 40 students from Grades 11 and 12 across four schools found that 62.5% of them were enrolled in private tutoring programs (Al-Mari & Al-Khamees, 2013). In Oman, a 2012 report documented a substantial expansion of private tutoring, particularly in Muscat, attributed to rising household incomes and expatriate teachers supplementing their salaries (The Sultanate of Oman, 2012). Notably, male students received a higher proportion of tutoring compared to female students due to factors such as the limited availability of female tutors and traditional norms that restricted female students from visiting tutors' homes or using public transportation.

In Qatar, a 2015 national survey conducted among 1803 students across 38 secondary schools revealed widespread private tutoring participation. Among Qatari students, 46% received private tutoring, compared to 28% for expatriate students. In international schools, the proportions were even higher, with 55% of Qatari students and 39% of expatriate students receiving private tutoring (Alemadi et al., 2016). A follow-up survey in 2018, encompassing 1639 students from 34 schools, demonstrated that the proportions remained relatively consistent, with 45% of students in government schools and 34% in international schools engaging in private tutoring (Sellami, 2019). In the context of Dubai, a 2017/18 national study involving 3929 parents of students in Grades 5, 9, 10, and 12 revealed that 27% of students were enrolled in private tutoring programs (Rocha & Hamed, 2018). Interestingly, the proportion of Emirati students receiving tutoring was higher (32%) compared to non-Emirati students (21%). In the specific context of Dubai, a recent study conducted by Bray and Ventura (2022) examined the educational landscape within private schools. The study found that a significant proportion of students in these schools seek supplementary tutoring. Specifically, approximately half of the students enrolled in private educational institutions in Dubai receive private tutoring. The phenomenon of private tutoring is not new in the Gulf States, but it has expanded considerably in recent decades. Private tutoring participation has also extended to intermediate and primary students.

THEORETICAL BACKGROUND

Private tutoring has become a prevalent supplementary educational practice outside traditional schooling, aimed at enhancing students' learning and academic performance. The assumption underlying private tutoring is that additional focused study time in specific subjects will lead to improved knowledge acquisition and better exam outcomes. However, its impact on student achievement, particularly at the primary level, remains a subject of ongoing debate and research, with findings in the literature being inconclusive (Dang & Roger, 2008; Bray, 2007; Mori & Baker, 2010).

Several studies have shown positive correlations between private tutoring and primary academic outcomes. For example, Banerjee et al. (2007) conducted randomized experiments in urban Indian schools over two years, involving grade 3 and 4 students, and found significant improvements in basic literacy and numeracy skills. Similarly, based on Vietnam Living Standards Surveys (VLSS) data from the late 1990s, Dang (2007) examined the significant role of private tutoring in Vietnamese households for primary and lower secondary students.

Conversely, other studies present a more nuanced view, indicating that private tutoring has limited or negligible effects on the academic achievements of primary students. For instance, research on Indonesian fourth graders by Suryadarma et al. (2006) found no significant impact on mathematics or dictation scores. Similarly, Ha and Harpham (2005) found that tutoring had no significant effect on writing and numeracy skills among eight-year-olds in Vietnam.

Furthermore, certain studies have suggested that private tutoring has a negative impact on academic achievement. Fang and Zhang (2018), analyzing data from China's National Collaboration and Innovation Center, reported a significant negative correlation between the amount of post-class study time and the academic performance of primary students.

Another aspect explored in the literature is the varying impact of private tutoring across different subject areas. For instance, Ha and Harpham (2005) observed positive effects on reading skills but no significant effects on writing and numeracy among eight-year-olds in Vietnam.

The empirical evidence on the impact of private tutoring varies considerably due to differing methodological approaches. A significant methodological concern in the literature is the need to control for endogeneity. Recent studies have increasingly addressed this issue through methods like instrumental variables and propensity score matching. For example, Suryadarma et al. (2006) employed instrumental variables to investigate the

impact of tutoring on mathematics and dictation scores among Indonesian fourth graders, finding no significant effect. In contrast, Keuren (2016) employed propensity score matching with World Bank data to demonstrate higher literacy, numeracy, and scores among Cambodian fourth graders who received tutoring compared to those who did not. Similarly, Kang (2007) employed Ordinary Least Squares (OLS) and Propensity Score Matching to examine the impact of tutoring on CSAT results in Korea, finding a slight positive effect on academic achievement.

In summary, the literature on the effects of private tutoring on primary students presents a complex picture with conflicting findings. While some studies suggest positive impacts on academic performance, others indicate neutral or negative associations. These findings call for caution in interpreting the impact of private tutoring as uniformly beneficial.

TIMSS DATA AND METHODS

TIMSS Data

The primary objective of the TIMSS (Trends in International Mathematics and Science Study) assessments is to establish a robust and valid measure of mathematics and science knowledge and skills. These assessments hold significant global importance within the education community and are designed to align with the curricula of participating countries. First conducted in 1994/1995, TIMSS initially included 45 countries, with Kuwait being the sole Arab participant at that time. Participation of Arab nations in TIMSS has expanded over time, covering both fourth and eighth-grade levels.

In the latest TIMSS assessment conducted in 2019, 64 countries and eight benchmarking participants were involved. Notably, nine Arab countries participated at the fourth-grade level, while eleven participated at the eighth-grade level. The TIMSS 2019 assessment included six Gulf States: Bahrain, Oman, Qatar, the United Arab Emirates, Saudi Arabia, and Kuwait. Dubai also participated as a benchmarking participant.

TIMSS assessments are strategically scheduled at the end of the academic year to facilitate meaningful comparisons across countries. This timing minimizes potential variations in instructional content and timing across different educational systems, enhancing cross-country comparability.

The selection of Gulf States as a sample for our study investigating the impact of private tutoring at the primary level using TIMSS data is justified for several reasons. First, the Gulf States have demonstrated a strong commitment to education, evidenced by significant investments and

educational reforms aimed at enhancing student performance and academic outcomes (Bray & Hajar, 2022). Second, the availability of comprehensive TIMSS data from these countries supports this selection. Finally, the Gulf States provide a diverse representation in terms of educational systems, cultural backgrounds, and socioeconomic contexts, offering a comprehensive perspective on the impact of private tutoring at the primary level.

Sample Design and Exclusion

Our study utilizes the TIMSS 2019 dataset, focusing specifically on fourth-grade students, a globally recognized assessment program known for its meticulous sampling procedures, which aim to ensure accuracy and reliability in estimates that represent the broader population. To maintain consistency across TIMSS cycles, rigorous exclusion criteria are applied at both school and student levels, as detailed by Martin et al. (2020).

At the school level, criteria exclude schools in remote areas or those with very small student populations, typically four students or fewer in the target grade. Schools that deviate significantly from mainstream education or cater exclusively to students with special needs are also excluded, ensuring the dataset reflects typical educational contexts.

International guidelines specifically dictate within-school exclusions for students with functional disabilities, intellectual disabilities, and non-native language speakers. These exclusions aim to standardize assessment conditions, ensuring results accurately reflect the abilities of the target population. Overall, our sample maintained an exclusion rate below 5.6%, upholding these guidelines while ensuring a robust and representative dataset (see Table A-1).

For analysis, we combined datasets from four Gulf States (Bahrain, Oman, Qatar, and the United Arab Emirates)[1]. Merging datasets offers advantages beyond maximizing information; it significantly increases the number of observations, providing a robust foundation for exploring the impacts of early-age private tutoring. This expanded dataset enables us to investigate the effects of private tutoring on fourth-grade students, a critical period for educational interventions that influence academic performance and learning outcomes.

Including fourth-grade students is pivotal, as it captures a formative stage where educational interventions, such as private tutoring, can have a profound impact on academic trajectories. Dataset merging enhances statistical power, bolstering study reliability and generalizability.

Our merged dataset comprises 14,822 observations. Students were divided into two groups: a treatment group receiving private tutoring and a control group without private tutoring. Across the sample, 18% of students received mathematics tutoring and 13% received science tutoring.

METHODS

Empirical Model and Technique

OLS regression is used first to estimate the average treatment effect of private tutoring. The variable treatment is introduced in **Equation 1** as a dummy variable.

$$P_{i,c,s} = \alpha_0 + \alpha_1 F_{i,c,s} + \alpha_2 T_{i,c,s} + \varepsilon_{i,c,s} \quad \textbf{(1)}$$

In our analysis, P_{ics} represents the score of student i in class c at school s, while F_{ics} denotes a vector of individual and family background characteristics. T_{ics} serves as the treatment variable, taking the value 1 for treated observations and 0 for control observations. To ensure robust statistical inferences, we apply students' sampling weights.

The Ordinary Least Squares (OLS) regression examines significant performance differences between treatment and control groups. However, the application of the OLS method to model (1) requires that several fundamental assumptions be satisfied to ensure the validity and reliability of the estimated coefficients. These assumptions include:

Linearity

The OLS model assumes a linear relationship between the dependent variable $P_{i,c,s}$ (student achievement) and the explanatory variables $F_{i,c,s}$ and $T_{i,c,s}$. This implies that the marginal effect of each explanatory variable on achievement is constant across the distribution of the dependent variable. However, this assumption may not be appropriate in our context. Specifically, the effect of tutoring is unlikely to be uniform across all students: those with lower prior achievement may benefit more from tutoring than higher-performing students.

To address concerns regarding the potential non-linearity in the relationship between prior achievement and tutoring, and to avoid imposing linear assumptions, we categorized student achievement into quartiles. We then included these quartile indicators and their interactions with the tutoring variable in the regression model. This approach allows the

estimated effect of tutoring to vary by achievement level and provides a more flexible dummy variable specification :

$$P_{i,c,s} = \alpha_0 + \alpha_1 F_{i,c,s} + \alpha_2 T_{i,c,s} + \Sigma_q \beta_q Q_{i,c,s,q} + \Sigma_q \gamma_q (T_{i,c,s} * Q_{i,c,s,q}) + \varepsilon_{i,c,s} \quad \textbf{(2)}$$

Where $Q_{i,c,s,q}$ are four student achievement quartiles (with quartile 4 as the reference group) and $T_{i,c,s} * Q_{i,c,s,q}$ represents the interaction between tutoring and achievement quartile.

Absence of Multicollinearity

OLS estimation assumes that the explanatory variables ($F_{i,c,s}$ and $T_{i,c,s}$) are not perfectly or highly correlated with one another. High multicollinearity inflates the standard errors of the coefficient estimates, making it difficult to identify the individual effects of each variable and reducing the reliability of inference.

To assess multicollinearity and better understand how well the conditions were controlled, we examined the correlation matrix between the main variables included in the analysis. In addition, we computed Variance Inflation Factors (VIFs) to quantify the extent of multicollinearity (see Table A-2). In summary, the correlation matrix, alongside the VIF test results (VIF = 2.4 < 5), supports the adequacy of our covariate selection and confirms that key assumptions regarding independence and the control of confounding factors are reasonably met, within the limits of observable characteristics. These diagnostic checks increase confidence in the robustness of the estimated effects.

Exogeneity of the Explanatory Variables:

The explanatory variables must be exogenous, meaning that they are not correlated with the error term $\varepsilon_{i,c,s}$. If this condition is violated, such as in the presence of omitted variables, measurement error, or simultaneity, the OLS estimates will be biased and inconsistent. In the context of our model, the variable "private tutoring" is endogenous, meaning that unobserved factors (e.g., motivation, parental involvement, teacher quality) influencing both tutoring investment and student performance may lead to biased estimates. If unaddressed, these factors contribute to the error term $\varepsilon_{i,c,s}$, rendering estimation results inconsistent due to correlation with the private tutoring variable. This selection bias is akin to challenges in educational literature, such as assessing whether religious or private schools outperform public schools (Bedi & Garg, 2000; Newhouse & Beegle, 2005).

To mitigate endogeneity in private tutoring, we employ propensity score matching (PSM). PSM is a quasi-experimental technique where each

treated unit is paired with a similar non-treated unit based on observed characteristics. The propensity score estimates the likelihood of enrolling in private tutoring, allowing treated units to be matched with untreated units as if assignment were randomized (Rubin, 2001). To adjust for potential confounding from pre-treatment variables (individual and family characteristics), we use a logit model to predict the propensity for treatment. Subsequently, Nearest Neighbor Matching is applied with replacement to pair cases in both groups based on their propensity scores (Frisco et al., 2007).

Variables

This section describes the outcome variables used in this study, as well as the covariates and the treatment variable.

Outcome variables

The dependent variables in our study are overall performance in mathematics and science, as assessed by TIMSS. Student achievement in TIMSS is measured using sets of five plausible values: Math1, Math2, Math3, Math4, and Math5 for mathematics, and Science1, Science2, Science3, Science4, and Science5 for science. Plausible values are imputed scores drawn from estimated ability distributions based on comprehensive student background data (Martin et al., 2016; Mislevy, 1991). These values are not intended as individual student scores but rather as estimates for students with similar response patterns and background characteristics within the sampled population. This methodology ensures accurate estimation of population characteristics (Martin et al., 2020). For further details on plausible values methodology, refer to Mislevy (1991).

Covariates

The core explanatory variables in our model encompass both individual and family background characteristics. Individual factors include student age, years in pre-primary education, language of instruction (1 for English, 0 for classical Arabic), and student status. Family factors consist of parental education and occupation, interrelation score, and possession of educational resources such as computers and internet access (coded as 1 for yes and 0 for no).

In our model, nominal variables were converted into dummy variables. For instance, "Language" took a value of 1 for students instructed in English and 0 for those in classical Arabic. Variables related to the possession of educational resources were coded as 1 if the response was affirmative and 0 if it was negative. Regarding parental education, occupation, and student status, specific categories were selected as reference points: "some primary"

for parental education, "never worked for pay" for parental occupation, and "any parent born in the country" for student status. The variable "interrelation" was constructed as a quantitative measure, ranging from 0 to 18, to capture parental involvement, support, and engagement in their children's education. This score serves as a quantitative indicator for our analysis.

Treatment variable

The treatment variable "Treat" is binary, taking a value of 1 if the student attended private tutoring outside of school in mathematics (and separately for science) within the past 12 months. It is coded as 0 if the student did not attend such tutoring sessions.

Descriptive Statistics

In the following sections, we present a detailed overview of the descriptive statistics for both the outcome variables and covariates used, as illustrated in Tables 1, 2, and 3. Continuous variables are typically summarized

Table 1. Mean score difference between treatment and non-treatment in mathematics achievements

Variables	Mean of Total sample	Mean of Control Group	Mean of Treatment Group	Effect Size (Cohen's d)
Math1	493.944	499.487	468.637	-0.326
Math 2	493.950	499.397	469.078	-0.319
Math 3	494.715	499.854	471.251	-0.305
Math 4	493.621	499.142	468.413	-0.325
Math 5	493.300	498.746	468.433	-0.321
Observations	14822	12159	2663	

Table 2. Mean score difference between treatment and non-treatment in science achievements

Variables	Mean of Total sample	Mean of Control Group	Mean of Treatment Group	Effect Size (Cohen's d)
Science 1	501.810	508.278	458.941	-0.491
Science 2	500.729	507.485	455.948	-0.507
Science 3	500.954	507.524	457.401	-0.495
Science 4	500.275	507.086	455.127	-0.506
Science 5	502.123	508.493	459.900	-0.478
Observations	14822	12879	1943	

Table 3. Descriptive statistics of covariates for all students

Variables	Mean/Proportion	Min	Max
Age	9.752	8	11
Inter-relation	16.204	0	18
Possess Computers	0.894	0	1
Possess Internet	0.886	0	1
Pre- Education	1.745	0	4
Language Test	0.522	0	1
Students' Status			
Any parent was born in the country	0.432	0	1
Only one parent was born in the country	0.088	0	1
Both of parents were born in the country.	0.480	0	1
Parents' Education			
Some primary	0.031	0	1
Lower Secondary	0.025	0	1
Upper secondary	0.166	0	1
Post-secondary	0.129	0	1
University or higher	0.648	0	1
Parents' occupation			
Never worked for pay	0.068	0	1
General laborer	0.016	0	1
Skilled worker	0.036	0	1
Clerical	0.160	0	1
Small business owner	0.073	0	1
Professional	0.646	0	1
Children's books			
0–10	0.325	0	1
11–25	0.288	0	1
26–50	0.212	0	1
51–100	0.111	0	1
More than 100	0.064	0	1
Observations	14822		

using means to depict the central tendency of the data. Meanwhile, categorical variables such as socioeconomic status are presented with proportions to illustrate their distribution among distinct categories or groups.

Within both the treated and control groups, the mean values of the outcome variables are close to the global average of approximately 500 points. Interestingly, when assessing the impact of private tutoring on mathematics and science performance, a negative trend becomes apparent.

While the effect size is relatively small in mathematics and moderate in science (for more detailed variability, see Borenstein, 2009; Hattie, 2009), the negative sign of the effect size indicates that participating in private tutoring leads to less favorable outcomes compared to not receiving such additional support.

The students included in the sample exhibit an average age of 9.75 years. In terms of educational background, students typically received two years of pre-primary education. Furthermore, a variable called "students' status" is utilized to distinguish between native and expatriate students. This variable considers whether both parents, one parent, or neither parent was born in the country, helping capture the students' demographic background and immigration status. Analyzing the sample composition, it is observed that approximately 50% of the students don't hold national citizenship, indicating a significant presence of expatriate students within the dataset. Additionally, when examining the language used during the testing process, it is revealed that 52% of the students took the tests in Arabic.

Among parents in the Gulf States, it is observed that a significant proportion possess a high level of education. Specifically, 64.8 percent of parents hold a university degree or higher, indicating a high educational attainment within these households. Moreover, a considerable portion of parents, specifically 64.6 percent, are employed in professional occupations. Compared to the full TIMSS population, the results show that the distribution of key background characteristics (such as parental education and occupation) is very similar to that of the final analytical sample used after selection. This suggests that the high proportion of highly educated and professionally employed parents is not a result of selection bias introduced by our methodology, but rather reflects the characteristics of the TIMSS population in the Gulf States (see Tables A-4 and A-5).

In terms of the home environment and available resources, the TIMSS data provides insights into the presence of children's books and other essential learning materials. The findings reveal that the Gulf States exhibit relatively low numbers of children's books in households. More than half of the students have fewer than 25 books available at home. This suggests a lower emphasis on reading and access to books for children within the region. However, it is noteworthy that parents in the Gulf States demonstrate the ability to afford basic resources for their children's education, as

evidenced by the high percentage (over 70 percent) of students who possess computers and internet connections.

To gain a more comprehensive understanding of the parent-child relationship in the Gulf States, a score was constructed to assess various aspects[2]. Notably, the average score obtained by the tested students in this regard is 16.2, indicating a supportive parental environment within the Gulf States.

RESULTS AND DISCUSSION

This section provides a detailed analysis of the results obtained through Ordinary Least Squares (OLS) regression and propensity score matching techniques. OLS regression was independently conducted five times for each assessment, and Rubin's rules (Rubin, 1987) were applied to aggregate these results, ensuring a comprehensive and reliable estimate. Table 4 summarizes the findings from the OLS regression analysis, while Table 5 presents results from propensity score matching.

Table 4. OLS Results

Variables	**Mathematics Achievement**		**Science Achievement**	
	Coefficient	***Standard Errors***	***Coefficient***	***Standard Errors***
Age	3.167**	(0.611)	2.837**	(0.686)
Inter-relation	0 .780**	(0.122)	0.833**	(0.152)
Possess Computers	4.166**	(0.985)	4.687**	(1.165)
Possess Internet	2.719**	(0.937)	4.714**	(1.094)
Pre- Education	0.664**	(0.229)	0.784**	(0.254)
Test's Language	-1.873**	(0.679)	-8.460**	(0.835)
Students' Status				
Only one parent was	-6.064**	(0.950)	-8.290**	(1.083)
born in the country	-6.198**	(0.656)	-9.799**	(0.777)
Both of parents were born in the country.				
Parents' Education				
Lower Secondary	1.816	(3.133)	12.169**	(3.891)
Upper secondary	7.787**	(2.555)	19.974**	(3.169)
Post-secondary	8.076**	(2.583)	20.770**	(3.180)
University or higher	10.800**	(2.526)	24.126**	(3.113)

(Continued)

Table 4. (Continued)

Variables	Mathematics Achievement		Science Achievement	
	Coefficient	*Standard Errors*	*Coefficient*	*Standard Errors*
Parents' occupation				
General laborer	-2.978	(2.867)	0.703	(3.387)
Skilled worker	-0.083	(1.912)	2.961	(2.267)
Clerical	1,646	(1.383)	1.119	(1.631)
Small business owner	5.018**	(1.507)	5.499**	(1.738)
Professional	3.977**	(1.303)	5.768**	(1.523)
Children's books				
11–25	1.369**	(0.651)	2.212**	(0.745)
26–50	3.323**	(0.710)	4.071**	(0.795)
51–100	5.993**	(0.925)	7.066**	(0.973)
More than 100	10.559**	(1.257)	9.342**	(1.248)
Quarter				
Q1	-223.085**	(1.010)	-233.416**	(1.116)
Q2	-138.646**	(0.765)	-134.066**	(0.717)
Q3	-79.379**	(0.734)	-73.727**	(0.671)
Treat*Quarter				
Treat*Q1	-5.925**	(1.575)	-12.496**	(2.184)
Treat*Q2	-1.718*	(0.835)	-2.587**	(0.981)
Treat*Q3	-0.559	(0.853)	0.225	(1.021)
Treat*Q4	-5.607**	(1.806)	-5.889**	(2.110)
Constant	543.527**	(7.098)	541.847**	(8.199)

*Notes: ** p<.01, * p<.05,*

Table 5. ATET Nearest Neighbor Results

ATE TREAT (1 vs. 0)	*Overall Performance in Mathematics*		*Overall Performance in Science*	
	Coefficients	*Standard Errors*	*Coefficient*	*Standard Errors*
For the First Quartile (Q_1)	-6.443**	(1.870)	-11.738**	(2.327)
For the Second Quartile (Q_2)	-1.794*	(0.834)	-3.567*	(0.991)
For the Third Quartile (Q_3)	-1.503	(0.820)	-1.865	(1.005)
For the Fourth Quartile (Q_4)	-4.535**	(1.930)	-8.473**	(2.251)

*Notes: ** p<0..01, * p<.0.05*

The OLS results reveal significant heterogeneity in the effect of tutoring depending on students' initial performance levels (quartiles). The coefficients on the quartile indicators Q_1, Q_2, *and* Q_3 are all strongly negative and

highly significant, confirming that lower-achieving students perform substantially worse compared to their higher-achieving peers in the absence of tutoring. Specifically, in mathematics achievement, students in the first quartile score, on average, 223 points less than those in the top quartile, followed by gaps of 139 and 79 points for the second and third quartiles, respectively. Similar disparities are observed in science achievement, with corresponding score differences of 233, 134, and 73 points.

Crucially, the interaction terms $Treat^*Q_1$, $Treat^*Q_2$, and $Treat^*Q_3$ provide insight into how the effect of tutoring varies across performance groups. The interaction coefficient for the lowest-performing students in mathematics is negative ($Treat^*Q_1$= -5.9) and statistically significant, indicating that the marginal effect of tutoring is slightly weaker for this group compared to the top quartile. However, despite this relative difference, the overall impact of tutoring remains strong and significant for students in the first quartile. This finding supports the idea that low-achieving students are more likely to receive tutoring than middle-achieving students, and while the effect may be slightly attenuated, they still benefit substantially. Likewise, the interaction term for the second quartile ($Treat^*Q_2$ = -1.73) is also significant, suggesting a modestly lower effect compared to top performers. The effect for the third quartile ($Treat^*Q_3$ = -0.58) is not significant and is statistically indistinguishable from that of the reference group (quartile 4), suggesting comparable benefits from tutoring for higher and upper-middle achievers. Similar results are observed in science achievement. OLS results highlight a U-shaped pattern in tutoring effectiveness: students at both ends of the performance distribution, particularly the lowest and highest achievers, tend to benefit more from tutoring, whereas those in the middle show relatively smaller gains.

We now turn to the socioeconomic variables. Consistent with prior studies (Ammermüller et al., 2005; Chiu & Khoo, 2005), family-related factors have a significant and positive impact on students' performance. This highlights the crucial role of family factors in influencing academic outcomes. Additionally, enrollment in preschool education emerges as a significant predictor of overall performance, confirming earlier research findings (Holla et al., 2021) and highlighting the enduring benefits of early education in fostering academic success. An intriguing observation specific to the Gulf States is the performance gap between native and expatriate students. Native students tend to achieve lower performance levels compared to their expatriate peers. Lastly, the findings in Table 4 indicate that students who took tests in Arabic outperformed those who

took them in English, aligning with previous studies (Bouhlila & Hentati, 2022) and emphasizing the importance of language of instruction in academic achievement.

One important limitation of our methodological approach lies in the potential endogeneity of the tutoring variable. Since the data are cross-sectional, and tutoring and achievement are measured simultaneously, it is difficult to establish a clear causal direction from tutoring to academic performance. In fact, there is a strong possibility of reverse causality: students who perform poorly may be more likely to receive tutoring, rather than tutoring leading to improved performance. Consequently, OLS estimates may be biased due to this endogenous relationship. While we attempt to address nonlinearity and heterogeneity of effects by interacting tutoring with achievement quartiles, this specification does not resolve the core endogeneity issue.

To address the endogeneity of private tutoring, we employ Propensity Score Matching (PSM), a quasi-experimental method widely used in the literature. PSM estimates the propensity score, defined as the probability of receiving treatment given a set of observed characteristics. Based on these scores, treated students are matched with untreated students who have similar observable features, thereby mimicking a randomized experimental design (Rubin, 2001). In our implementation, we first estimate the propensity scores using a logit model, where treatment status is regressed on pretreatment covariates (e.g., individual and family background variables). Then we use Nearest Neighbor Matching with replacement to pair each treated observation with a control observation that has the closest propensity score (Frisco et al., 2007). This approach helps mitigate bias due to observable differences and allows for a more credible estimation of the causal effect of private tutoring.

When estimating the effect of tutoring separately for each achievement quartile using propensity score matching, the results consistently indicate that students who received tutoring tended to perform worse than their matched peers who did not receive tutoring. In Q_1 and Q_4, the negative effects are both strong and statistically significant (-6.59 and -8.71 points for mathematics achievement, and -11.73 and -8.47 points for science achievement, respectively). In Q_2, the effect is small but still significant (-1.79 points for mathematics and -3.65 for science). While in Q_3, the effect is not statistically significant. These findings do not necessarily imply that tutoring causes a drop in performance. Rather, they are consistent with a negative selection effect, where students who are struggling, even among high achievers, are more likely to receive tutoring. This highlights

the difficulty of identifying a causal effect without addressing the problem of reverse causality and unobserved heterogeneity.

While PSM helps address bias from observed variables, it does not resolve endogeneity due to reverse causation. As such, causal interpretation remains limited, and we refrain from making strong claims about the effectiveness of tutoring on achievement. Instead, our findings should be interpreted as associational, reflecting patterns in tutoring participation and student outcomes in the context of the Gulf education systems.

CONCLUSION AND RECOMMENDATIONS

This study provides a comprehensive analysis of the relationship between private tutoring and student academic performance using both Ordinary Least Squares (OLS) regression and Propensity Score Matching (PSM) techniques. The findings reveal a complex and heterogeneous picture of tutoring effectiveness across different student performance levels.

The OLS results suggest a U-shaped pattern of effectiveness, with students in the lowest and highest achievement quartiles appearing to benefit the most from tutoring. However, the propensity score matching results paint a more cautious picture, indicating that students who receive tutoring often perform worse than their matched peers, particularly in the lowest and highest quartiles. These findings are consistent with a negative selection bias: students who struggle academically, regardless of their overall performance level, are more likely to seek tutoring, possibly reflecting parental concern or remedial strategies rather than tutoring per se driving academic improvement.

A major limitation lies in the potential endogeneity of the tutoring variable. While Propensity Score Matching (PSM) helps address observable selection biases, it cannot fully eliminate bias arising from unobserved heterogeneity or reverse causality. In this study, we did not employ an instrumental variable (IV) approach, as the TIMSS dataset does not provide suitable instruments for the tutoring variable. Consequently, the evidence presented should be interpreted as indicative of associations rather than definitive causal effects. To more accurately assess the effectiveness of tutoring programs, future research should consider more rigorous identification strategies to isolate the true causal impact of tutoring on learning outcomes. Moreover, the use of longitudinal data or experimental/quasi-experimental designs is essential to better establish causal relationships. In addition, qualitative exploration or in-depth case studies could offer further insight

into parental decision-making processes and the institutional factors shaping tutoring practices.

In light of the findings, several recommendations can be offered to parents, educators, and policymakers. For parents, it is important to approach private tutoring strategically, recognizing that its effectiveness is not guaranteed for all students. Instead of relying on tutoring as a default solution, parents should carefully assess their children's specific needs and ensure that any tutoring they pursue is high-quality, aligned with school curricula, and tailored to individual learning gaps, rather than focusing solely on exam preparation.

Educators should play a proactive role in identifying students who genuinely require additional support and guide families toward appropriate interventions. Schools could also establish better communication and coordination with tutoring providers to ensure consistency in instructional approaches and methods.

Policymakers should consider more targeted educational policies. Instead of promoting generalized tutoring programs, efforts should focus on supporting low-achieving students through structured, well-monitored, and publicly subsidized initiatives. In addition, early intervention policies, such as diagnostic assessments and in-school support systems, should be strengthened to reduce the need for private tutoring altogether. Regulation of the tutoring sector is also essential to ensure equitable access and consistent quality, especially in contexts where private tutoring is widespread and driven by market forces. Finally, there is a need for enhanced data systems and rigorous research, employing experimental and quasi-experimental methods, to better comprehend the conditions under which tutoring is truly effective. These combined efforts can help ensure that academic support strategies are both equitable and evidence-based.

Notes

1. Saudi Arabia and Kuwait were excluded from our sample due to the problem of missing data.

2. There is no problem of multicollinearity since the variance inflation factor (VIF) of the model =2,52<10.

References

Albuhi, F. S., & Alsadah, H. B. (1994). Private tutoring in the educational stages in Bahrain: Intensity, causes and ways to stop it. *The Educational Journal*. *8*(32), 23–85.

Alemadi, D., Sellami, A., & Al-Emadi, A. (2016). Qatar education study 2015: Students' motivation and parental participation report. *Social and Economic Survey Research Institute (SESRI),* Qatar University.

Al-Mari, H., & Al-Khamees, N. (2013). Private tutoring: reasons, solutions and its impact on the health of secondary school students in the State of Kuwait. *Journal of Reading and Knowledge*, *134*, 21–56.

Al-Salhi, M. H., Malak, B. M., & Al-Kandari, L. H. (2009). Private tutoring at the secondary school level in Kuwait: Its nature, causes and solutions. *Paper for the Ninth Scientific Conference: The Challenges of Education in the Arab World* (10–11 November), Faculty of Education, Minia University, pp. 1–38.

Ammermüller, A., Heijke, H., & Wößmann, L. (2005). Schooling quality in eastern Europe: educational production during transition. *Economics of Education Review*, *24*(5), 579–599.

Banerjee, A., Cole,S., Duflo, E. & Leigh Linden. (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *Quarterly Journal of Economics*, *122*(3), 1235–1264. https://doi.org/10.3386/w11904

Beaton, A., Mullis, I., Martin, M., Gonzalez, E., Kelly, D. & Smith, T. (1996). Mathematics Achievement in the Middle School Years. *IEA's third international mathematics and science study*, Boston College, Center for the Study of Testing, Evaluation and Educational Policy.

Bedi, A., & Garg, A. (2000). The Effectiveness of Private versus Public Schools: The Case of Indonesia. *Journal of Development Economics*, *61*(2), 463–94. https://doi.org/10.1016/S03043878(00)000651

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta analysis*.New York: Russell Sage Foundation. P. 279–293.

Bouhlila, S. D. (2017). Parents' education and literacy skills: Evidence on inequality of socioeconomic status in Arab countries. *World Development Perspectives, 5*, 34–43. https://doi.org/10.1016/j.wdp.2017.02.006

Bouhlila, D.S., & Hentati, I. (2023). The Effect of the Language of Instruction on Primary Students' Performance. *Journal of Education in Muslim Societies*, *4*(2), 53–80. https://doi.org/10.2979/jems.4.2.04

Bray, M. (2003). *Adverse Effects of Private Supplementary Tutoring: dimensions, implications and government responses*, Comparative Education Research Centre, The University of Hong Kong.

Bray, M. (2007). *The Shadow Education System: private tutoring and its implications for planners.* Paris, UNESCO, International Institute for Educational Planning.

Bray, M. (2009). *Confronting the shadow education system: What government policies for what private tutoring*? Paris: International Institute for Educational Planning, UNESCO.

Bray, M., & Hajar, A. (2022). Complexity of the Contexts: Features of Private Tutoring and Units for Comparison in the GCC Countries of the Middle East. *ECNU Review of Education*, *7*(1), 42–65. https://doi.org/10.1177/20965311221113350

Bray, M., & Ventura, A. (2022). Multiple systems, multiple shadows: Diversity of supplementary tutoring received by private-school students in Dubai. *International Journal of Educational Development, 92*, 102624. https://doi.org/10.1016/j.ijedudev.2022.102624

Brehm, W. C., & Silova, I. (2014). Hidden privatization of public education in Cambodia: Equity implications of private tutoring. *Journal for Educational Research Online*, *6*(1), 94–116. https://doi.org/10.25656/01:8842

Chapman, D.W. & Miric, S.L. (2009). Education Quality in the Middle East. *International Review of Education, 55*, 311–344. https://doi.org/10.1007/s11159-009-9132-5

Chiu, M. M., & Khoo, L. (2005). Effects of resources, inequality, and privilege bias on achievement: Country, school and student level analyses. American Education. *Research Journal*, *42*, 575603. https://doi.org/10.3102/00028312042004575

Dang, H. (2007). The determinants and impact of private tutoring classes in Vietnam. *Economics of Education Review, 26*(6), 684–699. https://doi.org/10.1016/j.econedurev.2007.01.003

Dang, H. & Rogers. F. (2008). The growing phenomenon of private tutoring: Does it deepen human capital, widen inequalities, or waste resources? The World Bank Research Observer, 23(2), P. 161–200. https://doi.org/10.1093/wbro/lkn009

Fang, C., Hu, Y., & Zhang, P. (2018). Can pu- pils of primary schools benefit from the study time after class. *Journal of Educational Science of Hunan Normal University*, *17*(1):69–77.

Frisco, ML., Muller, C. & Frank, K. (2007). Family structure change and adolescents' school performance: A propensity score approach. *Journal of Marriage and Family, 69*,721–741.

Guill, K., & Bos, W. (2014). Effectiveness of private tutoring in mathematics with regard to subjective and objective indicators of academic achievement. Evidence from a German secondary school sample. *Journal for Educational Research Online*, *6*(1), 34–67. https://doi .org/10.25656/01:8840

Ha, T. T., & Harpham, T. (2005). Primary education in Vietnam: Extra classes and outcomes. *International Education Journal, 6*(5), 626–634.

Hattie, J. (2009). *Visible Learning - A synthesis of over 800 meta-analyses relating to achievement*. Routledge, London, New York, ISDN: 978-0-203-88733-2.

Holla, A., Bendini, M., Dinarte, L. & Trako, I. (2021). Is investment in preprimary education too low? Lessons from (quasi) experimental evidence across countries. *The World Bank Research Observer, 36*(2), 1–31. https://doi.org/10.1093/wbro/lkab004

Ireson, J., & Rushforth, K. (2014). Why do parents employ private tutors for their children? Exploring psychological factors that influence demand in England. *Journal for Educational Research Online*, *6*(1), 12–33.

Kang, C.H. (2007). The Effect of Private Tutoring Expenditures on Academic Performance: Evidence from a Nonparametric Bounding Method. National University of Singapore, Department of Economics *Working Papers, wp0706.*

Martin, M. O., Mullis, I. V. S., Foy, P. & Hooper, M. (2016). *TIMSS achievement methodology. In Methods and procedures in TIMSS 2015*. pp. 12.1–12.9.

Martin, M. O., von Davier, M., & Mullis, I. V. S. (Eds.). (2020). Methods and procedures: TIMSS 2019 technical report. TIMSS & PIRLS International Study Center, Boston College. Récupéré de https://timssandpirls.bc.edu/timss2019/methods/

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, *56*(2), 177–196. https://doi.org/10.1007/BF02294457

Mori, I. & Baker, D. (2010). The origin of universal shadow education: What the supplemental education phenomenon tells us about the postmodern institution of education. *Asia Pacific Education Review, 11,* 36–48. https://doi.org/10.1007/s12564-009-9057-5

Newhouse, D. & Beegle, K. (2005). The effect of school type on academic achievement: Evidence from Indonesia. *Journal of Human Resources, 41*(3), 529–557.

OECD. (2001).*Knowledge and Skills for Life: first results from the OECD programme for international student assessment* (PISA) 2000, Paris, OECD Publications. https://doi.org /10.1787/9789264195905-en

The Sultanate of Oman. (2012). Oman Country report presented at the symposium on private tutoring in the Arab states: Problems and solutions [in Arabic]. League of Arab States.

Rocha, V., & Hamed, S. (2018). Parents' perspectives on paid private tutoring in the United Arab Emirates. *UNESCO Regional Center for Educational Planning*.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons. https://doi.org/10.1002/9780470316696

Rubin, D.B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology, 2*,169–188. https://doi.org/10.1023/A:1020363010465

Sellami, A. (2019). Qatar education study 2018: Private tutoring report. *Social & Economic Survey Research Institute (SESRI)*, Qatar University.

Suryadarma, D., Asep, S., Sumarto, S. & Halsey Rogers, F. (2006). Improving Student Performance in Public Primary Schools in Developing Countries: Evidence from Indonesia. *Education Economics*, *14*(4): 401–429. https://doi.org/10.1080/09645290600854110

APPENDIX

Table A-1. Coverage and Exclusion Rates

Country	Coverage	Overall exclusion	Schools	Students
Bahrain	100%	0.8%	0.4%	0.4%
Oman	100%	2.2%	1.4%	0.8%
Qatar	100%	2.2%	1.2%	1%
UAE	100%	5.6%	2.6%	3%

Source: TIMSS (2019)

Table A-2. Results of Multicollinearity Test

	Mathematics Achievement		Science Achievement	
	VIF	1/VIF	VIF	1/VIF
Age	1.056	0.947	1.058	0.945
Inter	1.092	0.916	1.096	0.912
Computer	1.098	0.911	1.099	0.910
Internet	1.087	0.920	1.088	0.919
PreEduc	1.064	0.939	1.066	0.938
2.Country2	1.192	0.839	1.200	0.833
3.Country2	1.926	0.519	1.957	0.511
2.ParEduc2	2.148	0.466	2.148	0.465
3.ParEduc2	7.471	0.134	7.474	0.134
4.ParEduc2	6.517	0.153	6.518	0.153
5.ParEduc2	12.302	0.081	12.328	0.081
2.ParOcup2	1.251	0.800	1.250	0.800
3.ParOcup2	1.551	0.645	1.551	0.645
4.ParOcup2	3.166	0.316	3.167	0.316
5.ParOcup2	2.132	0.469	2.133	0.469
6.ParOcup2	4.871	0.205	4.873	0.205
2.childrenBooks	1.438	0.695	1.438	0.696

(Continued)

Table A-2. (Continued)

	Mathematics Achievement		Science Achievement	
	VIF	**1/VIF**	**VIF**	**1/VIF**
3.childrenBooks	1.447	0.691	1.447	0.691
4.childrenBooks	1.353	0.739	1.352	0.740
5.childrenBooks	1.25	0.800	1.246	0.802
LangTestB	1.827	0.547	1.857	0.539
qm 1	2.273	0.440	2.186	0.457
qm 2	1.954	0.512	1.827	0.547
qm 3	1.801	0.555	1.673	0.598
TUMQ1	1.269	0.788	1.225	0.816
TUMQ2	1.185	0.844	1.137	0.879
TUMQ3	1.144	0.874	1.093	0.915
TUMQ4	1.109	0.901	1.056	0.947
Mean VIF	**2.428**		**2.412**	

Source: Author

Table A-3. Variables' List, coding, and meaning

List of Variables	Coding	Meaning
Overall Performance Math1-Math5	asmmat 01–05	The 1st to 5th plausible value of Overall Performance in Mathematics
Overall Performance Science1-Science5	asssci 01–05	The 1st to 5th plausible value of Overall Performance in Science
Treat	asbh08aa/b	Dummy variable that takes the value 1 if the student had attended tutoring lessons not provided by the school during the last 12 months in mathematics (or science) and 0 if not.
Age	asdage	Quantitative variable which indicates student's age.
Pre Education	asbh04b	A quantitative variable indicating the number of years in pre-primary education for each student.
Education	asdhedup	Categorical variable which reflects parents' education level as follows: some primary, upper secondary, post-secondary, and university or higher.The category 'some primary' is considered the reference category.

(Continued)

Table A-3. (Continued)

Occupation	asdhoccp	Categorical variable which reflects parents' occupation as follows: never worked for pay, general laborer, skilled worker, clerical, small business owner, professional. The category never worked for pay is considered the reference category.
Inter-relation	asbh01a-q	Quantitative variable that indicates a score for Inter-Relation, which is a sum of 18 dummy variables (asbh01a-q).
Internet	asbg05d	Dummy variable which takes the value 1 for « Yes » and 0 for « No ».
Computer	asbg05a	Dummy variable which takes the value 1 for « Yes » and 0 for « No ».
Children books	asbh11	A quantitative variable indicating the number of children's books.
Country	asbg06a+ asbg06b	Categorical variable which reflects parents' origin as follows: any parent born in country, only one parent born in the country, both parents born in the country. The category 'any parent born in the country' is considered the reference category.
Langage Test	lang_sa	Dummy variable that takes the value of 1 for students receiving instruction in English and 0 for those receiving instruction in classical Arabic

Source: TIMSS (2019)

Table A-4. Distribution of the population by parents' education level

Parents' highest education level	**Freq.**	**Percent**	**Cum.**
University or higher	9746	65.300	65.300
Post-secondary but not university	1952	13.080	78.380
Upper secondary	2493	16.700	95.090
Lower secondary	394	2.640	97.730
Some primary, lower secondary or no sch	339	2.270	100.000
Total	14924	100	

Source: TIMSS (2019)

Table A-5. Distribution of the population by parents' occupation

Parents' highest education level	**Freq.**	**Percent**	**Cum.**
Professional	9,702	65.010	65.010
Small business owner	1,113	7.460	72.460
Clerical	2,393	16.030	88.500
Skilled worker	554	3.710	92.210
General laborer	245	1.640	93.850
Never worked for pay	918	6.150	100.000
Total	14925	100	

Source: TIMSS (2019)