BILINGUAL PUBLISHING GROUP

Journal of International Education and Practice

https://journals.bilpubgroup.com/index.php/jiep

ARTICLE

Does Smaller Mean Better? Evidence from the China Education Panel Survey on Class Size and Student Achievement

Jianjiang Luo [1,2] **, *Jiawei Gu*** [3*]

[1] *Faculty of Education, Northeast Normal University, Changchun 130024, China*

[2] *The High School Attached to Northeast Normal University, Changchun 130021, China*

[3] *School of Education, Renmin University of China, Beijing 100872, China*

ABSTRACT

This study investigates the causal effect of class size on student academic achievement in China's compulsory education system, leveraging the institutional feature of randomized student assignment to establish a quasi-experimental framework. Using nationally representative data from the China Education Panel Survey (CEPS) and applying a two-level hierarchical linear model, the research effectively addresses endogeneity arising from non-random class allocation. Results reveal that large classes exert a statistically significant and negative impact on students' academic performance, even after controlling for individual, family, and school-level characteristics. Specifically, moving from medium to large classes reduces achievement by approximately 0.03 standard deviations—a magnitude comparable to the effect of parental education. The analysis further identifies that class-size distribution is primarily determined by institutional and demographic factors, such as urbanization, school ranking, and parental education levels, indicating both structural and demand-side sources of educational inequality. Robustness checks confirm the linearity and stability of these findings. These findings provide credible causal evidence that class size is not merely a contextual variable but a critical determinant of educational outcomes and a key mechanism through which inequality is produced and mediated. The results carry substantial policy implications for resource allocation and equity, particularly for large, rapidly developing economies grappling with intense demographic pressure on their urban education systems.

*CORRESPONDING AUTHOR:

Jiawei Gu, School of Education, Renmin University of China, Beijing 100872, China; Email: 2005gujiawei@163.com

ARTICLE INFO

Received: 10 July 2025 | Revised: 24 August 2025 | Accepted: 1 September 2025 | Published Online: 8 September 2025

DOI: https://doi.org/10.30564/jiep.v8i2.12076

CITATION

Luo, J., Gu, J., 2025. Does Smaller Mean Better? Evidence from the China Education Panel Survey on Class Size and Student Achievement. Journal of International Education and Practice. 8(2): 14–29. DOI: https://doi.org/10.30564/jiep.v8i2.12076

Keywords: Class Size; Academic Performance; Educational Equity; Multilevel Modeling; China Education Panel Survey (CEPS)

1. Introduction

Students, parents, and educational administrators widely believe smaller classes are more conducive to effective learning. In a small-class setting, students tend to receive more individualized attention, are more likely to be remembered by their teachers, and feel more comfortable in the classroom, which encourages them to ask questions and participate in discussions. However, a critical question remains: Does small-class instruction truly enhance academic performance? Much empirical research in education economics suggests that the answer is not always an intuitive "yes."

Class Size Reduction (CSR) has long been regarded as a key measure for improving student achievement, with decades of research and policy experimentation devoted to the topic. Even today, CSR remains among the most debated issues in global education research and reform [1]. Policymakers worldwide have invested heavily in class-size reduction (CSR) to enhance school effectiveness [2], yet its actual impact remains a subject of intense debate [3].

From an economic perspective, the assumption that "smaller is better" also raises practical policy dilemmas. Implementing CSR is often among the most expensive educational interventions, requiring the construction of additional classrooms and hiring extra teachers, thereby placing substantial pressure on public finances. Brewer, Krop, Gill, and Reichardt [4] estimated that the annual operational costs of nationwide CSR initiatives could range from approximately 2 billion to over 11 billion, with expenditures likely to increase further alongside rising teacher salaries and related expenses. Thus, decisions regarding class size involve educational effectiveness and significant economic trade-offs.

This study contributes to the international debate by exploiting a unique institutional feature within China's compulsory education system: the nationwide policy of "random student assignment". In many Chinese middle schools, students admitted under a "neighborhood enrollment" policy are subsequently allocated to specific classes through a formal, often computerized, random lottery. This institutional practice creates a natural experiment. By identifying and isolating a subsample of students subject to this random assignment process from a nationally representative dataset—the China Education Panel Survey (CEPS)—we construct a quasi-experimental design that effectively mitigates self-selection bias.

This study addresses a question of global relevance: What is the net causal effect of class size on student achievement in a context free from parental and student sorting?

Investigating this question within China offers a particularly valuable lens for several reasons. First, China's education system is characterized by exceptionally large average class sizes, providing a context to test the class-size effect in a high-variation setting that is underrepresented in the Western-centric literature. Second, the sheer scale of the system means that even a modest effect size has profound implications for the human capital development of a significant portion of the world's student population. Ultimately, by examining the mechanisms that determine class size, we gain insight into how educational inequality manifests in one of the world's largest and most rapidly urbanizing economies, providing valuable lessons applicable to other developing nations facing similar challenges of resource allocation and demographic pressure.

This paper makes three distinct contributions to the international literature. First, methodologically, it introduces and validates a quasi-experimental approach based on China's random assignment policy, providing a new source of credible causal evidence on the class size effect. Second, empirically, it not only confirms a significant negative causal effect of large classes but also contextualizes its magnitude: the effect size is comparable to that of foundational socioeconomic variables, such as parental education, challenging the recent narrative that class size is a second-order policy lever. Third, theoretically, by analyzing the determinants of class size, we illuminate the interplay between institutional policies, school-level sorting, and educational inequality, providing a richer model for understanding how resource stratification occurs within schools.

2. Literature Review

The relationship between class size and student achievement has long been the subject of intense scholarly debate. The classroom learning environment is a critical factor influencing secondary school students' academic achievement [5,6]. As learning occurs through interactions with teachers and peers, the number of students in the classroom—namely, class size—can fundamentally shape students' gains from learning activities. Although small-class instruction is widely regarded as a reform measure that can improve learning outcomes, empirical findings and theoretical interpretations vary considerably, and no consensus has been reached. These divergences are evident in the direction of the positive, adverse, or nonsignificant effects, as well as in the research methods, data sources, and explanatory mechanisms.

2.1. Positive Effects of Small Classes

A substantial body of research highlights the benefits of small-class instruction in creating higher-quality learning environments. Smaller classes increase teacher–student interaction, facilitate delayed-recall learning, and foster critical thinking [7]. Teachers in small classes are more likely to remember students' names, strengthening students' sense of responsibility and motivation [8]. Lippman [9] even argued that class size and academic performance are significantly linked, with smaller classes clearly more advantageous. Based on a randomized design, evidence from the Tennessee STAR project demonstrated improvements in mathematics and reading outcomes in small elementary school classes [10]. Similar conclusions were drawn from the UK CSPAR project and quasi-experimental studies in Israel [11]. Further international research reinforces these findings. European studies have consistently shown a correlation between smaller class sizes and improved student achievement. For instance, research in France and Denmark identified a positive, albeit modest, relationship between smaller classes and academic performance. At the same time, Swedish experiments demonstrated measurable improvements in mathematics and Swedish language test scores due to reduced class sizes [12]. Danish investigations also indicated a more pronounced impact on primary school students, with effects diminishing in higher grades [13]. Furthermore, a meta-analysis of 77 studies by Glass and Smith concluded that reducing class size has a significant positive impact on student achievement. Such empirical evidence has prompted policymakers worldwide to increase educational funding to reduce class size, with many countries promoting maximum class size legislation and incentive policies [14].

2.2. Challenges to the Small-Class Hypothesis

However, not all studies support the assumption that smaller classes are inherently better. In his systematic reviews of the "education production function," Hanushek found that only 9 out of 112 studies showed significant positive effects, and some even suggested that larger classes could be more beneficial [15]. Research in China also indicates that larger classes positively correlate with student performance in certain regions, a finding that contradicts conventional intuition [16]. Similarly, other recent studies suggest that an increase in class size does not always lead to a reduction in students' academic achievement [17,18]. The Educational Research Service challenged the meta-analysis of Glass and Smith, contending that their conclusions were overly dependent on limited data sources and thus lacked generalizability [19].

2.3. Insignificant and Nonlinear Effects

Some studies suggest that the relationship between class size and achievement is statistically insignificant. Increasingly, scholars have noted that the relationship may be nonlinear. Research by Borland et al. [20], Ding and Lehrer [21], Angrist et al. [22], and Kedagni et al. [23] indicates that huge and tiny classes can harm performance, suggesting that the "optimal" class size may vary by student characteristics and learning environments. Lazear [24] proposed that the ideal class size depends on differences in student attentiveness, underscoring the context-specific nature of class-size effects.

2.4. Mechanisms of Class-Size Effects

Scholars from various perspectives, including peer effects, teacher characteristics, and classroom interactions, have primarily analyzed how class size influences student achievement. A core mechanism is the provision of individualized support. Given the diversity of learners, such support is crucial for effective learning [25], yet it becomes increasingly challenging as class size grows. In a large classroom, teachers can only provide timely and appropriate guidance to a limited number of students, rather than monitoring each individual's progress [26].

This aligns with findings on classroom interaction: Finn et al. found that smaller classes significantly increase student engagement and participation [27]. Moreover, studies show that smaller classes strengthen the positive peer effects for lower-achieving students [28] and amplify the benefits of highly educated teachers [29]. However, other research suggests that teacher qualifications and class size may have independent effects [30]. Seeing class size reduction as a panacea oversimplifies the challenge; merely reducing class sizes without adjusting instructional strategies may not lead to the expected improvements in student learning.

Moreover, the effects of class size exhibit heterogeneity.

A critical line of inquiry concerns the link between class size and achievement gaps. Theoretical work suggests that the benefits of smaller classes should be most pronounced for groups that typically struggle in school [31]. Substantial empirical evidence supports this idea, indicating that smaller classes tend to yield greater advantages for primary school students and disadvantaged groups [32,33]. Conversely, some studies suggest larger classes may be more beneficial at the high school level [34].

An important reason for these divergent and context-dependent findings is methodological. Many studies may suffer from omitted variable bias if they fail to control for class composition, which is often correlated with class size. For example, Dobbelsteen, Levin, and Oosterbeek [35] demonstrate that controlling for the number of peers with similar IQs can significantly alter class size estimates. Similarly, Boozer and Rouse emphasize the need to control for the overall 'ability' of the class [36]. In many Western contexts, student placement into classes is not random due to school choice and tracking systems, leading to systematic differences in student characteristics across class sizes [37].

In this regard, China's compulsory education policy offers a more substantial research advantage: the transition from primary to junior secondary school generally follows a non-selective, neighborhood-based enrollment system, which enhances the credibility of research conclusions compared with the "voting with their feet" school-choice practices prevalent in Europe and the United States.

2.5. Summary and Research Gaps

Existing research reveals considerable disagreement regarding the extent to which class size influences student academic achievement. While some studies suggest that smaller classes enhance learning under certain conditions, others find positive associations between larger classes and achievement or no significant effects. More recent evidence suggests that the relationship is not linear but context-dependent, with threshold effects that vary across educational stages, student groups, and institutional environments. These divergences reflect differences in research design and sample characteristics, as well as a deeper theoretical debate over whether class size represents a universally decisive factor or one whose impact depends on specific contexts.

Despite decades of inquiry, current research faces notable limitations. First, ethical and data access constraints have limited the use of experimental and quasi-experimental designs, making it challenging to address endogeneity. Crucially, even when such designs are used, their generalizability can be questioned. A significant challenge for non-experimental studies is the endogeneity caused by omitted variables, particularly the failure to account for class composition. Second, many studies rely on cross-sectional data, which fail to capture dynamic learning processes; even longitudinal datasets, such as CEPS, have shortcomings, including the absence of subject-specific achievement measures at baseline. Third, as the literature suggests, the complex relationship between class size and student academic achievement remains largely unexplored. Relatively few studies systematically examine the mechanisms linking class size to outcomes, such as the interplay of student behavior, teacher instructional adjustments, and school-level practices.

In sum, divergences in findings can be traced to at least four sources: (1) identification strategies (cross-sectional versus experimental or quasi-experimental designs); (2) measurement issues (choice of achievement indicators and the conflation of class size with student–teacher ratios); (3) sample differences (elementary versus secondary education, or variation across countries and regions); and (4) institutional heterogeneity (resource allocation, admission systems, and parental school choice). Consequently, class-size effects will likely be complex and context-specific rather than universal.

Building on this literature, the present study extends prior work by drawing on nationally representative CEPS data, applying a multilevel modeling framework that accounts for student, family, and school-level influences, and explicitly analyzing class-size allocation mechanisms. The data used in this study enable a more comprehensive description of classroom composition than some previous studies, thereby addressing a key methodological concern. This approach provides more rigorous evidence on the relationship between class size and academic achievement in the Chinese context.

3. Model and Methodology

3.1. Operational Definition

Class size refers to the actual number of students in each classroom. It is essential to distinguish between class size and the student–teacher ratio. The latter is typically calculated by dividing the total number of students by the total number of teachers, but the two measures are not always equivalent in practice. For example, in a large class with additional teaching assistants or extracurricular tutors, the student–to–teacher ratio may appear low despite a large class size. This study treats class size as the core analytical concept rather than the student–teacher ratio.

3.2. Methodology

Regarding research methodology, Ordinary Least Squares (OLS) regression is a widely used tool in educational research, but it faces challenges when applied to hierarchically structured data. Students within the same class often share learning environments and peer effects, while students within the same school experience similar institutional practices and resources. As a result, their academic performances are not independent observations, violating the assumptions of OLS. Although clustered standard errors can somewhat mitigate bias, OLS cannot adequately capture the multi-level dependencies inherent in educational data. If OLS is applied inappropriately, it may underestimate standard errors, overstate statistical significance, and even produce misleading conclusions.

Hierarchical Linear Modeling (HLM) offers clear advantages in this context. It allows the partitioning of variance into within-school and between-school components, thereby explicitly accounting for the nested structure of students within classes and schools. By permitting intercepts to vary across schools, HLM captures unobserved heterogeneity in baseline achievement levels, while still estimating the fixed effects of explanatory variables. This design reduces bias and provides more precise estimates of the independent effect of class size on student outcomes. Moreover, HLM allows for testing cross-level interactions, such as whether the impact of parental education differs between key and non-key schools.

To further mitigate bias from sample self-selection, the analysis restricts the sample to schools that adopt random class-assignment procedures. This ensures that high-ability students are not disproportionately concentrated in smaller classes, thereby improving the representativeness and validity of the results.

3.3. Data Source

This study draws on two waves of the China Education Panel Survey (CEPS), designed and implemented by the China Survey and Data Center at Renmin University of China. The baseline survey was conducted in the 2013–2014 academic year and included two cohorts: seventh and ninth graders. The follow-up survey conducted in 2014–2015 successfully tracked the whole baseline cohort of seventh graders, with 9449 students retained, resulting in a follow-up rate of 91.9%. CEPS provides comprehensive data, including cognitive ability tests, personality assessments, and detailed information on students' personal, family, classroom, and school backgrounds. Notably, the dataset offers precise measures of class size.

3.4. Model Specification

Based on the educational production function, student performance is jointly determined by factors such as students, families, teachers, and schools. The model is as follows:

$$Y_{ij} = f(I_{ij}, F_{ij}, S_{ij}, B_{ij})$$

Consider a linear model:

$$Y = \theta I + \vartheta F + \mu S + \rho B + \varepsilon$$

Y represents the student's academic performance, and F represents their individual characteristics, F represents the student's family background characteristics, S represents the characteristics of the school, B represents the class size, and ε is the random interference term.

The structure of observational data inherently exhibits hierarchical characteristics: individual students are nested within classes, and classes are nested within schools. This means that students within the same school may show a stronger correlation in academic performance than students from different schools. Therefore, individual-level variables (such as personal effort) are influenced by school-level variables (such as teaching resources and school atmosphere). Under this data structure, the observed sample no longer satisfies the basic assumptions of "independent and identically distributed" and "constant variance" required by the classical linear regression model. Forcing the use of Ordinary Least Squares (OLS) would lead to biased parameter estimates and incorrect statistical inferences.

Given that the sample data in this study only includes one or two classes from each school, constructing a complete three-level model (student, class, and school) lacks sufficient variation at the class level, making it inappropriate. Thus, this study adopts a two-level random intercept model (Two-Level Random Intercept Model). Specifically, student-level and family characteristic variables are placed at the first level (student level), and school characteristic variables (including the key explanatory variable "class size") are placed at the second level (school level). The model assumes that the effect of class size on student scores is constant across different schools (i.e., a fixed slope), but allows for random variation of student average scores between schools (i.e., random intercepts).

In educational research, it is reasonable to assume that students within the same school share a standard baseline achievement level due to similar teaching philosophies, institutional resources, and peer environments. A random-intercept specification allows the intercept to vary across schools, thereby capturing unobserved heterogeneity at the school level. This means that each school can have its starting point (baseline achievement), while the slopes of explanatory variables remain fixed across schools. Compared with a simple OLS model, this approach avoids conflating school-level differences with individual-level effects, reduces bias in coefficient estimates, and provides more accurate standard errors. Furthermore, the random-intercept model allows us to partition variance into within-school and between-school components, which is essential for understanding how much of the outcome variation is attributable to differences among schools versus individual students.

The specific form of the hierarchical econometric model is as follows:

Level 1 (Student level):

$$Y_{ij} = u_{01} + u_{1j} I_{ij} + u_{2j} F_{ij} + \epsilon_{ij}$$

Here, Y_{ij} represents the average standardized score across three subjects (Chinese, mathematics, and English) for student i in school j. To address subject-level score disparities, individual subject scores were standardized and then averaged.

- I_{ij}: student-level variables (gender, age, baseline cognitive ability score).
- F_{ij}: family-level variables (parents' highest level of education and occupational status).

Level 2 (School level):

$$u_{01} = \lambda_{00} + \lambda_{01} S_j + \lambda_{02} B_j + \theta_{0j}$$

- S_{ij}: school-level variables, including public/private status, urban location, and designation as a "key school."
- B_{ij}: class-size categories, measured with dummy variables (small class = fewer than 35 students; medium class = 36–45 students; large class = more than 45 students).

This study's definition of class size integrates international experience with national policy standards. Internationally, data from OECD countries' PISA surveys show that class sizes are generally smaller than 35 students. However, in the context of China, policy documents serve as a more critical reference. On one hand, current construction standards set the upper limit for junior high school class size at 50 students; on the other hand, policies such as the Opinions on Building a Quality and Balanced Public Education Service System aim to reduce the average class size for junior high schools to below 40 students by 2027, while eliminating "overlarge classes" of more than 46 students.

Considering the above standards and policy trends, this study divides class size into three categories:

1. **Small classes (≤35 students)**, based on international practices and China's long-term education development goals;
2. **Medium classes (36-45 students)**, covering the range from the current typical class size to the national threshold for "overlarge classes";
3. **Large classes (>45 students)**, directly corresponding to the "overlarge class" phenomenon that needs to be addressed according to policy documents.

This classification aligns with the evolution of China's educational policies and provides clear operational definitions for empirical analysis.

3.5. Data Processing

To ensure the unbiasedness and consistency of the model estimates, this study implemented stringent measures during the data processing phase to address potential endogeneity issues, particularly the self-selection bias caused by non-random groupings.

The core endogeneity challenge stems from the fact that students are not randomly assigned to different classes. Suppose schools assign students to classes based on their entrance scores or other potential abilities. In that case, high-achieving students may systematically end up in specific classes (e.g., honor classes), which may also have better teacher resources (such as higher academic qualifications). This non-random allocation results in a correlation between teachers' educational qualifications and students' abilities (a critical omitted variable), leading to biased estimates of the impact of teacher qualifications on student outcomes.

To overcome this self-selection issue, the study utilized information from the China Education Panel Survey (CEPS) questionnaire to select a sample that approximates a "quasi-random experiment carefully." The selection criteria were as follows:

School Level:

In the school leadership questionnaire, we retained schools that met the following two conditions:

1. The school used "random or even distribution" as the criterion for class assignment.
2. After entering the eighth grade (or ninth grade), the school "did not" reassign students to new classes.

Class Level:

In the class teacher questionnaire, we further filtered the sample, retaining only those classes where the class teacher explicitly answered that the class was "not" assigned based on overall grades or subject-specific performance.

Through this rigorous screening process, we obtained a subsample in which the class assignment closely resembled random allocation. This significantly weakened the systematic relationship between student abilities and class characteristics (such as teacher qualifications), effectively controlling for self-selection bias due to non-random class assignment, thus providing a more reliable data foundation for subsequent causal inference.

Additionally, to ensure the comparability of academic performance across different schools and years, this study adopted the basic idea of the Analysis of Covariance (ANCOVA) model. We standardized all students' seventh and eighth-grade scores within their respective "school-grade" dimensions, adjusting them to a uniform scale with a mean of 70 and a standard deviation of 10 (T-score). This approach not only standardized the measurements but also enabled the model to estimate the net effects between variables more accurately.

3.6. Variables

This study incorporates variables at three analytical levels—student, family, and school —to estimate the effects of class size on student academic achievement. The operationalization of these variables is based on theoretical considerations and the availability of data in the CEPS dataset. The detailed operational definitions and measurement indicators for all variables are presented in **Table 1**.

At the student level, the analysis includes gender, age, and baseline cognitive ability scores. Gender is coded as a binary variable (female = 1, male = 0). Age is the difference between the survey year and birth year, serving as a proxy for potential grade repetition or early/late school entry. Baseline cognitive ability is measured using standardized test scores provided by CEPS, which capture students' foundational cognitive skills before subsequent achievement gains. These variables are necessary to account for individual heterogeneity in academic outcomes.

Table 1. Variable Description.

VARIABLE	DEFINITION/NOTES
STUDENT-LEVEL	
GENDER	Female = 1, Male = 0
AGE	Calculated as survey year minus birth year
BASELINE COGNITION	Standardized cognitive ability score
FAMILY-LEVEL	
PARENTS' EDUCATION	Graduate and above = 3; Bachelor = 2; High school/vocational = 1; Other = 0
PARENTS' OCCUPATION	Occupational status index as proxy for social capital (ranging from 0 to 4)
SCHOOL-LEVEL	
PUBLIC SCHOOL	Yes = 1, No = 0
URBAN SCHOOL	Yes = 1, No = 0
SCHOOL RANKING	"Best/Above average" = 2; "Average" = 1; "Below average/Worst" = 0
CLASS SIZE	Large = 3; Medium = 2; Small = 1

Note: Occupations in CEPS are coded initially into 12 categories. The classification above is a simplified scheme. Class size is categorized as: small (<35), medium (36–45), and large (>45).

At the family level, two indicators are used: parents' highest educational attainment and parents' occupational status. Parental education is measured categorically (graduate and above = 3; bachelor's degree = 2; high school or vocational = 1; other = 0). This measure reflects families' educational capital, which has consistently been shown to be one of the strongest predictors of children's academic performance. Parents' occupational status is coded according to CEPS occupational classifications and transformed into a simplified index ranging from 0 to 4. This serves as a proxy for social and cultural capital, capturing families' social status and access to resources. While household income is often used in similar studies, CEPS provides more reliable measures of parental education and occupation than income data, which are frequently incomplete or unreliable.

At the school level, variables include school type (public vs. private), urban vs. rural location, and school ranking (categorized as "best/above average," "average," and "below average/worst"). These variables capture institutional contexts that may shape student outcomes independently of class size. Class size is categorized into three groups: small (<35 students), medium (36–45 students), and large (>45 students). This categorical classification enables comparisons across different thresholds, while robustness checks also employ class size as a continuous variable.

This multi-level variable design ensures that the analysis adequately controls student heterogeneity, family background, and school-level institutional characteristics. By separating effects at different levels, the study aims to provide a more precise estimate of the independent role of class size.

3.7. Descriptive Statistics

Class-size categories present descriptive statistics to provide an initial overview of the data. **Table 2** summarizes student-level characteristics across small, medium, and large classes, while **Table 3** presents the distributions of family backgrounds.

Table 2 shows that large classes are the most prevalent in the sample (60.5%), particularly in urban and key schools. Interestingly, students in large courses record lower average achievement scores (0.644) than those in small (0.669) or medium classes (0.664), despite having higher baseline cognitive scores (0.403 vs. 0.343 in small classes). This suggests that performance differences cannot be attributed solely to student ability but may reflect the constraints of oversized classrooms on learning.

Table 3 highlights patterns of family background across class sizes. Parents with higher educational attainment and occupational status are overrepresented in small and large classes, while students from "average" families tend to be concentrated in medium-sized classes. This dual distribution suggests two distinct mechanisms: small-class enrollment reflects parental preference and ability to access favorable educational environments. In contrast, large-class enrollment is often driven by supply-side constraints in urban key schools where demand far exceeds capacity.

Table 2. Descriptive Statistics of Class Size Groups.

Indicator	Small Class	Medium Class	Large Class
Mean score	0.6692	0.6635	0.6438
Female ratio	0.4837	0.4854	0.4793
Age	14.49	14.51	14.51
Baseline score	0.3425	0.2202	0.4027
Public school	0.9453	0.9428	0.9366
Urban school	0.4775	0.4251	0.7082
Key school	0.6505	0.6363	0.8981
Size	932	2274	4912

Table 3. Family Background by Class Size.

Variable	Small	Medium	Large
Parents' occupation			
0	0.0343	0.0387	0.0370
1	0.4613	0.4262	0.4458
2	0.2854	0.3459	0.2792
3	0.0531	0.0497	0.0628
4	0.1659	0.1395	0.1785
Parents' education			
0	0.4884	0.5027	0.4592
1	0.3429	0.3956	0.3851
2	0.1433	0.0852	0.1382
3	0.0254	0.0164	0.0175

Together, these descriptive findings illustrate that class size is not randomly assigned. Instead, it is shaped by institutional arrangements (urbanization, school ranking, key school concentration) and family resources (particularly parental education). This non-random distribution underscores the need for multilevel modeling, as simple comparisons of average outcomes could conflate individual, family, and institutional influences with the actual effect of class size.

4. Empirical Results

4.1. Baseline Regression: Two-Level Hierarchical Random-Intercept Model

The baseline regression reveals that large classes have a negative impact on student achievement, even after controlling for a comprehensive set of individual, family, and school-level factors. As shown in **Table 4**, this finding is consistent with descriptive statistics and suggests that class size exerts an independent and adverse effect on academic performance.

At the individual level, girls outperform boys on average, consistent with prior evidence of gender gaps in early adolescence. Older students tend to perform less well, which may reflect difficulties in adapting academically or the impact of grade repetition. Strong positive effects of baseline cognitive ability confirm the cumulative nature of academic skills.

Table 4. Baseline Regression of Class Size and Average Student Performance.

Variable	Coefficient
Class size	−0.0310***
	(−4.11)
Gender	0.0633***
	(22.84)
Age	−0.00687**
	(−2.97)
Baseline cognitive score	0.115***
	(55.39)
Parents' occupational status	0.00147
	(1.00)
Parents' education level	0.0130***
	(5.31)
School ranking	0.0376***
	(4.33)
School location (urban)	0.0319**
	(2.72)
School type (public)	0.0709**
	(2.85)
N	7013

Note: The values in parentheses represent the standard errors, with * indicating $p < 0.1$, ** indicating $p < 0.05$, and *** indicating $p < 0.01$.

At the family level, parental education exerts a significant positive influence, while parental occupational status does not. This suggests that educational attainment captures a family's ability to provide learning support and navigate the school system more directly than occupational prestige.

At the school level, higher rankings, urban locations, and public-school status are positively associated with achievement, highlighting the role of institutional quality and resource allocation.

In terms of effect size, the coefficient of class size (−0.031) implies that moving from a medium-sized class to a large class corresponds to a decrease of about 0.03 standard deviations in achievement. While this effect may appear modest, it is comparable to the impact of parental education (0.013) and school ranking (0.038), and about half the magnitude of the gender gap (0.063). These comparisons show that class size reduction can yield improvements in academic achievement of a similar order to widely recognized determinants, making it a non-negligible factor in educational outcomes.

Moreover, the baseline regression results suggest potential heterogeneity in the effects of class size. Smaller classes may benefit disadvantaged students or those with lower baseline abilities. In contrast, the negative consequences of large classes may be partially offset in key schools where student selection, better teachers, and stronger institutional resources compensate for scale disadvantages. While these nuances are not directly captured in the present model, they point to important directions for future research.

4.2. Multinomial Logistic Regression on Class-Size Choice

The multinomial logistic regression highlights the determinants of class-size allocation. As shown in **Table 5**, the results indicate that school-level factors are the most significant. Urban schools and schools with higher rankings are significantly more likely to operate large classes, sharply reducing the probability that students in these contexts are placed in small or medium-sized classes. This reflects the supply-side pressures of concentrated demand in high-quality schools, particularly in urban areas, where enrollment pressure drives oversized classrooms.

Table 5. Determinants of Class-Size Allocation (Baseline: Large Class).

Variable	Small vs. Large	Medium vs. Large
School type	0.235	0.234
	(1.40)	(1.95)
School ranking	−1.153***	−0.760***
	(−18.85)	(−14.56)
School location (urban)	−0.917***	−1.136***
	(-10.89)	(−18.70)
Parents' occupation	−0.00365	0.0487
	(−0.09)	(1.67)
Parents' education	0.354***	0.134**
	(5.78)	(2.93)

Note: The values in parentheses represent the standard errors, with * indicating $p < 0.1$, ** indicating $p < 0.05$, and *** indicating $p < 0.01$.

At the family level, parental education significantly increases the likelihood that children are assigned to small or medium-sized classes, suggesting that educational attainment equips families with both the awareness and the strategies to secure access to preferred learning environments. By contrast, parental occupation is not statistically significant, indicating that occupational prestige does not necessarily translate into educational advantage in the Chinese context. This distinction highlights the significance of cultural and educational capital in shaping school-choice outcomes, surpassing the influence of purely economic or occupational status.

Taken together, these results point to a dual mechanism of class-size formation. On the one hand, institutional arrangements—particularly in urban and key schools—structurally generate large classes, regardless of individual family demand. On the other hand, families with higher educational attainment are more likely to navigate the system successfully to secure smaller classes for their children. This dual process creates a layered inequality: students in key schools benefit from superior resources despite being in large classes, while students from well-educated families are more likely to gain the advantages of small classes.

4.3. Robustness Checks

Two robustness checks were conducted: (1) replacing class-size dummies with actual class size (continuous values), and (2) including a squared term for class size to test for nonlinearity.

The rationale for these robustness checks is twofold. First, by replacing dummy variables with a continuous measure of class size, we can verify whether the negative association observed in the baseline regression is an artifact of categorization or a consistent linear relationship. Second, by introducing a squared term for class size, we test for the possibility of nonlinear effects, as some international studies have suggested the presence of threshold effects—where both extremely large and extremely small classes might harm learning outcomes. As reported in **Table 6**, both robustness checks yield results that reaffirm the main findings.

Table 6. Robustness Tests.

Model	(1) Baseline	(2) Class Size (Continuous)	(3) Squared Term	(4) Combined Model
Class size dummy	−0.0310***			
	(−4.11)			
Class size (cont.)		−0.0023***		−0.0002
		(−4.74)		(−0.07)
Class size squared			−0.00002***	−0.0002
			(−4.77)	(−0.85)

Note: The values in parentheses represent the standard errors, with * indicating $p < 0.1$, ** indicating $p < 0.05$, and *** indicating $p < 0.01$.

The results confirm the robustness of the baseline findings. When actual class size (continuous) is used in place of dummy variables, the adverse effect remains statistically significant (−0.0023, $p < 0.01$), indicating that the adverse impact is not sensitive to the method of measuring class size. This reinforces the conclusion that larger classes are consistently associated with lower academic performance.

The squared term analysis further suggests that nonlinearity may exist when considered in isolation: the negative and significant coefficient (−0.00002, $p < 0.01$) indicates that the detrimental effect of class size may accelerate as classes grow larger. However, neither remains substantial when the linear and squared terms are included simultaneously (Model 4). This implies that the relationship between class size and achievement is best characterized as approximately linear within the observed range in the Chinese junior middle school context. Unlike some cross-national studies based on PISA data, which report U-shaped or threshold effects, our findings suggest that the main problem lies in oversized classes rather than nonlinear trade-offs across the entire distribution.

Taken together, these robustness checks strengthen confidence in the baseline results. The negative impact of large classes is not an artifact of variable specification, nor is it driven by nonlinear patterns that would call for more nuanced interpretation. Instead, the evidence indicates that reducing oversized classes should be a clear policy priority, especially in urban and key schools where class sizes are most inflated.

5. Discussion

Our analysis, based on a quasi-experimental design using CEPS data, yields two primary findings. First, after mitigating selection bias through a “random assignment” subsample, we identify a statistically significant and substantively meaningful negative causal effect of large class sizes on junior middle school student achievement. Second, this effect appears to be essentially linear, with no evidence of a stable nonlinear relationship. These results, however, enter a crowded and contentious field of international research, demanding a nuanced interpretation. We organize our discussion around three key questions: (1) How do our findings reconcile with the “near-zero effect” conclusion of recent meta-analyses? (2) What unique mechanisms within the Chinese context explain our results? (3) What are the broader theoretical implications for understanding educational inequality?

5.1. Reconciling a Significant Effect with the "Near-Zero" Consensus

A central challenge to our findings comes from recent, comprehensive meta-analyses, most notably Opatrny[38]. Their synthesis of over 1700 estimates, after correcting for publication bias and model uncertainty, concluded that the average effect of class size is "virtually zero" across all identification strategies, with the Tennessee STAR experiment standing as a rare exception. Why, then, does our study detect a significant negative effect? We propose three complementary explanations.

First, and most critically, is the substantive magnitude of the effect in context. While the estimated coefficient may appear modest in absolute terms (a decrease of ~0.03 standard deviations in achievement when transitioning to a large class), its significance becomes striking when benchmarked against other variables. Our models show that the negative impact of being in a large class is comparable in magnitude to the positive impact of having a parent with a college degree versus a high school diploma. This is a powerful finding. While a meta-analysis averages effects to near-zero, our study demonstrates that in a real-world policy setting, class size can be as influential as one of the most widely recognized determinants of student success—parental socioeconomic status. This reframes the debate: class size is not a trivial or secondary variable, but a core structural factor in educational production.

Second, our study's quasi-experimental design, leveraging institutionalized random assignment, offers a high degree of internal validity. It is plausible that the "true" effect is non-zero, but that its signal is diluted by the noise from less robust identification strategies in the literature. Our study, therefore, contributes a high-quality data point from a non-Western context that supports the conclusion that when selection bias is credibly addressed, a significant effect emerges.

Third, context specificity matters. The meta-analysis averages effects across different countries, educational stages, and baseline class sizes. Our study focuses on junior middle schools in China, a system characterized by exceptionally large baseline class sizes. It is highly plausible that the marginal effect of adding a student is non-constant; the negative impact may only become pronounced after a certain threshold is crossed, a threshold that is routinely exceeded in the Chinese context but perhaps not in many Western systems from which a large portion of the literature is drawn.

5.2. Heterogeneity and Compensatory Mechanisms in the Chinese Context

A deeper look into our data reveals a crucial paradox: China's elite "key schools" often have larger classes yet maintain high academic outcomes. This suggests the presence of powerful countervailing mechanisms, which help to explain why the negative effects of large classes are not uniformly devastating across the board. Our findings highlight a critical source of heterogeneity: the negative impact of large class sizes is likely most pronounced for students in non-selective, resource-deprived schools.

Elite schools appear to "compensate" for the disadvantages of large classes through several channels. First, their selective admission policies create a homogenous student body of high-achieving, highly motivated individuals, which mitigates classroom management challenges and facilitates a faster instructional pace. Second, these schools often implement institutional-level strategies such as extending official instructional time and fostering a culture of intensive "shadow education" (after-school tutoring). Third, and perhaps most importantly, they attract and retain higher-quality teachers by offering better salaries, prestige, and opportunities for career advancement. These combined factors create an enriched educational ecosystem that can partially mitigate the negative consequences of overcrowding for students.

This implies a significant and previously underexplored equity dimension. For students in ordinary, non-selective schools who lack access to these compensatory resources, the burden of being in a large class likely falls much harder. Therefore, a uniform policy of merely tolerating large classes, based on an assumption of zero effect, would disproportionately harm the most vulnerable student populations. Our research suggests that future studies should explicitly model this heterogeneity, examining the interaction between class size and school-level resources.

5.3. Theoretical Implications for Educational Stratification

Finally, our analysis of the determinants of class size reveals how it functions as a mechanism of educational stratification. We found that higher parental education is associated with a greater likelihood of securing a spot in a smaller class. This suggests that even within a system of ostensibly "random" assignment, advantaged families find ways to navigate the system to their benefit, possibly through residential choices or informal influence.

Simultaneously, the concentration of large classes in urban and key schools points to a structural form of inequality. These institutions, magnets for population and ambition, become victims of their own success, leading to overcrowding that taxes their resources. The result is a complex, multilayered system of stratification where class size is both a cause and a consequence of inequality. It is a direct disadvantage for those within it, a problem that elite schools must actively compensate for, and an outcome of parental maneuvering and regional resource imbalances. This enriches the classic models of educational stratification by introducing class size as a key mediating variable between family background, school type, and student achievement.

6. Conclusions

6.1. Main Findings

Using data from the China Education Panel Survey (CEPS) and a two-level hierarchical random-intercept model, this study systematically examined the effects of class size on student achievement. The results demonstrate that large classes have a significantly negative impact on student performance, and this effect remains robust even after controlling for individual, family, and school characteristics. Although the estimated effect size is moderate (a reduction of about 0.03 standard deviations when moving from medium to large classes), it is comparable to the impact of parental education, a widely recognized determinant of academic success. This highlights that class size is not a trivial factor but one with meaningful consequences for student achievement.

The distribution of class sizes is primarily shaped by school-level supply mechanisms—urban and key schools are more likely to operate with large classes—while parental education, particularly parents' educational attainment, significantly increases the likelihood that children will be enrolled in small or medium-sized classes. Furthermore, institutional advantages in key schools, such as selective admissions, stronger teacher resources, and extended instructional hours, can partially mitigate the disadvantages of large class sizes. Finally, evidence suggests that the class size effect may vary by educational stage, with lower-grade students likely to benefit more from small class sizes. In contrast, due to greater autonomy and differentiated learning strategies, older students may exhibit lower sensitivity to feedback.

6.2. Policy Implications

Based on the above findings, the focus of policy should shift from broad calls to evidence-based, precise interventions. First, given the substantial negative impact of large class sizes, policymakers should elevate "reducing large class sizes" from a routine task to a core strategic agenda in the allocation of educational resources. This means that when formulating fiscal budgets and development plans, priority must be given to funding for the construction of new schools, as well as the expansion and renovation of existing ones. Additionally, a dynamic teacher allocation mechanism that aligns with changes in school-age populations should be established to address the supply and demand imbalance of "personnel" and "spaces" at the root.

Second, in response to the structural imbalance in class size distribution, resource allocation should move away from a "one-size-fits-all" approach and towards differentiated, targeted "gap-filling." Through measures such as fiscal transfers, priority should be given to urban areas with a shortage of school places and regions with high population inflows. Simultaneously, by promoting measures such as the "unified district system" and "teacher rotation," high-quality educational resources can be more easily shared between schools, thereby alleviating the siphoning effect of key schools and the associated issue of large class sizes.

Furthermore, the grade-level differences in the class size effect at the junior secondary stage provide a clear window for optimizing resource allocation. Under conditions of limited resources, an internal differentiated resource allocation strategy should be implemented, guiding schools to prioritize smaller class sizes for the first year of junior secondary school, providing students with a supportive "soft landing" environment during this key transitional period. Schools can explore "dynamic class size" management, where class sizes are strictly controlled in the lower grades and moderately relaxed in the higher grades. Spare teaching resources can be utilized for tiered instruction and personalized tutoring to optimize resource utilization.

Ultimately, the policy's goal should be to shift from "controlling class size" to "enhancing classroom quality." This requires a "dual-track strategy combining both hard and soft measures": on one hand, continue to reduce physical class sizes through hardware investment, while on the other, increase investment in teachers as the core resource, including providing targeted pedagogical training and optimizing salary and incentive systems. For schools where reducing class sizes is challenging in the short term, resources should be focused on enhancing teachers' classroom management efficiency and teaching innovation capabilities. By establishing evidence-based decision-making and evaluation mechanisms, policymakers can continuously optimize resource allocation, ultimately driving China's education reform towards a profound transformation from external scale improvements to internal quality enhancements.

6.3. Limitations and Future Directions

Despite offering new empirical evidence on the relationship between class size and academic achievement, this study has several limitations. First, the generalizability of our findings is limited because our predominantly urban sample differs from rural schools, where infrastructure factors often influence class size. Consequently, we urge caution in extrapolating our results and emphasize the need for future research in rural contexts to achieve a comprehensive national picture. Second, the data, drawn primarily from public schools during the 2014–2015 academic year, underrepresent private schools and limit the temporal scope of the findings. Third, the analysis did not fully incorporate detailed measures of family economic status or the structure of teacher resources, which may underestimate their roles in class-size allocation and student outcomes. Finally, while the two-level hierarchical model effectively distinguishes between student- and school-level effects, there remains scope for methodological refinement. Future research incorporating approaches such as regional fixed effects, multilevel random slopes, regression discontinuity, or instrumental variables could further strengthen causal identification.

Future research should draw on updated datasets and richer multidimensional variables to more comprehensively explore the heterogeneity of class-size effects and the underlying mechanisms. In particular, further work could investigate stage-specific effects (primary, junior middle, and high school) and assess how institutional advantages in key schools interact with class size to shape outcomes. Such efforts would provide more solid and nuanced evidence to inform the design and evaluation of education policies.

Author Contributions

Conceptualization, J.G.; methodology, J.G.; software, J.L.; validation, J.G., J.L.; formal analysis, J.G.; investigation, J.L.; resources, J.L.; data curation, J.G.; writing—original draft preparation, J.G.; writing—review and editing, J.G.; visualization, J.L.; supervision, J.G.; project administration, J.L.; funding acquisition, J.G. Both authors have read and agreed to the published version of the manuscript.

Funding

This study was funded by the 2024 Hainan Provincial Philosophy and Social Sciences Planning Project: Research on the Impact of ChatGPT on Free Trade Port Education Models and Countermeasures (No.: HNSK(JD)24-14), and by the 2025 Hainan Provincial Higher Education Teaching Reform Research Project: Digital Innovation and Practice of Teaching Paradigms in the Context of Free Trade Port Construction (Hnjg2025-42).

Institutional Review Board Statement

The empirical analysis in this study is based exclusively on the China Education Panel Survey (CEPS), a publicly accessible secondary dataset provided by the National Survey Research Center at Renmin University of China. The dataset is fully anonymized, contains no personally identifiable information, and is released for academic use under strict data-protection procedures. As the study involves no direct contact with participants and uses only de-identified public data, it qualifies for exemption from human-subjects ethics review under commonly accepted institutional guidelines. Accordingly, no additional ethical approval was required for this research.

Informed Consent Statement

Not applicable. This article contains no studies with human participants performed by any authors.

Data Availability Statement

The data used in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no competing interests.

References

[1] Webb, C.E., 2003. The Effects of a Class Size Reduction on Elementary Student Reading Achievement in a Midwestern Urban School District [PhD. thesis]. The University of Iowa: Iowa City, IA, USA. pp. 22–45.

[2] Rice, J.K., 1999. The Impact of Class Size on Instructional Strategies and the Use of Time in High School Mathematics and Science Courses. Educational Evaluation and Policy Analysis. 21(2), 215–229. DOI: https://doi.org/10.3102/01623737021002215

[3] Greenwald, R., Laine, R., Hedges, L., 1994. An Exchange: Part I: Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes. Educational Researcher. 23(3), 5–14.

[4] Brewer, D.J., Krop, C., Gill, B.P., et al., 1999. Estimating the Cost of National Class Size Reductions Under Different Policy Alternatives. Educational Evaluation and Policy Analysis. 21(2), 179–192. DOI: https://doi.org/10.3102/01623737021002179

[5] Howie, S.J., 2005. Contextual Factors at the School and Classroom Level Related to Pupils' Performance in Mathematics in South Africa. Educational Research & Evaluation. 11(2), 123–140. DOI: https://doi.org/10.1080/13803610500110703

[6] Winne, P.H., Nesbit, J.C., 2010. The Psychology of Academic Achievement. Annual Review of Psychology. 61(1), 653–678. DOI: https://doi.org/10.1146/annurev.psych.093008.100348

[7] Raimondo, H.J., Esposito, L., Gershenberg, I., 1990. Introductory Class Size and Student Performance in Intermediate Theory Courses. The Journal of Economic Education. 21(4), 369–382. DOI: https://doi.org/10.2307/1182537

[8] Siegfried, J.J., Kennedy, P.E., 1995. Does Pedagogy Vary with Class Size in Introductory Economics? The American Economic Review. 85(2), 347–351.

[9] Lippman, S.A., 1990. On the Optimality of Equal Class Sizes. Economics Letters. 33(2), 193–196. DOI: https://doi.org/10.1016/0165-1765(90)90168-Z

[10] Hanushek, E.A., 1999. Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects. Educational Evaluation and Policy Analysis. 21(2), 143–163. DOI: https://doi.org/10.3102/01623737021002143

[11] Angrist, J.D., Lavy, V., 1999. Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. The Quarterly Journal of Economics. 114(2), 533–575. DOI: https://doi.org/10.1162/003355399556061

[12] Jepsen, C., 2015. Class Size: Does It Matter for Student Achievement? IZA World of Labor. Article No. 190. DOI: https://doi.org/10.15185/izawol.190

[13] Nandrup, A.B., 2015. Do Class Size Effects Differ Across Grades? Education Economics. 24(1), 83–95. DOI: https://doi.org/10.1080/09645292.2015.1099616

[14] Li, W., Konstantopoulos, S., 2016. Class Size Effects on Fourth-Grade Mathematics Achievement: Evidence from TIMSS 2011. Journal of Research on Educational Effectiveness. 9(4), 503–530. DOI: https://doi.org/10.1080/19345747.2015.1105893

[15] Hanushek, E.A., 1999. The Evidence on Class Size. In: Mayer, S.E., Peterson, P. (eds.). Earning and Learning: How Schools Matter. Brookings Institution: Washington, DC, USA. pp. 131–168.

[16] Li, X., Li, M., 2021. The Effect of Large and Small Class Size in Compulsory Education in China. Rangsit Journal of Educational Studies. 8(1), 46–58.

[17] Shen, T., Konstantopoulos, S., 2021. Estimating Causal Effects of Class Size in Secondary Education: Evidence from TIMSS. Research Papers in Education. 36(5), 507–541. DOI: https://doi.org/10.1080/02671522.2019.1697733

[18] Sohn, K., 2016. A Review of Research on Project STAR and Path Ahead. School Effectiveness and School Improvement. 27(2), 116–134. DOI: https://doi.org/10.1080/09243453.2014.994643

[19] Educational Research Service, 1980. Class Size Research: A Critique of Recent Meta-Analyses. Phi Delta Kappan. 62(4), 239–241.

[20] Borland, M.V., Howsen, R.M., Trawick, M.W., 2005. An Investigation of the Effect of Class Size on Student Academic Achievement. Education Economics. 13(1), 73–83. DOI: https://doi.org/10.1080/0964529042000325216

[21] Ding, W., Lehrer, S.F., 2010. Estimating Treatment Effects from Contaminated Multiperiod Education Experiments: The Dynamic Impacts of Class Size Reductions. The Review of Economics and Statistics. 92(1), 31–42. DOI: https://doi.org/10.1162/rest.2009.11453

[22] Angrist, J.D., Lavy, V., Leder-Luis, J., et al., 2019. Maimonides Rule Redux. American Economic Review: Insights. 1(3), 309–324. DOI: https://doi.org/10.1257/aeri.20180120

[23] Kedagni, D., Krishna, K., Megalokonomou, R., et al., 2021. Does Class Size Matter? How, and at What Cost? European Economic Review. 133, 103664. DOI:

https://doi.org/10.1016/j.euroecorev.2021.103664

[24] Lazear, E.P., 2001. Educational Production. The Quarterly Journal of Economics. 116(3), 777–803. DOI: https://doi.org/10.1162/00335530152466232

[25] Tetzlaff, L., Hartmann, U., Dumont, H., et al., 2022. Assessing Individualized Instruction in the Classroom: Comparing Teacher, Student, and Observer Perspectives. Learning and Instruction. 82, 101655. DOI: https://doi.org/10.1016/j.learninstruc.2022.101655

[26] Laitsch, D., Nguyen, H., Younghusband, C.H., 2021. Class Size and Teacher Work: Research Provided to the BCTF in Their Struggle to Negotiate Teacher Working Conditions. Canadian Journal of Educational Administration and Policy. 196, 83–101. DOI: https://doi.org/10.7202/1078519ar

[27] Finn, J.D., Pannozzo, G.M., Achilles, C.M., 2003. The "Why Is" of Class Size: Student Behavior in Small Classes. Review of Educational Research. 73(3), 321–368. DOI: https://doi.org/10.3102/00346543073003321

[28] Levin, J., 2001. For Whom the Reductions Count: A Quantile Regression Analysis of Class Size and Peer Effects on Scholastic Achievement. Empirical Economics. 26(1), 221–246. DOI: https://doi.org/10.1007/s001810000054

[29] Barrett, N., Toma, E.F., 2013. Reward or Punishment? Class Size and Teacher Quality. Economics of Education Review. 35, 41–52. DOI: https://doi.org/10.1016/j.econedurev.2013.03.001

[30] Blatchford, P., Bassett, P., Goldstein, H., et al., 2003. Are Class Size Differences Related to Pupils' Educational Progress and Classroom Processes? Findings from the Institute of Education Class Size Study of Children Aged 5–7 Years. British Educational Research Journal. 29(5), 709–730. DOI: https://doi.org/10.1080/0141192032000133668

[31] Bosworth, R., Caliendo, F., 2007. Educational Production and Teacher Preferences. Economics of Education Review. 4, 487–500. DOI: https://doi.org/10.1016/j.econedurev.2005.04.004

[32] Krueger, A.B., Whitmore, D.M., 2001. The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project Star. Economic Journal. 111(468), 1–28. DOI: https://doi.org/10.1111/1468-0297.00586

[33] Bonesronning, H., 2003. Class Size Effects on Student Achievement in Norway: Patterns and Explanations. Southern Economic Journal. 69(4), 952–965. DOI: https://doi.org/10.1002/j.2325-8012.2003.tb00542.x

[34] Addonizio, M.F., Phelps, J.L., 2000. Class Size and Student Performance: A Framework for Policy Analysis. Journal of Education Finance. 26(2), 135–156.

[35] Dobbelsteen, S., Levin, J., Oosterbeek, H., 2002. The Causal Effect of Class Size on Scholastic Achievement: Distinguishing the Pure Class Size Effect from the Effect of Changes in Class Composition. Oxford Bulletin of Economics and Statistics. 64(1), 17–38. DOI: https://doi.org/10.1111/1468-0084.00003

[36] Boozer, M., Rouse, C., 2001. Intraschool Variation in Class Size: Patterns and Implications. Journal of Urban Economics. 50(1), 163–189. DOI: https://doi.org/10.1006/juec.2001.2216

[37] Hoxby, C.M., 2000. The Effects of Class Size on Student Achievement: New Evidence from Population Variation. The Quarterly Journal of Economics. 115(4), 1239–1285. DOI: https://doi.org/10.1162/003355300555060

[38] Opatrny, M., Havranek, T., Irsova, Z., et al., 2023. Publication Bias and Model Uncertainty in Measuring the Effect of Class Size on Achievement. IES Working Papers 19/2023. IES FSV, Charles University: Prague, Czech Republic. Available from: https://ies.fsv.cuni.cz/sites/default/files/uploads/files/wp_2023_19_opatrny%2C%20havranek%2C%20irsova%2C%20scasny.pdf