Grading Systems and Student Effort: Evidence from China's Gaokao Reform

By RUI ZHOU*

China's College Entrance Examination (Gaokao) is the main channel for college admissions. The Gaokao consists of main subjects and selective subjects, both of which adopted absolute grading before the Gaokao reform. The Gaokao reform changed the grading system of selective subjects from absolute to relative. This study focuses on its potential to reduce student effort. Specifically, we document evidence of this effect through a theoretical model, questionnaire survey, and online experiment. The results consistently show that this change in grading systems induces less student effort, entirely driven by the effort reduction in selective subjects, with no change in main subjects. This paper is relevant to policy design, suggesting that changing the grading system from absolute to relative could serve as an indirect yet effective tool for alleviating the intense competition among high school students.
Keywords: Grading System, Student Effort, Absolute Grading, Relative Grading, Gaokao

Education is crucial to the formation of human capital of the society (Schultz 1961, 1960; Heckman 2005). A long line of research has focused on screening and matching mechanisms in the admission process, including studies on optimal test design (Berger and Veerkamp 1994; Van der Linden 2015) and the stability of matching systems (Kojima and Pathak 2009; Ha, Kang and Song 2020; Jiang 2019). Another line of research has examined strategies to enhance student effort, primarily through interventions within the existing grading frameworks, such as experimenting with class sizes (Czibor et al. 2020), providing information feedback (Azmat and Iriberri 2010), and finding the optimal allocation of prizes (Moldovanu and Sela 2008).

However, little attention has been devoted to the potential effects of grading systems on student effort, particularly the comparison between absolute and relative grading systems. And among the existing studies, most are theoretical (Becker and Rosen 1992; Dubey and Geanakoplos 2010; Landeras 2009), and some of their predictions are mixed (for example, Becker and Rosen 1992 and Dubey and Geanakoplos 2010). Moreover, many remain untested in empirical contexts.

This paper combines theoretical and empirical approaches to examine how students respond to grading systems in terms of effort exerted within the context of the Gaokao (China's College Entrance Examination) reform[1]. We distinguish between two grading systems: absolute grading and relative grading. An absolute grading system, also referred to as "criterion-referenced," evaluates students based on how their performances compare

* Rui Zhou: Peking University (ruizhou@stu.pku.edu.cn).
[1] Further details on the Gaokao and Gaokao reform are provided in Section I.

against an objective criterion. In contrast, a relative grading system, commonly known as "grading on a curve," classifies students into different grades based on rank order, and then gives converted scores corresponding to each grade. For clarity, we refer to scores provided by absolute grading as *raw scores* and those provided by relative grading as *converted scores*. The fundamental distinction between these two systems lies in the step of conversion in relative grading, which transforms raw scores into converted scores. A notable attribute of relative grading is its *granularity*. We mean granularity by the number of grade intervals into which students are classified. A higher granularity means more grade intervals and smaller score differences between adjacent grades. There are two types of subjects in the Gaokao: main subjects and selective subjects[2]. Main subjects consistently use absolute grading. Selective subjects adopted absolute grading before the Gaokao reform but switched to relative grading afterward.

When comparing the effort incentives under absolute and relative grading systems, intuition itself cannot give a clear answer: Under relative grading, within each grade interval where students receive the same converted score, there is some heterogeneity in ability. Intuitively, high-ability students within a grade may have incentives to shirk while still remaining in the same interval. In contrast, students at the boundaries of a grade have stronger incentives to work harder: those who have just achieved a higher grade aim to secure it, while those near the very upper boundary of a lower grade strive to move up. The granularity of the relative grading system also matters. Higher granularity reduces the score difference between adjacent grades, and decreases the number of students within each grade in the meantime. The former weakens incentives at grade boundaries, while the latter discourages the high-ability students within a grade from shirking. Therefore, we cannot arrive at a definitive conclusion through reasoning alone. Addressing this question requires both theoretical and empirical analysis.

We follow the work of Krishna et al. 2022 to model the student effort decisions by drawing on theories on rank-order tournaments (Lazear and Rosen 1981) and and large contests (Olszewski and Siegel 2016; Bodoh-Creed and Hickman 2018; Olszewski and Siegel 2019). Specifically, we formalize the setting as a tournament that replicates the incentive structure of Gaokao and rewards students based on rank order. The model predicts that (a) effort in subjects with a consistent grading system remains unaffected by the grading system changes in other subjects, and (b) the total effort in subjects that adopt relative grading systems decreases as the granularity increases. Additionally, by calibrating parameters according to real-world relative grading systems implemented in the Gaokao, model simulations suggest that (c) the change from absolute to relative grading reduces overall student effort.

Empirically, we conducted a questionnaire survey of over 1,000 high school students across multiple Gaokao years and provinces to gather directly relevant data. The results align with the theoretical predictions, showing that the Gaokao reform significantly reduces students' study time by approximately 5.93 hours per week. This effect is totally contributed by reduced effort in selective subjects, with no change in study time in main subjects. Notably, unlike the findings from Andreoni and Brownback 2017, we find no evidence of heterogeneity of such effect across different ability groups: both high-ability

[2] Further details on main and selective subjects are provided in Section I.

and low-ability students response similarly to the grading system change in terms of study time. Additionally, the results suggest improvements in students' mental health and more engagement in extracurricular activities following the Gaokao reform.

Building on these findings, we designed and conducted an online experiment to establish causal relationships. The experiment involved two tasks: mental arithmetic, which resembled the "main subject," and typing, which resembled the "selective subject." Subjects were randomly assigned to groups in which typing, the "selective subject," was graded under different grading systems. The comparisons of effort exerted across treatment arms are consistent with the theoretical predictions: relative grading systems induces reduced effort in typing compared to absolute grading. This effect is more pronounced under the relative grading system with higher granularity. Furthermore, the experiment extends the research context to more general comparisons between absolute and relative grading systems, as well as comparisons among relative grading systems with different levels of granularity.

This paper contributes to the literature in four ways. First, it extends the theoretical analysis of absolute and relative grading systems. Prior studies primarily examine contexts where students focus solely on subjects that adopt either an absolute or a relative grading system (Krishna et al. 2022; Paredes 2017; Dubey and Geanakoplos 2010; Landeras 2009), which corresponds to the "selective subjects" in our analysis. Our work enriches this framework by accounting for another type of subjects with a consistent grading system, namely the "main subjects." Thus, we analyze not only the effect of grading system changes on student effort in the directly affected subjects, but also whether the reduced effort in these subjects is redirected towards subjects with a consistent grading system, and how the total effort changes as a result. Moreover, we further Dubey and Geanakoplos 2010's work on why coarse grading schemes (i.e., relative grading with low granularity) may be advantageous by investigating how granularity moderates the effects on student effort both theoretically and empirically.

Second, this paper provides empirical evidence for the relevant theoretical studies. Most existing studies examining the relationship between grading systems and student effort are theoretical (Paredes 2017; Dubey and Geanakoplos 2010; Landeras 2009). Their predictions are mixed (e.g., Becker and Rosen 1992 and Dubey and Geanakoplos 2010) and few have been empirically tested. Our research situates this question within the context of the Gaokao reform, an unprecedented large-scale social experiment, and supports the idea that relative grading system induces less student effort.

Third, within the literature focusing on the Gaokao and Gaokao reform, most studies investigate whether and how the reform influences subject selection (Zhou, Shan and Qin 2024; Liu and Wei 2023; Zhong and Wang 2019), while few explore its potential impact on student effort. To my knowledge, this paper is the first to examine the Gaokao reform's effect on student effort.

Lastly, this paper offers significant implications for policy design by proposing an indirect approach to mitigate the intense competition among high school students. In response to the mounting academic pressure of high school students (Yu, Li and Wang 2016; Fu et al. 2023), the Chinese government has implemented a series of regulatory policies to reduce students' burdens from schools (State Council 2021). However, these measures are actually thinking within the current incentive structure and prove largely ineffective (Yang and Wen

2022; Zhou and Qi 2022; Du 2022). In contrast, changing the grading system is thinking out of the box: it fundamentally alters the incentive structure under which students make their effort decisions. The empirical evidence in this paper highlights the potential of relative grading systems in alleviating the fierce competition.

The rest of the paper proceeds as follows. Section I introduces the context of Gaokao and Gaokao reform. Section II describes the model and provides model predictions. Section III reports the results of a questionnaire survey conducted with high school students. Section IV details the design and findings of the online experiment. Finally, Section V concludes with a discussion of alternative explanations and directions for future research.

I. Gaokao and Gaokao Reform

The Chinese College Entrance Examination, commonly referred to as *Gaokao*, is the primary channel for high school students to gain admission to colleges. Gaokao is held annually, and we refer to *Gaokao year* as the year in which a student takes the Gaokao. Gaokao involves two types of subjects: *main subjects* and *selective subjects*. Main subjects include Chinese, Mathematics, and Foreign Language[3]. The pool of selective subjects is typically divided into two categories: arts subjects and science subjects. Arts subjects consist of History, Geology, and Politics. Science subjects include Physics, Chemistry, and Biology[4]. Students are required to select three from the six subjects. Gaokao has a maximum total score of 750, with each main subject contributing 150 points and each selective subject contributing 100 points.

However, the subject selection typically does not allow full freedom. In the *traditional Gaokao*, the universal Gaokao mode before the Gaokao reform, students had to choose either all 3 arts subjects or science subjects. After 2017, a series of Gaokao reforms have been gradually implemented. The primary content of Gaokao reform is to allow greater flexibility in subject selection (State Council 2014). For instance, some provinces now allow students to freely choose any three selective subjects as they wish.

A key change accompanying the Gaokao reform is the introduction of relative grading system for selective subjects. Previously, under the traditional Gaokao mode, all subjects, including the selective subjects, were graded under an *absolute grading system*, which assigns *raw scores* based on how the students' performances compare against the object criteria. This practice was viable because the only distinction in subject selection is between the arts and science bundles. Students choosing either bundle face different sets of exams and are evaluated separately during college admissions. As a result, there were no issues regarding the comparability of scores across students with different subject selections.

However, the issue of score comparability across selective subjects arises after the Gaokao reform. The expanded possibilities of subject selection make it impractical to stick to the absolute grading system and open separate admission channels for every possible subject selection. To address this challenge, the *relative grading system* is introduced. Basically, it works by first classifying students into grade intervals based on the rank order of their

[3]The majority of students choose English, though some opt for other languages such as Japanese or Spanish. Regardless of the specific choice, grading rubrics are standardized to ensure score comparability across different languages.

[4]In Zhejiang province, an additional selective subject, General Technology, is available.

raw scores, and then assigning a *converted score* corresponding to each grade. After the Gaokao reform, selective subjects are graded under the relative grading system. Relative grading ensures the additivity of scores across subjects and the comparability of scores across students with different subject selection. Additionally, because there is no such issue of score comparability for main subjects, they consistently adopt absolute grading.

In practice, the relative grading systems implemented in different provinces differ in their *granularity*. By granularity, we mean the number of grade intervals a relative grading system classifies raw scores into. Higher granularity means more grade intervals and smaller differences in converted scores between adjacent grades. Following the Gaokao reform, two granularity levels are implemented: 1-point or 3-point score difference between adjacent grades. In the following discussions, we refer to the former as *granular* relative grading and the latter as *coarse* relative grading.

II. Model

We follow the framework of Krishna et al. 2022 to model student effort decisions. Without loss of generality, assume there is only one main subject and one selective subject. The main subject always adopts absolute grading, while the selective subject adopts either absolute or relative grading. For a representative student, achieving a raw score of t_M in the main subject and t_m in the selective subject requires effort levels of $c_M(t_M)$ and $c_m(t_m)$, respectively. Assume that both $c_M(\cdot)$ and $c_m(\cdot)$ are twice continuously differentiable with non-negative first and second derivatives, i.e., $c'_M(\cdot), c''_M(\cdot), c'_m(\cdot), c''_m(\cdot) \geq 0$, and satisfy $c_M(0) = c_m(0) = 0$.

The student's total Gaokao score is then given by $t_M + \phi(t_m)$, where the *transformation function* ϕ transforms the raw score t_m into the converted score $\phi(t_m)$. Specifically, when the selective subject adopts absolute grading, $\phi(t) = t$. That is, absolute grading is actually a special case of relative grading where the transformation is the identity function.

To characterize the heterogeneity of student ability, assume that each student belongs to a specific type x, where $x \in [0,1]$. Let $g_1(x)$ $(g_1(x) \geq 0)$ represent the marginal utility of the Gaokao score, meaning that each additional point in the Gaokao score increases the student's utility by $g_1(x)$ units. Let $g_2(x)$ $(g_2(x) \geq 0)$ denote the student's learning ability, such that the effort required for a student of type x to achieve a score of t in a subject is $\frac{c(t)}{g_2(x)}$. Without loss, assume $g_1(x) \cdot g_2(x) = x$ for any $x \in [0,1]$, as x can always be rearranged to satisfy this relationship. Additionally, assume that $g_1(x)$ is monotonically increasing in x, i.e., $g'_1(x) \geq 0$.

Let the cumulative distribution function of x be $F(x)$ and the density function be $f(x)$, such that $f(x) = F'(x)$. Assume that the distribution of x is symmetric, i.e., $f(x)$ is symmetric about $x = \frac{1}{2}$. Additionally, suppose $f(x)$ is monotonically increasing on $\left[0, \frac{1}{2}\right]$, i.e., $f'(x) \geq 0$ for all $x \in \left[0, \frac{1}{2}\right]$. These assumptions imply that the distribution of student ability follows a bell-shaped curve, with most students concentrated around the median and relatively fewer at both ends.

Under the above assumptions, the utility function of a representative student of type x is defined as:

$$U_{total}(t_M, t_m; x) = g_1(x) \cdot (t_M + \phi(t_m)) - \frac{c_M(t_M) + c_m(t_m)}{g_2(x)} \tag{1}$$

The total effort exerted by all students is given by:

$$C_{total} = \int_0^1 \frac{c_M(t_M) + c_m(t_m)}{g_2(x)} f(x)\mathrm{d}x \tag{2}$$

A. *Independence of Optimal Effort Choices In Two Subjects*

PROPOSITION 1: *The optimal choices of t_M and t_m are separable. Specifically, the optimal study time on the main subject is independent of the grading system of the selective subject, and the optimal study time on the selective subject is independent of the grading system of the main subject.*

Intuitively, Proposition 1 suggests that a student maximizes their utility from the total Gaokao score as if separately maximizing the utility from the main subject and the selective subject. This result holds because, for a given type x, the student's utility is linear in the total Gaokao score, which is the sum of the main subject's score and the selective subject's converted score, and the effort costs to achieve these scores are also separable. Proof details are provided in Appendix A.A1.

By Proposition 1, the optimal choice of t_M is constant because the grading system of the main subject remains unchanged. This directly implies that the total effort in the main subject is always constant. Consequently, any change in the grading system of the selective subject affects total student effort solely through changes in effort in the selective subject. Therefore, we can focus exclusively on the selective subject. For clarity, we introduce the following notations:

$$U := g_1(x)\phi(t_m) - \frac{c_m(t_m)}{g_2(x)}, \quad C := \int_0^1 \frac{c_m(t_m)}{g_2(x)} f(x)\mathrm{d}x, \quad t := t_m \tag{3}$$

B. *Effect of Granularity on Total Student Effort*

Under relative grading, a student's score is determined by the rank order of her raw score, i.e., the relative position of her raw score among all students. Thus, the converted score can be expressed as $\phi(t, \boldsymbol{t}_-)$, where $\boldsymbol{t}_-$ represents the target raw scores of the selective subject chosen by all other students. Hence, the utility function U is given by:

$$U(t, \boldsymbol{t}_-; x) = g_1(x)\phi(t, \boldsymbol{t}_-) - \frac{c(t)}{g_2(t)} \tag{4}$$

It is challenging to directly solve such a multi-player game for a Nash equilibrium. However, Olszewski and Siegel 2016's work on large contests demonstrates that this game can be approximated by a single-principal mechanism that directly assigns scores based on student type x from high to low. Specifically, almost all students of type x select the following target score:

$$t(x) = c^{-1}\left(x \cdot y(x) - \int_0^x y(\tilde{x})\mathrm{d}\tilde{x}\right) \tag{5}$$

where the function $y(x)$ specifies score assignments ordered by student types, rather than following the transformation rule of the relative grading system.

Let $G(y)$ be the cumulative distribution function of the assigned scores. Then, $y(x)$ is an implicit function determined by the following relationship:

$$G(y) = F(x). \tag{6}$$

This relationship indicates that the percentile of a student's type within the population equals the percentile of the score assigned to that student among all assigned scores. Define

$$G^{-1}(z) = \inf\{y \mid G(y) \geq z\} \tag{7}$$

for $z \in [0,1]$. Then, $y(x) = G^{-1}(F(x))$.

Consider a relative grading system with a full score of β [5] and a total of n grades ($n \in \mathbb{Z}$, $n \geq 2$). For simplicity, we assume in subsequent discussions that the score difference between any two adjacent grades is constant and that each grade interval contains the same number of students.

Denote the target raw score chosen by a type x student at equilibrium as $t_n(x)$, the total student effort in the selective subject as C_n, and the cumulative distribution function of the transformed scores as G_n, with $y_n(x) = G_n^{-1}(F(x))$. Then we have:

$$G_n^{-1}(z) = \beta\frac{i}{n-1},\ \forall i \in 0,1,\ldots,n-1, z \in \left[\frac{i}{n},\frac{i+1}{n}\right). \tag{8}$$

Naturally, we assume without loss that the student of the highest type is assigned with the full score, i.e., $G_n^{-1}(1) = \beta$. Then, C_n can be expressed as:

$$\begin{aligned} C_n &= \int_0^1 \frac{c(t_n(x))}{g_2(x)} f(x)\mathrm{d}x \\ &= \int_0^1 g_1(x)y_n(x)f(x)\mathrm{d}x - \int_0^1 \frac{f(x)}{g_2(x)}\mathrm{d}x\int_0^x y_n(\tilde{x})\mathrm{d}\tilde{x} \end{aligned} \tag{9}$$

PROPOSITION 2: *C_n is monotonically decreasing with respect to n.*

[5]Theoretically, if we focus solely on the selective subject, the full score can always be normalized to 1 and is thus unnecessary. However, in this setting, the main subject also contributes to the total score and adopts a different grading system, making it necessary to define the full score of the selective subject. In Gaokao's practice, while each selective subject nominally contributes 100 points to the total Gaokao score of 750, some relative grading systems ensure that the lowest converted score is not 0, but 40 instead. In such cases, the effective full score—the range of scores that influences student incentives—is only 60.

Proposition 2 implies that, within relative grading systems, as granularity increases, the total student effort in the selective subject decreases. The proof of Proposition 2 is provided in Appendix A.A2.

C. *Total Effort Under Absolute vs. Relative Grading*

Finally, we are going to compare the total effort under absolute grading versus that under relative grading. As shown in Equation (9), the total effort in the selective subject under relative grading system with n grade intervals is given by:

$$C_n = \int_0^1 g_1(x)y_n(x)f(x)\mathrm{d}x - \int_0^1 \frac{f(x)}{g_2(x)}\mathrm{d}x \int_0^x y_n(\tilde{x})\mathrm{d}\tilde{x}.$$

When the selective subject adopts absolute grading, the transformation function is actually $\phi(t) = t$, which means the raw score is exactly the transformed score. Under absolute grading, denote the optimal choice of t as t_0, and the total student effort in the selective subject as C_0. Then, the utility function for a type x student is given by:

$$U(t;x) = g_1(x)t - \frac{c(t)}{g_2(x)} \tag{10}$$

By first-order condition, the optimal choice of t_0 for a student of type x satisfies:

$$t_0(x) = c'^{-1}(x) \tag{11}$$

Thus, the total effort in the selective subject under absolute grading, C_0, is:

$$C_0 = \int_0^1 \frac{c(t_0(x))}{g_2(x)} f(x)\mathrm{d}x = \int_0^1 c(c'^{-1}(x))\frac{f(x)}{g_2(x)}\mathrm{d}x \tag{12}$$

The total effort under absolute grading (C_0) and relative grading (C_n) cannot be directly compared due to several factors, including the undetermined full score β, the number of grades n, and the functional forms of the cost function $c(\cdot)$ and the type distribution $f(\cdot)$. Thus, we consider different choices β and n based on real-world Gaokao grading systems, together with appropriate specifications of $c(\cdot)$ and $f(\cdot)$. Then, we run simulations to predict whether the relative grading system reduces total student effort compared to absolute grading.

In particular, we set $c(t) = \frac{1}{2}t^2$. For the type distribution $f(\cdot)$, we consider two distinct forms: (a) uniform distribution on $[0,1]$, denoted as $f(x) := f_{\mathscr{U}}(x) = 1$; and (b) trimmed normal distribution derived from $\mathscr{N}\left(\frac{1}{2}, \sigma^2\right)$ restricted to $[0,1]$[6], denoted as $f(x) := f_{\mathscr{N},\sigma}(x)$. And we consider $f_{\mathscr{N},\frac{1}{2}}$ ($\sigma = \frac{1}{2}$) and $f_{\mathscr{N},\frac{1}{6}}$ ($\sigma = \frac{1}{6}$), respectively.

Table 1 summarizes the simulation results, where $\left(1 - \frac{C_n}{C_0}\right) \cdot 100\%$ represents the percentage decrease in total effort in the selective subject when changing from absolute to

[6] Denote the probability density function of $\mathscr{N}\left(\frac{1}{2}, \sigma^2\right)$ as $z_\sigma(\cdot)$, and its cumulative density function as $Z_\sigma(x)$. By trimmed normal distribution, we mean for any x in $[0,1]$, the probability density at x is $z_\sigma(x) + 2Z(0)$.

TABLE 1—SIMULATION RESULTS OF STUDENT EFFORT UNDER ABSOLUTE AND RELATIVE GRADING SYSTEMS

f	n	β	C_0	C_n	$\left(1-\frac{C_n}{C_0}\right)\cdot 100\%$
f_U	20	0.6	0.2500	0.1571	37.16%
		0.7		0.1833	26.68%
		0.8		0.2095	16.21%
		1		0.2618	-4.74%
	70	0.6		0.1521	39.16%
		0.7		0.1774	29.02%
		0.8		0.2028	18.88%
		1		0.2535	-1.40%
$f_{\mathcal{N},\frac{1}{2}}$	20	0.6		0.1630	34.82%
		0.7		0.1901	23.96%
		0.8		0.2173	13.09%
		1		0.2716	-8.63%
	70	0.6		0.1578	36.89%
		0.7		0.1841	26.37%
		0.8		0.2104	15.85%
		1		0.2630	-5.18%
$f_{\mathcal{N},\frac{1}{6}}$	20	0.6		0.2157	13.71%
		0.7		0.2517	-0.67%
		0.8		0.2876	-15.05%
		1		0.3595	-43.82%
	70	0.6		0.2112	15.53%
		0.7		0.2464	1.45%
		0.8		0.2816	-12.63%
		1		0.3520	-40.79%

Notes: This table presents the results of model simulations comparing total student effort under absolute and relative grading systems. f is the distribution of student types, n denotes the number of grades into which the relative grading system divides students, and β is the full score under the relative grading system. C_0 refers to the total effort under absolute grading, while C_n represents the total effort under relative grading with n grades. The column $\left(1-\frac{C_n}{C_0}\right)\cdot 100\%$ shows the percentage decrease in total student effort in the selective subject when transitioning from absolute grading to relative grading, with negative values indicating an increase in effort. The distribution of student types f is specified as f_U or $f_{\mathcal{N},\sigma^2}$, where f_U denotes a uniform distribution on $[0,1]$, and $f_{N,\sigma}$ represents a trimmed normal distribution $\mathcal{N}(0.5,\sigma^2)$ restricted to $[0,1]$, with $\sigma=\frac{1}{2}$ or $\sigma=\frac{1}{6}$. The cost function is specified as $c(t)=\frac{1}{2}t^2$. The full score β takes values of 0.6, 0.7, 0.8, and 1, respectively. The number of grade intervals n under the relative grading system is set to 20 or 70 to align with real-world grading practices in the Gaokao system.

relative grading. The results indicate that relative grading systems generally reduces total effort compared to the absolute grading system across most parameter settings. For instance, when $(n,\beta) = (20,0.6)$—a scenario reflecting the Gaokao reform implemented in Beijing—the model predicts a 13.71% reduction in student effort under relative grading.

Therefore, based on the model and simulation results, we derive the following predictions. First, the change from absolute to relative grading in selective subjects affects student effort exclusively in those subjects, leaving effort in main subjects unchanged. Second, greater granularity in relative grading systems leads to lower total effort. Third, the relative grading systems implemented following the Gaokao reform are expected to reduce total student effort.

III. Questionnaire Survey

We designed and distributed a questionnaire survey targeting students in progress of their high school education and those who have completed it. The primary objective was to collect data on key variables that are otherwise rarely available in existing databases, such as time allocation between main and selective subjects, or proficiency levels in individual subjects. We nonetheless obtained some preliminary and suggestive results from Chinese Family Panel Studies (CFPS), as detailed in Appendix B.

The contents of the questionnaire are available on Appendix C.C1. The questionnaire was distributed online via social networks. To address potential sample biases, we conducted a heterogeneity analysis with regard to ability (see Section III.B). Eventually, we collected 1196 samples in total, of which 760 were valid[7]. The sample spans numerous provinces across China and a wide range of Gaokao years, which allows us to estimate Gaokao reform's impact on student effort. Specifically, the sample includes three types of grading system. The first is absolute grading under the traditional Gaokao mode; the two other are relative grading systems differing in granularity. We refer to the one with 1-point score different between adjacent grades as *granular relative grading*, and the other with 3-point score difference as *coarse relative grading*. Figure C1 illustrates the distribution of Gaokao year and Gaokao grading system. Figure C2 shows the geographical distribution of the sample across China.

A. Main Results

As predicted by the model, under reasonable assumptions, the study time in selective subjects is expected to decrease following the Gaokao reform, while the study time in main subjects remains unchanged. As a result, the total study time is expected to decrease. To test these predictions, we estimate the following two equations using ordinary least squares (OLS):

$$y_{it} = \beta_0 + \beta_1 \cdot reform + \delta C + \varepsilon \tag{13}$$

[7] The data filtering criteria are detailed in Appendix C.C2. In principle, most of the excluded samples are those whose college admission channel is not through Gaokao, or whose Gaokao year is too distant to have a sufficient sample size for that year.

$$y_{it} = \gamma_0 + \gamma_1 \cdot granular + \gamma_2 \cdot coarse + \delta C + \varepsilon, \tag{14}$$

where y represents the study time per week (in hours) on main subjects (*main*), selective subjects ($selective$), or the total study time ($total$). The variables $reform$, $granular$, and $coarse$ are dummies indicating whether the respondent experienced a relative grading system following Gaokao reform, a granular relative grading system, or a coarse relative grading system for selective subjects, respectively. C is the matrix of control variables, including key factors such as Gaokao year, student ranking, sleep time, self-reported pressure, and the degree of imbalance between subjects.

The regression results presented in Table 2 align with the model's predictions. Specifically, the Gaokao reform significantly reduces study time on selective subjects by 5.96 hours per week ($p < 0.001$). Meanwhile, the time spent on major subjects remains unchanged, with coefficients both economically and statistically insignificant. As a result, the total study time decreases by 5.93 hours per week ($p < 0.001$), driven entirely by the reduction of the study time in selective subjects. There is some heterogeneity in the effects of granular and coarse relative grading systems, but the differences are not significant.

TABLE 2— IMPACT OF GAOKAO REFORM ON WEEKLY STUDY TIME

Dependent Variables:	$selective$		$main$		$total$	
	(1)	(2)	(3)	(4)	(5)	(6)
$reform$	-5.959		0.029		-5.930	
	(1.301)		(1.482)		(1.861)	
$granular$		-6.356		1.015		-5.341
		(1.481)		(1.656)		(2.088)
$coarse$		-5.720		-0.564		-6.284
		(1.356)		(1.575)		(2.024)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	760	760	760	760	760	760

Notes: This table presents the regression results for the impact of the Gaokao reform on weekly study time from estimating Equation (13) and (14). The dependent variables are weekly study time on selective subjects (Columns 1–2), main subjects (Columns 3–4), and total weekly study time (Columns 5–6), measured in hours. The variable $reform$ is a dummy indicating whether the respondent experienced the Gaokao reform, while *granular* and *coarse* are dummies indicating whether granular or coarse relative grading systems, respectively, were implemented. Control variables include Gaokao year, student ranking, sleep time, self-reported pressure, and the degree of imbalance between subjects, among others. Heteroskedasticity-robust standard errors are reported in parentheses.

Even though the grading system change is limited to selective subjects, the focus on any potential change in the study time in main subjects has important implications. Theoretically, while relative grading systems reduce student effort in selective subjects, the saved time could potentially be reallocated to main subjects, which might prevent a reduction in total study time as a result. However, our findings indicate that students did not redirect the saved time from selective subjects to main subjects. The findings are consistent with the model's prediction, which in turn supports the assumption that utility gains and effort costs from main and selective subjects are separable. Therefore, there is no spillover effect between these two types of subjects. We can then conclude that the total study time decreases following the transition from absolute to relative grading, entirely driven by the study time decrease in selective subjects.

B. Heterogeneity Analysis

Existing literature highlights the heterogeneous impacts of grading systems on students with different abilities (Paredes 2017; Andreoni and Brownback 2017; Dubey and Geanakoplos 2010). Specifically, as noted by Paredes 2017, high-ability students tend to reduce their effort under relative grading, whereas low-ability students tend to increase their effort. Therefore, we expect to observe such heterogeneous impacts on study time in selective subjects across ability groups. However, for main subjects, as proposed by Proposition 1, the effort choice for main subjects is independent of the grading system adopted by selective subjects. Thus, we expect to observe no such heterogeneity in study time for main subjects.

To examine this heterogeneity, we classify students into high-ability and low-ability groups based on the median ranking in their most recent school exam. Students ranked above the median are categorized as high-ability, while those below the median are considered low-ability. We then use OLS to estimate Equation (13) and (14) separately for the two groups. The regression results are presented in Table 3.

Interestingly, the Gaokao reform has similar effects on study time for selective subjects among both high- and low-ability students. For high-ability students, the findings align with Paredes 2017: they exert less effort in selective subjects. The estimated coefficients for high-ability students are close to those reported in the main results (see Table 2). However, the response of low-ability students differs from Paredes 2017: their study time in selective subjects decrease in a manner similar to that of the high-ability students. This deviation may arise from contextual differences.

In practice, under Gaokao's relative grading system, every test taker is guaranteed a minimum converted score of 40 even for submitting a blank answer, which is obviously not the case under absolute grading. Additionally, the relative grading system enforces a predetermined score distribution with a higher mean and lower variance compared to that under absolute grading, making the distinction between high- and low-ability students less pronounced. Consequently, low-ability students face relatively mild consequences for bad performances in selective subjects. This motivates them to shirk on selective subjects and redirect the saved time to other activities that gives them higher utility.

For study time on main subjects, we again observe no such heterogeneity: none of the coefficients for $reform$, $granular$, or $coarse$ is statistically significant. The results show that the study time devoted to main subjects by both high-ability and low-ability students remains unaffected by the grading system change in selective subjects.

Combining the main results and the heterogeneity analysis, we conclude that students, regardless of their ability, reduce their study time in selective subjects following the Gaokao reform, while their study time in main subjects remains unchanged. Consequently, the total study time decreases across all students.

However, the survey method has its inherent limitations. First, no student can experience the Gaokao twice; each individual undergoes either the absolute grading system or the relative one. Thus, it is impossible to construct panel data or derive a perfect counterfactual for each student. This limitation is inevitable for any analysis using survey data. In light of this,

TABLE 3— HETEROGENEITY IN THE IMPACT OF GAOKAO REFORM ON STUDY TIME BY STUDENT ABILITY

Panel A. Study Time on Selective Subjects (select)

Dependent Variable:		*select*		
Sample	High-Ability	Low-Ability	High-Ability	Low-Ability
	(1)	(2)	(3)	(4)
reform	-6.221	-5.069		
	(1.867)	(1.942)		
granular			-6.325	-5.792
			(2.101)	(2.232)
coarse			-6.140	-4.745
			(1.918)	(2.015)
Control Variables	Yes	Yes	Yes	Yes
Observations	378	382	378	382

Panel B. Study Time on Main Subjects (main)

Dependent Variable:		*main*		
Sample	High-Ability	Low-Ability	High-Ability	Low-Ability
	(5)	(6)	(7)	(8)
reform	-1.100	2.059		
	(1.765)	(2.547)		
granular			0.563	2.662
			(1.979)	(2.808)
coarse			-2.387	1.788
			(1.928)	(2.634)
Control Variables	Yes	Yes	Yes	Yes
Observations	378	382	378	382

Notes: This table presents the heterogeneity analysis of the Gaokao reform's impact on weekly study time (in hours) for selective subjects (*select*) and main subjects (*main*), divided by student ability. Students are categorized into high-ability and low-ability groups based on their rankings in the most recent school exam: those ranked above the school median are classified as high-ability, while those at or below the median are classified as low-ability.

Panel A reports the heterogeneity analysis of the Gaokao reform's impact on weekly study time (in hours) for selective subjects. Columns 1–2 present the estimates of *reform*, a dummy variable indicating whether the respondent experienced the Gaokao reform, for high-ability and low-ability groups, respectively. Columns 3–4 show the estimates for *granular* and *coarse*, indicating granular and coarse relative grading systems, for the same ability groups. Panel B reports the heterogeneity analysis of the Gaokao reform's impact on weekly study time (in hours) for main subjects. Columns 5-6 present the estimates of *reform* for high-ability and low-ability groups, respectively. Columns 7-8 present the estimates for *granular* and *coarse*. Control variables include year, student rankings, sleep time, self-reported pressure, family support, and other factors. Heteroskedasticity-robust standard errors are reported in parentheses.

the option is to rely on cross-sectional data with the implicit assumption that, conditional on the factors we control for, students from different Gaokao years are comparable. Second, it is inherently difficult to control for all relevant variables to avoid omitted variable bias, even though we have tried to include a wide range of background variables. Nonetheless, despite all these limitations, the analysis of the questionnaire survey offers empirical evidence on how changes in the grading system affect student effort. In addition, we conducted further analysis in Appendix C.C3 to explore its broader implications for potential improvements in students' mental health.

IV. Experiment

As discussed in the previous section, the inherent limitations of survey data may obscure the causal identification of the effects of grading systems on student effort. Thus, we conducted an online experiment to identify the impact of grading systems on effort decisions. This experiment also extends the research context beyond the Gaokao to more general comparisons between absolute and relative grading systems.

A. Experimental Design

The experimental design replicates the incentive structure of the Gaokao. Specifically, subjects are to complete two tasks: *mental arithmetic* and *typing*. Mental arithmetic serves as the "main subject" that always adopts absolute grading system, and its raw score directly contributes to the total score. Typing is the "selective subject" that adopts either absolute or relative grading, varing by group assignment.

Subjects are randomly assigned to the control group or two treatment groups. In the control group, typing adopts absolute grading. In the treatment groups, typing adopts relative grading with one of the two levels of granularity: the *granular group* adopts the one with higher granularity (*granular relative grading*), classifying students into 11 grades; the *coarse group* adopts the one with smaller granularity (*coarse relative grading*), classifying students into only 5 grades. Figure 1 and 2 demonstrate the grading rules of the granular and coarse relative grading, respectively. Table 4 summarizes the grading rules across groups.

FIGURE 1. GRADING RULE FOR THE GRANULAR RELATIVE GRADING IN THE TYPING TASK

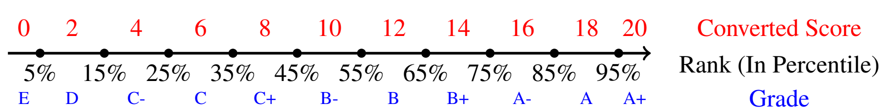
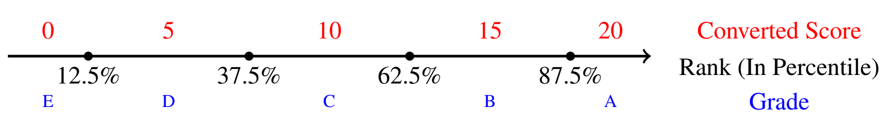


FIGURE 2. GRADING RULE FOR THE COARSE RELATIVE GRADING IN THE TYPING TASK



A total of 255 individuals registered for the experiment, of whom 117 fully completed it. Specifically, 42 subjects were assigned to the control group, 39 to the granular group, and 36 to the coarse group. Subjects participated in the experiment through online platforms[8].

[8] The website for the mental arithmetic task is https://www.preplounge.com/en/mental-math/add/1. The website for the typing task is https://dazi.91xjr.com.

TABLE 4—GRADING SYSTEMS ACROSS EXPERIMENTAL GROUPS

Group Assignment	Control Group	Granular Group	Coarse Group
Grading System			
Mental Arithmetic	Absolute	Absolute	Absolute
Typing	Absolute	Granular Relative	Coarse Relative
No. of Subjects	42	39	36

Notes: This table summarizes the grading systems adopted in each experimental group. Details of granular and coarse relative grading systems are shown in Figure 1 and 2, respectively.

They were allowed to attempt both tasks as many times as they wish, potentially aiming for higher scores.

For the mental arithmetic task, the context is the addition of two two-digit numbers. Each trial begins with a 1-minute countdown. For every correct answer provided, the subject earns 1 point toward their raw score and an additional 2 seconds in the remaining time. For incorrect answers, 1 point is deducted from the raw score, without any deduction to the remaining time. The system records the raw score and the answer time for each trial.

For the typing task, the text to be typed remains the same across all trials and subjects. In each trial, the subject has 1 minute to type the text[9]. The website records the typing speed (in characters per minute, *cpm*) for each trial and the total number of trials. The highest typing speed is taken as the subject's score in the typing task.

This experiment was conducted online during the winter vacation. After receiving the experiment manual[10], subjects had one week to freely decide how to participate under the given rules. The incentives for the subjects were structured as follows: all participants received a basic reward of 3 yuan, which fairly compensates for the opportunity cost of participation. On this basis, some participants received additional rewards based on their exceptional performance within their group. Specifically, the top 2% received an additional reward of 100 Chinese yuan, and those ranked between the top 2% and 10% received an additional reward of 20 yuan. Participants ranked below the top 10% did not receive additional rewards. The subjects were not informed of the number of subjects in their group. This significant disparity in rewards was designed to simulate the large gap between ideal and less-than-ideal outcomes in real-world tournaments.

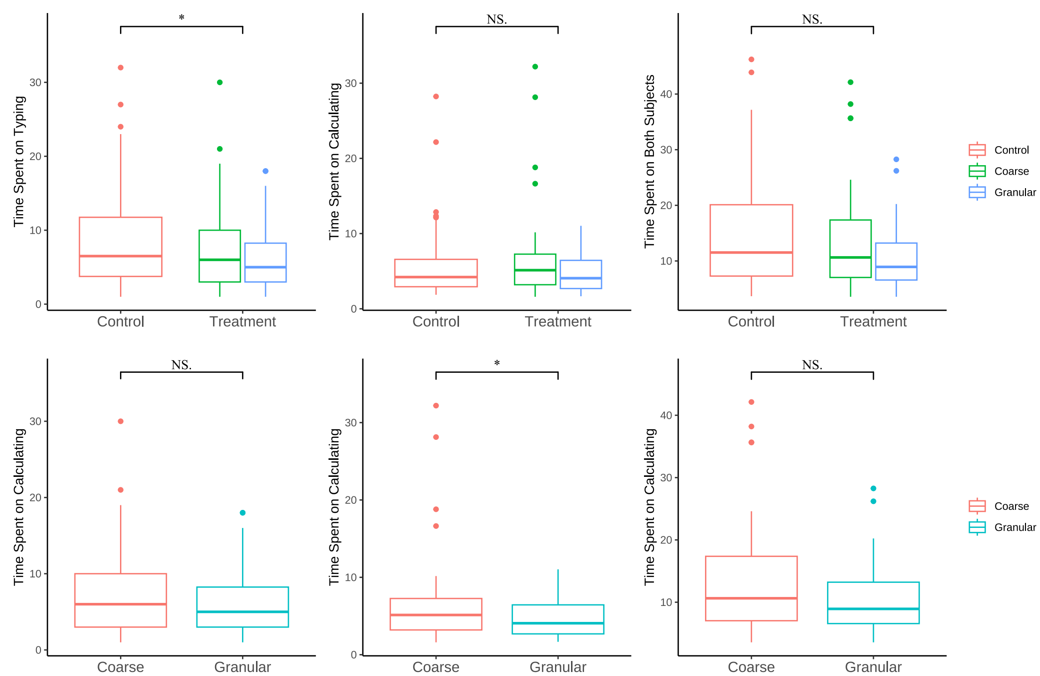
B. Experimental Results

We are interested in the comparisons between control and treatment groups, and the comparisons between the coarse and granular groups. Specifically, we compare the average time invested on typing, mental arithmetic, and the total time by treatment arm. Table 5 presents the results. Panel A compares the control group with the two treatment groups combined. Panel B compares the granular group with the coarse group. Figure 3 visualizes those re-

[9]The website prevents the subjects from simply doing "copy and paste."
[10]For details of the experiment manual, see Appendix D.D1.

sults. Recall that typing serves as the "selective subject," and mental arithmetic serves as

FIGURE 3. COMPARISON OF INVESTED TIME: CONTROL VS. TREATMENT, COARSE VS. GRANULAR



the "main subject." The results are consistent with our previous findings. For typing, the "selective subject," relative grading systems significantly reduce the time subjects spend on it, with a decrease of 2.362 minutes compared to absolute grading ($p = 0.085$). Given that the control group spends an average of 9.000 minutes on typing, this reduction represents a substantial decrease of 26.2% in effort for the "selective subject."

For mental arithmetic, the "main subject," the difference of -0.378 minutes is neither economically meaningful nor statistically significant ($p = 0.726$), indicating that the grading system adopted by typing, the "selective subject," does not influence effort choices on the "main subject." As a result, the total time for the treatment groups is 2.740 minutes less than that of the control group ($p = 0.178$), corresponding to a notable 17.9% decrease in the control group's average total time.

We observe similar patterns in the comparisons between the granular and coarse groups[11]. Additional comparisons between the control group and the granular group, and the comparisons between the control group and the coarse group, are detailed in the Table D3.

The lack of statistical significance for some coefficients is less than ideal. This can be

[11] Two outliers in the coarse group spent extremely large amounts of time on mental arithmetic, substantially driving up the group mean. Excluding these outliers brings the mean back to a normal level of 5.575 minutes.

TABLE 5—MAIN COMPARISONS OF INVESTED TIME: CONTROL VS. TREATMENT AND COARSE VS. GRANULAR

	Panel A. Control and Treatment Groups			
	Mean			
	Treatment	Control	Difference	p-value
Time in Typing	6.638	9.000	-2.362	0.085
	(0.655)	(1.195)	(1.360)	
Time in Mental Arithmetic	5.887	6.265	-0.378	0.726
	(0.637)	(0.870)	(1.076)	
Total Time	12.525	15.265	-2.740	0.178
	(1.034)	(1.742)	(2.022)	
Number of Subjects	75	42		
	Panel B. Coarse and Granular Treatment Group			
	Mean			
	Granular	Coarse	Difference	p-value
Time in Typing	6.111	7.212	-1.101	0.412
	(0.732)	(1.115)	(1.334)	
Time in Mental Arithmetic	4.806	7.066	-2.259	0.086
	(0.429)	(1.224)	(1.297)	
Total Time	10.918	14.278	-3.360	0.112
	(1.012)	(1.830)	(2.090)	
Number of Subjects	39	36		

Notes: This table compares the average time spent (in minutes) on typing, mental arithmetic, and the average of the total time across different treatment arms. Panel A compares the control group with two treatment groups combined. Panel B compares the granular and coarse groups. The "Difference" column reports the mean differences. p-values are calculated using a two-sample t-test allowing for unequal variances. Standard errors are reported in parentheses.

attributed to two main reasons. First, although the experiment included a total of 117 subjects, each group ends up with around 40 participants after the random assignment. The limited sample size increased the variability in the data, making it harder for the estimates to be statistically significant. Second, the experiment could have been designed in a way that amplifies the differences between treatment arms to distinguish and identify the effects more clearly. For example, we could have increased number of grades the granular relative grading categorizes students into, and decreased that in the coarse relative grading. Nonetheless, the experiment successfully broadens the context beyond Gaokao and offers valuable insights into the more general comparison of grading systems and their influence on effort choices.

V. Conclusions and Discussions

The Gaokao reform, which allows more flexibility in subject selection and changes the grading system for selective subjects from absolute to relative, has sparked extensive debate about whether it actually promotes subject selection (Zhou, Shan and Qin 2024; Liu and

Wei 2023; Zhong and Wang 2019). Despite a robust discussion on the Gaokao and Gaokao reform, existing research lacks both theoretical and empirical analyses of student effort under different grading systems. This paper takes a first step in addressing this gap by examining how changes in the grading system influence student effort within the Gaokao context.

The primary finding of this paper is that relative grading systems reduce the study time in selective subjects without increasing the study time in main subjects. Consequently, it leads to an overall decrease in the total study time. Furthermore, this effect is more pronounced in relative grading systems with higher granularity, i.e., systems that categorize students into more grades.

Our study contributes to the existing literature in three key aspects. First, it extends the theoretical analysis of grading systems by examining both selective subjects, which adopt absolute or relative grading, and main subjects, which consistently adopts absolute grading. This perspective enables us to analyze not only the direct effects of grading system changes on effort in selective subjects, but also whether the reduced effort in selective subjects is reallocated to effort in main subjects. Furthermore, we investigate the granularity of relative grading systems, showing that higher granularity has a more pronounced effect in reducing student effort.

Second, to our knowledge, this paper is the first to empirically investigate the impact of the Gaokao reform on students' effort decisions. By integrating CFPS data analysis, the questionnaire survey, and the online experiment, we provide robust evidence that the Gaokao reform reduces student effort for both the high- and low-ability students. We also present results suggesting that the Gaokao reform alleviates academic pressure and enables students to engage more in extracurricular activities.

Third, our findings have strong implications for educational policy design. In the context of intense competition among Chinese high school students (Yu, Li and Wang 2016; Fu et al. 2023), we suggest that adopting relative grading systems is more effective than regulatory measures in mitigating student competition. This is because students always make their effort decisions in response to the grading systems that provide the right incentives. Rather than enforcing regulations within the current incentive structure, changing the grading system can shift the equilibrium in a subtly way towards a more socially desirable outcome.

There are potential challenges to the interpretation of our findings. The first alternative explanation is that the empirical results from the CFPS data and questionnaire survey are actually estimating the impact of the Gaokao reform, rather than the impact of changes in grading systems. For instance, one could argue that the reduction in study time is driven by the increased flexibility in subject selection following the Gaokao reform, because students are more likely to excel in subjects they voluntarily choose. However, this concern can be addressed by the experiment, where the controlled experiment design ensures that the differences in effort can only be attributed to the adoption of different grading systems.

The second challenge is that our results may be confounded by other shocks or policies implemented during the same period. For instance, in 2021, the Chinese government introduced a regulatory policy aimed at alleviating students' academic burden by imposing restrictions on schools' course schedules, exam frequency, and related activities (State

Council 2021). Additionally, factors such as the COVID-19 pandemic could have significantly influenced students' study modes (Aristeidou and Cross 2021; Yates et al. 2021). However, these factors do not invalidate of our empirical results. On the one hand, the analysis of CFPS data employed a difference-in-differences (DiD) method, which effectively controls for nationwide shocks and policies under proper assumptions of parallel trend. On the other hand, and most importantly, our experiment was specifically designed to construct a controlled context that eliminates potential confounding factors.

Of course, there are certain limitations to this study that should be noted in future research. First, there is some homogeneity in the sample of the questionnaire survey and the subjects recruited for the online experiment, as both were distributed through our social network. In the experiment, the control and treatment group settings mitigate this homogeneity to ensure accurate identification. However, in the survey data, this homogeneity may somewhat confound the estimation. While it does not affect the estimates themselves, it influences the standard errors of the estimates. To address this, we applied heteroskedasticity-robust standard errors throughout all regression analyses. Nevertheless, the optimal approach would be to distribute the questionnaires, whether for data collection or subject recruitment, anonymously[12] and through online platforms rather than via social networks.

Second, certain aspects of the experimental design could be further improved. One notable limitation is the sample size, as the relatively small number of participants in each group limits the statistical power of the analysis. Expanding the sample size in future experiments would enhance the reliability and generalizability of the findings. Also, the experiment design could be refined to better amplify the differences between treatment arms, which helps distinguish and identify the effect of interest. Notably, conducting the experiment online brings in some challenges, because subjects may face greater learning cost of getting familiar with the designated platforms due to the lack of in-person instructions from the organizers, and they are more likely to be distracted. Conducting the experiment offline would help address these challenges, although some existing studies suggest little differences in response behavior and data quality between online and offline experiments (Clifford and Jerit 2014; Wenz 2021). Despite all these, offline experiments can definitely help test and establish the robustness of our experimental results.

There are many areas left for future work. First, with more comprehensive data, further studies could explore the impact of grading systems in the United States or other regions worldwide. This is because regional heterogeneity matters (Fang et al. 2013; Zhao, Kuh and Carini 2005). Thus, students from different countries or with different cultural backgrounds may have different responses to the grading systems. Such research would be particularly valuable for informing educational policy design in diverse contexts. For example, in some cases, the focus may be on alleviating intense competition among students, while in others, the focus might be to motivate greater student effort.

Second, within the context of the Gaokao reform that aims to encourage more diverse subject selection, future research could evaluate its actual policy outcomes. Although there

[12]By "anonymous," we mean that respondents are unaware of who is conducting the questionnaire survey or organizing the experiment. Throughout the research, we did not disclose the research purpose in the questionnaire survey or online experiment, nor did we collect personal information. This was clearly communicated to all respondents and subjects.

are existing studies debating the policy's effects, many are grounded in normative arguments and lack rigorous methodologies or empirical evidence (e.g., Liu and Wei 2023; Zhong and Wang 2019). One interesting topic is, are students genuinely diversifying their subject choices based on their interests, or are they just strategically selecting subjects in which they are most competitive? This arises because, under the relative grading system, students compete with peers to earn their transformed scores, which creates a trade-off between selecting subjects which they feel most interested or are most competitive in. There might be some heterogeneity in how students evaluate this trade-off. Moreover, their beliefs about how others treat this trade-off constitute another key factor. These dynamics constitute a Bayesian game involving hundreds of thousands of students that need further theoretical and empirical exploration.

REFERENCES

Andreoni, James, and Andy Brownback. 2017. "All pay auctions and group size: Grading on a curve and other applications." *Journal of Economic Behavior & Organization*, 137: 361–373.

Aristeidou, Maria, and Simon Cross. 2021. "Disrupted distance learning: the impact of Covid-19 on study habits of distance learning university students." *Open Learning: The Journal of Open, Distance and e-Learning*, 36(3): 263–282.

Azmat, Ghazala, and Nagore Iriberri. 2010. "The importance of relative performance feedback information: Evidence from a natural experiment using high school students." *Journal of Public Economics*, 94(7-8): 435–452.

Becker, William E, and Sherwin Rosen. 1992. "The learning effect of assessment and evaluation in high school." *Economics of Education Review*, 11(2): 107–118.

Berger, Martijn PF, and Wim J Veerkamp. 1994. "A Review of Selection Methods for Optimal Test." *Journal of the American Statistical Association*, 78: 90–98.

Bodoh-Creed, Aaron L, and Brent R Hickman. 2018. "College assignment as a large contest." *Journal of Economic Theory*, 175: 88–126.

Clifford, Scott, and Jennifer Jerit. 2014. "Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies." *Journal of Experimental Political Science*, 1(2): 120–131.

Czibor, Eszter, Sander Onderstal, Randolph Sloof, and C Mirjam van Praag. 2020. "Does relative grading help male students? Evidence from a field experiment in the classroom." *Economics of Education Review*, 75: 101953.

Dubey, Pradeep, and John Geanakoplos. 2010. "Grading exams: 100, 99, 98,… or a, b, c?" *Games and Economic Behavior*, 69(1): 72–94.

Du, Xiao. 2022. "The Difficulties of After-class Service under the Vision of "Double Reduction" Policy and Its Solutions." *Jinan Journal (Philosophy and Social Sciences Edition)*, 44(03): 121–132.

Fang, Zheng, Leslie W Grant, Xianxuan Xu, James H Stronge, and Thomas J Ward. 2013. "An international comparison investigating the relationship between national culture and student achievement." *Educational Assessment, Evaluation and Accountability*, 25: 159–177.

Fu, Xiaolan, Wei Zhang, Xuefeng Chen, and Sheyan Chen. 2023. *Report on National Mental Health Development in China (2021–2022).* Psychological Health Blue Book.

Ha, Wei, Le Kang, and Yang Song. 2020. "College matching mechanisms and matching stability: Evidence from a natural experiment in China." *Journal of Economic Behavior & Organization*, 175: 206–226.

Heckman, James J. 2005. "China's human capital investment." *China Economic Review*, 16(1): 50–70.

Jiang, Ming. 2019. "When do stable matching mechanisms fail? The role of standardized tests in college admissions." *The Role of Standardized Tests in College Admissions (March 31, 2019).*

Kojima, Fuhito, and Parag A Pathak. 2009. "Incentives and stability in large two-sided matching markets." *American Economic Review*, 99(3): 608–627.

Krishna, Kala, Sergey Lychagin, Wojciech Olszewski, Ron Siegel, and Chloe Tergiman. 2022. "Pareto improvements in the contest for college admissions." National Bureau of Economic Research.

Landeras, Pedro. 2009. "Student effort: standards vs. tournaments." *Applied Economics Letters*, 16(9): 965–969.

Lazear, E. P., and S. Rosen. 1981. "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy*, 89(5): 841–864.

Liu, Haifeng, and Huafeng Wei. 2023. "Survey and Analysis of High School Students' Understanding and Intention to Subjects' Selection in the College Entrance Examination under the Background of the New College Entrance Examination Reform." *Education Development Research*, 43(12): 1–9, 17.

Moldovanu, Benny, and Aner Sela. 2008. "The optimal allocation of prizes in contests." In *40 Years of Research on Rent Seeking 1*. 615–631. Springer.

Olszewski, Wojciech, and Ron Siegel. 2016. "Large contests." *Econometrica*, 84(2): 835–854.

Olszewski, Wojciech, and Ron Siegel. 2019. "Bid caps in large contests." *Games and Economic Behavior*, 115: 101–112.

Paredes, Valentina. 2017. "Grading system and student effort." *Education Finance and Policy*, 12(1): 107–128.

Schultz, Theodore W. 1960. "Capital formation by education." *Journal of political economy*, 68(6): 571–583.

Schultz, Theodore W. 1961. "Investment in human capital." *The American economic review*, 51(1): 1–17.

State Council. 2014. "Implementation Opinions of the State Council on Deepening the Reform of the Examination and Enrollment System." Issued on September 4, 2014. Classification: Science and Education / Education.

State Council. 2021. "Opinions on Further Reducing the Burden of Homework and Off-campus Training for Students in Compulsory Education." Issued on July 24, 2021. Commonly referred to as the "Double Reduction" policy.

Van der Linden, Wim J. 2015. "Optimal test assembly." *Handbook of test development*, 507–530.

Wenz, Alexander. 2021. "Do distractions during web survey completion affect data quality? Findings from a laboratory experiment." *Social Science Computer Review*, 39(1): 148–161.

Yang, Xiaowei, and Yan Wen. 2022. "The Current Situation, Challenges and Future Focus of the Implementation of the "Double Reduction" Policy." *Journal of Xinjiang Normal University (Philosophy and Social Sciences Edition)*, 43(04): 25–38+2.

Yates, Anne, Louise Starkey, Ben Egerton, and Florian Flueggen. 2021. "High school students' experience of online learning during Covid-19: the influence of technology and pedagogy." *Technology, Pedagogy and education*, 30(1): 59–73.

Yu, Guoliang, Tianran Li, and Qing Wang. 2016. "Changes in High School Students' Mental Health: A Cross-Temporal Meta-Analysis." *Educational Research*, 37(10): 113–122.

Zhao, Chun-Mei, George D Kuh, and Robert M Carini. 2005. "A comparison of international student and American student engagement in effective educational practices." *The Journal of Higher Education*, 76(2): 209–231.

Zhong, Bing-lin, and Xin-feng Wang. 2019. "The Realistic Dilemma, Value Basis and Strategic Choice of the New College Entrance Examination of China." *Journal of Educational Studies*, 15(5): 62–69.

Zhou, Hongyu, and Yanlei Qi. 2022. "Implementation of the "Double Reduction" Policy: Focus, Difficulties and Suggestions." *Journal of Xinjiang Normal University (Philosophy and Social Sciences Edition)*, 43(01): 69–78.

Zhou, Jianhua, Zhengyi Shan, and Hongxia Qin. 2024. "Has the New College Entrance Examination Promoted Students' Independent Subject Selection: Empirical Analysis Based on CatBoost Regression Tree Model." *Journal of East China Normal University (Educational Sciences Edition)*, 42(03): 12–25.

MATHEMATICAL APPENDIX

A1. Proof of Proposition 1

PROOF:

Let $U_M(t_M;x) := g_1(x)t_M - \frac{c_M(t_M)}{g_2(x)}$ and $U_m(t_m;x) := g_1(x)\phi(t_m) - \frac{c_m(t_m)}{g_2(x)}$. We can rewrite $U_{total}(t_M,t_m;x)$ as

$$\begin{aligned} U_{total}(t_M,t_m;x) &= g_1(x)\cdot(t_M+\phi(t_m)) - \frac{c_M(t_M)+c_m(t_m)}{g_2(x)} \\ &= \left(g_1(x)t_M - \frac{c_M(t_M)}{g_2(x)}\right) + \left(g_1(x)\phi(t_m) - \frac{c_m(t_m)}{g_2(x)}\right) \\ &= U_M(t_M;x) + U_m(t_m;x) \end{aligned} \tag{A1}$$

This implies that the utility from the total Gaokao score can be split into two parts: the main subject and the selective subject.

To maximize the student's utility, t_M satisfies the first-order condition:

$$\begin{aligned} 0 &= \frac{\partial U_{total}}{\partial t_M} \\ &= \frac{\partial U_M}{\partial t_M} + \frac{\partial U_m}{\partial t_M} \\ &= \left(g_1(x) - \frac{c'_M(t_M)}{g_2(x)}\right) + 0 \end{aligned} \tag{A2}$$

Solving the first-order condition yields $t_M = c'^{-1}_M(x)$, which depends only on x and $c_M(\cdot)$. Similarly, we know that t_m satisfies $x\phi'(t_m) = c'_m(t_m)$, which only depends on x, c_m and $\phi(\cdot)$, namely the grading system on selective subjects.

Q.E.D.

A2. Proof of Proposition 2

PROOF:

Denote $I_1(n)$ and $I_2(n)$ as

$$I_1(n) := \int_0^1 g_1(x)y_n(x)f(x)\mathrm{d}x,\ I_2(n) := \int_0^1 \frac{f(x)}{g_2(x)}\mathrm{d}x \int_0^x y_n(\tilde{x})\mathrm{d}\tilde{x}.$$

It suffices to prove that $I_1(n)$ is monotonically decreasing with respect to n, and $I_2(n)$ is monotonically increasing with respect to n.

For $I_1(n)$, taking the difference between $I_1(n)$ and $I_1(n+1)$ yields:

(A3)
$$\begin{aligned}\Delta I_1(n) =& I_1(n) - I_1(n+1)\\ =& \int_0^1 g_1(x) y_n(x) f(x)\mathrm{d}x - \int_0^1 g_1(x) y_{n+1}(x) f(x)\mathrm{d}x\\ =& \int_0^1 g_1(G_n^{-1}(F(x)) - G_{n+1}^{-1}(F(x))) f(x)\mathrm{d}x\\ =& \int_0^1 g_1(F^{-1}(x))(G_n^{-1}(x) - G_{n+1}^{-1}(x))\mathrm{d}x\end{aligned}$$

Let $h(x) = g_1(F^{-1}(x))$, then $h(x) > 0$ and is monotonically increasing with respect to x.

(A4)
$$\begin{aligned}\Delta I_1(n) &= \int_0^1 h(x)(G_n^{-1}(x) - G_{n+1}^{-1}(x))\mathrm{d}x\\ &= \sum_{i=1}^{n-1} \int_{\frac{i}{n+1}}^{\frac{i}{n}} h(x)\beta\left(\frac{i-1}{n-1} - \frac{i}{n}\right)\mathrm{d}x + \sum_{i=1}^{n-1} \int_{\frac{i}{n}}^{\frac{i+1}{n+1}} h(x)\beta\left(\frac{i}{n-1} - \frac{i}{n}\right)\mathrm{d}x\\ &\geq \sum_{i=1}^{n-1} \int_{\frac{i}{n+1}}^{\frac{i}{n}} h\left(\frac{i}{n}\right)\beta\left(\frac{i-1}{n-1} - \frac{i}{n}\right)\mathrm{d}x + \sum_{i=1}^{n-1} \int_{\frac{i}{n}}^{\frac{i+1}{n+1}} h\left(\frac{i}{n}\right)\beta\left(\frac{i}{n-1} - \frac{i}{n}\right)\mathrm{d}x\\ &= \beta\sum_{i=1}^{n-1} h\left(\frac{i}{n}\right)\left(\left(\frac{i-1}{n-1} - \frac{i}{n}\right)\left(\frac{i}{n} - \frac{i}{n+1}\right) + \left(\frac{i}{n-1} - \frac{i}{n}\right)\left(\frac{i+1}{n+1} - \frac{i}{n}\right)\right)\\ &= \beta\sum_{i=1}^{n-1} h\left(\frac{i}{n}\right)\left(\frac{i(i-n)}{(n-1)n^2(n+1)} + \frac{i(n-i)}{(n-1)n^2(n+1)}\right)\\ &> 0\end{aligned}$$

Thus, $I_1(n)$ is monotonically decreasing with respect to n.

For $I_2(n)$, taking the difference between $I_2(n)$ and $I_2(n+1)$ yields:

(A5)
$$\begin{aligned}\Delta_{\text{latter}} &= \int_0^1 \frac{f(x)}{g_2(x)}\mathrm{d}x \int_0^x y_n(\tilde{x})\mathrm{d}\tilde{x} - \int_0^1 \frac{f(x)}{g_2(x)}\mathrm{d}x \int_0^x y_{n+1}(\tilde{x})\mathrm{d}\tilde{x}\\ &= \int_0^1 \frac{f(x)}{g_2(x)}\mathrm{d}x \int_0^x \left(y_n(\tilde{x}) - y_{n+1}(\tilde{x})\right)\mathrm{d}\tilde{x}\end{aligned}$$

It suffices to show that $\int_0^x (y_n(\tilde{x}) - y_{n+1}(\tilde{x}))\mathrm{d}\tilde{x} \leq 0$ holds for any $x \in [0,1]$.

For any $n \geq 2$, G_n^{-1} is symmetric about $\left(\frac{1}{2}, \frac{\beta}{2}\right)$, and F is symmetric about $\left(\frac{1}{2}, \frac{1}{2}\right)$, hence

y_n is symmetric about $\left(\frac{1}{2}, \frac{\beta}{2}\right)$. Thus, for all $a \in \left[0, \frac{1}{2}\right)$:

$$
\begin{aligned}
&\int_{\frac{1}{2}-a}^{\frac{1}{2}+a} (y_n(\tilde{x}) - y_{n+1}(\tilde{x}))\mathrm{d}\tilde{x} \\
&= \int_{\frac{1}{2}-a}^{\frac{1}{2}} (y_n(\tilde{x}) - y_{n+1}(\tilde{x}))\mathrm{d}\tilde{x} + \int_{\frac{1}{2}}^{\frac{1}{2}+a} (y_n(\tilde{x}) - y_{n+1}(\tilde{x}))\mathrm{d}\tilde{x} \\
&= \int_{\frac{1}{2}-a}^{\frac{1}{2}} (y_n(\tilde{x}) - y_{n+1}(\tilde{x}))\mathrm{d}\tilde{x} + \int_{\frac{1}{2}-a}^{\frac{1}{2}} (y_n(1-\tilde{x}) - y_{n+1}(1-\tilde{x}))\mathrm{d}\tilde{x} \\
&= \int_{\frac{1}{2}-a}^{\frac{1}{2}} (y_n(\tilde{x}) - y_{n+1}(\tilde{x}))\mathrm{d}\tilde{x} + \int_{\frac{1}{2}-a}^{\frac{1}{2}} ((\beta - y_n(\tilde{x})) - (\beta - y_{n+1}(\tilde{x})))\mathrm{d}\tilde{x} \\
&= \int_{\frac{1}{2}-a}^{\frac{1}{2}} (y_n(\tilde{x}) - y_{n+1}(\tilde{x}))\mathrm{d}\tilde{x} + \int_{\frac{1}{2}-a}^{\frac{1}{2}} (y_{n+1}(\tilde{x}) - y_n(\tilde{x}))\mathrm{d}\tilde{x} \\
&= 0
\end{aligned} \tag{A6}
$$

Thus, it suffices to consider the case where $x \in \left[0, \frac{1}{2}\right]$.

For all $i \in \{1, 2, \ldots, n-1\}$, note that when $\tilde{x} \in \left(F^{-1}\left(\frac{i}{n+1}\right), F^{-1}\left(\frac{i}{n}\right)\right)$, we have:

$$
y_n(\tilde{x}) - y_{n+1}(\tilde{x}) = \beta \cdot \left(\frac{i-1}{n-1} - \frac{i}{n}\right) < 0 \tag{A7}
$$

And when $\tilde{x} \in \left(F^{-1}\left(\frac{i}{n}\right), F^{-1}\left(\frac{i+1}{n+1}\right)\right)$, we have:

$$
y_n(\tilde{x}) - y_{n+1}(\tilde{x}) = \beta \cdot \left(\frac{i}{n-1} - \frac{i}{n}\right) > 0 \tag{A8}
$$

Thus, it suffices to show that for all $i \in \{1, 2, \ldots, n-1\}$ and $F^{-1}\left(\frac{i+1}{n+1}\right) \leq \frac{1}{2}$, we have:

$$
\int_{F^{-1}\left(\frac{i}{n+1}\right)}^{F^{-1}\left(\frac{i}{n}\right)} (y_n(\tilde{x}) - y_{n+1}(\tilde{x}))\mathrm{d}\tilde{x} + \int_{F^{-1}\left(\frac{i}{n}\right)}^{F^{-1}\left(\frac{i+1}{n+1}\right)} (y_n(\tilde{x}) - y_{n+1}(\tilde{x}))\mathrm{d}\tilde{x} \leq 0 \tag{A9}
$$

Rewrite the left-hand side of Equation (A9) yields:

$$\begin{aligned}
LHS &= \left(F^{-1}\left(\frac{i}{n}\right) - F^{-1}\left(\frac{i}{n+1}\right)\right)\beta\left(\frac{i-1}{n-1} - \frac{i}{n}\right) \\
&\qquad + \left(F^{-1}\left(\frac{i+1}{n+1}\right) - F^{-1}\left(\frac{i}{n}\right)\right)\beta\left(\frac{i}{n-1} - \frac{i}{n}\right) \\
&= \frac{\beta}{n(n-1)}\left(iF^{-1}\left(\frac{i+1}{n+1}\right) + (n-i)F^{-1}\left(\frac{i}{n+1}\right) - nF^{-1}\left(\frac{i}{n}\right)\right) \\
&= \frac{\beta}{n-1}\left(\frac{i}{n}F^{-1}\left(\frac{i+1}{n+1}\right) + \left(1-\frac{i}{n}\right)F^{-1}\left(\frac{i}{n+1}\right)\right. \\
&\qquad \left. - F^{-1}\left(\frac{i}{n}\cdot\frac{i+1}{n+1} + \left(1-\frac{i}{n}\right)\cdot\frac{i}{n+1}\right)\right) \\
&\leq 0 = RHS
\end{aligned} \tag{A10}$$

The last inequality in Equation (A9) is true from the assumption that $F''(x) = f'(x) \geq 0$ holds for any $x \in \left[0, \frac{1}{2}\right]$.

Q.E.D

Analysis of CFPS Data

The China Family Panel Studies (CFPS)[13], initiated in 2010, is a nationally representative, biennial longitudinal survey of Chinese communities, families, and individuals. In its 2010 baseline survey, the CFPS successfully interviewed nearly 15,000 families and approximately 30,000 individuals within these families. The sample was drawn using a scientific stratification method, ensuring diversity in geographical distribution and richness in social contexts. CFPS collects comprehensive data across multiple domains, including education, economics, health, and family dynamics. As of now, the CFPS data has been updated through 2022.

This study focuses on the 2017 Gaokao reform implemented in two pivotal regions, Shanghai and Zhejiang, to examine how the shift from absolute grading to relative grading in selective subjects affects student effort. Specifically, we treat the reform as a quasi-natural experiment and employ the Difference-in-Differences (DiD) method to identify its impact on high school students' effort choices. Because a mainstream of Gaokao reform in other provinces was implemented in year 2021, we restrict our analysis to CFPS data up to 2020.

We estimate the following equation using ordinary least squares (OLS):

$$y_{it} = \beta_0 + \beta_1 \times reform_{it} + \beta_2 \times post_{it} + \beta_3 \times (reform_{it} \times post_{it}) + \gamma \boldsymbol{C} + \varepsilon, \tag{B1}$$

where y can be the study time measured in hours per week (*week*), per weekday (*weekday*), or per weekend (*weekend*). The variable $reform$ is dummy variable indicating if the obser-

[13] The website for CFPS is available at https://www.isss.pku.edu.cn/cfps/en/, providing a more detailed introduction to the CFPS.

vation is from a region affected by the 2017 Gaokao reform, and *post* is a dummy indicating if the observation is from a year following the reform. C includes control variables such as Gaokao year, sleep time, self-reported pressure, school ranking, family support, and others. The coefficient β_3 estimates the impact of the relative grading system on student effort.

TABLE B1— IMPACT OF THE 2017 GAOKAO REFORM ON STUDY TIME

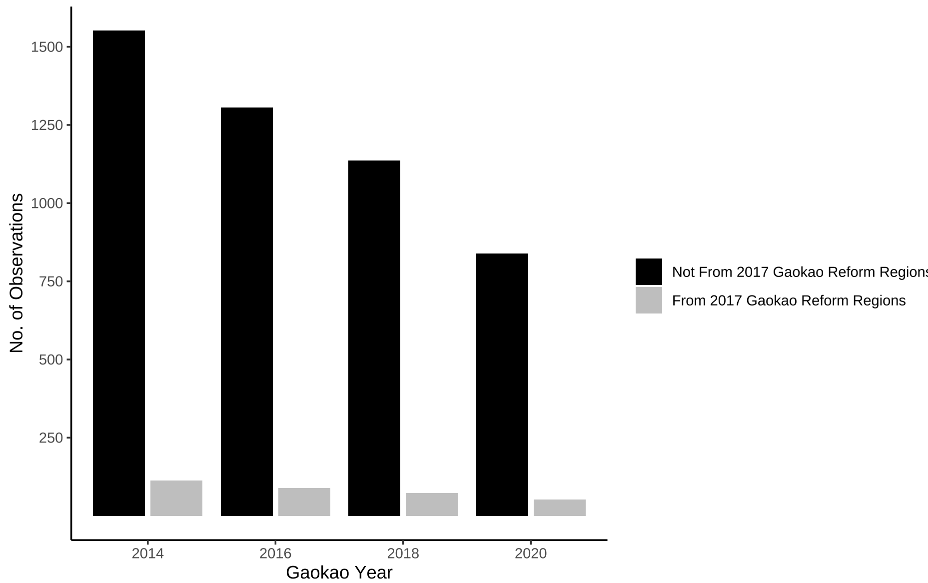
Dependent Variables:	*week*	*weekday*	*weekend*
	(1)	(2)	(3)
$reform \times post$	-6.269	-1.066	-0.469
	(11.240)	(1.666)	(1.706)
Control Variables	Yes	Yes	Yes
Observations	1,697	1,697	1,697

Notes: This table reports the regression results estimating the impact of the 2017 Gaokao reform on students' study time using CFPS data. The dependent variables are weekly study time (*week*), study time per weekday (*weekday*), and study time per weekend (*weekend*), measured in hours. The key variable of interest is $reform \times post$, which captures the effect of the 2017 Gaokao reform on student study time. $reform$ is a dummy indicating observations from regions affected by the reform, and *post* is a dummy indicating observations from a year after the reform. Control variables include year factors, sleep time, self-reported pressure, school ranking, family support, and others. Heteroskedasticity-robust standard errors are reported in parentheses.

The regression results in Table B1 align with the model's predictions. The coefficients of $reform \times post$, the estimate of the 2017 Gaokao reform on student effort, indicate that the reform leads to an average decrease of 4.36 hours in students' weekly study time. Specifically, the reform reduces study time by 0.604 hours per weekday and 0.670 hours per weekend. While these coefficients are not statistically significant, their magnitudes are economically meaningful. The lack of statistical significance can largely be attributed to the limited sample size of students affected by the Gaokao reform.

Although the CFPS theoretically tracks tens of thousands of families biennially, only 3% of the sample consists of high school students. Furthermore, since the 2017 Gaokao reform was implemented in only two pivotal regions, there is a significant disparity in sample size between reform and non-reform regions, as illustrated in Figure B1. This increases the variability of the estimates. Despite these limitations, the results are suggestive of the answer to our research question, indicating that the Gaokao reform reduces student effort.

FIGURE B1. SAMPLE SIZES OF HIGH SCHOOL STUDENTS IN 2017 GAOKAO REFORM AND NON-REFORM REGIONS



APPENDIX FOR QUESTIONNAIRE SURVEY

C1. Contents of the Questionnaire Survey

Survey on High School Student Study Mode for the Gaokao[14]

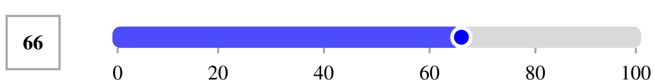
Thank you for participating in this survey!

This survey asks about aspects related to your subject choices and preparation for the Gaokao, such as your performances in different subjects and your time allocation on them. We will not ask for sensitive information such as specific scores or rankings.

Completing this survey is expected to take approximately **5 minutes**.

During the survey, you may encounter questions involving sliders:

[Sample Slider Question] How would you rate your proficiency in a given subject? / What is the proportion (%) of time you allocate to a given activity?



You may drag this slider to familiarize yourself with it.

For such slider questions, you can either directly input a number on the left or drag the slider to the desired position.

Click “Next Page” to begin.

[Page Break]

[14]This survey was designed and implemented using the Wenjuanxing platform at https://www.wjx.cn.

1) Which province did (or will) you take the Gaokao in? [Single Choice]

   Options: Shanghai, Zhejiang, Beijing, Tianjin, Shandong, Hainan, Hebei, Liaoning, Jiangsu, Fujian, Hubei, Hunan, Guangdong, Chongqing, Heilongjiang, Jilin, Anhui, Jiangxi, Guangxi, Guizhou, Gansu, Shanxi, Inner Mongolia, Shaanxi, Ningxia, Qinghai, Sichuan, Yunnan, Henan, Xinjiang, Tibet, Other (e.g., Hong Kong, Macao, Taiwan, overseas).

2) Which year did (or will) you take the Gaokao? [Single Choice]

   Options: 2015 or earlier, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026 or later.

3) Select your Gaokao mode[15]

   (3-1) Please determine whether the following statement is true: Your Gaokao policy follows the traditional Gaokao mode (in which you have to choose either the arts or science bundle). [Single Choice]

   Options: True, False.

   (3-2) Please determine whether the following statement is true: Your Gaokao policy follows the "3+3" mode[16] (If "False" in (3-1)). [Single Choice]

   Options: True, False.

   (3-3) Please determine whether the following statement is true: Your Gaokao policy follows the 3+1+2 mode[17] (If "False" in (3-1)) . [Single Choice]

   Options: True, False.

4) Subject Selection[18]

   (4-1) What was your subject stream selection? [Single Choice]

   Options: Arts, Science.

   (4-2) What were your selected subjects? [Multiple Choice, Choose 3]

   Options: History, Geography, Politics, Physics, Chemistry, Biology, (Technology).

[15] If it can be deduced from the Gaokao year and Gaokao province the respondent gives in the previous two questions that this province has implemented Gaokao reform at her Gaokao year, then (3-1) is displayed. If the province is going to implement "3+3" mode after Gaokao reform that had not been implemented at her Gaokao year, then (3-2) is displayed. If the province is going to implement "3+3" mode after Gaokao reform that had not been implemented at her Gaokao year, then (3-3) is displayed. These guarantees that any valid answer for this question must be "False."
Based on the respondent's reported province and year, the system determines the Gaokao policy in place at that time. Depending on the policy, one of the above questions is displayed, ensuring attentive respondents select "False" as the answer.

[16] "3+3" mode allows students to freely choose the three selective subjects out of six as they wish. The grading system under "3+3" mode is relative grading whose score difference between adjacent grades is 3. This corresponds to the *coarse* relative grading system.

[17] "3+1+2" mode requires students to choose one from physics and history, and choose 2 subjects from chemistry, biology, geology and politics. The grading system under "3+3" mode is relative grading whose score difference between adjacent grades is 1. This corresponds to the *granular* relative grading system.

[18] If the respondent selects "True" for (3-1) or "False" for (3-2) or (3-3), (4-1) is shown; otherwise, (4-2) is displayed. The option "Technology" is only available for respondents from Zhejiang province.

5) During the school days in your senior year of high school, how much time did you spend on average each day preparing for the Gaokao? ______hours ______minutes.

   This includes time spent in class, completing assignments, and additional practice.

6) During the weekends[19] in your senior year of high school, how much time did you spend on average each day preparing for the Gaokao? ______hours ______minutes.

7) Of the total study time mentioned above: The proportion spent on core subjects (Chinese, Mathematics, and Foreign Language) was approximately ______%; The proportion spent on selective subjects was approximately ______%. Please ensure that the two percentages add up to 100%.

The following questions will assess your proficiency and time allocation across six Gaokao subjects[20]. A good standard for your proficiency in each subject can be the percentile of this subject's score among the school in the most recent exam.

8) Regarding Chinese [Input a number between 0 and 100]

   How proficient do you think you are in Chinese? ______

   What proportion (%) of your study time is allocated to Chinese? ______

9) Regarding Mathematics [Input a number between 0 and 100] How proficient do you think you are in Mathematics? ______

   What proportion (%) of your study time is allocated to Mathematics? ______

10) Regarding Foreign Language [Input a number between 0 and 100]

    How proficient do you think you are in Foreign Language? ______

    What proportion (%) of your study time is allocated to Foreign Language? ______

11) Regarding History [Input a number between 0 and 100]

    How proficient do you think you are in History? ______

    What proportion (%) of your study time is allocated to History? ______

12) Regarding Geography [Input a number between 0 and 100]

    How proficient do you think you are in Geography? ______

    What proportion (%) of your study time is allocated to Geography? ______

13) Regarding Politics [Input a number between 0 and 100]

    How proficient do you think you are in Politics? ______

    What proportion (%) of your study time is allocated to Politics? ______

[19] Specifically, this refers to days that students do not need to go to school. It may not necessarily be the weekends, because some school have classes in weekends in practice. On the other hand, it includes holidays, winter or summer vacations, etc.

[20] Questions 11 to 17 will selectively appear only if the respondent indicates they have selected the corresponding subjects.

14) Regarding Physics [Input a number between 0 and 100]
    How proficient do you think you are in Physics? ______
    What proportion (%) of your study time is allocated to Physics? ______

15) Regarding Chemistry [Input a number between 0 and 100]
    How proficient do you think you are in Chemistry? ______
    What proportion (%) of your study time is allocated to Chemistry? ______

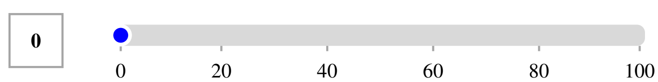
16) Regarding Biology [Input a number between 0 and 100]
    How proficient do you think you are in Biology? ______
    What proportion (%) of your study time is allocated to Biology? ______

17) Regarding Technology [Input a number between 0 and 100]
    How proficient do you think you are in Technology? ______
    What proportion (%) of your study time is allocated to Technology? ______

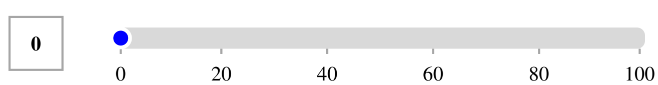
18) Do you believe you are significantly stronger in some subjects and weaker in others (i.e., a significant proficiency gap)? [Single Choice] Options: Yes, No.

19) What is your average daily sleep time (including naps)? ______hours ______minutes.

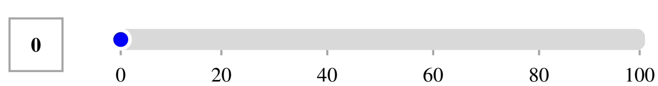
20) How often do you utilize fragmented time for studying (e.g., memorizing classical texts while waiting in line or learning vocabulary on the bus)? (0 indicates never using fragmented time to study, and 100 indicates always use fragmented time to study.)



21) To what extent do you think your family supported your Gaokao preparation in your senior year (e.g., better nutrients, extracurricular educational expenses, companionship)? (0 indicates no support, and 100 indicates fully support.)



22) How would you rate the amount of homework assigned by your school during your senior year? (0 indicates no homework, 100 indicates overwhelming amount of homework.)



23) During your senior year, how many exams did your school organize? ______times.

This includes exams such as weekly tests, monthly tests, midterms, finals that provide scores and rankings.

24) In your senior year, what percentage (%) of the students of your Gaokao year in your school do you believe you were academically stronger than? (0 indicates the weakest in the school, and 100 indicates the strongest.)

0

0 20 40 60 80 100

25) In your senior year, what percentage (%) of the students of your Gaokao year in your Gaokao province do you believe you were academically stronger than? (0 indicates the weakest in the province, and 100 indicates the strongest.)

0

0 20 40 60 80 100

26) What percentage (%) of schools in your Gaokao province do you think your school's overall Gaokao performance was stronger than? (0 indicates the weakest in the province, and 100 indicates the strongest.)

0

0 20 40 60 80 100

27) What was your level of academic pressure during your senior year? (0 indicates no pressure at all, and 100 indicates overwhelming pressure.)

0

0 20 40 60 80 100

28) In your senior year, what percentage (%) of senior students in your province do you think you face more pressure than? (0 indicates experiencing the least pressure in the province, and 100 indicates the greatest.)

0

0 20 40 60 80 100

29) Considering the difficulty of the Gaokao in your province, the number of competing students and their overall ability, and the number of available university seats, what percentage (%) of your peers nationwide do you think you faced more pressure for college admission than? (0 indicates experiencing the least pressure in the country, and 100 indicates the greatest.)

0

0 20 40 60 80 100

C2. Survey Sample

For a response to be considered valid, it must satisfy the following five data filtering criteria:

1) The respondent's Gaokao year and Gaokao province must match their reported Gaokao mode. (By design, a valid response corresponds to a "False" answer in the question asking the respondent to verify their Gaokao mode.)
2) The reported proportions of time spent on main subjects and selective subjects must sum to 100%.
3) The respondent is not from Hainan, Hong Kong, Macao, Taiwan, or other regions where the Gaokao mode or college admission process differs significantly from the majority.
4) The respondent's Gaokao year must not be 2015, 2016, 2024, or 2025, as there are extremely few respondents from these years. Since year factors are controlled in the analysis, including such samples would obscure the year effect.
5) The respondent's answers must be logically reasonable. For example, one cannot report studying for 24 hours per day.

Initially, we received 1,196 responses, of which 760 met these criteria. Figure C1 presents the distribution of respondents by Gaokao year and grading system in selective subjects. Figure C2 illustrates the geographical distribution of respondents across China.

FIGURE C1. DISTRIBUTION OF RESPONDENTS BY GAOKAO YEAR AND GRADING SYSTEM IN SELECTIVE SUBJECTS

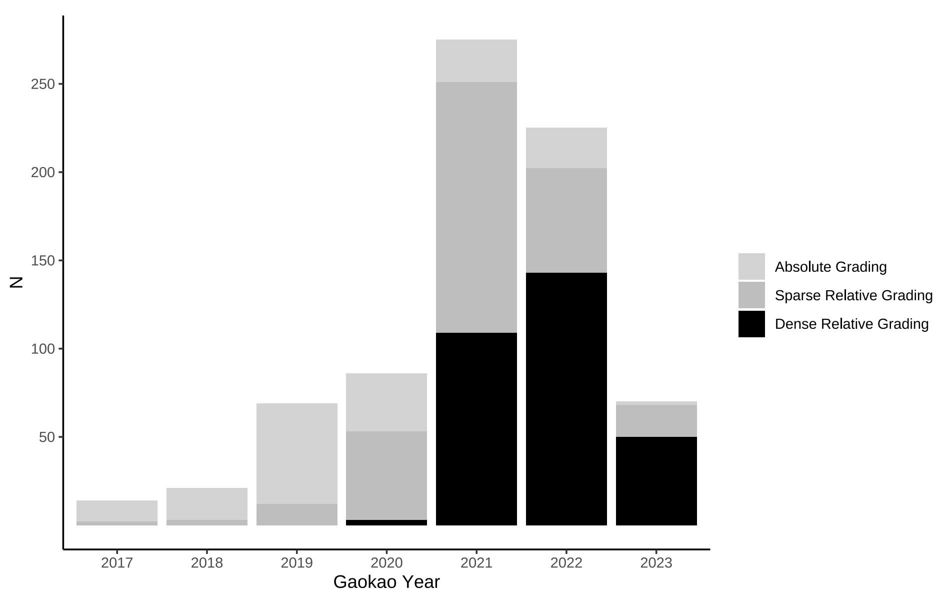
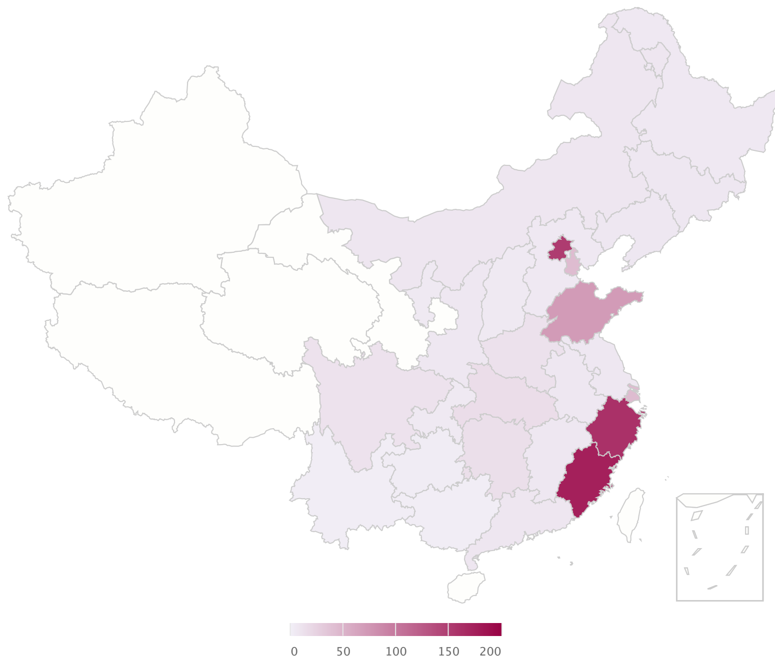


FIGURE C2. GEOGRAPHICAL DISTRIBUTION OF SURVEY RESPONDENTS ACROSS CHINA



C3. Additional Analysis

We have demonstrate through the main results (see Section III.A) and heterogeneity analysis (see Section III.B) that the Gaokao reform reduces student effort. Building on these findings, we further explore its broader implications by examining how the reform affects students' self-reported pressure levels and sleep time. Specifically, we set y in Equation (13) and (14) as self-reported pressure level (*pressure*, ranging from 0 to 100) and daily sleep time (*sleep*, measured in hours), and estimate these equations using ordinary least squares (OLS). The regression results are presented in Table C1.

For self-reported pressure levels, the regression results reveal that the Gaokao reform significantly alleviates student pressure by around 5.41% ($p = 0.031$). This finding highly correlates with, and strongly supports our earlier results that the Gaokao reform leads to reduced student effort. The implication is that the relative grading system not only successfully makes students physically spend less time studying, but also makes them mentally feel less stressful.

For sleep time, the results show that the Gaokao reform, as well as both types of relative grading systems, has no significant impact. This finding, however, may not be a bad thing. It suggests that the time saved from reduced effort in selective subjects is likely redistributed to other daytime extracurricular activities, such as sports or music, which contributes to students' well-being. Notably, while the coefficient for Gaokao year (*year*) is negative and statistically significant, its magnitude is negligible. This indicates that high

TABLE C1— IMPACT OF GAOKAO REFORM ON PRESSURE LEVEL AND SLEEP TIME

Dependent Variables:	*pressure*		*sleep*	
	(1)	(2)	(3)	(4)
reform	-5.409		0.055	
	(2.509)		(0.093)	
granular		-4.185		0.043
		(2.923)		(0.109)
coarse		-6.136		0.062
		(2.611)		(0.098)
year	1.926	1.712	-0.085	-0.083
	(0.789)	(0.821)	(0.031)	(0.032)
Control Variables	Yes	Yes	Yes	Yes
Observations	760	760	760	760

Notes: This table presents the regression results for the impact of the Gaokao reform on self-reported pressure levels and daily sleep time. The dependent variables are self-reported pressure levels (*pressure*, ranging from 0 to 100) in Columns 1–2, and sleep time per day (*sleep*, in hours) in Columns 3–4. The variable *reform* indicates whether the respondent experienced the Gaokao reform, while *granular* and *coarse* represent granular and coarse relative grading systems, respectively. Control variables include Gaokao year, student ranking, self-reported pressure, the degree of imbalance between subjects, and other relevant factors. Heteroskedasticity-robust standard errors are reported in parentheses.

school students have already allocated nearly all their available time to studying. Although they attempt to reduce sleep time to study more, this strategy appears to have reached its limit. Thus, engaging in extracurricular activities becomes crucial for maintaining students' mental health, which highlights the broader implications of this finding.

APPENDIX FOR EXPERIMENT

D1. Experiment Manual

Thank you for participating this experiment!

We are conducting a study to assess fundamental office skills. The experiment includes two tasks: *typing* and *mental arithmetic*. Your task is to test your typing speed and mental arithmetic skills. The experiment data and results will only be used for our research purpose. If you have any concerns, please feel free to contact the organizers.

To encourage you to perform at your best, we will additionally reward participants ranked in the top 2%–10% for their overall performance across both skills with **20** CNY per person, and those ranked in the top 2% with **100** CNY per person.

For both tasks, we take the best score for each subject, and you are allowed to make multiple attempts. It takes some time to get adapted to the specific contexts, so we encourage you to try multiple times.

Below is for Control group

Your overall score will be calculated as follows:

Overall Score = Best Typing Speed (cpm) + Best Mental Arithmetic Score (points)

(*Please answer this question for your understanding*) If a participant's best typing speed is 20 cpm and their best mental arithmetic score is 15 points, what is their overall score? ______

Note: We provide the correct answer of *35* after they submit the answer.

———————————— Below is for Dense group ————————————

Your overall score will be calculated as follows:

We will rank all participants' fastest typing speeds. Rankings will be divided into 11 categories, with the top 5%, 5%–15%, 15%–25%, ..., 85%–95%, and the bottom 5% receiving grades of A+, A, A-, B+, B, B-, C+, C, C-, D, and E, respectively. These grades correspond to scores of 20, 18, 16, ..., 2, and 0. This score will be added to your best mental arithmetic score (points) to determine your overall score. See the table below for details.

TABLE D1—GRADING RULE FOR GRANULAR GROUP

Rank	Grade	Score
Top 5%	A+	20
5%–15%	A	18
15%–25%	A-	16
25%–35%	B+	14
35%–45%	B	12
45%–55%	B-	10
55%–65%	C+	8
65%–75%	C	6
75%–85%	C-	4
85%–95%	D	2
Bottom 5%	E	0

(*Please answer this question for your understanding*) If a participant's best typing speed is 20 cpm, their best mental arithmetic score is 21 points, and their typing speed ranks in the 30% percentile among all participants, what is their overall score? ______

Note: We provide the correct answer of *35* after they submit the answer.

———————————— Below is for Coarse group ————————————

Your overall score will be calculated as follows:

We will rank all participants' fastest typing speeds. Rankings will be divided into 5 categories: Top 1/8, 1/8–3/8, 3/8–5/8, 5/8–7/8, and Bottom 1/8, which will receive grades of A, B, C, D, and E, respectively. These grades correspond to scores of 20, 15, 10, 5, and 0. This score will be added to your best mental arithmetic score (points) to determine your overall score. See the table below for details.

(*Please answer this question for your understanding*) If a participant's fastest typing speed is 25 cpm, their highest mental arithmetic score is 20 points, and their typing speed ranks in the 2/8 fraction (i.e., 1/4), what is their overall score? ______

Note: We provide the correct answer of *35* after they submit the answer.

————————————Below is the same for all groups————————————

TABLE D2—GRADING RULE FOR COARSE GROUP

Rank	Grade	Score
Top 1/8	A	20
1/8–3/8	B	15
3/8–5/8	C	10
5/8–7/8	D	5
Bottom 1/8	E	0

Please proceed to the following websites as instructed. If you encounter any technical or operational issues during the experiment, please feel free to contact the organizers.

1) Typing Test

   For the typing test, please use the website https://dazi.91xjr.com. Once you enter the site, you will see the following interface... (details omitted)[21].

2) Mental Arithmetic Test

   For the mental arithmetic test, please use the website https://www.preplounge.com/en/mental-math/add/1. Once you enter the site, you will see the following interface... (details omitted)[22]

D2. Additional Results

Table D3 presents the comparisons of invested time between the control and granular groups, as well as between the control and coarse groups. The results generally align with the patterns observed in Table 5. Notably, the effect is more pronounced in the comparisons between the control and granular groups. In contrast, the coarse group shows less distinction from the control group. These findings are consistent with the model predictions: the granular relative grading system differs more markedly from the absolute grading system, as it is more effective in reducing effort compared to the coarse relative grading system.

[21] The details in the manual include instructions for using the typing test website. As it is lengthy and not essential, it is omitted.

[22] The details in the manual include instructions for using the mental arithmetic test website. As it is lengthy and not essential, it is omitted.

TABLE D3—ADDITIONAL COMPARISONS OF INVESTED TIME: CONTROL VS. GRANULAR AND CONTROL VS. COARSE

	Panel A. Control and Granular Groups			
	Mean			
	Granular	Control	Difference	*p*-value
Time in Typing	6.111	9.000	-2.889	0.043
	(0.732)	(1.195)	(1.360)	
Time in Mental Arithmetic	4.806	6.265	-1.459	0.137
	(0.429)	(0.870)	(0.9705)	
Total Time	10.918	15.265	-4.347	0.034
	(1.012)	(1.742)	(2.015)	
Number of Subjects	39	42		

	Panel B. Control and Coarse Groups			
	Mean			
	Coarse	Control	Difference	*p*-value
Time in Typing	6.111	7.212	-1.788	0.278
	(0.732)	(1.195)	(1.635)	
Time in Mental Arithmetic	7.066	6.265	0.801	0.595
	(1.224)	(0.870)	(1.501)	
Total Time	14.278	15.265	-0.987	0.697
	(1.830)	(1.742)	(2.526)	
Number of Subjects	36	42		

Notes: This table presents comparisons of average time invested (in minutes) in typing, mental arithmetic, and total activities. Panel A reports comparisons between the granular group and the control group. Panel B compares the coarse group with the control group. The "Difference" column shows the mean differences between the groups. *p*-values are computed using a two-sample *t*-test, allowing for unequal variances.