Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Sentencing imbalance and the absence of reference points: the crime of exam substitution in China

Lifeng Li

lilifengjlu@163.com

Jilin University

Ting Song

Jilin University

Xue Xia

Northeast Normal University

Xiaoye Li

Jilin University

Article

Keywords:

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11183241/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Sentencing imbalance and the absence of reference points: the crime of exam substitution in China

Abstract: As China increasingly criminalizes minor wrongdoing, precise and proportionate sentencing has become a pressing concern, yet little empirical attention has been paid to micro-crimes that lack statutory reference points. The crime of taking an exam on behalf of another person is a case in point: it is punishable by criminal detention, public surveillance, or a fine but, unlike other exam-cheating offenses, provides for neither statutory sentencing circumstances nor an aggravated tier. Drawing on a nationwide sample of 1,882 defendants from the PKU Law database, this article combines regression models with multilevel variance decomposition across court levels. It finds that mitigating circumstances operate at the lenient end of sentencing—probation, a fine alone, or a shorter custodial term—and that sentencing imbalance is concentrated in basic-level courts, where the court level contributes the largest variance component (intraclass correlation coefficient $\approx$ 0.54, rising to 0.63 for probation). This micro-crime thus serves as a boundary case for the theory of jurisdictional uniformity, tracing the observed court-level imbalance to the absence of objective reference points.

I. Introduction

In China, examinations have long functioned as the principal mechanism for allocating scarce social resources—above all access to education and public office—through a meritocratic tradition rooted in the Confucian civil service examinations (Elman, 2000, Miles, 2016, Ho, 1962). Examinations are still widely regarded as the fairest means of evaluating and selecting talent (Jin and Ball, 2020, Liu, 2013). Cheating on an examination therefore subverts this mechanism: it confers an illicit competitive advantage, undermines the fairness of resource allocation, and erodes public trust in the selection system (Wei and Yan, 2018). To protect the integrity of "national examinations prescribed by law," Amendment IX to the Criminal Law (effective 1 November 2015) inserted Article 284a, creating three offenses: organizing examination cheating, illegally selling or providing examination questions or answers, and taking an examination on behalf of another person. In September 2019 the Supreme People's Court and the Supreme People's Procuratorate jointly issued *the Interpretation on Several Issues concerning the Application of Law in Handling Criminal Cases such as Organizing Examination Cheating* (hereinafter "the Interpretation"), which clarified the scope of the examinations covered and the standards for convicting and sentencing the other examination-related offenses.

Unlike the other examination-related offenses, the crime of taking an examination on behalf of another person—a micro-crime whose maximum statutory penalty is criminal detention—receives no specific guidance from Article 7 of the Interpretation, whether for defining the crime or for punishing it (Liang, 2021). This offense has consequently received little empirical scrutiny.

This article therefore asks how discretionary circumstances actually operate on the sentencing of this micro-crime and why similar cases are sentenced inconsistently—whether the imbalance reflects identifiable circumstances or the absence of objective sentencing reference points. Three testing paths are pursued. First, sentence-length regressions test the direction in which mitigating circumstances operate on custodial or freedom-restricting sentences. Second, multilevel variance decomposition of the intraclass correlation coefficient at three levels (province, city, court) in a properly nested random-intercept model tests the level to which sentencing differences are attributable. Third, subgroup

comparisons test the reference-point hypothesis—the correlation between the absence of reference points and inter-court differences. A normative premise of the third path should be stated: other things being equal, substitution in a national education examination warrants more severe punishment than substitution in, for example, a motor-vehicle driving examination. Because driving-examination substitution directly endangers public road safety, this ranking is contestable; this article therefore treats it as a contestable baseline and reports the empirical pattern without relying on it for the core findings.

This article proceeds as follows. Section II reviews the existing scholarship on the offense's definition, punishment, and judicial application; Section III develops the theoretical argument—the reference-point hypothesis and its bearing on the theory of jurisdictional uniformity. Section IV describes the sample and the analytic strategy; Section V reports the empirical findings on sentencing imbalance and the mechanism that explains it; and Section VI discusses the implications and policy suggestions. Throughout, this article argues that the imbalance reflects the absence of objective sentencing reference points rather than the discretionary circumstances themselves.

II. Literature review

The limited scholarship on this offense has concentrated on two controversies—its definition and its punishment.

As to its definition, some scholars hold that strengthening the management and reform of the examination system under regulations such as *the Measures for Handling Violations in National Education Examinations* would suffice, and that stringent administrative measures alone can deter substitution (Zhao and Zhao, 2016). Relatedly, criminalizing substitution is widely seen as a "front-loading" of punishment (Guo, 2024, Husak, 2011, Yankah, 2011). Because the offense requires no special circumstances, it is complete as soon as a person takes a prescribed national examination on another's behalf — conduct indistinguishable from the substitution that attracts administrative punishment—so no clear standard separates a mere administrative violation from a crime. Most existing research has therefore concentrated on the scope of the examinations covered, noting that "national examinations prescribed by law" suffer from inconsistent legal bases, unclear modes of stipulation, doubtful legality of the organizing institutions, and an ambiguous catch-all clause (Zhang, 2025); considerable disputes also remain over the forms of the offense and its boundary with organizing examination cheating (Wang and Xu, 2019).

The controversy over punishment is greater. The statutory penalty is low, comprising only criminal detention or public surveillance, with a fine imposed concurrently or alone, and this configuration is widely thought to suffer from two structural defects. First, it lacks an aggravated tier for serious circumstances—"the criminalization of exam-substitution conduct is subject to no limitation on circumstances, nor is there any circumstance warranting an aggravation of the statutory penalty" (Song, 2020)—in sharp contrast to organizing examination cheating, which provides for such a tier. Second, unlike dangerous driving (also punishable only by criminal detention), the harm of substitution varies in degree and ought to be reflected in sentencing; yet the statutory range is narrow and lacks internal gradations, there are no targeted statutory sentencing circumstances, and the Interpretation enumerates none for this offense. Some judges do invoke general and discretionary circumstances—the importance of the examination, subjective culpability, the number of substitutions, the forms of the crime, and the form of organization (Shi, 2018)—but it is argued that "imposing criminal detention on professional substitute exam-takers cannot achieve proportionality between crime and punishment"

(Gui, 2017), which further undermines the legitimacy of the configuration.

Academic assessments of this configuration fall into three groups. Some affirm its leniency: the penalty is light and conforms to the principle of proportionality between crime, criminal responsibility, and punishment (Rui, 2015), and making criminal detention the maximum embodies the legislative trend toward lighter punishment of minor offenses (Dun and Huang, 2017). Others identify structural defects: the maximum penalty of criminal detention makes this a micro-crime whose lenient and severe ranges should be graded according to the punishability of the examination (He and Zhang, 2024), and the narrow statutory range produces inconsistent judgments in similar cases, so sentencing standards should be refined and the scale unified through guiding cases (Gui, 2021). Still others regard the configuration as imbalanced: treating professional and occasional substitute exam-takers alike violates the same principle of proportionality, so a heavier penalty should attach to professionals (Tian, 2017); imposing the same penalty on the substitute and on the person for whom the exam is taken fails to reflect their differing degrees of illegality (Yin and Yin, 2017); and circumstance limitations should be added to exclude conduct of minor harm from criminality (Huang, 2017).

These normative defects are transmitted to the judicial level, producing "inconsistent judgments in similar cases," with the gap in fine amounts reaching as much as tenfold (Gui, 2021). Equal punishment within joint offenses is likewise hard to achieve (Li, 2018), and regional differences persist, with sentencing in some regions displaying an indiscriminate character (Wang, 2018). The imbalance also appears in the direction in which sentencing circumstances operate: substituting in a national education examination, which should aggravate, is instead punished significantly more leniently; repentance, which should mitigate, correlates significantly with heavier sentences; and the directions for substitution in a professional-technical qualification examination and for returning illegal gains, though likewise contrary to expectation, are not significant (Yang, 2017). These anomalies suggest that the application of discretionary circumstances lacks unified rules, and that the imbalance is a pervasive structural problem rather than an isolated deviation in discretion.

III. Theoretical framework

The argument speaks to four broader literatures. First, the reference-point hypothesis applies reference-dependence and anchoring (Tversky and Kahneman, 1974, Tversky and Kahneman, 1981, Kahneman and Tversky, 1982) to the hierarchical structure of sentencing discretion: just as an unanchored numerical judgment shows larger dispersion, a court lacking objective numerical benchmarks shows larger inter-court variance. Second, the finding of inter-court disparity parallels the long-standing evidence of inter-judge disparity that motivated the United States Sentencing Guidelines, and supplying reference points is the functional analogue of a guideline's "anchor" rather than of rigid mandatory ranges(Mustard, 2001, Ulmer and Light, 2011, Ulmer and Johnson, 2017). Third, the policy suggestions in the concluding section connect with China's sentencing-standardization reform and its emphasis on sentencing guidelines, guiding cases, and similar-case retrieval (Lin et al., 2022, Hou and Truex, 2022, Liebman et al., 2020). Fourth, and most important for the present contribution, the finding bears directly on the recently advanced theory of jurisdictional uniformity (JUT)(Xiong et al., 2025), which synthesizes organizational and court-community accounts of sentencing (Dixon, 1995, Ulmer, 2019): scholars report negligible inter-court differences for serious offenses and generalize this into the claim that a uniform criminal law, sentencing guidelines, and the political mechanism jointly guarantee nationwide sentencing equilibrium, with intraclass correlation coefficients below 0.05 at every court tier for rape (Xiong et al., 2025). The present crime offers a boundary-condition test of JUT: within the

very same uniform legal and political system, a micro-crime whose statutory penalty has no internal gradation, no aggravated tier, and no quantified amount or head-count threshold exhibits a court-level ICC an order of magnitude above the JUT threshold. Legal and political uniformity, though necessary, is not sufficient for sentencing equilibrium, which holds only where objective reference points exist.

IV. Samples and Research Methods

(A) Data Sources

The sample was drawn from the PKU Law database by retrieving judgments for the crime of taking an exam on behalf of another person, and was deduplicated by two rules: each case number was recorded once, and where first- and second-instance judgments belonged to the same case, only the final-instance judgment was retained. This yielded 1,017 judgments and 2,639 defendants spanning 2016–2025. Among the 1,890 defendants whose charge-name field recorded only this crime, eight were excluded after case-by-case verification against the judgment notes: their probation had been revoked and they had in fact received combined punishment for multiple offences, so the recorded sentence was an aggregate rather than the penalty for this crime. Two of the eight were identified only by their sentence values (3.0 years; 1.25 years with a ¥408,000 fine), which exceed the statutory maximum and therefore reflect combined punishment even though the notes do not use that phrase. Two further defendants sentenced to two years of public surveillance (converted to 1.0 year), the lawful maximum, were retained. The analytic sample for this crime in its pure form is thus 1,882 defendants, reported as the sample size throughout. Descriptive statistics and regressions use all 1,882 defendants, while the hierarchical models drop the two defendants for whom no court could be attributed, leaving N = 1,880.

(B) Variable Coding

Four dependent variables are used: sentence length, fine, probation, and probation period. Sentence length is measured in years: criminal detention is converted by its number of months, public surveillance at two days to one day, and a sentence consisting solely of a fine is coded as 0. Where a defendant received combined punishment, only the penalty for this crime is recorded. This conversion is the only one with a statutory basis, but the resulting "sentence length" conflates length and severity: public surveillance, a non-custodial sanction legally lighter than detention, can convert to a value (up to 1.0 year) exceeding the maximum detention (0.5 year). This inversion affects only two of the 919 defendants who received a custodial or freedom-restricting sentence, and it does not drive any finding (Section V(C)). There are 21 independent variables in four categories—examination category, counterpart-offender status, criminal-fact circumstances, and statutory and discretionary circumstances—all coded as binary (1 if present) except illegal gains and the number of repeated substitutions. Table 1 shows the operational definitions.

Table 1: Operational definitions of the 21 independent variables.

Variable	Type	Assignment (Coding) Rule
National education examination	Examination category	Gaokao, postgraduate entrance, adult college entrance, self-study, etc. = 1
Civil service examination	Examination category	Civil service recruitment examination = 1
Professional-technical qualification examination	Examination category	Legal professional qualification, teacher qualification, etc. = 1
Other examinations	Reference category	Motor vehicle driving examination, etc. (reference; not entered into the regression)
Person for whom the exam is taken	Counterpart-offender status	Person for whom the exam is taken = 1; substitute exam-taker = 0
Number of repeated cheating	Criminal fact	Number of times of repeated substitution; a single instance = 0; "multiple times" without a number = 0
Illegal gains	Criminal fact	Amount (yuan), counted in ten-thousand yuan in the regression
Attempted crime / accessory	Criminal fact	1 if found
Minor role	Criminal fact	1 if the defendant's role in the joint offense was minor and could not be clearly distinguished as principal or accessory
No gain or attempted gain	Criminal fact	1 if the defendant obtained no illegal gain, or the gain was only attempted
Truthful confession / voluntary surrender / guilty plea / repentance / meritorious service / return of illegal gains	Circumstance	1 if present, otherwise 0
Prior record / first offender / occasional offender	Circumstance	1 if present, otherwise 0
Whether the fine has been paid / occupational prohibition	Circumstance	1 if present, otherwise 0

(C) Analysis Methods

Ordinary least squares (OLS) is used for sentence length, the log-fine, and the probation period, and binary logit for probation, with coefficients exponentiated to odds ratios. Because 51.2% of sentence-length observations are zero (fine only), a single OLS is misspecified for this zero-inflated, right-skewed outcome (Lambert, 1992): it conflates a binary margin (whether any custodial or freedom-restricting sentence is imposed) with a continuous margin (its length), so that it appears mitigating circumstances do not shorten sentences. The sentence-length analysis therefore rests on a two-part (hurdle) model (Cragg, 1971, Mullahy, 1986, Duan et al., 1983): a logit for whether any custodial or freedom-restricting sentence is imposed, and an OLS for its length conditional on such a sentence (N = 919). In this conditional equation, voluntary surrender and truthful confession do shorten the custodial or freedom-restricting sentence. The pooled OLS is retained only for comparability with the existing literature. All effect sizes are also converted to days, because at a scale where the median sentence is zero and detention is capped at six months, statistical significance is not practical significance (Cohen, 1988).

The level to which sentencing differences are attributable is estimated with a three-level nested random-intercept model (defendants within courts within cities within provinces) (Raudenbush and Bryk, 2002, Snijders and Bosker, 2012), fitted by restricted maximum likelihood (REML). The intraclass correlation coefficient (ICC) at each level is the ratio of that level's variance component to the total. The nested model replaces the earlier practice of computing three separate one-way ICCs and subtracting them, which is unsound because the one-way models have different denominators, are not

estimated simultaneously, and materially understated the court level while overstating the city level. It is preferred over separate one-way ANOVAs because the data form a strictly nested hierarchy (each court belongs to exactly one city, each city to exactly one province). REML is preferred over maximum likelihood because it yields approximately unbiased variance components; here the two estimators agree to two decimal places. The 0.05 ICC threshold (LeBreton and Senter, 2008) is adopted as the prima facie criterion for a group effect. Delta-method 95% confidence intervals are reported for the nested REML estimates, and cluster-bootstrap intervals for the subgroup ICCs. Because the ICC is a ratio of variances, its magnitude depends on the scale of the outcome, so the decomposition was repeated for five outcomes—sentence length, any custodial or freedom-restricting sentence, probation, the log-fine, and the probation period—as well as for a severity-ordered coding (fine < surveillance < detention) and with the two two-year-surveillance cases removed. The court level is the largest component for every custodial and probation outcome, and for the fine the individual level is largest.

To diagnose collinearity and heteroscedasticity, the variance inflation factor (Belsley et al., 1980) and HC1 heteroscedasticity-robust standard errors (White, 1980, MacKinnon and White, 1985) are reported, together with F-tests and likelihood-ratio χ^2 tests for overall model significance. Coding used the original judgment text as the sole basis and was completed in 77 batches with three rounds of retrospective correction; coding reliability is assessed in Section V(E).

All hypothesis tests are two-tailed at $\alpha = 0.05$, and exact p-values are reported (values below 0.001 in scientific notation). Given the large sample (N = 1,882), inference relies on the central limit theorem rather than on residual normality, and the fine amount is analysed in logarithms to reduce skewness. Because the models include 21 covariates across several outcomes, p-values are reported for every coefficient and interpreted with explicit attention to the multiple-comparison problem; no formal family-wise alpha adjustment is applied, because the analysis is exploratory and the substantive conclusions rest on the multilevel variance decomposition and the subgroup evidence rather than on any single coefficient.

V. An Empirical Examination of Sentencing Imbalance

(A) Sentencing Profile

The punishment mix is highly concentrated: 51.2% of defendants (963) are fined alone and 48.8% (919) receive criminal detention or public surveillance. By examination category, motor-vehicle driving examinations account for 51.8% (975 defendants), national education examinations 33.3% (627), professional-technical qualification examinations 13.5% (254), and civil service examinations 1.4% (26)—a concentration in driving examinations that sits awkwardly with a legislative intent aimed at national education examinations. Caseloads rose from 88 defendants in 2016 to a peak of 528 in 2018, then declined to just over ten cases a year by 2022–2025.

This post-2022 collapse bears on representativeness. Because the sample is drawn from published judgments, it captures only cases that were prosecuted and disclosed; minor offences of this kind are disproportionately diverted through relative non-prosecution, administrative handling, or non-publication. The published sample therefore plausibly over-represents the more serious cases, so the 51.2% fine-only share and the 38.4% probation rate understate the population's leniency, and the imbalance estimates are more likely a lower bound than an overstatement. This direction of selection bias cannot, however, be verified from the data at hand. Table 2 reports the descriptive statistics.

Table 2: Descriptive statistics of the analytic sample (N = 1,882).

Indicator	Value
Sentence length (years) - Mean (SD)	0.103 (0.126)
Sentence length (years) - Median	0
Sentence length (years) - IQR (P25-P75)	0-0.17
Sentence length (years) - Maximum	1
Fine (yuan) - Mean (SD)	5,204 (4,763)
Fine (yuan) - Median	5,000
Fine (yuan) - IQR (P25-P75)	2,000-6,000
Fine (yuan) - Maximum	40,000
Fine only	963 (51.2%)
Detention / surveillance	919 (48.8%)
Probation	723 (38.4%)
Probation period (months) - Mean (SD)	4.24 (2.04)
Probation period (months) - Median	4
Probation period (months) - IQR (P25-P75)	3-6

(B) Anomalies in the Operation of Mitigating Circumstances

The pooled OLS regression on sentence length is significant overall ($R^2 = 0.113$, adjusted $R^2 \approx 0.103$, $F(21, 1860) = 11.32$, $p < 1.0\times10^{-4}$) yet explains only about 11% of the variance. After excluding combined-punishment cases, prior record is no longer significant ($p > 0.05$), and the largest sentence-increasing effects shift to low-frequency circumstances—occupational prohibition (+0.212 years) and civil service examination (+0.071 years). No single discretionary circumstance can thus stably and substantially lengthen the sentence. The two variables that should mitigate—repentance and accessory status—carry the wrong sign (the accessory sample has only 20 defendants), and national education examination is significantly negative (-0.028, $p = 1.2\times10^{-5}$). Converted to days, the significant pooled coefficients are occupational prohibition +78 days, civil service examination +26, accessory +27, illegal gains +4 days per ¥10,000, and repeated substitution +1 day per occasion. Conditional on custody, voluntary surrender (-0.049 years, $t = -3.95$, $p = 8.4\times10^{-5}$, ≈ -18 days) and truthful confession (-0.025, $t = -2.14$, $p = 0.032$, ≈ -9 days) significantly shorten the sentence, while accessory lengthens it (+37 days). On a scale whose median sentence is zero these effects are statistically detectable but substantively modest; the policy magnitudes proposed in Section VI are therefore normative choices anchored to, rather than derived from, these coefficients. Table 3 reports the complete pooled regression results; Tables 4 and 5 report the two parts of the hurdle model. The first part shows sharply higher odds of custody for civil service examinations (OR 2.97, $p = 0.018$) and lower odds for national education examinations (OR 0.58), first offenders (OR 0.53), and defendants whose fine has been paid (OR 0.52).

Table 3: Pooled OLS regression of sentence length (years).

Independent variable	B	SE	t	p	~ Days
Intercept	0.1311	0.0098	13.35	<1.0e-4	-
Occupational prohibition	0.2124	0.0496	4.29	1.9e-5	78
Accessory	0.0747	0.0275	2.72	0.007	27
Civil service examination	0.0713	0.0239	2.98	0.003	26
Repentance	0.0229	0.0077	2.97	0.003	8
Illegal gains (ten-thousand yuan)	0.0097	0.0023	4.29	1.9e-5	+4 / 10,000 yuan
Number of repeated cheating	0.0031	0.0007	4.36	1.4e-5	+1 / occasion
Guilty plea	0.0138	0.0058	2.38	0.017	5
Person for whom the exam is taken	-0.0141	0.0063	-2.25	0.024	-5
First offender	-0.0291	0.0102	-2.85	0.004	-11
Occasional offender	-0.0246	0.0134	-1.84	0.066	-
National education examination	-0.0283	0.0064	-4.39	1.2e-5	-10
Whether the fine has been paid	-0.0294	0.0057	-5.19	2.3e-7	-11
Attempted crime	-0.0302	0.0129	-2.35	0.019	-11

Independent variable	B	SE	t	p	~ Days
Voluntary surrender	-0.0186	0.0087	-2.14	0.032	-7
Truthful confession	0.0052	0.0084	0.62	0.537	-
Prior record	0.0149	0.0166	0.9	0.37	-
Meritorious service	-0.048	0.0342	-1.4	0.161	-
Minor role	0.0208	0.0539	0.39	0.699	-
No gain or attempted gain	0.0131	0.0365	0.36	0.72	-
Return of illegal gains	0.0122	0.0113	1.08	0.279	-
Professional-technical qualification exam	-0.0163	0.0087	-1.87	0.061	-

Table 4: Binary logit of any custodial or freedom-restricting sentence (first part of the two-part model).

Independent variable	OR	z	p
Civil service examination	2.966	2.36	0.018
Minor role	1.602	0.51	0.612
Repentance	1.58	3.27	0.001
Prior record	1.496	1.36	0.173
Accessory	1.472	0.8	0.423
Truthful confession	1.451	2.53	0.011
Illegal gains (ten-thousand yuan)	1.366	1.69	0.091
Guilty plea	1.324	2.76	0.006
Number of repeated cheating	1.13	2.79	0.005
Voluntary surrender	1.106	0.66	0.51
Return of illegal gains	1.073	0.33	0.739
No gain or attempted gain	1.05	0.08	0.938
Person for whom the exam is taken	0.874	-1.2	0.229
Attempted crime	0.776	-1.12	0.264
Professional-technical qualification examination	0.689	-2.42	0.016
National education examination	0.575	-4.83	1.3e-6
Occasional offender	0.545	-2.38	0.017
First offender	0.529	-3.47	5.1e-4
Whether the fine has been paid	0.523	-6.47	9.8e-11
Meritorious service	0.497	-1.08	0.279
Occupational prohibition	omitted		

Table 5: OLS regression of sentence length conditional on custody (second part of the two-part model, N = 919).

Independent variable	B	SE	t	p
Intercept	0.2432	0.0136	17.87	<1.0e-4
Occupational prohibition	0.0936	0.0363	2.58	0.01
Accessory	0.1021	0.0307	3.33	9.2e-4
Civil service examination	0.0449	0.0218	2.06	0.04
Repentance	0.0041	0.0082	0.5	0.617
Illegal gains (ten-thousand yuan)	0.0047	0.0017	2.77	0.006
Number of repeated cheating	0.0018	0.0004	4.53	6.8e-6
Guilty plea	0.0016	0.0065	0.24	0.811
Person for whom the exam is taken	-0.014	0.0078	-1.81	0.071
First offender	-0.0002	0.012	-0.01	0.99
Occasional offender	-0.0034	0.0165	-0.2	0.838
National education examination	0.0015	0.0075	0.2	0.841
Whether the fine has been paid	-0.0004	0.0065	-0.06	0.953
Attempted crime	-0.0344	0.0163	-2.11	0.035
Voluntary surrender	-0.0486	0.0123	-3.95	8.4e-5
Truthful confession	-0.0254	0.0118	-2.14	0.032
Prior record	-0.0015	0.0145	-0.11	0.916
Meritorious service	-0.0094	0.0518	-0.18	0.856
Minor role	0.0087	0.0287	0.3	0.762
No gain or attempted gain	0.0163	0.0138	1.18	0.239
Return of illegal gains	0.0002	0.0105	0.02	0.982

Independent variable	B	SE	t	p
Professional-technical qualification examination	0.0099	0.0092	1.08	0.279

The leniency effects land at the lenient margins hierarchically. Guilty plea, repentance, voluntary surrender, and truthful confession concentrate at probation (odds ratios 1.59, 1.73, 2.32, and 2.30); first and occasional offenders are more often fined alone, appearing as a lower probation rate. One caveat is the endogeneity of these circumstances: the return of illegal gains already reflects selection into more serious offending. A basic check shows the selection to be present but uneven—defendants who plead guilty have 3.3 times the illegal gains of those who do not (¥1,494 vs. ¥455) and more repeated substitutions (0.79 vs. 0.33), whereas defendants who express repentance have lower gains (¥603 vs. ¥1,157) but more repeated substitutions (1.14 vs. 0.46). The positive pooled coefficient on repentance is better read as a pooling artefact than as punishment of mitigation, though neither association admits a fully causal reading without instruments or matching, which this article does not attempt. Table 6 reports the three-level variance decomposition, Table 7 reports the probation logit, and Table 8 reports the unconditional leniency rates by circumstance.

Table 6: Three-level nested random-intercept variance decomposition of sentence length.

Level	Groups	Empty model - variance	Empty model - incremental ICC	Empty model - 95% CI	Adjusted (21 covariates) - variance	Adjusted - incremental ICC
Province	31	0.000415	0.0225	[-0.030, 0.075]	0.000061	0.0035
City (prefecture)	189	0.001622	0.0879	[-0.032, 0.208]	0.001357	0.0785
Court	384	0.009918	0.5372	[0.421, 0.654]	0.010298	0.5959
Individual (residual)		0.006507	0.3525	[0.307, 0.398]	0.005566	0.3221

Table 7: Binary logit of probation (odds ratios).

Independent variable	OR	z	p
Minor role	5.506	1.43	0.152
Civil service examination	3.82	2.88	0.004
Voluntary surrender	2.319	4.79	1.7e-6
Truthful confession	2.295	4.86	1.2e-6
Repentance	1.731	3.85	1.2e-4
Guilty plea	1.591	4.38	1.2e-5
Accessory	1.197	0.37	0.711
Illegal gains (ten-thousand yuan)	1.067	1.58	0.114
Number of repeated cheating	1.02	1.17	0.242
No gain or attempted gain	0.891	-0.17	0.861
Professional-technical qualification examination	0.847	-1.02	0.306
Return of illegal gains	0.818	-0.99	0.323
Prior record	0.785	-0.82	0.41
Person for whom the exam is taken	0.769	-2.28	0.023
National education examination	0.744	-2.5	0.013
Whether the fine has been paid	0.699	-3.45	5.6e-4
Attempted crime	0.62	-1.91	0.056
Occasional offender	0.512	-2.45	0.014
First offender	0.484	-3.75	1.8e-4
Meritorious service	0.439	-1.02	0.308
Occupational prohibition	omitted		

The “return of illegal gains” anomaly is only partly explained by reverse causality. Those who returned gains did offend more seriously (mean gains ¥8,276 vs. ¥444, $t = 2.66$, $p = 0.008$; mean sentence 0.154 vs. 0.099 years, $t = 4.66$, $p = 3.4\times10^{-6}$), but stepwise controls for illegal gains, repeated

substitution, examination type, cooperative circumstances, and fine payment attenuate the restitution coefficient by roughly half, and in the full model (Table 3) it is no longer significant (0.0122, p = 0.279). The residual may reflect unmeasured severity or a genuine restitution paradox in which formally quantifying gains triggers a heavier assessment; either way, the clean "pure selection" reading is unsupported.

(C) Hierarchical imbalance

Three separately estimated one-way intraclass correlation coefficients—0.597 at the court level, 0.374 at the city level, 0.085 at the province level—are cumulative rather than incremental, and subtracting one-way ICCs is statistically unsound. A three-level nested random-intercept model (REML) yields the incremental decomposition in Table 6 (Fig. 1): province 0.0225 (2.2%), city 0.0879 (8.8%), court 0.5372 (53.7%), and individual residual 0.3525 (35.2%). The imbalance is therefore concentrated in basic-level courts: more than half of total variance lies between courts within the same city, against about a tenth at the city level. The 95% confidence interval for the court-level ICC is [0.42, 0.65], far above the 0.05 threshold, whereas the province- and city-level intervals include zero.

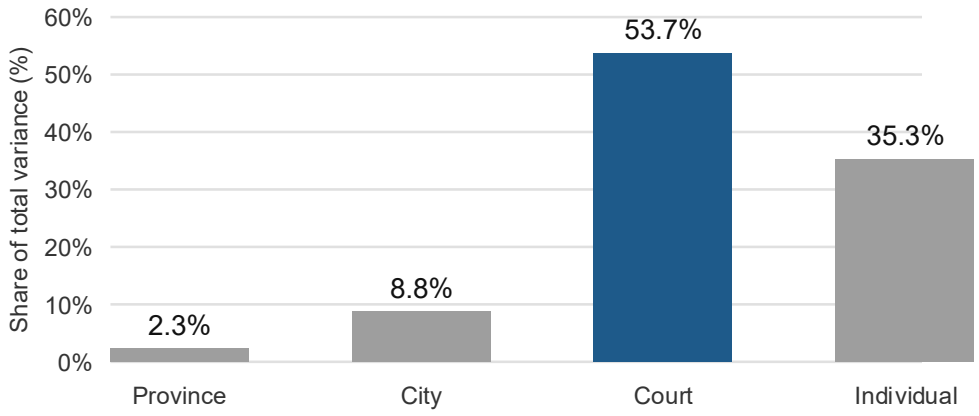


Figure 1: Incremental intraclass correlation coefficients from the three-level nested model (sentence length).

Table 8: Mean sentence, probation rate, and fine-only rate by discretionary circumstance.

Circumstance	N	Mean sentence (years)	Probation rate (%)	Fine-only rate (%)
Prior record (Yes)	55	0.126	38.2	36.4
Prior record (No)	1827	0.102	38.4	51.6
Return of illegal gains (Yes)	145	0.154	42.1	36.6
Return of illegal gains (No)	1737	0.099	38.1	52.4
Guilty plea (Yes)	1073	0.111	42.1	48.3
Guilty plea (No)	809	0.093	33.5	55
Repentance (Yes)	368	0.109	43.5	49.7
Repentance (No)	1514	0.102	37.2	51.5
Voluntary surrender (Yes)	888	0.085	38.3	55.6
Voluntary surrender (No)	994	0.119	38.5	47.2
Truthful confession (Yes)	863	0.115	40.1	46.8
Truthful confession (No)	1019	0.093	37	54.9

Circumstance	N	Mean sentence (years)	Probation rate (%)	Fine-only rate (%)
First offender (Yes)	325	0.07	26.2	67.4
First offender (No)	1557	0.11	41	47.8
Occasional offender (Yes)	154	0.056	20.1	74
Occasional offender (No)	1728	0.107	40	49.1

Note: voluntary surrender shows a nearly flat unconditional probation rate (38.3% vs 38.5%) despite its strong adjusted association with probation (OR 2.32, $p = 1.7 \times 10^{-6}$ in Table 7), because defendants who surrender voluntarily face more serious cases and are more often taken into custody, so the unconditional comparison understates the leniency.

Because the ICC is a ratio of variances, its magnitude is scale-dependent: a short, zero-inflated length turns modest absolute differences between courts into a large proportional ICC. The article therefore does not rest its conclusion on one scale. Table 9 reports the decomposition for five outcomes. The court level is the largest single component for the three custodial and probation outcomes—sentence length (0.537), any custodial sentence (0.577), and probation (0.631)—and remains so when the two two-year-surveillance cases are removed. Because the two-days-to-one conversion of public surveillance can rank a non-custodial sanction above the six-month detention ceiling, the custodial outcome is additionally re-estimated on a severity-ordered scale (fine < surveillance < detention) that involves no such conversion; the court level again emerges as the largest component (0.590), so the finding is robust to the measurement of the custodial outcome. By contrast, fine dispersion lies mainly within courts: the individual level (0.581) exceeds the court level (0.310). The fine is the dimension whose tenfold gap motivates the problem, and the decomposition shows this gap arises primarily between cases within the same court rather than between courts—an important qualification. The probation period is a partial exception: the court level is the largest single component (0.305), but province and city together account for 0.399; its N of 721 reflects the 723 probation defendants less the two cases lacking a court identifier—the same two dropped from the N = 1,880 three-level sample. The precise conclusion is that the basic-level court is the principal locus of imbalance on the custodial and probation dimensions, whereas fine-level dispersion is predominantly intra-court.

Table 9: Variance decomposition by level across five outcomes.

Outcome	Province	City	Court	Individual
Sentence length (years)	0.022	0.088	0.537	0.352
Any custodial / freedom-restricting sentence (0/1)	0.052	0.038	0.577	0.333
Probation (0/1)	0.068	0.016	0.631	0.285
log(fine + 1)	0.095	0.015	0.31	0.581
Probation period (months; N = 721)	0.112	0.287	0.305	0.295
Severity-ordered (fine < surveillance < detention)	0.058	0.036	0.59	0.316

(D) Mechanism

Combining individual-level and regional-level evidence, this article proposes the sentencing-reference-point hypothesis: courts' sentencing scales are uniform insofar as they can invoke objective reference points—statutory penalty tiers, threshold numbers of persons for escalation, and amount standards; the more reference points, the smaller the differences between courts. The offence of taking an exam on behalf of another has virtually no reference points: a single statutory range without internal gradations, no aggravated tier, no amount threshold, and no person-number threshold. This explains both the low explanatory power (≈ 0.11) and the large court-level variance component, which the observed circumstances do not reduce: adding the 21 covariates leaves the court-level component

unchanged or slightly larger (0.009918 to 0.010298, +3.8%; incremental ICC 0.537 to 0.596), while the province-level component collapses from 0.000415 to 0.000061 (−85%): the observed circumstances absorb provincial differences almost entirely yet leave court-level dispersion intact (Table 6).

That persistence is consistent with the reference-point reading, but a competing reading—unobserved severity heterogeneity clustering by court—cannot be excluded, since the observed covariates are only rough proxies for true severity. The subgroup evidence below discriminates between these readings in a way the persistence result alone cannot.

Subgroup comparison is more direct. Repeated substitution, delimited at five or more times, is a relatively clear quantified aggravating signal, and its court-level ICC (0.309) is clearly below that for single substitution (0.617). The national education examination, which should be heavier yet is realised as lighter, has a higher court-level ICC (0.783) than other examinations (0.670). Two caveats apply. First, the high ICC of the national education subgroup partly reflects compositional heterogeneity—it mixes high-stakes (gaokao, postgraduate) with low-stakes (adult college entrance, self-study) examinations—so an internally heterogeneous subgroup mechanically exhibits larger between-court variance even under consistent rules. Second, the subgroups carrying the mechanism argument are small: repeated substitution has 34 defendants in 14 courts (64.3% single-defendant), and gaokao/postgraduate has 135 defendants in 40 courts (35.0% single-defendant). Sparse clusters yield unstable ICCs, as the wide bootstrap confidence intervals show (0.309, [0.000, 0.763]; 0.726, [0.547, 0.906]); these are therefore descriptive observations rather than conclusive evidence.

Decomposing the national education subgroup resolves the ambiguity in the expected direction. The high-stakes gaokao/postgraduate subgroup (N = 135, mean sentence 0.117 years) has a court-level ICC of 0.726, while the low-stakes adult college entrance/self-study subgroup (N = 492, mean sentence 0.073 years) has the highest ICC of any subgroup (0.806), exceeding other examinations (0.670). Divergence concentrates where the severity ranking is objectively ambiguous rather than merely mis-weighted: courts sentence more consistently when a clear severity signal exists, and diverge most when no quantified yardstick exists at all. This is consistent with—but does not by itself establish—the reference-point account, and should be read as suggestive: an unobserved-severity explanation is not excluded, the adult/self-study examinations are regionally heterogeneous in ways that could themselves cluster unobserved severity by court, and the intermediate ranks do not follow a clean monotone gradient (other 0.670 < gaokao/postgraduate 0.726 < adult/self-study 0.806). The mechanism is therefore advanced as a tentative explanation to be tested on larger samples. Table 10 reports the subgroup ICCs with cluster structure and confidence intervals.

Table 10: Subgroup intraclass correlation coefficients with cluster structure and 95% confidence intervals.

Subgroup	N	Courts	Mean per court	One-defendant courts (%)	ICC	95% CI
Adult college entrance / self-study	492	123	4	11.38	0.806	[0.739, 0.872]
National education examination (overall)	627	156	4.02	17.31	0.783	[0.715, 0.852]
Gaokao / postgraduate entrance	135	40	3.38	35	0.726	[0.547, 0.906]
Other examinations (motor vehicle driving)	975	199	4.9	9.55	0.67	[0.593, 0.747]
Single exam substitution	1846	384	4.81	12.76	0.617	[0.556, 0.678]

Subgroup	N	Courts	Mean per court	One-defendant courts (%)	ICC	95% CI
Repeated exam substitution (>= 5 times)	34	14	2.43	64.29	0.309	[0.000, 0.763]

(E) Robustness checks and data validation

The results are estimated from published judgments and thus remain exposed to the selection concern that the most minor—and most diverted—cases are under-represented or undisclosed. Following the validation strategy of (Xiong et al., 2025), which addresses the same concern for China Judgments Online with a manually re-coded subsample and sensitivity analyses, four parallel checks are reported.

Coding reliability. Five coders independently coded the same 199 defendants and the five sets were compared pairwise. Across binary variables the mean Cohen's κ was 0.60 and the mean agreement 90.8%, with the core circumstances—probation, taking an exam on behalf of another, voluntary surrender, and truthful confession—all reaching $\kappa \geq 0.80$; continuous variables (sentence length, fine, probation period) exceeded 96% agreement. Organised count was excluded because the manual defined it ambiguously and the coders adopted three incommensurable conventions. An internal consistency audit flagged 63 defendants coded as both voluntary surrender and truthful confession, which are legally mutually exclusive; dropping these 63 does not alter the conclusions, so the dual coding does not drive the results.

Subsample sensitivity. Re-estimating the three-level model on random 80%, 60%, and 40% subsamples yields court-level ICCs of 0.538, 0.521, and 0.497. The modest decline in the smallest subsample reflects thinning within-court clusters (the median court contributes only about four defendants) rather than fragility: the court level remains a leading component, an order of magnitude above the 0.05 threshold throughout, and preserves the ordering court > city > province.

Year robustness. Restricting the sample to 2016–2021 leaves the court-level ICC essentially unchanged (0.541), and the 2018 peak year alone yields 0.615; the post-2022 collapse therefore does not drive the result.

Cross-offence placebo. Using the 587 defendants convicted of organising exam cheating (excluding the 30 jointly charged with the target offence)—an offence that, unlike the target crime, carries a statutory aggravated tier—the same decomposition yields a court-level ICC of 0.217 (Fig. 2), less than half the ≈ 0.54 for the target crime.

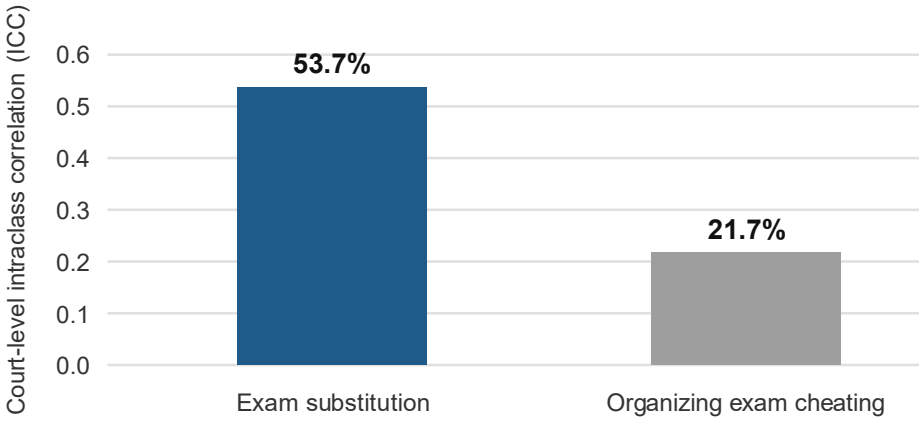


Figure 2: Cross-offence placebo—court-level ICC for organising exam cheating (N = 587) versus the target crime.

VI. Reflection and Suggestions

(A) The Confirmation and Falsification of the Proposition

The reference-point hypothesis is supported by individual-level and regional-level evidence. At the individual level, observable circumstances explain only about 11% of the variance in sentence length, and no single discretionary circumstance stably increases it. At the regional level, the court-level ICC is the largest component for the custodial and probation outcomes, does not decline when the 21 observed circumstances are controlled (+3.8%), and the fine-level dispersion is predominantly intra-court. At the mechanism level, the adult-college and self-study examination—whose severity ranking is objectively ambiguous—exhibits the greatest divergence (ICC = 0.806), whereas repeated substitution—a quantified reference—produces convergence (ICC = 0.309). The first two strands (the persistence of court-level dispersion across outcomes, and the ambiguity-driven divergence) support the judgment that the absence of reference points is a root cause; the third (the small repeated-substitution subgroup) is suggestive only. The hypothesis fails only at the custodial-length margin, where voluntary surrender and truthful confession still shorten the sentence once custody is imposed—the existing circumstances anchor the probation and fine-only margins but not the custodial-length margin. Two competing readings are acknowledged but not dispositive: the persistence result could reflect unobserved severity clustering by court, and the low R^2 could partly reflect the mechanical compression of a zero-inflated outcome; the subgroup pattern tips the balance toward the reference-point account.

Relative to the jurisdictional-uniformity theory (JUT), which predicts negligible inter-court dispersion under a uniform legal and political system, this micro-crime is a boundary case: the uniform system produces equilibrium only where objective reference points exist. The court-level ICC of about 0.54 here, against the sub-0.05 ICCs reported for rape, indicates that the binding constraint on balanced sentencing is the reference point rather than jurisdictional uniformity alone.

The conclusions of this article and the existing empirical research corroborate each other on five

points. First, the opposite direction of the repentance coefficient is consistent with the observed anomalous directions of repentance and the return of illegal gains (Li, 2018). Second, the inter-court imbalance corroborates the findings of inconsistent judgments in similar cases and the tenfold gap in fines (Gui, 2021). Third, the high proportion of motor-vehicle driving examinations is consistent with the statistic that driving examinations account for more than 40 percent (Zhang, 2025). Fourth, the absent reference points corroborate the observation of inconsistent judicial determinations (Song, 2020). Fifth, the absent aggravated statutory tier is consistent with the observation of no circumstance limitation and no statutory escalation. Two points require discrimination. First, the absence of sentencing differentiation in some regions is superficially contrary to the high court-level ICC: the former points to intra-case convergence that does not differentiate circumstances, the latter to inconsistent adjudication across courts, and the two belong to different dimensions and are complementary (Wang, 2018). Second, the proportion of actual detention sentences in earlier samples is higher than here, suggesting that the application of actual custodial sentences varies by time and region; this is a difference in samples and does not conflict with the conclusions (Huang, 2017).

(B) Examination of the Judicial Interpretation and Suggestions

Article 7 of the Interpretation provides no specific guidance for sentencing this crime, and its "serious circumstances" apply only to crimes with an aggravated tier, such as organizing exam cheating, and not to this crime. Based on the conclusion of absent reference points, the Interpretation should be revised—or a separate judicial interpretation issued—to supplement the sentencing reference points and establish graded-treatment rules. Each suggestion corresponds directly to an empirical finding.

First, establish a quantified aggravating reference of "the number of times of exam substitution + the amount of illegal gains." Substitution of five or more times has a court-level ICC of only 0.309, clearly lower than the 0.617 for single substitution; yet each additional instance increases the sentence length by only 0.0031 years (≈1 day) and each additional ten-thousand yuan of illegal gains by only 0.0097 years (≈4 days). The interpretation should therefore clarify that "repeated substitution" is aggravated by tiers delimited at three and five times, and clarify the aggravating range for "a relatively large amount of illegal gains" (a base of ten-thousand yuan is suggested). Because the observed effect sizes are small, these tier boundaries are normative anchoring choices informed by the convergence evidence, not magnitudes derived from the coefficients.

Second, establish benchmark-sentence gradations based on "examination importance." The national education examination is the core legal interest protected by this crime, yet its sentence-length coefficient is significantly negative (-0.028, $p = 1.2\times10^{-5}$) and its court-level ICC is as high as 0.783; the gaokao/postgraduate mean sentence is higher than that of the adult-college/self-study subgroup, indicating that the severity ranking objectively exists but is not consistently implemented. The normative premise—that national-education cheating warrants heavier punishment than driving-examination substitution—is contestable, because the latter directly endangers public road safety. The article defends the premise on an independent ground: national education examinations implicate equality of opportunity in education and the credibility of the national examination system, a legal interest distinct from road safety and precisely the one the legislator meant to protect. On that ground, the interpretation should set benchmark-sentence gradations according to "selective examinations > special passing examinations > general passing examinations," and at least clarify that substitution in the gaokao, the postgraduate entrance examination, and the civil service recruitment examination is aggravated above the base sentence.

Third, take the basic-level court—together with the city (prefecture) level—as the focal point for the unification of sentencing. The court-level share (0.54–0.63 across outcomes) is by far the largest, followed by the individual and city levels; the differences therefore lie principally among the 384 courts (376 basic-level and 8 intermediate courts hearing second-instance appeals) and among individual cases within the same court. This crime should therefore be included in the catalogue of minor offenses under the sentencing-standardization pilot, and—through guiding cases, similar-case retrieval, and the publication of sentencing medians and distributions by prefecture and by court—the inter-court and intra-court differences should be narrowed in a targeted manner, rather than by a blanket demand for "unified sentencing."

Fourth, clarify the "tiered placement" of mitigating circumstances so as to eliminate the misreading of a single indicator. The leniency effects of guilty plea, repentance, voluntary surrender, and truthful confession are concentrated in probation (odds ratios 1.59–2.32), with the latter two additionally shortening the custodial length once custody is imposed; the leniency for first and occasional offenders is concentrated in a fine only (67.4% and 74.0%); and the return of illegal gains is not significant for the actual custodial sentence ($p > 0.05$). Guilty plea, repentance, voluntary surrender, and truthful confession should therefore be linked to the conditions for probation; first and occasional offenders to a fine only and public surveillance; and the return of illegal gains to fine reduction—so as to avoid the mutual offsetting of circumstances at the sentence-length level and the judicial paradox that "returning illegal gains leads to a heavier sentence."

Fifth, set a reference base and upper and lower limits for fines. Fine amounts are highly dispersed (P25/P75 of 2,000/6,000 yuan, but with a significant range), and fines for national education examinations are instead lower. A fine range should therefore be set with illegal gains as the reference base (e.g., one to three times), a statutory base amount set for cases without gains, and "prepayment of fines" codified as a mitigating factor.

Sixth, provide complementary diversion for minor cases such as motor-vehicle driving examinations. This crime is highly concentrated in driving examinations (51.8%), more than half of the cases involve only a fine, and after 2022 there are only a little over 10 cases per year. Cases involving driving examinations, first and occasional offenders, attempt, and other clearly minor circumstances should therefore be diverted through relative non-prosecution and exemption from criminal punishment, so as to alleviate the labeling effect of "once convicted, a criminal record remains." This is a policy response at the complementary level, related to the main line of sentencing reference points but not directly tested by the regressions, and is therefore stated separately.

VII. Limitations

The research and conclusions have limitations. First, the ICC is estimated with a three-level nested random-intercept model (REML), which corrects the approximate error of the earlier one-way decomposition; the absolute ICC values are scale-dependent even though the level ranking is not. Second, low-variance variables—accessory (20 persons), meritorious service (13 persons), occupational prohibition (6 persons), and civil service examination (26 persons)—have extremely small samples and unstable estimates; occupational prohibition exhibits quasi-complete separation in the probation model (all six received probation) and is reported only descriptively. Third, the repeated-substitution subgroup has only 34 persons in 14 courts (64.3% single-defendant) and its ICC should be read as descriptive; the gaokao/postgraduate subgroup (135 persons, 35.0% single-defendant)

is likewise fragile. Fourth, the reverse causality between the return of illegal gains and case severity has not been eliminated by instruments or matching, and the same endogeneity concern applies symmetrically to guilty plea, repentance, voluntary surrender, and truthful confession. Fifth, the reliability assessment's independent re-coding covered a batch of 199 defendants rather than the full sample, and an internal audit flags 63 defendants coded as both voluntary surrender and truthful confession, which are legally mutually exclusive; dropping these 63 defendants leaves the coefficients unchanged, and if anything strengthens them (voluntary surrender -0.049 to -0.060; truthful confession -0.025 to -0.037). Sixth, city-level groupings are based on prefecture-level cities (189 groups) rather than all administrative city-level divisions.

BELSLEY, D. A., KUH, E. & WELSCH, R. E. 1980. *Regression diagnostics : identifying influential data and sources of collinearity / David A. Belsley, Edwin Kuh, Roy E. Welsch,* New York, Wiley.

COHEN, J. 1988. *Statistical Power Analysis for the Behavioral Sciences,* New York, Lawrence Erlbaum.

CRAGG, J. G. 1971. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica,* 39**,** 829–844.

DIXON, J. 1995. The Organizational Context of Criminal Sentencing. *American Journal of Sociology,* 100**,** 1157–1198.

DUAN, N., MANNING, W. G., MORRIS, C. N. & NEWHOUSE, J. P. 1983. A Comparison of Alternative Models for the Demand for Medical Care. *Journal of Business & Economic Statistics,* 1**,** 115–126.

DUN, N. & HUANG, Y. 2017. The Crime of Taking an Exam on Behalf of Another Person and the Legislative Trend of Micro-Crimes (tidai kaoshi zui yu weizui lifa qushi). *Journal of Baoding University (baoding xueyuan xuebao),* 30**,** 63–68.

ELMAN, B. A. 2000. *A Cultural History of Civil Examinations in Late Imperial China,* Berkeley, University of California Press.

GUI, Y. 2017. The Reasonable Limits of Criminalizing Exam Substitution (tikao ru xing de heli jiexian). *Journal of Jiangsu Police Institute (jiangsu jingguan xueyuan xuebao),* 32**,** 44–50.

GUI, Y. 2021. An Examination of the Current Situation and a Criminal-Law Analysis of the Crime of Taking an Exam on Behalf of Another Person: Based on 44 Criminal Judgments (tidai kaoshi zui de xianzhuang kaocha yu xingfa fenxi——yi 44 fen xingshi panjueshu wei li). *China Examinations (zhongguo kaoshi)***,** 43–50.

GUO, Z. 2024. Clarifying the Proper Logic of Misdemeanor Legislation in the Context of the Modernization of the Rule of Law (fazhi xiandaihua yujing xia qingzui lifa yingran luoji de kuoqing. *Journal of Criminal Investigation Police University of China (zhongguo xingjing xueyuan xuebao)(4)***,** 76–83.

HE, Q. & ZHANG, S. 2024. On the Judicial Limits of Protecting Collective Legal Interests: From the Perspective of the Crime of Taking an Exam on Behalf of Another Person (lun jiti fayi baohu zhi sifa xianzhi——yi tidai kaoshi zui wei

shijiao). *Journal of Taiyuan University of Technology (Social Sciences Edition) (taiyuan ligong daxue xuebao (shehui kexue ban)),* 42**,** 55–65.

HO, P.-T. 1962. *The Ladder of Success in Imperial China: Aspects of Social Mobility, 1368-1911,* New York and London, Columbia University Press.

HOU, Y. & TRUEX, R. 2022. Ethnic Discrimination in Criminal Sentencing in China. *The Journal of Politics,* 84**,** 2294–2299.

HUANG, C. 2017. Problems and Suggestions Concerning the Crime of Taking an Exam on Behalf of Another Person: An Empirical Analysis Based on 76 Exam-Substitution Cases (tidai kaoshi zui de wenti yu jianyi——jiyu 76 li tikao anjian de shizheng fenxi). *Journal of Fujian Agriculture and Forestry University (Philosophy and Social Sciences Edition) (fujian nonglin daxue xuebao (zhexue shehui kexue ban)) 20(4),* 20**,** 103–109.

HUSAK, D. 2011. Reservations About Overcriminalization. *New Criminal Law Review: An International and Interdisciplinary Journal,* 14**,** 97–107.

JIN, J. & BALL, S. J. 2020. Meritocracy, social mobility and a new form of class domination. *British Journal of Sociology of Education,* 41**,** 64–79.

KAHNEMAN, D. & TVERSKY, A. 1982. The Psychology of Preferences. *Scientific American,* 246**,** 160–173.

LAMBERT, D. 1992. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics,* 34**,** 1–14.

LEBRETON, J. M. & SENTER, J. L. 2008. Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods,* 11**,** 815–852.

LI, L. 2018. A Study on Accomplice Issues in the Crime of Taking an Exam on Behalf of Another Person (tidai kaoshi zui gongfan wenti yanjiu). *Journal of Guangdong Open University (guangdong kaifang daxue xuebao),* 27**,** 53–58.

LIANG, Y. 2021. The Consequences and Countermeasures of the Expansion of Micro-Crimes from the Perspective of Proactive Criminal Law (jiji xingfa guan shiye xia weizui kuozhang de houguo ji yingdui). *Political Science and Law (zhengzhi yu falv)***,** 35–48.

LIEBMAN, B. L., ROBERTS, M. E., STERN, R. E. & WANG, A. Z. 2020. Mass Digitization of Chinese Court Decisions: How to Use Text as Data in the Field of Chinese Law. *Journal of Law and Courts,* 8**,** 177–201.

LIN, X., LIU, S., LI, E. & MA, Y. 2022. Sentencing Disparity and Sentencing Guidelines: the Case of China. *Asian Journal of Criminology,* 17**,** 127–155.

LIU, Y. 2013. Meritocracy and the Gaokao: a survey study of higher education selection and socio-economic participation in East China. *British Journal of Sociology of Education,* 34**,** 868–887.

MACKINNON, J. G. & WHITE, H. 1985. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics,* 29**,** 305–325.

MILES, S. B. 2016. From Mobility to Stability: Civil Examinations and Social Status in Eighteenth-Century China. *The Eighteenth Century,* 57**,** 519–525.

MULLAHY, J. 1986. Specification and testing of some modified count data models. *Journal of Econometrics,* 33**,** 341–365.

MUSTARD, DAVID B. 2001. Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts. *The Journal of Law & Economics,* 44**,** 285–314.

RAUDENBUSH, S. W. & BRYK, A. S. 2002. *Hierarchical linear models : applications and data analysis methods / Stephen W. Raudenbush, Anthony S. Bryk,* Thousand Oaks, Sage Publications.

RUI, Q. 2015. Some Reflections on Criminalizing Exam Substitution: Based on Article 25 of Amendment IX to the Criminal Law (dui tikao ru xing de ruogan sikao——jiyu xingfa xiuzhengan (jiu) di ershihao tiao zhi guiding). *Journal of Xinjiang Police College (xinjiang jingcha xueyuan xuebao),* 35**,** 38–43.

SHI, W. 2018. Criminal Forms and Sentencing of the Crime of Taking an Exam on Behalf of Another Person (tidai kaoshi zui de fanzui xingtai he liangxing). *People's Judicature (Cases) (renmin sifa (anli))***,** 38–40.

SNIJDERS, T. A. B. & BOSKER, R. J. 2012. *Multilevel analysis : an introduction to basic and advanced multilevel modeling / Tom A.B. Snijders, Roel J. Bosker,* Los Angeles, Sage.

SONG, D. 2020. Several Issues in the Application of Law to Exam-Cheating Crimes (kaoshi zuobi fanzui falv shiyong de jige wenti). *China Examinations (zhongguo kaoshi)***,** 30–35.

TIAN, P. 2017. A Study on the Judicial Determination of the Crime of Taking an Exam on Behalf of Another Person (tidai kaoshi zui de sifa rending yanjiu). *Social Science Front (shehui kexue zhanxian)***,** 207–215+2.

TVERSKY, A. & KAHNEMAN, D. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science,* 185**,** 1124–1131.

TVERSKY, A. & KAHNEMAN, D. 1981. The Framing of Decisions and the Psychology of Choice. *Science,* 211**,** 453–458.

ULMER, J. & LIGHT, M. T. 2011. Beyond Disparity: Changes in Federal Sentencing After Booker and Gall? *Federal Sentencing Reporter,* 23**,** 333–341.

ULMER, J. T. 2019. Criminal Courts as Inhabited Institutions

Making Sense of Difference and Similarity in Sentencing. *Crime and Justice,* 48**,** 483–522.

ULMER, J. T. & JOHNSON, B. D. 2017. ORGANIZATIONAL CONFORMITY AND PUNISHMENT

FEDERAL COURT COMMUNITIES AND JUDGE-INITIATED GUIDELINES DEPARTURES. *The Journal of Criminal Law and Criminology (1973-),* 107**,** 253–292.

WANG, D. & XU, F. 2019. Judicial Determination of the Crime of Taking an Exam on Behalf of Another Person (tidai kaoshi zui de sifa rending). *People's Judicature (renmin sifa)***,** 73–77.

WANG, J. 2018. Judicial Proof Dilemmas in the Crime of Taking an Exam on Behalf of Another Person and an Exploration of Solutions: Starting from the Dual

Constraints of Substantive Law and Procedural Law (tidai kaoshi zui sifa zhengming kunjing ji lujing tanxi——cong shiti fa yu chengxu fa de shuangchong yueshu xia tanqi). *Hebei Law Science (hebei faxue),* 36**,** 45–60.

WEI, L. & YAN, C. 2018. Social Quality of China: Indicators, Reality, and Problems. *The International Journal of Social Quality,* 8**,** 78–102.

WHITE, H. L. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica,* 48**,** 817–838.

XIONG, M., XIA, Y. & YU, X. 2025. Sentencing equilibrium in rape cases: a legal and political explanation of jurisdictional uniformity in China. *Humanities and Social Sciences Communications,* 12**,** 59.

YANG, Q. 2017. Determination of the Offense and Application of Law in Surrogate Exam-Taking Cases(tikao an de zuiming rending yu falv shiyong). *People's Judicature (Cases) (renmin sifa (anli)***,** 54–58.

YANKAH, E. N. 2011. A Paradox in Overcriminalization. *New Criminal Law Review: An International and Interdisciplinary Journal,* 14**,** 1–34.

YIN, J. & YIN, M. 2017. Advocating Limited Decriminalization of Exam Substitution (tikao xingwei youxian feizuihua zhi tichang). . *Jiangxi Social Sciences (jiangxi shehui kexue),* 37**,** 206–213.

ZHANG, M. 2025. Discussion on the Scope of National Examinations Prescribed by Law: A Case Analysis Based on 933 Judgment Documents of the Crime of Surrogate Exam-Taking (falü guiding de guojia kaoshi fanwei tantao——jiyu 933 fen daiti kaoshi zui caipan wenshu de anli fenxi). *China Examinations (zhongguo kaoshi)***,** 61–70.

ZHAO, B. & ZHAO, Y. 2016. Characteristics and Deficiencies of the Amendment: A Brief Review of Amendment IX to the Criminal Law (xiufa tedian yu quehan xingfa xiuzhengan (jiu) jianping). *Seeker (qiusuo)***,** 4–10.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- examsubstitutiondatasetfinal.csv