Article

Chinese Journal of Sociology

Chinese Journal of Sociology
2023, Vol. 9(2) 219–249

DOI: 10.1177/2057150X231165145
journals.sagepub.com/home/chs
SAGE

Does cram school participation bring about negative emotions? Causal inference based on Chinese Education Panel Survey (CEPS) data

Shiyuan Li[1] and Aiyu Liu[2]

Abstract

This study focuses on the causal relationship between teenagers' participation in cram schools and their emotional well-being. Our analysis is based on Chinese Education Panel Survey data (CEPS 2013–2015). We construct instrumental variables and introduce lagged variables to mitigate possible endogeneity problems. The major findings are as follows. First, the participation of socioeconomically disadvantaged teenagers in cram schools significantly exacerbates their negative emotions. Teenagers of advantaged family socioeconomic status (SES) show better ability in emotion regulation and in avoiding specific negative emotions. Second, we explore class heterogeneity in emotion regulation. Specifically, we propose a preliminary theoretical framework based on a typology of parenting styles. We conjecture that parents of different social classes may respond differently to their children's negative emotions. And for children who attend cram schools, there might be intergroup disparities in their own emotion-regulation ability by their class origins. Empirically, we carry out a counterfactual analysis to further illustrate the association between class difference in emotion regulation and that in the depressive effect of cram school participation, providing supporting evidence for our

[1]Center for Social Research, Guanghua School of Management, Peking University, China
[2]Department of Sociology, Peking University, China

This article has been translated into English and first appeared in the Chinese (simplified character) language in the *Chinese Journal of Sociology* (Chinese version). For permission to reprint all or part of the article in Chinese (simplified characters), please contact CJS@oa.shu.edu.cn. For all other reuse requests, please email permissions@sagepub.co.uk for permission.

Corresponding author:
Aiyu Liu, Department of Sociology, Room 124, Science Building 5, Peking University, Beijing 100871, China.
Email: layyby@pku.edu.cn

theoretical framework. In sum, our study is helpful to further our understanding of the implications of emotional health inequality which may derive from the "shadow education" system.

Keywords
Social stratification, cram school, negative emotion, causal inference, instrumental variable

Introduction

With the economic reform and marketization of education, private tutoring—generally in the form of "cram schools"—has flourished in Chinese society. In recent decades, the market size and the participation rate of cram school courses for middle school students have grown rapidly (Sun and Tang, 2019; Xue and Ding, 2009).[1] Meanwhile, negative emotions such as depression among China's youth have experienced a substantial increase**.** In a national sample from the Chinese Education Panel Survey (CEPS), more than 30% of junior high school students reported that they had felt depressed sometimes or often in the past week (adjusted for sampling weight). Also, recent medical and psychological studies have found that the prevalence of depression among Chinese youths has reached a high level (Miao, 2020). Depression has been documented as one of the most common triggers leading to teen suicide (Bridge et al., 2006; Waldvogel et al., 2008). Furthermore, the negative emotions of classmates are contagious through class social networks, amplifying the consequences of individuals' negative emotions (Li and Lin, 2019). Therefore, it is of urgent relevance to explore possible mechanisms through which negative emotions are generated among China's youth.

The rising trend of teenagers' negative emotions might be closely related to the rapid expansion of private tutoring in recent years (Lin et al., 2008; McCarty et al., 2008; Yi, 2013). Several studies have found that taking cram school courses can lead to an increase in teenagers' negative emotions such as depression (Chen and Kuan, 2021; Kuan, 2018; Zheng et al., 2020). Such side effects of cram school participation on teenagers' emotional well-being can be understood as the "depressive effect" of private tutoring.

The issue mentioned above is one of the major motivations of the "double-reduction" policy, launched by the Ministry of Education of the People's Republic of China in July 2021.[2] The double-reduction policy stipulates that after-school tutoring institutions targeting academic subjects (i.e. cram schools) will now be under strict governance and regulation, and the establishment of tutoring institutions for students in compulsory education will no longer be approved. Meanwhile, the double-reduction policy emphasizes that the government will focus more on the physical and mental health of students and protect their right to rest. Evidently, the mental health of China's youth has become the subject of remarkable policy attention, and the Chinese government is determined to impose strict constraints on the future development of cram schools.

Promoting the mental health of teenagers is of substantial significance. First, from the perspective of economic growth, the health status of teenagers lays a foundation for the

accumulation of human capital (Azarnert, 2020; Becker, 2007; Hartog and Van den Brink, 2007), which is a building block for sustainable development; second, from the perspective of social stratification, negligence of teenagers' health problems may amplify the consequences of existing educational inequalities due to class differences in parenting styles and access to educational resources. Therefore, safeguarding and improving the mental health of teenagers has practical implications for promoting social equality.

So far, policymakers have noted that the over-expansion of cram schools can result in overloading of academic stress and a series of emotional health problems for China's youth. However, relatively few studies have examined the causal relationship between cram school participation and the worsening of emotional well-being. It still remains to be carefully evaluated whether and to what extent taking cram school courses brings about negative emotions in China's youth.

Considering that subjective perceptions, emotions and behavior patterns could be heterogenous across social groups (Duncan et al., 1988; Xie, 2006; Xie, 2007), the psychosocial consequences of cram school participation may also be differentiated by teenagers' social class origins. For example, even if they all attend similar cram schools, some teenagers may continuously accumulate negative emotions during the tutoring process, while others may have better emotion-regulation ability and or may even be "psychologically immune" to these negative emotions. Because of population heterogeneity (Xie, 2007), evaluating the causal effect of cram school participation requires a close look at whether teenagers from different family backgrounds (i.e. class origins) exhibit different or even opposing patterns in terms of emotion regulation. Practically, with better understanding of the patterns in the heterogeneous consequences of cram school participation, we can provide useful information with which the "double-reduction" policy can achieve more targeted and effective implementation. To date, there are no clear data about the existence of and the patterns of class heterogeneity in the depressive effect of cram school participation; as such, further empirical investigation is required.

This paper is organized as follows. First, we attempt to identify the causal effect of cram school participation on teenagers' emotional well-being using CEPS data. We adopt the instrumental variable approach to examine whether and to what extent taking cram school courses increases teenagers' negative emotions. Second, we explore whether and how the depressive effect varies by teenagers' family background (i.e., social class origin). Third, we develop a theoretical framework to explain the observed pattern of class heterogeneity in the depressive effect of cram school participation.

Two clarifications about the research questions are worth making at this stage. First, the private tutoring of interest in this paper is specifically that of academic tutorials centered around the subjects required for entrance examinations (e.g. Chinese, math, and English) in China's primary and secondary schools (Bray, 2013; Kuan, 2018; Zheng et al., 2020). Our focus does not involve non-academic extracurricular classes in subjects such as arts and sports. In this study, we use the terms "cram school" (*buxi ban*) and private tutoring interchangeably. Second, the negative emotions we study are limited to depressive mood, which can be understood as a depressed and temporarily negative emotional state. It is necessary to distinguish depressive mood from major depressive

disorder (MDD) or clinical depression. The latter is a mental disorder that is characterized by persistent low mood, loss of interest in daily life, and difficulty in self-regulation (De Zwart et al., 2019). When temporary depressive mood persists without resolution, it may further evolve into MDD. In our study, the analysis of emotional health mainly focuses on negative mood instead of mental health disorders.

Literature review and theoretical framework

The central questions in this paper are whether and in whom private tutoring brings about negative emotions. Concerning the first question, there are two competing perspectives. One suggests that taking cram school courses has a so-called placebo effect on teenagers' emotions. A teenager may experience peer pressure and other negative emotions when observing many of their classmates or friends attending cram schools. However, if they also join such classes, then their perceived peer pressure and other derived negative emotions can be alleviated (Sun and Tang, 2019; Zheng et al., 2020). In contrast, the other point of view argues that cram school participation will increase teenagers' negative emotions and can cause them to exhibit a depressive effect (Bray, 1999, 2013; Kuan, 2018; Zheng et al., 2020). In terms of the second question of who is more prone to respond to tutoring with negative emotions, few studies have examined whether there is population heterogeneity in the effect of cram school participation, and there are few detailed investigations about the patterns of the effect's heterogeneity across teenagers of different social class origins. The following literature review is motivated by these empirical questions (whether and in whom cram school participation brings about negative emotions).

The depressive effect and placebo effect

As mentioned above, there are two competing perspectives on the impact of cram school participation on emotional well-being. The first view is that it has a depressive effect, meaning that cram school participation may increase teenagers' depressive emotions. A possible mechanism is that private tutoring crowds out teenagers' free time and opportunities for parent–child interactions, leaving them trapped in a highly stressful and rigid time constraint, which can become a threat to emotional well-being. One of the most obvious consequences is sleep deprivation (Kuan, 2018), caused by teenagers sacrificing some of their standard sleeping time in order to complete extra assignments from cram schools. Lack of sleep or reduced quality of sleep are important triggers for the accumulation of negative emotions (Gruber et al., 2012; Gujar et al., 2011; Meijer et al., 2010; Pasch et al., 2010; Yoo et al., 2007).

Another view is that extracurricular tutoring participation has a placebo effect on emotional well-being. That is to say, private tutoring is not necessarily beneficial for academic achievement, but is effective in mitigating peer pressure and its relationship to negative emotions such as depression (Sun and Tang, 2019; Zheng et al., 2020). More specifically, a teenager's decision to participate in a cram school might be driven by social comparisons among peers. When the student observes their peers taking cram school courses,

they tend to join them out of fearing that their peers will outperform them. As a result, the student can gain a certain degree of psychological comfort by attending a cram school (Zheng et al., 2020).

To date, the placebo effect and the depressive effect have both been empirically examined to varying degrees. Among them, Kuan (2018) used data from the Taiwan Education Panel Study in 2001 and 2003 and found that attending a cram school significantly increased 9th-grade teenagers' depression. Looking at the relationship between cram school participation and depression between the 7th and 9th grades, Chen and Kuan (2021) used data from Taiwan Upper Secondary Database (TUSD) during the 2014 academic year and found that cram school participation exacerbated depression for two types of students, those who consistently took private tutoring classes in every semester between Grade 7 and Grade 9 and students who dropped out of private tutoring in Grade 8 or Grade 9, which corroborates the findings of Kuan's (2018) study.

On the other hand, Zheng et al. (2020) used data from CEPS and found that private tutoring significantly improves teenagers' self-confidence, while its depressive effect occurs only if the intensity of private tutoring is excessively high (Zheng et al., 2020). Similarly, Sun and Tang (2019) corroborated the placebo effect. They combined the difference-in-difference method with propensity score matching (PSM-DID), finding that cram school participation significantly mitigated teenagers' peer pressure and reduced other negative emotions such as depression.

Possible explanations for the empirical divergence

In addition to temporal and geographic differences in the empirical data used by previous studies, we conjecture that the different empirical strategies in coping with the endogeneity problem can also explain the divergence in observation between the placebo effect and the depressive effect. To identify the causal effect of cram school participation on teenagers' emotional well-being, the major endogeneity problems can be classified into two categories: one is omitted variable bias (or self-selection bias), and the other is reverse causality (or the reflection problem).

We first discuss omitted variable bias. This problem is caused by the fact that the treatment variable is not randomly assigned. In other words, the decision to attend cram schools is made by the teenagers themselves (or by their parents), rather than by a random assignment. Therefore, there may be unobserved confounding factors that not only influence teenagers' decision to attend cram schools but also affect their emotional well-being, resulting in omitted variable bias. For illustration, the disposition or temperament of a teenager is difficult to observe. For example, a teenager with a pessimistic disposition may be more inclined to believe that private tutoring is of limited use and thus is less willing to participate in it. Meanwhile, this teenager is more likely to feel more negative emotions such as depression. Thus, dispositional tendencies are likely to be a typical omitted variable, which will bias the estimation of the causal effect of cram school participation on emotional well-being.

Another problem is reverse causality. That is to say, the treatment variable (whether to attend a cram school) can be reversely influenced by the outcome variable (teenagers'

negative emotions). Several psychological and educational studies have documented that emotional well-being has an impact on teenagers' learning motivation (Chen and Li, 2000; Crosnoe et al., 2007; Goldston et al. 2007; McCarty et al., 2008; McLeod and Fettes, 2007; Needham et al., 2004). Specifically, teenagers of poorer emotional health status (a higher level of negative emotions) are less motivated to learn and possibly less willing to attend cram schools. Furthermore, the negative feedback of negative affect (outcome variable) on willingness to take private tutoring (treatment variable) may lead to a systematic underestimation of the depressive effect based on ordinary least squares (OLS) estimation. This is because that the accumulation of negative emotions during the tutoring process might discourage teenagers from attending cram schools in the future, thus diluting the magnitude of the depressive effect.

In terms of relevant studies discussing emotional consequences of cram school participation, the main analytical strategies used are propensity score matching (PSM) and the difference-in-difference (DID) approach. Nevertheless, only if all possible confounders are observable can PSM handle the problem of self-selection, and it is difficult to eliminate the bias brought by unobserved confounders and thereby effectively overcome the reverse causality problem. Moreover, since whether to attend cram schools is more an individual decision than an exogenous policy shock, the relevance and applicability of the DID approach is questionable.

To further resolve the endogeneity problem caused by omitted variable bias and reverse causality, we adopt the instrumental variable approach to achieve causal identification. Theoretically, the use of valid instrumental variables is a powerful method for dealing with omitted variable bias and the reverse causality problem (Angrist and Pischke, 2009; Angrist et al., 1996; Morgan and Winship, 2015), allowing us to better capture the causal effect of cram school participation on teenagers' emotional well-being.

Heterogeneous consequences of private tutoring

The discussion of whether private tutoring has a positive or negative effect on emotional well-being must also address the possible existence of population heterogeneity that is obscured by the average causal effect. Population heterogeneity can be understood as individual differences within a given population, which means that the same causes can lead to different outcomes (Xie, 2006: 48). In the context of our study, even if teenagers received the same quantity and quality of private tutoring, their achievement gains, perceptions and emotional experiences may not be identical. In other words, the psychosocial consequences of cram school participation are likely to vary from one person to another. Analysis of population heterogeneity not only helps to further clarify the differences between social groups obscured by the aggregate mean, but also helps to provide more detailed and precisely targeted guidelines for social policies and interventions.

In recent years, heterogeneity analysis of children's developmental outcomes (Kim, 2011) has received increasing attention. For example, Xie et al. (2020) linked data from the US Census with experimental data from the Perry Preschool Program, an example of early-childhood intervention, and found that the program was more effective

in improving the non-cognitive skills of children whose families are more socio-economically disadvantaged. The positive effects of the program were also more durable and profound among children of lower family SES. Another example comes from a recent study of marriage and childrearing in the West. Brand et al. (2019) found that the impact of parental divorce on children's development is not uniformly disruptive, and that the negative effects of parental divorce on children's educational achievement are only significant in the subgroup of children whose family SES is relatively higher.

For studies involve private tutoring, emotional well-being can also be understood as one important developmental outcome. Kuan (2018) analyzes the heterogenous effects of private tutoring on teenagers' emotional well-being. He divides his sample into five groups according to their predicted probability of attending cram schools (i.e. propensity scores), and estimates the effects of cram school participation for each group of teenagers. The depressive effect is found to be statistically significant only among one of the five groups, whose participation probability was comparatively high. Moreover, Chen and Kuan (2021) recently found that the severity of the depressive effect might depend on patterns of participation. However, the heterogeneity analysis of these two studies mainly focused on the tutoring process itself, and did not directly investigate the effect of heterogeneity across teenagers' social class origins.

Theoretical framework

Capturing and examining the pattern of effect heterogeneity by social class has important implications for social inequality and can provide the basis for developing more detailed and precise policy guidelines.

Beginning with class differences in parenting style, we provide plausible explanations for potential class heterogeneity in the psychosocial consequences of cram school participation. The class division in parenting styles has received continuous attention since the mid-twentieth century (Dornbusch et al., 1987; Kohn, 1963; Lareau, 2002, 2011). In the context of Chinese society, Huang (2018) has classified four types of parenting styles by two dimensions, the frequency of parent–child communication and strictness of parents' requirement. These four types are the "authoritarian style" (high frequency of communication and strict requirements), the "authoritative style" (low frequency of communication and strict requirements), the "indulgent style" (high frequency of communication and soft requirements), and the "neglectful style" (low frequency of communication and soft requirements). On this basis, he finds that Chinese parents generally have high academic expectations and requirements for their children, and that the key difference in parenting styles lies in the frequency of parent–child communication. In detail, higher SES parents are more likely to adopt authoritarian parenting styles, while lower SES parents tend to use authoritative parenting styles. More importantly, the authoritarian parenting style is found to be more productive in the development of non-cognitive skills. Children who received authoritarian parenting styles have better self-efficacy, interpersonal skills, and social integration (Huang, 2018).

The typology of family parenting styles and its relevant empirical evidence provide ideas for exploring the pattern of class heterogeneity in the psychosocial consequences of private tutoring. Parents' timely detection of and guidance with regard to their children's negative emotions derived from cram school participation, as well as the child's own emotion-regulation ability, may serve as key factors in explaining the heterogenous effects of private tutoring.

Specifically, higher SES parents are inclined to adopt a parenting style which is more responsive and communicative, enabling them to quickly detect their children's negative emotions and guide their effective regulation or adjustment of their emotional status. Moreover, children's emotion-regulation ability can be viewed as a specific type of non-cognitive skill, and the parenting style adopted by higher SES parents may be more conducive to the development of such ability.

In contrast, lower SES parents may focus more on their children's competence to achieve desirable academic rewards from private tutoring. These parents are less likely to detect and effectively respond to subtle changes in their children's emotional status when the children are exposed to negative experiences during the tutoring process. Meanwhile, the parenting style of lower SES parents is also less helpful for cultivating their children's own emotion-regulation skills. Therefore, negative emotions induced by cram school participation could be more severe and persistent for children of lower SES origins.

The theoretical framework outlined above provides the basis for an empirical prediction that the heterogenous effects of private tutoring on emotional well-being may follow a class-heterogenous pattern, in which the depressive effect is more pronounced in socioeconomically disadvantaged children, whereas children from more advantaged class origins are more likely to be "immune" to such emotional consequences. Moreover, since the long-term accumulation of negative emotions may have a continuing negative impact on teenagers' future academic performance and mental development, private tutoring may contribute to the maintenance and perpetuation of education stratification through the reproduction of health inequalities.

Analytical strategies to detect effect heterogeneity

From a methodological perspective, analysis of effect heterogeneity can be achieved by different approaches. First, it is a common practice to partition the full sample based on the values of propensity scores (e.g. predicted probabilities of attending cram schools) into several subgroups and then compare the group-specific causal effects.

Nevertheless, considering that propensity score is a summary measure of family background (e.g. parental income, education, occupation, etc.), the information on family class origin conveyed by propensity scores is not clear enough. For example, the associations between various indicators of teenagers' family background and their probabilities of attending cram schools do not necessarily follow a uniform and homogeneous pattern. Some family background indicators may be positively associated with participation probabilities, while others may be negatively associated. In addition, the propensity scores used for partition are actually estimates based on sample data rather than their true

values, thus the patterns of effect heterogeneity captured by propensity scores may be subject to sampling error and sample selection bias (Breen et al., 2015). Therefore, if we use estimated propensity scores as a proxy variable for teenagers' family background (or social origins), misleading results can be generated.

Instead, we use several objective indicators of family background (rather than estimated propensity scores) to divide the full sample into subgroups, and examine the pattern of class heterogeneity of depressive effect using subgroup regression. Specifically, the family background indicators are family economic conditions (five categories), highest level of parental education, and father's current occupation (categorical).

Data, measures, and methods

Data

We use data from two waves of CEPS, covering the baseline survey (2013–2014) and follow-up survey (2014–2015) data. The survey adopted probability-proportional-to-size (PPS) sampling, in which 112 schools, 438 classes, and about 20,000 junior high school students were sampled in the baseline survey; the second phase of the survey successfully followed up nearly 10,000 7th-grade students.

The sample in our study consists of the 7th-grade students in the baseline survey (2013–2014), and this cohort received the follow-up survey during 2014–2015. Our analyses are mainly based on the follow-up survey data, while statistically controlling for important variables from the baseline data as lagged variables. Respondents who are not missing on any of the variables used in this study entered in our sample, resulting in an effective sample size of 6535.

Measures

Treatment variable. Our treatment variable is students' participation in private tutoring. It is operationalized as whether or not the student respondent had taken a Mathematical Olympiad, ordinary math, Chinese or English course in a cram school during the preceding year (participated = 1, otherwise = 0).

In our sample, 38.69% (2528) of the full sample attended cram schools during the academic year 2014–2015, and 61.31% (4007) of the full sample did not. By comparing the differences in cram school participation rates of teenagers from different family backgrounds (such as family economic conditions, parental education, and father's current occupation), we found that the participation rates are positively related with teenagers' family SES.[3]

Outcome variable. Our outcome variable is the emotional health status of the student respondent. It is operationalized as a standardized score of negative emotions. We use item response theory (IRT) with a graded response model (GRM) to measure negative emotions (Chen and Kuan, 2021; Kuan, 2018). This approach assumes that the relationship between the true status of respondents' negative emotions and their response probabilities to the items in the emotion scale obeys a logistic function form. Then, using the

maximum likelihood estimation (MLE) method, the observed pattern of responses can be transformed into the IRT score of negative emotions for each respondent.[4] In the following analysis, we refer to the outcome score as the negative emotion score for brevity.

In our sample, teenagers' negative emotion scores during the 2014–2015 academic year range from −1.53 to 2.75, with mean and median values close to 0 and a standard deviation (s.d.) of around 0.9. Generally, a higher negative emotion score indicates more severe negative emotions that a teenager experiences.

We examined mean differences in negative emotion scores among children from different family backgrounds, and found that the severity of negative emotions is negatively associated with the family SES. In other words, higher SES children tend to have lower negative emotion scores.[5]

Control variables. The control variables are divided into following categories: student-level, parent-level, class-level characteristics, survey time, and county-level fixed effects.

1. Student-level variables. The following variables measure students' characteristics. First, demographic characteristics include teenagers' gender, age, ethnicity, number of siblings, type of household registration (*hukou*), migration status, and self-reported physical health status. Second, we use several variables to proxy students' academic ability, including their standardized cognitive test score in the 7th grade, standardized math, Chinese and English scores in the midterm examinations in the 7th grade,[6] and self-reported difficulty in studying math, Chinese and English courses in the 6th grade. Third, teenagers' living arrangement variables include whether living with both parents, whether boarding at school, and the type of residential community. Fourth, parent–child relationship variables include the teenagers' self-reported closeness to their parents, the family member who is responsible for their daily life care, and the family member who is responsible for guiding and supervising their learning at home.
2. Parent-level variables. Several measures are adopted at the family level. First, family economic status is measured by parent-reported financial conditions[7] and the relative rank of family income.[8] Second, parental education is measured by the highest level of education of either parent. Third, family cultural resources are measured by the number of books at home, whether the student owns an individual desk, and whether the student owns a computer and has access to the internet. Fourth, father's current occupation is measured by eleven categories. Fifth, strictness of parents' requirements for their children's daily life and academic behavior are measured using several items.[9] Sixth, parent–child communication is measured by the frequency of parents' discussing daily life and academic problems with their children.[10]

In our analysis of effect heterogeneity of cram school participation, we use family economic conditions, the highest level of education of either parent and father's occupation type as indicators for teenagers' social class origins. According to the classical theory of social stratification, income, education, and occupation are viewed as the core components of socioeconomic status (Blau and Duncan, 1967). And from the methodological

view, compared to continuous propensity scores, variables such as parental education and occupational status are multicategory variables that are naturally associated with the meaning of social groups, which enables easier identification of teenagers' social class origins. Admittedly, however, these criteria have shortcomings, as they are more cumbersome and complicated for modeling and interpretation.

3. Class-level and school-level variables.[11] Several measures are adopted to assess educational characteristics of each respondent. First, we use data on whether respondents are questioned, praised or criticized by their Chinese, mathematics, and English teachers, as well their homeroom teachers. Second, data on respondents' perception of class atmosphere, the friendliness of their classmates and self-reported frequency of taking part in class activities are used. Third, teenagers' self-reported fondness for their school, closeness to other students in the same school and desire to transfer to another school are incorporated.
4. Survey time fixed effect. That is the dummy variable indicating whether the student was surveyed in fall or spring semester. It is helpful to control for unobserved heterogeneity that varies only over survey time and not across individuals.
5. School fixed effects. When using ordinary least squares (OLS) estimation, a set of dummy variables are specified for each school in order to control for unobservable heterogeneity at the school level.
6. County fixed effects. Considering that our instrumental variables are perfectly collinear with school fixed effects,[12] we specify a set of dummy variables for each county when using two-stage least squares (TSLS) estimation.

Empirical strategy

Baseline model. We first estimate a baseline regression model, including all the control variables measured in the current academic year (2014–2015). The regression equation is shown in equation (1)[13]

$$Y_{isct} = \beta_1 \cdot cramsch_{isct} + X'_{isct}\delta + \mu_s + \lambda_t + \varepsilon_{isct} \quad (1)$$

First, the outcome variable Y_{isct} is the negative emotion score of a teenager i, who studies at s school, c county during t semester (i.e. surveyed at fall or spring semester in a given academic year). Second, the treatment variable $cramsch_{isct}$ is a dummy indicating whether the teenager attended a cram school in the preceding year. Third, X_{isct} stands for the set of student-level, parent-level and class-level control variables mentioned before. Fourth, μ_s stands for school fixed effects, and λ_t is the survey time fixed effect. Finally, ε_{isct} is the error term containing all the unobserved factors which could influence the teenager's negative emotions.

In equation (1), β_1 represents the causal effect[14] of cram school participation on teenagers' negative emotions. If the error term ε_{isct} is uncorrelated with the treatment variable $cramsch_{isct}$, then the sample regression coefficients $\hat{\beta}_{1,ols}$ obtained using OLS is a consistent estimator for the true causal effect.[15]

To mitigate the endogeneity problem, we add several measures of prior academic ability and the lagged outcome variable (prior negative emotions) as controls in the baseline model. Specifically, we include the teenagers' midterm exam scores in the previous school year (the 7th grade), the standardized scores of the cognitive ability test in the 7th grade and their self-reported learning difficulty in the 6th grade, which can be considered as reasonable proxies for prior academic ability. In regression equation (2), the column vector $X_{isc,t-1}$ represents the measures of prior academic ability.

In addition, we add the negative emotion scores measured in the previous academic year $Y_{isc,t-1}$ as a control variable. If we assume that only the outcome variable measured at the previous period can affect the treatment variable in the current academic year, then once the lagged outcome variable is controlled for, the outcome variable measured at the current academic year no longer has a reverse effect on the teenagers' decision to participate in private tutoring. Therefore, it is helpful to reduce the bias derived from reverse causality problem (Chen and Fan, 2010; Manski, 1991).

$$Y_{isct} = \beta_1 \cdot cramsch_{isct} + X'_{isct}\delta + \mu_s + \lambda_t + \rho Y_{isc,t-1} + X'_{isc,t-1}\gamma + u_{isct} \tag{2}$$

Instrumental variable approach. We use the instrumental variable approach to further resolve the endogeneity problem. Our instrumental variable IV_{sct} is the cram school participation rate of other teenagers who share a similar residential location to that of the focal teenager i. We refer to these other teenagers as the *location-like group* for the focal teenager i. In detail, the location-like group refers to a group of teenagers who study in the same county as but at different schools to that of the focal teenager i. The idea of constructing this instrumental variable originates from existing empirical studies (Zhang, 2013; Zheng et al., 2020) and we have made further improvements.

The formula of the instrumental variable is shown in equation (3). Note that the denominator is the size of the location-like group, or the total number of other teenagers who live in the same county but study at different schools compared to the focal teenager i. The numerator counts the number of people attending cram schools among the location-like group of the teenager i.

$$IV_{sct} = \overline{Cramsch}_{c-c_s} = \frac{\sum_{c=1}^{N_c}\sum_{i=1}^{N_s} cramsch_{isc} - \sum_{i=1}^{N_s} cramsch_{isc}(s = s_j, c = c_k)}{\sum_{c=1}^{N_c}\sum_{i=1}^{N_s} student_{isc} - \sum_{i=1}^{N_s} student_{isc}} \tag{3}$$

The key prerequisite for causal identification is the validity of the instrumental variable**,** which relies on the instrument relevance and exogeneity assumptions.

We first discuss the instrument relevance assumption. It means that the population (partial) correlation coefficient between the instrumental variable (the location-like group's participation rate in cram schools) and the treatment variable (whether the focal teenager participated in a cram school) is not zero. In our study, the instrumental variable essentially measures the county-level rate of participating in cram schools. To some extent, it can be understood as the result of the supply and demand equilibrium of the regional private tutoring market: on the one hand, this variable reflects the demand for private tutoring among local

teenaged students; on the other hand, it also reveals the scarcity or abundance of the supply of local tutoring institutions. From the demand side, regional demand for private tutoring is likely to transmit through within-county interpersonal networks. From the supply side, the provision of the local private tutoring can also affect the opportunities for and willingness of individuals to participate in private tutoring (e.g. number of cram schools and tuition fees). This suggests that our instrumental variable and treatment variable are highly likely to be correlated. Empirically, the F-statistic from the first-stage regression is much larger than 10, indicating that there is no evidence of the weak instrumental variable problem.

The exogeneity assumption requires that the instrumental variable is uncorrelated with the error term ε_{isct}. That is to say, given the control variables, the participation rate of an individual teenager's location-like group has nothing to do with other unobserved factors that affect the individual's negative emotions. Three arguments are provided to demonstrate the exogeneity of our instrumental variable.

First, the instrumental variable reflects decisions of other teenagers within a county to participate in cram schools, which may be influenced by the economic, social, and cultural circumstances across different regions, leading to the correlation of instrumental variables with unobserved regional heterogeneity. Therefore, we specify county fixed effects, which largely reduces the potential bias introduced by the unobserved regional heterogeneity. Second, the instrumental variable captures macro-level characteristics of the private tutoring market, thus it is unlikely to have a direct effect on individual teenager's emotional well-being. Third, and most importantly, when constructing the instrumental variable, we intentionally exclude those teenagers who study at the same school as that of the focal teenager i. According to recent studies on peer effect, the respondent's classmates or students from the same school can be viewed as their peer group (Du and Guo, 2019; Du and Yuan, 2016), and the social network consisting of the peer group is highly likely to transmit negative emotions (Li and Lin, 2019). In other words, private tutoring behaviors of the peer group may directly affect individual teenagers' own emotional status. For example, observing the classmates attending cram schools may bring about peer pressure in a teenager who has yet to participate in private tutoring, which in turn increases their negative emotions.

For this reason, our instrumental variable intentionally excludes the participating status of the peer group who share the same class or school with the focal teenager i. We argue that the peer group that are closer to the teenager i are more likely to exert a substantial and direct impact on their emotional well-being, while those who study at different schools and are more distant from them are less likely to have a direct impact on their emotional well-being, since this would require an intermediate channel (i.e. the within-school peer group) to exert their influence on the focal teenager i. When constructing the instrumental variable, this channel has been blocked. Meanwhile, the focal teenager i may have close friends who study at different schools. But considering the random sampling design of CEPS, respondents from other schools are unlikely to exert a sizable peer effect on focal teenager i.

When it comes to estimation, we use TSLS to obtain $\hat{\beta}_{1,TSLS}$. The corresponding regression equations are shown in equations (4) and (5). To further secure the

exogeneity of the instrumental variable, a set of lagged variables, such as measures of teenagers' prior academic ability $X_{isc,t-1}$, are added to the above two equations. Theoretically, if the instrumental variable IV_{sct} satisfies the relevance and exogeneity assumptions ($\pi_1 \neq 0$ and $Cov(IV_{sct}, \varepsilon_{isct}) = 0$), then $\hat{\beta}_{1,TSLS}$ can approximate the true causal effect.

$$Cramsch_{isct} = \pi_1 \cdot IV_{sct} + X'_{isct}\alpha + \mu_c + \lambda_t + \kappa Y_{isc,t-1} + v_{isct} \quad (4)$$

$$Y_{isct} = \beta_{1,TSLS} \cdot \widehat{cramsch}_{isct} + X'_{isct}\delta + \mu_c + \lambda_t + \rho Y_{isc,t-1} + \varepsilon_{isct} \quad (5)$$

Results

Descriptive statistics

We divide the full sample into a treated group and a control group, based on whether or not respondents attended cram schools in the preceding year. Table 1 presents the sample means and standard deviations of the outcome variable (negative emotion scores) and selected control variables for each group.[16]

The findings are as follows. First, the treated group (who attended cram schools) have significantly lower negative emotion scores than the control group (who did not attend cram schools). That is to say, teenagers who did not attend cram schools tend to show more negative emotions. However, since we have not considered teenagers' family background, previous emotional status, and academic performance, all of which may influence their current cram school participation decision and be associated with current negative emotions, thus the descriptive results require further clarification.

Second, the sample means of the control variables shows that teenagers who attend cram schools are more likely to come from higher SES backgrounds and to have non-rural *hukou*, fewer siblings, and higher cognitive ability scores and midterm exam scores. In contrast, teenagers who do not take cram school courses are more likely to come from a lower social class and to have rural *hukou*, more siblings, and poorer cognitive ability test scores and midterm exam scores.

Third, the between-group comparison of migration status is worth noting. The percentage of teenagers who have local *hukou* status is slightly lower among the treated group (i.e. teenagers who took cram school courses). And there is no significant difference in the inter-provincial migration rate between the treated and control group. This may result from the selection criteria of our sample, which includes only the respondents who have no missing values on all control variables. In fact, we found that respondents with one or more missing variables (especially their *hukou* type) are more likely to reside in areas with a larger migrant population.[17] Therefore, some respondents who were migrants might have failed to enter our sample due to missing values. To avoid this potential sample selection bias, we additionally control for the teenagers' migration status and type of residential area.[18]

Table 1. Descriptive statistics of selected variables.

	Students who attended cram schools during the past year (N = 2528)		Students who did not attend cram schools during the past year (N = 4007)	
	Mean	S.d.	Mean	S.d.
Student characteristics				
Negative emotion score (Current academic year)	−0.10	0.873	0.06	0.899
Negative emotion score (Last academic year)	−0.19	0.833	−0.03	0.822
Gender (Male = 1)	0.46	0.498	0.51	0.500
Age	14.43	0.634	14.72	0.802
Number of siblings	0.58	0.709	0.92	0.822
Ethnicity (Han = 1)	0.93	0.260	0.84	0.363
Hukou (Rural = 1)	0.49	0.500	0.71	0.456
Migration status				
Local non-migrant = 1	0.88	0.322	0.90	0.297
Cross-province = 1	0.04	0.203	0.04	0.197
Cognitive ability test score (Current academic year)	0.47	0.727	0.14	0.813
Cognitive ability test score (Last academic year)	0.20	0.806	−0.14	0.844
Standardized test results in the 2013 midterm—Chinese	72.65	7.885	71.03	9.344
Standardized test results in the 2013 midterm—English	72.62	8.363	70.78	9.919
Standardized test results in the 2013 midterm—math	72.91	7.960	70.90	9.712
Family characteristics				
Family economic status				
Very poor	0.03	0.172	0.06	0.239
Somewhat poor	0.13	0.333	0.27	0.445
Moderate	0.77	0.423	0.63	0.484
Somewhat rich	0.07	0.252	0.04	0.190
Very rich	0.01	0.084	0.00	0.050
Highest years of parental education	11.71	3.248	9.83	2.663
Years of schooling that parents expect their children to receive	17.71	3.101	16.61	3.595
Family cultural resources (Standardized factor score)	0.22	0.882	−0.53	1.069
Living arrangements (Living with both parents = 1)	0.72	0.449	0.62	0.485

Note: The results of descriptive statistics have been adjusted for sampling weights. The current academic year refers to the 2014–2015 academic year, and the previous academic year refers to the 2013–2014 academic year. S.d. = standard deviation.

Main effect of cram school participation

To further examine the causal effect of the treatment variable (cram school participation) on emotional well-being, we move to regression analysis. Table 2 presents the results of estimated regression coefficients based on different methods and model specifications. Column (1) shows the estimated coefficient obtained from the baseline OLS model. Controlling for individual-level covariates and school and survey time fixed effects, the estimated effect of attending cram schools on teenagers' negative emotion scores is not significantly different from zero. Next, we include lagged variables to mitigate the reverse causality problem and the omitted variable bias. The coefficients in column (2) suggest that the effect size of attending cram schools increases, and the estimate becomes positively significant (p value < 0.05). To some extent, the change of coefficients indicates that the possible feedback effect of the outcome variable may lead to underestimation of the causal effect.

To better deal with the endogeneity problem, we consider the instrumental variable approach. First, we use the current period data (surveyed during the 2014–2015 school year) to construct the participation rate of the respondent's location-like group as our first instrumental variable, and use two-stage least squares (TSLS) estimation to estimate the coefficient of cram school participation. As column (3) in Table 2 shows, the F-statistic of the first-stage regression is much larger than 10 (the empirical criterion), providing strong evidence that our instrumental variable may satisfy the instrument relevance assumption and there is no weak instrumental problem for now.[19] Moreover, the TSLS estimation in column (3) shows that the estimated effect of cram school participation on negative emotion scores is significantly positive (p value < 0.05). Holding control variables fixed, attending cram schools can on

Table 2. Main effect of cram school participation on negative emotions.

	(1) Full sample OLS	(2) Full sample OLS	(3) Full sample TSLS	(4) Full sample TSLS
Attended cram schools? (Yes = 1)	0.039 (0.026)	0.052** (0.023)	0.355** (0.139)	0.357*** (0.138)
Individual-level covariates	YES	YES	YES	YES
Lagged variables	NO	YES	YES	YES
School FE	YES	YES	NO	NO
County FE	NO	NO	YES	YES
Survey time FE	YES	YES	YES	YES
First-stage F-statistic			181.29	102.33
Hansen's J statistic				0.01
N	6535	6535	6535	6535

Note: Cluster-robust standard errors by classes are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1. FE = fixed effects. OLS = ordinary least squares estimation. TSLS = two-stage least squares estimation.

average increase a teenager's negative emotion score by nearly 0.38 s.d., providing supporting evidence for the depressive effect.

Considering our first instrumental variable is constructed based on the data from the current academic year, it may suffer from reverse causality from the outcome variable (i.e. teenagers' negative emotion scores during 2014–2015). Therefore, we use CEPS baseline survey data (2013–2014) to construct the second instrumental variable, which can be viewed as the lagged term of the participation rate of the respondent's location-like group. We then use both instrumental variables to conduct TSLS estimation. The results are shown in column (4) of Table 2, yielding an estimated effect that is close to the that in column (3). In addition, the Hansen's J statistic of the over-identification test is about 0.01 and the corresponding p value exceeds 0.9, suggesting there is no statistical evidence implying the violation of the instrument exogeneity assumption (Chen, 2014).

Note that the effect size of TSLS estimation is up to 7 times larger than that of OLS estimation (with lagged variables included). We provide two possible explanations. First, the OLS estimation may have underestimated the true depressive effect due to omitted variable bias and the reverse causality problem. Second, when the causal effect is heterogeneous and differs from one individual to another, the instrumental variable approach actually identifies the local average treatment effect (LATE) rather than the population average treatment effect (Angrist and Pischke, 2009; Angrist et al., 1996). That is to say, the TSLS estimation for depressive effect reflects the average treatment effect for a specific group of teenagers whose decision to participate in private tutoring can be truly influenced by the behaviors of their location-like group.[20] In other words, these teenagers' participation decisions are more susceptible to external influences, and also their emotional status could be more likely to be influenced by their social environment. Consequently, the negative emotions induced by private tutoring may be relatively stronger among this group of teenagers. It also suggests that our findings may not be directly generalizable to all teenagers. Instead, it is more applicable to those whose participation decisions are more sensitive to those of their location-like group. However, despite the quantitative difference between the TSLS and the OLS estimations, they are qualitatively consistent in terms of the direction and significance of the estimated effects.

Next, we examine whether the depressive effect is moderated by the intensity of private tutoring (Sun and Tang, 2019; Zheng et al., 2020). We transform teenagers' average hours spent taking cram school courses on the weekend into four dummy variables (no cram school courses, 0–2 h, 2–4 h, and more than 4 h). The last three dummy variables (the reference group is teenagers who did not take cram school courses) are treated as three independent variables. We generate their associated instrumental variables using the same construction approach as before, and run a TSLS regression. We then conduct a Wald test in which the null hypothesis is that there is no difference among the coefficients corresponding to these three dummy variables, and the chi-square statistic shows that the three coefficients are not significantly different from one another (p value > 0.1). Using a similar approach, we transform the number of cram school courses that a teenager took into five dummy variables (no courses, one course, two courses, three courses, and four courses). And we treat the last four dummy variables (the reference group is teenagers who did not take cram school

courses) as independent variables and run a TSLS regression, with other settings similar to the former test. The chi-square statistic shows that there is still no difference among the coefficients of these four dummy variables (p value > 0.1)[21].

In general, our findings suggest that taking cram school course can bring about a significant depressive effect on teenagers' emotional well-being. The effect size is roughly 0.38 standard deviations of the negative emotion score and the results remain robust across different lengths of time spent in cram school courses and number of cram school courses taken.

Effect heterogeneity

This section examines the moderating effect of teenagers' family background (i.e. their social class origins) on the depressive effect of private tutoring.

First, we partitioned the full sample into several subgroups based on the teenagers' family SES. For each group we ran a TSLS regression to obtain the group-specific estimate of the depressive effect and then examined the differences in the group-specific effect estimates. Our approach is more flexible than adding interaction terms into the main effect model, as we do not need to assume that the associations between control variables and the outcome variable are homogenous across different subgroups.

We use the highest level of parental education and father's current occupation to classify different subgroups and conduct subgroup regression analysis on this basis.[22] Table 3 shows the results from the subgroup regressions by parental education, with the same model settings to that in the main effect analysis (i.e. the model setting in column (4) of Table 2). By comparing the coefficients in the last three columns of Table 3, the depressive effect of cram school participation is found to be significantly positive only for teenagers whose parental highest level of education is junior high school degree or lower. In contrast, the estimated effects of cram school participation are not statistically significant (p value > 0.1) for teenagers whose parental highest level of education is a technical secondary school degree or senior high school degree, and for teenagers whose parents' highest level of education is a college degree or higher. These two groups of teenagers seem to be "psychologically immune" to the negative emotions derived from private tutoring.

Similarly, Table 4 presents the results of the subgroup regressions by father's current occupation. Comparing the coefficients in the last five columns of Table 4, the depressive effect of cram school participation is found to be significantly positive only for teenagers whose fathers' occupational status is relatively lower (e.g. ordinary workers, peasants, and urban elementary workers such as cleaners, guards, and housekeepers). In contrast, the estimated effects of cram school participation are not significant (p value > 0.1) for teenagers whose fathers' occupational status is comparatively higher (e.g. government officials, cadres and administrators of public institutions and enterprises/corporations, and professionals). These findings again suggest that teenagers of higher SES origins tend to be "psychologically immune" to the negative emotions induced by private tutoring.

Table 3. Results of subgroup regressions by parental education.

	(1) Full sample	(2) Junior high school or lower	(3) Technical secondary school or high school	(4) Junior college degree or higher
Attended cram schools? (Yes = 1)	0.357*** (0.138)	0.561*** (0.160)	0.231 (0.252)	0.313 (0.332)
Individual-level covariates	YES	YES	YES	YES
Lagged variables	YES	YES	YES	YES
County FE	YES	YES	YES	YES
Survey time FE	YES	YES	YES	YES
First-stage F-statistic	102.33	48.35	29.21	17.29
N	6535	3130	1807	1598

Note: Cluster-robust standard errors by classes are in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; column (1) shows results from the full-sample regression; column (2) show results from the subsample regression for teenagers whose parents' highest education level is junior high school or lower; column (3) shows results from the subsample regression for teenagers whose parents' highest education level is technical secondary school or senior high school (including vocational high school and ordinary high school); column (4) shows results from the subsample regression for teenagers whose parents' highest education level is junior college or higher. FE = fixed effects.

Table 4. Results of subgroup regressions by father's current occupation.

	(0) Full sample	(1) Higher SES	(2) Higher SES	(3) Lower SES	(4) Lower SES	(5) Lower SES
Attended cram schools? (Yes = 1)	0.357*** (0.138)	0.081 (0.247)	0.346 (0.225)	0.500*** (0.180)	0.536** (0.214)	0.818** (0.347)
Individual-level covariates	YES	YES	YES	YES	YES	YES
Lagged variables	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES
Survey time FE	YES	YES	YES	YES	YES	YES
First-stage F-statistic	100.43	27.23	33.16	44.02	42.38	18.92
N	6535	1052	1282	3166	1841	1049

Note: 1. Cluster-robust standard errors by classes are in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; column (1) shows results from the subsample regression for teenagers whose fathers are government officials/cadres, administrators of public institutions or enterprises/corporations, or professionals; column (2) shows results from the subsample regression for teenagers whose fathers are government officials/cadres, administrators, professionals, or ordinary clerical workers; column (3) shows results from the subsample regressions for teenagers whose fathers are technical workers (e.g. drivers, electricians, plumbers, or mechanics), ordinary workers (e.g. porters or production line workers), peasants, and urban elementary workers (e.g. cleaners, guards, housekeepers); column (4) shows results from the subsample regression for teenagers whose fathers are ordinary workers, peasants, and urban elementary workers; column (5) shows results from the subsample regression for teenagers whose fathers are ordinary workers or peasants.

Explanation for class heterogeneity

In the theoretical framework proposed above we conjectured that the class differences in parenting styles and the associated differences in teenagers' emotion-regulation ability may help to explain the heterogeneous effect of cram school participation on teenagers' emotional well-being. To provide supporting evidence, we explore the pattern of class heterogeneity in teenagers' emotion-regulation status.

Specifically, we restrict the sample to teenagers who have attended cram school courses in the past year. We find that there is a salient difference in emotion-regulation status across teenagers of different social class origins, and this heterogenous pattern is similar to that of the depressive effect.

First, we consider the frequency of parent–child discussion about children's worries and troubles. One variable is parent-reported frequency of discussing their children's worries and troubles with their children. The other two variables are child-reported frequency of discussing their worries and troubles with their father and with their mother, respectively. These three are ordinal variables ranging from 1 (never) to 3 (often). They can reflect the extent to which parents actively and quickly respond to their children's negative emotions. Second, we consider the extent to which teenagers agreed that they could usually recover from negative feelings. This is an ordinal variable ranging from 1 (strongly disagree) to 4 (strongly agree), which captures teenagers' own emotion-regulation ability. A similar variable is the extent to which they agreed that they could quickly recover from negative emotions with the help of others. This variable may reflect teenagers' emotion-regulation ability and how likely they are to receive emotional support from others.

Additionally, we generate a composite measure by integrating these five variables, the standardized score of emotion-regulation status with a zero mean and a unit variance. Note that a higher score indicates a better status of emotion regulation.

Table 5 presents group-specific means of teenagers' emotion-regulation status among those who attended cram schools in the past year, by parental highest education. It is notable that teenagers' emotion-regulation status exhibits a positive gradient by parental education. That is to say, teenagers whose parents have higher levels of education tend to exhibit better emotion regulation on average. For example, the average emotion-regulation status score of teenagers whose parents' highest level of education is a master's degree or higher is nearly 0.6 standard deviations higher than that of teenagers whose parents' highest level of education is elementary school or lower, and about 0.4 standard deviations higher than that of teenagers whose parents' highest level of education is a junior middle school degree. The last row of Table 5 reports the results of F-tests comparing means of different groups. The robust F-statistics show that there are significant between-group differences in means (p value < 0.01) of parent–child discussion, ability to recover from negative feelings with others' help, and standardized scores of teenagers' emotion-regulation status.

Similarly, by comparing group-specific means of emotion-regulation status by fathers' current occupation, we find a similar pattern of class differences. As Table 6 shows, teenagers whose fathers have a higher occupational status (e.g. government officials/cadres or

Table 5. Emotion-regulation status by parental education.

	(1) Frequency of parent–child discussion about children's worries and troubles	(2) Children can recover from negative feelings by themselves	(3) Children can recover from negative feelings with other's help	(4) Standardized score of emotion-regulation status
Elementary school or lower	2.082	2.994	2.677	−0.061
Junior high school degree	2.182	3.115	2.798	0.109
Technical secondary school degree	2.229	3.010	2.727	0.129
Ordinary high school degree	2.261	3.120	2.765	0.188
Vocational high school degree	2.322	3.113	2.782	0.125
Junior college degree	2.452	3.073	2.860	0.219
Bachelor's degree	2.488	3.232	2.969	0.398
Master's degree or higher	2.530	3.402	3.175	0.500
F-statistic	18.74***	1.63	4.38***	6.16***

Note: F-statistics are obtained from cluster-robust variance-covariance matrix; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The group means have been adjusted by sampling weights. The sample was limited to the teenagers who attended cram schools during the past year ($N = 3421$).

administrators) have better emotion-regulation status than that of teenagers whose fathers belong to lower occupational classes (e.g. ordinary workers, peasants, or urban elementary workers). In detail, the average standardized scores of emotion-regulation status for teenagers whose fathers are government officials/cadres is around 0.45 standard deviations higher than that of teenagers whose fathers are ordinary workers or peasants. Moreover, results of F-tests indicate that, except for the mean difference in the variable that teenagers can recover from negative feelings by themselves being only marginally significant at the 10% level, the mean difference in the other three variables are all significant at the 1% level.

To further validate the pattern of difference in emotion-regulation status by teenagers' social class origins, we conducted two OLS regressions. The parents' highest education and father's occupational category are core independent variables in each of the models, with the standardized score of emotion-regulation status being the outcome variable and demographic characteristics, survey time and school fixed effects controlled. We find that the mean differences in emotion-regulation status across teenagers of different social class origins are still significant (p value < 0.05 for the joint F-test), with a positive gradient by parental education and father's occupational status.

Next, we examine whether the observed pattern of class difference in emotion-regulation status can explain the observed pattern in the heterogenous effect of cram school participation on teenagers' emotional well-being. This can be understood as a special form of causal inference. Here, the explanatory variable is the class difference

Table 6. Emotion-regulation status by fathers' current occupation.

	(1) Frequency of parent–child discussion about children's worries and troubles	(2) Children can recover from negative feelings by themselves	(3) Children can recover from negative feelings with other's help	(4) Standardized score of emotion-regulation status
Government official/cadre	2.501	3.242	2.945	0.520
Cadre/official/administrator of public institutions, enterprises/corporations	2.414	3.151	2.933	0.321
Scientist, engineer, university professor or other professionals	2.453	3.274	2.914	0.528
Doctor, lawyer, high school/ primary school teacher	2.385	3.130	2.910	0.246
Ordinary clerical staff	2.298	3.091	2.888	0.212
Self-employed worker	2.290	3.166	2.856	0.198
Technical worker	2.320	3.200	2.748	0.157
Ordinary worker	2.161	3.077	2.778	0.082
Peasant	2.178	2.986	2.768	0.059
Urban elementary worker	2.039	2.743	2.628	−0.267
Retired, jobless, unemployed or laid-off	2.025	2.955	2.681	−0.294
F-statistic	4.62***	1.65*	2.05**	3.54***

Note: F-statistics are obtained from cluster-robust variance-covariance matrix; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The group means have been adjusted by sampling weights. The sample was limited to the students who attended cram schools during the recent year ($N = 3421$).

in teenagers' emotion-regulation status, and the outcome variable is the class difference in the depressive effect induced by private tutoring. However, the causal identification in this case is not so straightforward as conventional ones. Note that both the explanatory variable (class difference in emotion regulation) and the outcome variable (class difference in the depressive effect) have only one corresponding observation, making it infeasible to estimate the causal effect by regressing the outcome variable on the explanatory variable. Therefore, we adopt another approach called counterfactual analysis to answer this question.

Basically, the counterfactual analysis is conducted to see how the class heterogeneity in the depressive effect will change, or whether the effect heterogeneity by social class origins will be reduced, if we set the emotion-regulation status of lower SES teenagers who attended cram schools to the same level as that of higher SES teenagers who also attended cram schools. In such a counterfactual scenario, the de facto class difference in emotion regulation between the lower and higher SES teenagers attending cram

schools is eliminated. On this basis, we can evaluate whether the de facto difference in the depressive effect between higher and lower SES teenagers will be reduced or even disappear due to the equalization of class difference in emotion-regulation status.

We draw on the counterfactual analysis framework used in representative studies (Liu, 2019; VanderWeele, 2015) and adapt it accordingly to the specificity of our study, which proceeds with the following steps.[23] First, we use the full sample to estimate a TSLS regression model with the specification similar to the main effect model (see column (4) in Table 4). Importantly, we additionally include two variables into this model, which are teenagers' emotion-regulation status and the interaction term between emotion-regulation status and the treatment variable (cram school participation).[24] The formula is listed below, where Y_i is the negative emotion score, D_i is the treatment variable, M_i is the emotion-regulation status, and X_i stands for the control variables. The estimated coefficient of the interaction term β_3 is -0.384 (p value < 0.01); and the estimated coefficient of the emotion-regulation status β_2 is -0.202 (p value < 0.01).

$$E(Y_i|D_i, M_i, X_i) = \beta_1 D_i + \beta_2 M_i + \beta_3 (D_i \cdot M_i) + X_i'\gamma$$

Second, we estimate a TSLS regression model separately for two groups of teenagers, one being those coming from lower SES backgrounds and the other coming from higher SES background, with the same model settings as in the first step. We set the values of the emotion-regulation status (M_i) at the group-specific sample means, denoted as $\bar{m}_{Low}$ and $\bar{m}_{High}$. We then estimate the depressive effect separately for these two groups, which we term "observed depressive effect". We denote the observed depressive effect for the lower SES teenagers as $\delta_{Low,act}$ and that for the higher SES teenagers as $\delta_{High,act}$. Next, we obtain the "actual difference" in the observed depressive effect between the higher and lower SES teenagers, denoted as Δ_{actual}.

$$\delta_{Low,act} = E(Y_i|D_i = 1, SES_i = low, M_i = \bar{m}_{Low}) - E(Y_i|D_i = 0, SES_i = low, M_i = \bar{m}_{Low})$$

$$\delta_{High,act} = E(Y_i|D_i = 1, SES_i = high, M_i = \bar{m}_{High}) - E(Y_i|D_i = 0, SES_i = high, M_i = \bar{m}_{High})$$

$$\Delta_{actual} = \delta_{High,act} - \delta_{Low,act}$$

Third**,** we set the emotion-regulation status of lower SES teenagers who attended cram schools at $\bar{m}_{High}$, the observed sample mean of higher SES teenagers. Then we estimate a TSLS model for the lower SES teenagers, and obtain the "counterfactual depressive effect" for this group of teenagers, denoted as $\delta_{Low,ct}$. By subtracting the counterfactual depressive effect of the lower SES teenagers from the actual depressive effect of the higher SES teenagers, we obtain the "counterfactual difference" in the depressive effect between the higher and lower SES teenagers, denoted as $\Delta_{counter}$.

$$\delta_{Low,ct} = E(Y_i|D_i = 1, SES_i = low, M_i = \bar{m}_{High}) - E(Y_i|D_i = 0, SES_i = low, M_i = \bar{m}_{Low})$$

$$\Delta_{counter} = \delta_{High,act} - \delta_{Low,ct}$$

Fourth, we compute the difference between the *actual* difference and the "counterfactual difference" in the depressive effect, which can reflect the size of reduction in the observed class difference in the depressive effect if the emotion-regulation status of lower SES teenagers (who attended cram schools) is set at a level equivalent to that of higher SES teenagers. In our linear model setting, such difference-in-difference of the depressive effect can be simply calculated by the following equation.

$$\Delta_{counter} - \Delta_{actual} = \delta_{Low,act} - \delta_{Low,ct} = (\beta_2 + \beta_3)(\bar{m}_{Low} - \bar{m}_{High})$$

In other words, if the lower SES teenagers who attended cram schools equal to higher SES teenagers in terms of their emotion-regulation status, holding other things constant, the actual difference in the depressive effect between the higher and lower SES teenagers would be reduced by $(\beta_2 + \beta_3)(\bar{m}_{Low} - \bar{m}_{High})$.[25]

Finally, by dividing the difference-in-difference of the depressive effect by the actual difference of the depressive effect, we can obtain the percentage reduction of the class difference in depressive effect, which is induced by the counterfactual manipulation.

$$\left|\frac{\Delta_{counter} - \Delta_{actual}}{\Delta_{actual}}\right| = \left|\frac{(\beta_2 + \beta_3)(\bar{m}_{Low} - \bar{m}_{High})}{\delta_{High,act} - \delta_{Low,ct}}\right|$$

Empirically, we carry out the counterfactual analysis to examine the difference in the depressive effect for teenagers who have less- or better-educated parents, and for those whose fathers have lower or higher occupational status. The results are shown in Table 7. The coefficients in the first row of Table 7 indicate that the difference in the depressive effect between teenagers of higher and lower SES origins would be reduced by about 42% if the emotion-regulation status of teenagers who attended cram schools and whose highest parental education is a junior high school degree or lower is set at a level equivalent to that of teenagers whose highest parental education is junior college or higher. We further limit the between-group comparison to those whose highest parental education is a junior high school degree and those whose highest parental education is

Table 7. Results of counterfactual analysis.

	(1) Actual difference	(2) Counterfactual difference	(3) Difference in difference = [(2)− (1)]	(4) Percentage reduction= | (3)/ (1) |
Education stratification	−0.270	−0.157	0.113	41.9%
Education stratification	−0.151	−0.025	0.126	83.5%
Occupation stratification	−0.386	−0.287	0.099	25.6%
Occupation stratification	−0.192	−0.081	0.111	57.7%

Note: The counterfactual scenario here refers to that the emotion-regulation status of the lower SES teenagers who attended cram schools became as good as that of the higher SES teenagers.

a bachelor's degree or higher. The coefficients are reported in the second row, suggesting that the class difference in the depression effect would be reduced by about 84% as a result of this counterfactual manipulation.[26] Similarly, the coefficients in the third row indicate that the difference in the depressive effect would be reduced by about 26% if the emotion-regulation status of teenagers who attended cram schools and whose fathers were ordinary workers, peasants, or urban elementary workers was set at a level equivalent to that of teenagers whose fathers were government officials/cadres, administrators, and professionals. Finally, if we limit the comparison to those whose fathers were ordinary workers and those whose fathers were government officials/cadres and administrators, the class difference in the depressive effect would be reduced by about 58% after counterfactual adjustment.

The results of counterfactual analysis suggest that class difference in emotion-regulation status can explain away a substantial portion of the class difference in the depressive effect, which provides supporting evidence for the proposed theoretical framework.

Conclusion and discussion

Along with the increasing popularity of private tutoring in recent decades, the trend of teenagers' deteriorating emotional well-being has become an important topic in social science research. This paper focuses on the causal relationship between cram school participation and negative emotions of junior high school students in the Chinese mainland. Using CEPS survey data and the instrumental variable approach, we identified the depressive effect of cram school participation on teenagers' emotional well-being. Moreover, we examined the pattern of effect heterogeneity by teenagers' social class origins, and provided plausible explanations and quantitative validation for the observed heterogeneity in depressive effect.

The main findings are as follows. First, both the OLS and TSLS estimates show that taking cram school courses significantly increases teenagers' negative emotion scores. The estimated depressive effect based on the instrumental variable approach is about 0.38 standard deviations. Second, we divided the full sample into different subgroups according to the highest level of parental education and father's current occupation, and used subgroup regressions to explore the effect heterogeneity by teenagers' social class origins. We find that the depressive effect is significant only for lower SES teenagers, whereas higher SES teenagers show better ability in emotion regulation, and are more likely to be "psychologically immune" to negative emotions derived from private tutoring. Third, we proposed a theoretical framework based on a typology of parenting styles, and hypothesized that parents of different social classes may respond differently to their children's negative emotions, and the children's own emotion-regulation ability may also vary by their class origins. We specifically examined the class differences in the emotion-regulation status of teenagers who attended cram schools, and carried out a counterfactual analysis to demonstrate the empirical association between class difference in emotion-regulation status and that in the depressive effect.

Our findings help to bridge the empirical divide in existing research. First, in terms of the causal effect of private tutoring on negative affect, we compared the OLS and TSLS

estimates and argue that the previously documented placebo effect is questionable due to the fact that the reverse causality has not been fully considered, which could result in a sizeable underestimation of negative impact of private tutoring (Sun and Tang, 2019).[27] Second, contrary to one recent study suggesting that the depressive effect is not salient unless the intensity of private tutoring is excessively high (Zheng et al., 2020), we find no evidence that the depressive effect is moderated by the time length and number of cram school courses. In other words, our findings reveal that the threshold beyond which private tutoring induces negative emotions is relatively lower than that found in existing studies, a cautionary finding when it comes to developing policies to address the side effects of private tutoring.

Furthermore, we extend our knowledge of social class heterogeneity in the depressive effect, which is helpful to better understand how education and health stratification can be shaped by private tutoring. In detail, the observed pattern of effect heterogeneity implies that inequality in children's developmental outcomes will persist even if the opportunity gap for attending cram schools across teenagers from different class origins is gradually being narrowed. This is because the negative emotions derived from the tutoring process may be specific to those of lower SES origins, and the long-term accumulation of negative emotions will further bring about multiple consequences for teenagers' academic performance and mental development. Consequently, private tutoring may contribute to the reproduction of health inequality, as well as maintaining or even perpetuating the existing patterns of education inequality.

A number of limitations need to be noted. First, we cannot fully demonstrate the causal relationship between class difference in parenting style and the depressive effect of private tutoring. Second, it is hard to effectively capture the difference in the type and quality of private tutoring using available data, such as the division between enrichment and remediation courses, the class sizes of cram schools, and the teaching skills of tutors. Differences in the types of private tutoring in which teenagers participated could be one other explanation for the observed heterogeneity in depressive effect. Third, the measurement of social class origin needs to be improved, since we only considered parental education and father's current occupation. Fourth, teenagers from different family backgrounds may have different understandings of the same item of the emotion scale, and the results of the between-group comparison should be interpreted with caution.

Acknowledgements

Special thanks to the two anonymous reviewers of *Chinese Journal of Sociology* (Chinese version) for their valuable comments. This research was presented at the China Workshop organized by Professor Yu Xie, and special thanks to the teachers and students who provided valuable suggestions for this paper. The authors take sole responsibility for the views expressed in this article.

Contributorship

Shiyuan Li carried out research design and data analysis, Shiyuan Li and Aiyu Liu wrote the manuscript together. Both authors read and approved the final manuscript.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research was supported by National Social Science Fund of China "Precarious work in transforming society and the reshaping of gender ideology" (grant no.:21BSH080).

Notes

1. In China, the market size of the private tutoring industry for primary and secondary school students exceeded 800 billion *yuan* in 2016, with more than 137 million students participating and 7 million to 8.5 million teachers in tutoring institutions, according to the Status Investigation Report of the Private Educational Industry and Teachers, released by The Chinese Society of Education (2016).
2. See the official website of the Ministry of Education of the People's Republic of China at:http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/202107/t20210724_546576.html (accessed 25 September 2021).
3. Due to space limitations, the corresponding descriptive statistics are not presented in the paper.
4. In the CEPS (2014–2015) questionnaire, student respondents were asked how often they have experienced the following depression-related feelings in the past seven days: (1) dejection; (2) too depressed to focus on anything; (3) unhappy; (4) not enjoying life; (5) having no passion to do anything; (6) sad, sorrowful. For each of the items, the response options are in five-point order ranging from never to always. For robustness check, the standardized factor scores were also obtained using principal component analysis. The sample correlation coefficient of the two measures was above 0.95 and the analytical results were generally consistent.
5. Due to space limitations, the corresponding descriptive statistics are not presented.
6. These are mid-term exam scores (in the 7th grade) and have been officially standardized by the CEPS program to be comparable across individuals.
7. This is parent-reported family financial conditions that ranges from 1 (very poor) to 5 (very rich).
8. This is the parent-reported family income rank compared with that of others in their community, which ranges from 1 (very low) to 5 (very high).
9. This is the sum of the frequency of following items: to what extent parents care about: (1) Their children's homework and examinations; (2) behavior at school; (3) whom they make friends with; (4) dress style; (5) time spent on the internet; (6) time spent watching TV (do not care = 0, not strict = 1, very strict = 2).
10. This is measured by summing up the frequency of parent-initiated discussions with their children about: (1) What is happening at school; (2) the child's relationship with friends; (3) the child's relationship with teachers; (4) the child's worries and troubles (never = 1, sometimes = 2, often = 3).
11. We use IRT scores to summarize the atmosphere of class and the characteristics of teachers and classmates.
12. This will be discussed later.
13. For brevity, the intercept is included in the vectors of control variables.
14. In this paper, the terms "causal effect" and "treatment effect" are used interchangeably.

15. The difference between average treatment effect and local average treatment effect will be discussed in the next section.
16. Due to space limitations, some categorical variables are transformed into continuous variables (e.g. years of parental education) in the subsection on descriptive statistics.
17. According to CEPS documentation, the national supplemental sample was obtained from 10 counties (districts) out of 120 counties with large migration populations. Available at: http://ceps.ruc.edu.cn/xmwd/cysj.htm for details (accessed 25 September 2021).
18. We also performed robustness checks. For example, sample source (national core sample, Shanghai sample, and national supplemental sample) was included as an additional control variable, while no longer controlling for *hukou* type. The sample size for the regression analysis rose to 8153. The results remained consistent and the precision of the estimates increased.
19. For the sake of brevity, the values of the first-stage F-statistic will not be interpreted specifically in the following if there is no problem of weak instrumental variables.
20. We follow the convention informed by Stock and Watson (2015:506–508) to interpret the LATE.
21. For brevity, the corresponding regression results are not shown.
22. It should be noted that there is no direct measure of household income in the CEPS data. Respondents only reported their family economic conditions (five or three categories), and the distribution of this variable is heavily skewed, with more than 60% of respondents answering that their household economic conditions were moderate and less than 5% answering that their household conditions were rich or very rich. If we run a subgroup regression for teenagers whose family economic conditions are rich, the sample size would be smaller than 500 and induce the weak instrumental variable problem. Therefore, we did not use family economic conditions to divide the full sample when conducting subgroup regression analysis.
23. For simplicity, we only focus on the comparison of two subgroups of distinct social class origins.
24. Note that the purpose of the interaction term here is to simulate the counterfactual scenario to be used in the next step. This is because the zero coefficient of interaction term of emotion regulation and treatment variable means that the depressive effect will not change regardless of the level of emotion-regulation status. Therefore, the non-zero coefficient of the interaction term is necessary to confirm that class difference in emotion-regulation status can explain class difference in the depressive effect.
25. Here, we assume that the interaction effect of emotion-regulation status and cram school participation does not dependent on the students' class origins, and thus estimating the coefficient of interaction term directly using the full sample. We also considered other methods (e.g. full-sample regression only or subsample regression only) to calculate the above three indicators, and obtained similar results. For brevity, the results of other methods are not shown.
26. The larger decline in this case may also be due to the fact that the actual difference in the depressive effect between the two groups of teenagers was relatively small.
27. Another possible explanation is that the instrumental variable approach identifies the LATE rather than the average treatment effect.

References

Angrist JD, Imbens GW and Rubin DB (1996) Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91(434): 444–455.

Angrist JD and Pischke JS (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Azarnert LV (2020) Health capital provision and human capital accumulation. *Oxford Economic Papers* 72(3): 633–650.
Becker GS (2007) Health as human capital: Synthesis and extensions. *Oxford Economic Papers* 59(3): 379–410.
Blau PM and Duncan OD (1967) *The American Occupational Structure*. New York: The Free Press.
Brand JE, Moore R, Song X, et al. (2019) Parental divorce is not uniformly disruptive to children's educational attainment. *Proceedings of the National Academy of Sciences* 116(15): 7266–7271.
Bray M (1999) *The Shadow Education System: Private Tutoring and Its Implications for Planners*. Paris: UNESCO.
Bray M (2013) Benefits and tensions of shadow education: Comparative perspectives on the roles and impact of private supplementary tutoring in the lives of Hong Kong students. *Journal of International and Comparative Education* 2(1): 18–30.
Breen R, Choi S and Holm A (2015) Heterogeneous causal effects and sample selection bias. *Sociological Science* 2: 351–369.
Bridge JA, Goldstein TR and Brent DA (2006) Adolescent suicide and suicidal behavior. *Journal of Child Psychology and Psychiatry* 47(3–4): 372–394.
Chen IC and Kuan PY (2021) The heterogeneous effects of participation in shadow education on mental health of high school students in Taiwan. *International Journal of Environmental Research and Public Health* 18(3): 1222.
Chen Q (2014) *Advanced Econometrics and Stata Applications*, 2nd edition. Beijing: Higher Education Press.
Chen XY and Li BS (2000) Depressed mood in Chinese children: Development significance for social and school adjustment. *International Journal of Behavioral Development* 24(4): 472–479.
Chen YS and Fan XG (2010) The endogeneity problem in quantitative analysis: A review of estimating causal effects of social interaction. *Chinese Journal of Sociology* (Chinese version) 30(4): 91–117 (in Chinese).
Crosnoe R, Riegle-Crumb C and Muller C (2007) Gender, self-perception, and academic problems in high school. *Social Problems* 54(1): 118–138.
De Zwart PL, Jeronimus BF and De Jonge P (2019) Empirical evidence for definitions of episode, remission, recovery, relapse and recurrence in depression: A systematic review. *Epidemiology and Psychiatric Sciences* 28(5): 544–562.
Dornbusch SM, Ritter PL, Leiderman PH, et al. (1987) The relation of parenting style to adolescent school performance. *Child Development* 58(5): 1244–1257.
Du YH and Guo YZ (2019) Peer effect of junior high school students: A natural experiment based on puberty maturity. *Education and Economy* 149(3): 40–48 (in Chinese).
Du YH and Yuan YZ (2016) The review of the research on peer effect in education: Concept, model, and method. *China Economics of Education Review* 1(3): 77–91 (in Chinese).
Duncan OD, Stenbeck M and Brody CJ (1988) Discovering heterogeneity: Continuous versus discrete latent variables. *American Journal of Sociology* 93(6): 1305–1321.
Goldston DB, Walsh A, Arnold EM, et al. (2007) Reading problems, psychiatric disorders, and functional impairment from mid to late adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry* 46(1): 25–32.
Gruber R, Cassoff J, Frenette S, et al. (2012) Impact of sleep extension and restriction on children's emotional lability and impulsivity. *Pediatrics* 130(5): e1155–e1161.
Gujar N, Yoo SS, Hu P, et al. (2011) Sleep deprivation amplifies reactivity of brain reward networks, biasing the appraisal of positive emotional experiences. *The Journal of Neuroscience* 31(12): 4466–4474.

Hartog J and Van den Brink HM (2007) *Human Capital: Advances in Theory and Evidence*. New York: Cambridge University Press.

Huang C (2018) Parenting styles and the development of non-cognitive skills among Chinese adolescents. *Chinese Journal of Sociology* (Chinese version) 38(6): 216–240 (in Chinese).

Kim HS (2011) Consequences of parental divorce for child development. *American Sociological Review* 76(3): 487–511.

Kohn ML (1963) Social class and parent–child relationships: An interpretation. *American Journal of Sociology* 68(4): 471–480.

Kuan PY (2018) Effects of cram schooling on academic achievement and mental health of junior high students in Taiwan. *Chinese Sociological Review* 50(4): 391–422.

Lareau A (2002) Invisible inequality: Social class and childrearing in black families and white families. *American Sociological Review* 67(5): 747–776.

Lareau A (2011) *Unequal Childhoods: Class, Race, and Family Life*. Berkeley: University of California Press.

Li CH and Lin WD (2019) Are negative emotion contagious? Based on perspective of class social network. *China Economic Quarterly* 18(2): 597–616 (in Chinese).

Liu AR (2019) Can non-cognitive skills compensate for background disadvantage? The moderation of non-cognitive skills on family socioeconomic status and achievement during early childhood and early adolescence. *Social Science Research* 83: 102306.

Lin HC, Tang TC, Yen JY, et al. (2008) Depression and its association with self-esteem, family, peer and school factors in a population of 9586 adolescents in southern Taiwan. *Psychiatry and Clinical Neurosciences* 62(4): 412–420.

Manski CF (1991) Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* 60(3): 531–542.

McCarty CA, Mason WA, Kosterman R, et al. (2008) Adolescent school failure predicts later depression among girls. *Journal of Adolescent Health* 43(2): 180–187.

McLeod JD and Fettes DL (2007) Trajectories of failure: The educational careers of children with mental health problems. *American Journal of Sociology* 113(3): 653–701.

Meijer AM, Reitz E, Deković M, et al. (2010) Longitudinal relations between sleep quality, time in bed and adolescent problem behaviour. *Journal of Child Psychology and Psychiatry* 51(11): 1278–1286.

Miao RK (2020) A meta-analysis of the relationship between negative life events and depressive symptoms of Chinese middle school students. *Journal of Shanghai Educational Research* 394(3): 47–50 (in Chinese).

Morgan SL and Winship C (2015) *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, 2nd edition. New York: Cambridge University Press.

Needham BL, Crosnoe R and Muller C (2004) Academic failure in secondary school: The interrelated role of health problems and educational context. *Social Problems* 51(4): 569–586.

Pasch KE, Laska MN, Lytle LA, et al. (2010) Adolescent sleep, risk behaviors, and depressive symptoms: Are they linked? *American Journal of Health Behavior* 34(2): 237–248.

Stock JH and Watson MW (2015) *Introduction to Econometrics*, 3rd edition. New York: Pearson.

Sun LX and Tang JJ (2019) The effectiveness of private tutoring: An estimation based on China education panel survey. *Peking University Education Review* 17(1): 123–141 (in Chinese).

The Chinese Society of Education (2016) Status investigation report of the private educational industry and teachers. Available at: http://www.gov.cn/shuju/2016-12/27/content_5153561.htm (accessed 25 September 2021).

VanderWeele TJ (2015) *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York: Oxford University Press.

Waldvogel JL, Rueter M and Oberg CN (2008) Adolescent suicide: Risk factors and prevention strategies. *Current Problems in Pediatric and Adolescent Health Care* 38(4): 110–125.

Xie Y (2006) *Sociological Methodology and Quantitative Research*, 1st edition. Beijing: Social Science Academic Press.

Xie Y (2007) Otis Dudley Duncan's legacy: The demographic approach to quantitative reasoning in social science. *Research in Social Stratification and Mobility* 25(2): 141–156.

Xie Y, Near C, Xu H, et al. (2020) Heterogeneous treatment effects on children's cognitive/non-cognitive skills: A reevaluation of an influential early childhood intervention. *Social Science Research* 86: 102389.

Xue HP and Ding XH (2009) A study on additional instruction for students in cities and towns in China. *Educational Research* 348(1): 39–46 (in Chinese).

Yi CC (ed.) (2013) *The Psychological Well-being of East Asian Youth*. Dordrecht: Springer.

Yoo SS, Gujar N, Hu P, et al. (2007) The human emotional brain without sleep—a prefrontal amygdala disconnect. *Current Biology* 17(20): R877–R878.

Zhang Y (2013) Does private tutoring improve students' national college entrance exam performance? — a case study from jinan, China. *Economics of Education Review* 32: 1–28.

Zheng XD, Wang CC, Shen Z, et al. (2020) Associations of private tutoring with Chinese students' academic achievement, emotional well-being, and parent–child relationship. *Children and Youth Services Review* 112: 104934.