

The Influence of Ethnic Identity on the Academic Performance of Chinese College Students: An Empirical Study Based on the Administrative Data of a University

Zerong Wang ^{a, *}

^a School of Economics, Zhejiang University of Finance and Economics, Hangzhou, China

ABSTRACT

The policy of ethnic equality and solidarity aims to narrow the ethnic gap and develop harmoniously. At the level of higher education, it should promote the equity of ethnic education. This paper uses the administrative data of undergraduate students from 2008 to 2015 in a liberal arts university in southwest China to study the differences, causes and dynamic trends in the scores of college students from different nationalities. The results show that compared with Han students, the average GPA of ethnic minority students is 0.13 points lower (equivalent to 0.23 and 0.3 standard deviations of the scores of ethnic minority and Han students), the fourth grade scores are 0.28 standard deviations lower, and 7.3% lower. This difference is more significant in the western region, especially in Xinjiang, Yunnan, Hubei and other provinces inhabited by ethnic minorities; Kazak and uygur are the nationalities with the biggest gap with han students; The difference between girls is more significant than that between boys. The "peer effect" and cultural distance between roommates are important reasons for this difference. Further analysis shows that although the ethnic differences in academic performance are gradually expanding, the growth rate tends to converge. The above conclusions are helpful to clarify the object and time of making national higher education policy, and also have important reference significance for primary and secondary education.

KEYWORDS

Ethnic identity; Academic performance; Peer effect; Cultural distance

*Corresponding author: Zerong Wang E-mail address: wongzerong@zufe.edu.cn

ISSN 2972-3515 doi: 10.58567/rea01010001 This is an open-access article distributed under a CC BY license (Creative Commons Attribution 4.0 International License)

1. Introduction

In the past two decades, the proportion of ethnic minority students in various schools at all levels in China has risen steadily. By 2020, the proportion of minority students in schools at all levels has significantly exceeded that of the national minority population.¹This indicates that the equalization of ethnic education in China has changed from fair access to a new stage of fair education output (Fang and Wang, 2019). How to promote the equality of ethnic education while giving consideration to the construction of high-quality education system needs to start from narrowing the ethnic differences in students' academic performance. On the one hand, the main goal of the previous ethnic education policy was to improve the enrollment rate of ethnic minority students, thus ignoring the ethnic differences in academic performance. Narrowing ethnic differences in academic performance, as an inherent requirement of the equality policy of ethnic education and the construction of high-quality education system, is of great significance for promoting the equalization of public education, improving the quality of higher education and deepening education reform mentioned in the 14th Five Year Plan. On the other hand, the stage of higher education is the final stage of the accumulation of knowledge-based human capital. National equality in this stage determines social harmony and stable economic development (Pilkington, 2018). So, is there any difference in academic performance between ethnic minority students and Han students in China's higher education? If so, what are the ways to narrow this difference? How will the difference change over time? The study of these problems can provide useful enlightenment for the correct understanding of the differences, mechanisms and changing trends of China's ethnic education.

The social and economic effects of national identity have always been concerned. The existing literature is mainly based on the research of life cycle theory. It is found that there are significant differences between different ethnic groups in terms of fertility, upbringing methods, children's enrollment rate, academic performance, labor income, employment level, elderly care services and health level throughout the life course (Niemesh&Shester, 2020; Cassidy, 2020; Henry et al., 2020; Hirata&Soares, 2020; Bone et al., 2020; Sutton et al., 2021). Although based on the theory of human capital and the framework of educational production function, relevant studies on the educational effects of national identity have examined the ethnic differences in students' academic performance (Kuhfeld et al., 2021), determinants (Nguyen et al., 2020) and time trends (Fahle et al., 2020). However, there are the following shortcomings: (1) These studies focus on the analysis of ethnic differences in academic performance in the basic education stage, ignoring this difference in the higher education stage; (2) These documents mainly define and classify the concept of nationality based on the color of skin, emphasize the achievement differences among Asian, white, black and Hispanic students, and ignore the ethnic differences based on specific customs and linguistic and cultural concepts, such as the differences between minority students such as Zhuang, Manchu, Uygur and Han students. Third, the current literature mainly discusses the differences in ethnic education achievements based on the background of western countries, lacking empirical evidence from a large country with a multi-ethnic population such as China.

Based on the above policy and literature background, this paper empirically analyzes the marginal impact, mechanism and dynamic trend of ethnic identity on students' academic performance by using the administrative data of 2010-2015 undergraduate students in a liberal arts financial college. The study found that the GPA, major ranking, and CET-4 scores of ethnic minority students were significantly lower than those of Han students. On the one hand, the difference in scores between ethnic minority students and Han students is mainly caused by female students, Kazak and Uyghur nationalities, Xinjiang Uygur Autonomous Region, Yunnan Province and Hubei Province, and samples from the western region. On the other hand, the improvement of roommate's average score and the reduction of cultural distance will help to narrow the difference. Although the ethnic differences in academic

¹ According to the data of the seventh national population census in 2020, the minority population accounts for 8.89%.

performance gradually expand with the term and year, the growth rate gradually tends to converge.

The rest is arranged as follows: The second part introduces data and variables. The third part is the empirical model setting. The fourth part is the benchmark results. The fifth part is robustness test. The sixth part is the mechanism analysis. The seventh part is the further analysis. Finally, summarize.

2. Data description

2.1. Data source

The main data of this paper is from the administrative data of 2010-2015 undergraduate students of a liberal arts finance and economics college in the west.² From the Academic Affairs Office, the School Work Department, the Logistics Department, the Admissions Office and the official website of the school, the author obtained the undergraduate's school scores, college entrance examination results, personal characteristics, accommodation information, parents' characteristics, college and department characteristics, and the school's admission scores in liberal arts and science in each province in the calendar year. Part of the liberal arts and science admission score data comes from the college entrance examination volunteer filling reference system.3The initial sample consisted of 22898 undergraduates from 2010 to 2015. Although this paper is unable to obtain a nationally representative sample, it is feasible to use undergraduate administrative data to study this problem, because: (1) the sample size is large enough to cover 31 provinces nationwide; (2) As far as the characteristics of ethnic distribution are concerned, the administrative data of undergraduates are basically close to the results of previous national population censuses. As shown in Table 1, the proportion of ethnic minorities in the initial sample, analysis sample, "The Fifth Population Census", "The Six Population Census" and "The Seventh Population Census" of undergraduate administrative data is 9.96%, 9.36%, 8.41%, 8.49% and 8.89% respectively, which is very close. In addition, the proportion of minority students in different years is basically consistent with the previous census data. See Table A1 for the specific results.

	Number of ethnic minorities (10000 people)	Number of Han people (10000 people)	Proportion of ethnic minorities (%)
	(1)	(2)	(3)
Raw sample	0.282	2.548	9.96
Analysis sample	0.085	0.823	9.36
The Fifth Population			
Census	10643	115940	8.41
The Six Population	11379	122593	8 4 9
Census	11577	122375	0.49
The Seventh			
Population Census	12547	128631	8.89

Table 1. Comparison of ethnic distribution.

Notes: The data of all previous national censuses are from the National Bureau of Statistics.

The data processing process is mainly divided into four steps: (1) Delete students' personal privacy information. Extract variables such as province, city, birth year and gender from ID card number, and delete personal privacy information such as name and ID number; (2) Data consolidation. Merge various data tables (mainly

² The author signed a confidentiality agreement with the university, so it is inconvenient to disclose the name of the unit.

 $^{^{\}scriptscriptstyle 3}~$ The website of China Education Online Volunteer Filling Reference System is:

https://www.eol.cn/e_html/gk/zytbsj/index. shtml, In addition, the admissions office of the university only provides the college entrance examination results of self enrollment.

including academic achievement table, personal characteristics table, family situation questionnaire and college entrance examination result table) according to index variables such as student number, province and major; (3) Missing value processing. Variables including missing values include: CET-4 scores, ethnic identity, father's age, mother's age, father's party, mother's party, average scores of liberal arts and science admissions. The number of missing values is 191, 2496, 10425, 12596, 10425, 12596, 15125 and 10478 respectively. Samples containing missing values are deleted. (4) Exception handling. To mitigate the estimation error caused by outliers. The author excluded the samples whose GPA was less than or equal to 0 (that is, the number of failed samples was only 112), and the samples whose parents were younger than 30 or older than 60 years old. The final sample consisted of 9084 undergraduate students of 2010 to 2015, with the number of minority and Han students being 850 and 8234 respectively.

2.2. Variable definition

The variables in this paper mainly include academic achievements, students' individual characteristics, parents' characteristics and college and department characteristics. The specific definitions and calculation methods of each variable are as follows:

First of all, the academic performance of the explanatory variables in this paper is measured by the average grade point (GPA), major ranking and CET-4 scores of students in school. Among them, GPA is the score obtained by students after weighted average of all courses in school, with the value range of 0-5 points. Major ranking refers to the percentage ranking of students in their majors, which is defined as major ranking/total number of majors. The higher the major value, the lower the ranking. CET4 is a national standardized English test administered and implemented by the Examination Center of the Ministry of Education, with a full score of 710.⁴ In order to facilitate the analysis, the author standardized the scores of CET-4 at the level of grade and major. After standardization, the average score of CET-4 is 0, and the standard deviation is 1. It is worth noting that GPA and major are absolute and relative indicators of academic performance, respectively. Due to differences in the style of writing questions and the subjectivity of marking papers among different teachers, these two indicators are not comparable among different subjects, majors, colleges and grades. The unified proposition of CET-4 by experts from the Ministry of Education can effectively alleviate this problem. Therefore, the advantage of CET-4 is that it has strong comparability among different schools and departments, but the disadvantage is that CET-4 can only reflect students' English level and cannot comprehensively reflect students' professional quality (except English majors). The advantages and disadvantages of GPA and major are just opposite. Therefore, the above three indicators are considered as dependent variables in this paper.

This paper focuses on the explanatory variable of national identity, which is defined as a binary dummy variable. If the student's ethnic identity is minority, the value is 1; otherwise, the value is 0.

The control variables mainly include students' individual characteristics, parents' characteristics and college and department characteristics. Students' personal characteristics include age, gender and whether they are independent candidates. The student's age is obtained by subtracting the year of birth from 2015. This parameter defines whether the gender is male. If the gender is male, the value is 1. Otherwise, the value is 0. Independent enrollment candidates are also defined as binary dummy variables. If it is self enrollment, the value is 1; otherwise, the value is 0. The personal ability of students is measured by the college entrance examination results, but because the school only provides the college entrance examination results for self-enrollment, it is unable to obtain the college entrance examination results for general enrollment. Therefore, this paper uses the average score of the school from 2010 to 2015 in each province and each major liberal arts and sciences to supplement this data. The

⁴ The score of CET-4 is measured by the highest score of all the CET-4 exams that students take.

Wang

rationality of this is that the number of students enrolled in each specific major in each province is very small each year, and the college entrance examination results of the enrolled students are relatively close. For example, in 2015, the highest, lowest and average scores of Fujian's liberal arts and science finance were 622, 619 and 620 for liberal arts and 650, 639 and 645 for science; Accounting: liberal arts 620, 618, 619, science 649642, 647. Admittedly, there may still be some deviation in such treatment. In addition, as part of the robustness analysis, this paper also uses the highest and lowest scores of liberal arts and sciences admitted by the university in each province from 2010 to 2015 to impute the college entrance examination results. It is worth noting that the scores of ethnic minority students in the college entrance examination are the scores of the college entrance examination minus the policy bonus points. And after controlling the college entrance examination results, the problem of self selection before enrollment has also been alleviated.

Parents' characteristics mainly include father's age, mother's age, father's political outlook, mother's political outlook and other variables. The father's age and the mother's age are obtained by subtracting the year of birth from 2015. The political outlook of the father (or mother) includes the following five options: Party members, probationary party members, members of the Communist Youth League, the masses and other parties. It is defined as a binary dummy variable. If the father (or mother) is a party member in political status, the value is 1. In other cases, the value is 0. In addition, we further controlled the fixed effects of major, college, grade, city level and the interactive fixed effects of city * grade.

2.3. Descriptive statistical analysis

Table 2 reports the results of descriptive statistical analysis of the main variables. The following conclusions can be drawn: (1) The academic performance of minority students is significantly lower than that of Han students. The average GPA, major and CET-4 scores of ethnic minority students were 2.98 points, 59.5% and -0.36 standard deviations, which were significantly lower than the 3.24 points, 47.6% and 0.1 standard deviations of Han students, and the above differences were significant at the 1% level. (2) The proportion of minority male students is significantly lower than that of Han students. The probability of enrollment through independent enrollment is about 5% lower than that of Han students. The average admission score of liberal arts students is also significantly lower than that of Han students. In view of the significant differences between minority students and Han students in these observable personal characteristics and parental characteristics, in order to exclude the interference of these factors on the estimation results, this paper controls these variables in the subsequent regression analysis.

					Han		
	Ethnic 1	ninorities st	udents		students		
	Ν	Mean	Std.	Ν	Mean	Std.	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(2)-(5)
GPA	850	2.980	0.557	8234	3.238	0.44	-0.258***
Rank	850	0.595	0.274	8234	0.476	0.285	0.119***
CET-4	833	-0.357	1.133	8180	0.098	0.925	-0.454***
Age	850	20.97	2.017	8234	20.91	2.015	0.061
Male	850	0.300	0.459	8234	0.336	0.473	-0.036**
Self-enrollment	850	0.046	0.209	8234	0.097	0.296	-0.051***
Age of father	850	46.41	4.000	8234	45.69	3.612	0.723***
Age of mother	850	44.63	3.568	8234	43.93	3.157	0.703***
Party of Father	850	0.399	0.490	8234	0.399	0.49	-0.000
Party of mother	850	0.164	0.370	8234	0.185	0.388	-0.021
Average score of liberal	850	-0.086	0.637	8234	0.010	1.001	-0.097***

 Table 2. Group Descriptive Statistics Results.

arts Average score of science	950	0.046	0 502	0724	0.020	1 015	0.026	
Average score of science	030	-0.040	0.373	0234	-0.020	1.015	-0.020	_
Notes: *** n<0.01 ** n<0.05 *	* n<0 1							

3. Model setting and benchmark results

3.1. Empirical Model Setting

With reference to the research methods of Blatt and Votruba Drzal (2021) and Wu Jia et al. (2021), in order to examine the impact of students' national identity on academic performance, the econometric model is set as follows:

 $Y_{imsgp} = \alpha + \delta minority_i + \beta X_{imsgp} + \gamma_m + \tau_s + \pi_g + \varphi_{p+\varphi_p*\pi_g} + \varepsilon_{imsgp}$ (1) Where, dependent variable Y_{imsgp} refers to the GPA, ranking or CET-4 scores of m major students *i* from grade *g*, college *s* in *p* city.⁵ This paper focuses on the variable of minority, which indicates whether the student's national identity is ethnic minority. δ It is the parameter of interest in this paper, which means the difference in GPA, major and CET-4 scores between ethnic minority students and Han students under the condition that other factors remain unchanged. *X* is the control variable, including students' individual characteristics (gender, age, enrollment method, college entrance examination results, etc.) and parents' characteristics (parents' age, political outlook, etc.). γ , τ , π and φ represent the fixed effects at the professional, college, grade and municipal levels respectively, $\varphi * \pi$ is the interaction fixed effect between city and grade. E is random disturbance term.

It is worth noting that, on the one hand, students' national identity may be manipulated artificially. The Measures for the Administration of the Registration of Ethnic Composition of Chinese Citizens issued in 2016 clearly stipulates that parents of intermarriage families can choose their children's ethnic identity independently, and children under 18 can only change their ethnic identity when their parents' marriage relationship changes. At the age of 18-20, there is a chance to change the national identity, which cannot be changed after the age of 20. In order to enable children to enjoy the policy of bonus points in the college entrance examination, parents of intermarriage families have incentives to set or change their children's ethnic identity to a minority. On the other hand, teachers may add extra points when grading ethnic minority students, which leads to higher academic performance of ethnic minority students than the real level. The above two reasons will lead to a negative bias in the OLS estimation, which also means that the true academic performance of ethnic differences may be greater. Therefore, this paper is a conservative estimate of this topic.

3.2. Benchmark results

Table 3 reports OLS regression results. The explanatory variables of columns (1) - (3), (4) - (6) and (7) - (9) models are GPA, major and CET-4 scores, respectively. This paper examines the robustness of the model and the impact of different explanatory variables on the benchmark results by adding control variables step by step. First of all, the models in columns (1), (4) and (7) only control the major fixed effect, college fixed effect, grade fixed effect, municipal fixed effect, and city * year interactive fixed effect. The results show that the GPA, major and CET-4 scores

⁵ It should be noted that when the explanatory variable is CET-4, the regression sample does not include students from the School of Foreign Languages and Economics. This is because the National English Test includes Band 4 and Band 6 (full score 710) for non English majors and Band 4 and Band 8 (full score 100) for English majors. The students of the School of Foreign Languages and Economics can take both kinds of examinations. There are many students whose scores are lower than 100 in Band 4. The author cannot tell whether the reason for the low scores is that they failed (which is less likely) or that they took the TEM 4.

of ethnic minority students are significantly lower than those of Han students and are significant at the 1% level. On this basis, the models in columns (2), (5) and (8) further control the individual characteristics of students, that is, the variables such as age, gender, enrollment methods and college entrance examination scores. Finally, the models (3), (6) and (9) further control the parents' characteristics, that is, the father's age, the mother's age, the father's political outlook and the mother's political outlook. The results still show that the academic performance of ethnic minority students is significantly lower than that of Han students, that is, GPA, major and CET-4 scores are significantly lower than that of Han students. It is worth noting that after controlling the characteristics of departments, grades, regions and time-varying regions, controlling students' individual characteristics and parents' characteristics has little influence on the results.

It can be seen from the regression results in columns (3), (6) and (9) that the estimated results of other factors affecting academic performance (such as age, gender, enrollment method, college entrance examination results, parents' age and political identity) are basically consistent with the existing literature (Wang Chunchao and Xiao Aiping, 2019; Ma Liping and Bu Shangcong, 2019; Wu Yuxiao and Zhang Fan, 2020; Blatt&Votruba Drzal, 2021). Under the condition of keeping other factors unchanged, girls' academic performance is better than boys'; The older the students are, the worse their academic performance is; Self enrollment score is lower than that of general enrollment; The higher the score of the college entrance examination, the better the students' academic performance at the undergraduate stage; Father's Party membership, father's age and CET-4 scores are significantly and positively correlated, which may be related to the fact that party member fathers and older fathers pay more attention to their children's English learning. Interestingly, the influence of parents' age on GPA and major is opposite. The older the father, the worse the child's academic performance; The older the mother, the better the child's academic performance, which is consistent with the findings of Veldkamp et al. (2020).

When examining the impact of national identity on academic performance, we believe that there is less possibility of endogenous problems. The reasons are as follows: First, two-way causality is unlikely. Students' national identity depends on their parents' national identity at birth, and their academic achievements at the university stage cannot affect their choice of national identity at birth. Second, the possibility of measurement error is low. The data in this paper are from the school administration data. The key variables such as GPA, major, CET-4 scores and national identity have been repeatedly confirmed by the marking teachers, counselors and students themselves, and there is little possibility of measurement error. Third, there are some missing variables. Limited by data, factors such as parents' income, teachers' quality, school resources, regional economy and education level were omitted in the regression analysis. However, this paper controls the major fixed effect, college fixed effect, grade fixed effect, city fixed effect, and city * grade interactive fixed effect to alleviate the problem of missing variables. And it can be seen from the results in Table 3 that the coefficient and significance of the variables of concern almost remain unchanged after adding students' individual characteristics and parents' characteristics. This also shows that the problem of missing variables in this paper may have little impact on the benchmark results. To sum up, the author thinks that there is little possibility of endogenesis.

	GPA			Rank			CET-4		
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Ethnic minorities	- 0.128***	- 0.129***	- 0.130***	0.072***	0.072***	0.072***	- 0.286***	- 0.274***	- 0.281***
	(0.025)	(0.022)	(0.022)	(0.013)	(0.011)	(0.011)	(0.047)	(0.044)	(0.044)
Age		- 0.026***	- 0.026***		0.017***	0.017***		- 0.111***	- 0.108***
		(0.005)	(0.005)		(0.003)	(0.003)		(0.018)	(0.018)
Male		- 0.291***	- 0.291***		0.189***	0.189***		- 0.454***	- 0.457***

Table 3. OLS estimation of national identity on academic performance.

		(0.020)	(0.020)		(0.012)	(0.012)		(0.030)	(0.030)
Self- enrollment		- 0.060***	- 0.060***		0.048***	0.047***		-0.197**	- 0.217***
		(0.020)	(0.020)		(0.010)	(0.010)		(0.075)	(0.076)
Average Score of Liberal Arts		0.038*	0.039*		-0.023*	-0.023*		0.134***	0.131**
		(0.022)	(0.022)		(0.012)	(0.012)		(0.049)	(0.049)
Average Score of Science		0.053**	0.053**		-0.037**	-0.036**		0.019	0.023
		(0.026)	(0.026)		(0.014)	(0.014)		(0.047)	(0.048)
Age of Father			-0.004**			0.003***			0.011**
			(0.001)			(0.001)			(0.005)
Age of Mother			0.004**			-0.003**			0.000
			(0.002)			(0.001)			(0.004)
Party of Father			-0.005			0.004			0.105***
Deuter of			(0.010)			(0.008)			(0.020)
Mother			0.002			0.005			0.040
			(0.014)			(0.009)			(0.026)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Ν	9084	9084	9084	9084	9084	9084	8584	8584	8584
R2	0.278	0.357	0.358	0.19	0.277	0.278	0.234	0.29	0.295

Notes: The report in brackets is the clustering standard error adjusted at the major level*** p<0.01, ** p<0.05, * p<0.1.

4. Robustness check

4.1. Change variable measurement indicators

First, replace the dependent variable calculation method and measurement indicators. After drawing the nuclear density map of GPA, major and standardized CET-4 scores, it is found that the distribution patterns of these three indicators are left biased distribution, uniform distribution and normal distribution. If the dependent variable is not distributed properly, outliers may cause estimation errors. In order to verify the robustness of the results, this paper refers to the research method of Schindler and Westcott (2021) to make inverse hyperbolic sine (IHS) transformation for GPA and major. In addition, CET-6 is used as a substitute for CET-4.⁶ Repeat the same regression as columns (3), (6) and (9) of Table 3, and report the results in columns (1) - (3) of Table 4. The results show that the benchmark conclusion is still valid, that is, the GPA, major and English scores of ethnic minority students are significantly lower than those of Han students.

Secondly, replace the key variables to measure college entrance examination scores. Pratiwi et al (2020) found that the average score in the first semester of college is an effective indicator of students' personal ability before entering college. In order to verify the robustness of the benchmark results, GPA in the first semester was used as a measure of ability and the benchmark model was re-estimated. The results were reported in columns (4) - (6) of Table 4. The results show that the conclusion that the scores of minority students are significantly lower than those of Han students is still valid. It should be noted that if the GPA in the first semester is used as a measure of ability, the sample size will decrease significantly.⁷ Therefore, this paper uses the college entrance examination results as

⁶ In order to retain a larger sample size, the English scores in the benchmark analysis were not measured by CET-6. The data shows that the missing value of CET-6 is 1975 more than that of CET-4, which indicates that CET-4 is a more popular English proficiency test for undergraduates.

⁷ The results of taking the professional ranking of the first semester as the ability measurement indicator are consistent with this, and will not be reported due to space limitations.

a measure of ability in the benchmark analysis.

Finally, we manually collected the highest and lowest scores of liberal arts and sciences in each province in the calendar year. In order to verify the robustness of the results, replace the college entrance examination scores of general enrollment with the highest scores of liberal arts and sciences in each province and repeat the benchmark analysis. The corresponding regression results are reported in columns (1) - (3) of Table 5. The results show that the benchmark conclusion is still valid. In addition, this paper also replaces the college entrance examination results of general enrollment with the lowest scores for admission of liberal arts and sciences in all provinces and majors. The results are reported in Table A2 (1) - (3). The conclusions from the benchmark analysis are still stable.

	IHS Transformation		Standardization	Ability changed to GPA in the first semester		
Variables	GPA	Rank	CET-6	GPA	Rank	CET-4
	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic minorities	-0.045***	0.063***	-0.162***	-0.039**	0.020*	-0.201***
	(0.008)	(0.010)	(0.057)	(0.016)	(0.011)	(0.051)
Controls	YES	YES	YES	YES	YES	YES
Ν	9084	9084	6665	7331	7331	6975
R2	0.356	0.275	0.241	0.707	0.663	0.397

Table 4. Robustness test - measures replacing key variables.

Notes: The report in brackets is the clustering standard error adjusted at the major level; The control variables are the same as those in columns (3), (6) and (9) of Table 3. *** p<0.01, ** p<0.05, * p<0.1.

4.2. Change of Empirical Model Setting

First of all, in order to exclude the possibility that the benchmark results come from artificially selected clustering levels, this paper conducts clustering adjustment on standard errors at the college level and repeats the benchmark analysis. The corresponding regression results are reported in columns (4) - (6) of Table 5. The results show that the size and significance of the coefficient of the variables of concern have not changed, which verifies the robustness of the benchmark results.

	The hig lib	hest score for admi eral arts and scien	Clustered to college level			
Variables	GPA (1)	Rank (2)	CET-4 (3)	GPA (4)	Rank (5)	CET-6 (6)
Ethnic minorities	-0.131***	0.075***	-0.286***	-0.130***	0.072***	-0.281***
	(0.021)	(0.013)	(0.046)	(0.022)	(0.013)	(0.046)
Controls	YES	YES	YES	YES	YES	YES
Ν	9026	9026	8526	9084	9084	8584
R2	0.358	0.279	0.296	0.358	0.278	0.295

Table 5. Robustness test - replacing variable indicators and improving model.

Notes: The reports in brackets in columns (1) - (3) are robust standard errors clustered to the major level; The reports in columns (4) - (6) are robust standard errors clustered to the school level; The control variables are the same as those in columns (3), (6) and (9) of Table 3. *** p<0.01, ** p<0.05, * p<0.1.

Secondly, the benchmark model did not control the parents' income due to data limitations. Some studies have found that health and employment status are highly correlated with income (Barkindo et al., 2018; Schneider, 2018). In order to investigate whether the omission of parents' income will lead to estimation bias, the author further controls father's health status, mother's health status, father's employment status and mother's employment status on the basis of the benchmark model. The employment status of father (or mother) is defined as a binary dummy

variable. If the employment status of father (or mother) is "working", the value is 1, and the value is 0 in other cases. The father (or mother)'s health status includes three options: health, serious illness and disability, which are defined as binary dummy variables. If the father (or mother)'s health status is "healthy", the value is 1, and the value is 0 in other cases. Repeat the benchmark analysis, and the corresponding regression results are reported in columns (1) - (3) of Table 6. The regression results still support the conclusion that the academic performance of minority students is significantly lower than that of Han students.

Finally, the national culture, educational policy and economic level of different regions may change over time. Omission of these characteristics may lead to estimation errors. Although the author has controlled grade fixed effects, municipal fixed effects and their interactive fixed effects in the benchmark analysis, controlling regional characteristics at different levels may produce different results. In order to exclude the possibility of this subjective choice of significant results, the author replaced the regional fixed effects with provincial fixed effects and provincial * grade fixed effects and repeated the benchmark analysis. The corresponding regression results are reported in columns (4) - (6) of Table 6. The results still show that the benchmark conclusion remains unchanged.

Variables	Join parents	s for employmen	t and health	Control the fixed effect of provincial and provincial grade			
Variables	GPA (1)	Rank (2)	CET-4 (3)	GPA (4)	Rank (5)	CET-6 (6)	
Ethnic minorities	-0.146***	0.072***	-0.305***	-0.158***	0.086***	-0.318***	
	(0.030)	(0.014)	(0.052)	(0.013)	(0.007)	(0.038)	
Controls	YES	YES	YES	YES	YES	YES	
Ν	5993	5993	5684	9084	9084	8618	
R2	0.351	0.280	0.309	0.254	0.170	0.181	

Table 6. Stability test - change of empirical model setting.

Notes: The report in brackets is the clustering standard error adjusted at the major level; Columns (1) - (3) control municipal fixed effects and municipal grade fixed effects, and columns (4) - (6) control provincial fixed effects and provincial grade fixed effects. Other control variables are the same as those in columns (3), (6) and (9) of Table 3. *** p<0.01, ** p<0.05, * p<0.1.

4.3. Replacement of regression samples

First of all, in order to eliminate the interference of outliers on the benchmark analysis, the author deleted the first and last 1% of the observed values of GPA, major, CET-4, liberal arts college entrance examination and science college entrance examination. Columns (1) - (3) of Table 7 report the regression results after censoring; Secondly, samples from regions with developed education and economy (such as Beijing, Tianjin, Shanghai and Chongqing) may interfere with the estimation results. These regions have more abundant educational resources and better academic performance. In order to exclude the influence of municipalities' samples on the regression results, this paper deleted the samples of four municipalities and re-estimated the benchmark model. The results are shown in columns (4) - (6) of Table 7. Finally, Guangxi Zhuang Autonomous Region, Yunnan Province and Guizhou Province are not only the provinces with the largest number of ethnic minorities in China, but also the top three provinces in terms of the proportion of ethnic minorities in the national population. So, is the benchmark result mainly driven by the sample of these three provinces? In order to verify the robustness of the results, the author removed the samples from the three provinces and re-estimated the model (1). The corresponding results are reported in columns (4) - (6) of Schedule A2. To sum up, no matter how the regression sample is changed, the minority coefficient is positive and significant at the level of 1%, thus verifying the robustness of the benchmark results. In addition, in order to ensure the comparability between minority and Han students with a small number of students, this paper uses the tendency score matching method (PSM) to estimate and find that the conclusions are unchanged.

	Delete the te	op and bottom	1% samples	Remov	Remove four municipalities			
Variables	GPA	Rank	CET-4	GPA	Rank	CET-6		
	(1)	(2)	(3)	(4)	(5)	(6)		
Ethnic minorities	-0.106***	0.075***	-0.168***	-0.171***	0.091***	-0.345***		
	(0.017)	(0.011)	(0.039)	(0.013)	(0.007)	(0.036)		
Controls	YES	YES	YES	YES	YES	YES		
Ν	8692	8763	8239	7931	7931	7508		
R2	0.346	0.277	0.287	0.260	0.168	0.187		

However, due to space limitations, this part of the results were not reported.

Notes: The report in brackets is the clustering standard error adjusted at the major level; The control variables are the same as those in columns (3), (6) and (9) of Table 3. *** p<0.01, ** p<0.05, * p<0.1.

5. Heterogeneity analysis

How to explain the fact that the academic performance of minority students is significantly lower than that of Han students? This section explains the benchmark conclusion based on the heterogeneity analysis results. It can be seen from the results in Table 3 that academic performance is mainly affected by individual and regional characteristics. Therefore, the impact of ethnic identity on academic performance may also be affected by gender, ethnic category, provinces and regions, and thus shows heterogeneity in sub-sample regression. In order to test the heterogeneity of the above factors, this paper uses the sub-sample regression method to verify.

5.1. Gender heterogeneity

We divided the sample into female and male samples and repeated the benchmark regression analysis. Columns (1) - (3) and (4) - (6) of Table 8 reported the OLS estimation results of female and male samples respectively. The results of Column (1) and Column (4) show that the GPA difference between minority girls and Han girls is significantly lower than that between minority boys and Han boys. However, the results of Column (2), Column (3) and Column (5), Column (6) show that the difference in major and grade four performance between ethnic minority girls and Han girls is significantly greater than that between ethnic minority boys and Han boys. One possible explanation is that the school is a liberal arts college, and the gender ratio is seriously unbalanced. The proportion of male students in the data is only 32%. The competition between ethnic minority girls and Han girls is more intense, which not only leads to a more convergent GPA difference than the male sample, but also makes the difference in major between ethnic minority girls greater than that of male students (Condon et al., 2017). In addition, the undergraduate administration data show that the average grade four scores of boys and girls are -0.22 and 0.19 standard deviations, respectively. It can be seen from the "Matthew Effect" that girls' performance is better, and their polarization may be more serious. To sum up, if GPA is used to measure academic performance, ethnic differences in academic performance are mainly caused by differences among girls.

		Female			Male	
Variables	GPA	Rank	CET-4	GPA	Rank	CET-6
	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic minorities	-0.128***	0.085***	-0.293***	-0.147**	0.065	-0.211*
	(0.021)	(0.012)	(0.058)	(0.056)	(0.039)	(0.116)
Controls	YES	YES	YES	YES	YES	YES
Ν	5850	5850	5500	2660	2660	2520

Table 8. Heterogeneity Analysis by Sex.

R2	0.331	0.257	0.285	0.382	0.296	0.339
Notes: The report in b	prackets is the clus	stering standard	error adjusted at	the major level; '	The control varial	bles are the same

as those in columns (3), (6) and (9) of Table 3. *** p<0.01, ** p<0.05, * p<0.1.

5.2. Ethnic heterogeneity

This article only reserves the ethnic groups whose number of minority students is greater than or equal to 30. According to the number of students, they are Tujia, Zhuang, Hui, Manchu, Miao, Mongolian, Yi, Uygur, Qiang and Kazak from high to low, and the corresponding number of students is 118, 108, 87, 75, 62, 56, 48, 36, 36 and 34 respectively. Then, the above samples of minority and Han students were retained, and ten new samples were obtained and the benchmark regression was repeated. Table 9 reports the corresponding regression results, and the following conclusions can be drawn:

First, the difference in academic performance between Tujia, Zhuang, Hui, Manchu, Miao and Mongolian students and Han students is lower than the benchmark results. This may be because these ethnic groups have a higher degree of sinicization and a deeper understanding of the Han language and culture. Second, there is no statistically significant difference in academic performance between the students of Hui, Manchu and Miao nationalities and the students of Han nationality, indicating that the students of these three nationalities have reached the same level of knowledge as the students of Han nationality. Only the Hui nationality may be affected by their own language and culture, and there is a significant difference between them and the Han nationality in their CET-4 scores, but this difference is far lower than the benchmark result of 0.28. Third, the differences between Kazak, Uygur, Yi and Qiang students and Han students are significantly greater than the benchmark results. The difference between Yi, Qiang and Han students is significantly greater than the benchmark results, but the values are very close. The GPA, major and CET-4 scores of Kazakh students were 0.96 standard deviations, 39.4% and 1.95 standard deviations lower than those of Han students, respectively, which were 7.4, 5.5 and 6.9 times of the benchmark difference. The GPA, major and CET-4 scores of Uyghur students were 0.9 standard deviations, 25.7% and 2.31 standard deviations lower than those of Han students, respectively, which were 6.9, 3.6 and 8.2 times of the benchmark difference. The reason for these amazing results may be that Kazaks and Uighurs live in the north and south of Xinjiang respectively, and have their own languages - Kazakh and Uighur, which are quite different from Chinese (Wang Jian et al., 2011). Ethnic differences in language and culture will further lead to ethnic differences in academic performance (Xu Yayun et al., 2017). To sum up, the difference in scores between Kazakh, Uyghur and Han students is the main reason for the benchmark difference.

Table 9. Heterogeneity Analysis by Nationality.

	GPA	Rank	CET-4
	(1)	(2)	(3)
Tujia nationality	-0.121***	0.090***	-0.144
Zhuang Nationality	-0.110*	0.065*	-0.087
Hui nationality	-0.061	0.019	-0.188**
Manchu	-0.031	0.019	-0.007
Miao nationality	-0.003	0.029	-0.131
Mongolian	-0.088*	0.090**	-0.076
Yi Nationality	-0.157*	0.084*	-0.525**
Uygur ethnic group	-0.896***	0.257***	-2.308***
Qiang Nationality	-0.146	0.115**	-0.307**
Kazak	-0.959***	0.394***	-1.951***

Notes: The report in brackets is the clustering standard error adjusted at the major level; The control variables are the same as those in columns (3), (6) and (9) of Table 3. *** p<0.01, ** p<0.05, * p<0.1.

5.3. Provincial heterogeneity

Only the top ten provinces in terms of the number of minority students are retained in this paper: Inner Mongolia Autonomous Region, Liaoning Province, Hubei Province, Hunan Province, Guangxi Zhuang Autonomous Region, Chongqing City, Sichuan Province, Guizhou Province, Yunnan Province and Xinjiang Uygur Autonomous Region. Then OLS regression is carried out for the samples of these ten provinces, and the estimated results are reported in Table 12. We found that: (1) There was no significant difference in academic performance between ethnic minority students and Han students from Inner Mongolia Autonomous Region, Liaoning Province, Hunan Province and Guizhou Province. This may be because the main ethnic minorities in these provinces are Mongolian, Manchu, Tujia and Miao. According to Table 9, these four ethnic groups are the four ethnic groups with the lowest difference in scores compared with Han students. (2) The estimated result that the native place is Chongqing is significantly lower than the benchmark result, which may be because the Tujia people account for the largest proportion of the minority population in Chongqing. Although the difference in GPA between ethnic minority students and Han students in Sichuan Province is also lower than the benchmark result, the difference in major ranking and CET-4 scores is significantly higher than the benchmark result. The reason may be that Qiang and Yi people account for the largest proportion of minority population in the sample of Sichuan Province. It can be seen from Table 9 that the ranking differences between the two ethnic students and Han students are greater than the benchmark results, while the differences in English scores are far greater than the benchmark results. The difference between the scores of ethnic minority students and Han students from Guangxi Zhuang Autonomous Region is close to the benchmark results. (3) The difference in academic performance between ethnic minority students from Xinjiang Uygur Autonomous Region, Yunnan Province and Hubei Province and Han students far exceeds the benchmark results. Xinjiang Uygur Autonomous Region has the largest number of Uyghur and Kazak people, Yunnan Province has the largest number of Yi people, and Hubei Province has the largest number of Tujia people (the degree of Hanization of Hubei Tujia people may be much lower than that of Hunan Tujia people). To sum up, the differences in academic performance between ethnic minority students and Han students are mainly driven by the samples from Xinjiang Uygur Autonomous Region, Yunnan Province and Hubei Province.

	GPA	Rank	CET-4
	(1)	(2)	(3)
Inner Mongolia Autonomous Region	0.000	0.029	0.107
Liaoning Province	0.043	-0.005	-0.553
Hubei province	-0.309*	0.179**	-0.087
Hunan Province	-0.142	0.081	-0.052
Guangxi Zhuang Autonomous Region	-0.134*	0.086*	-0.161
Chongqing City	-0.056*	0.047*	-0.055
Sichuan Province	-0.079*	0.076**	-0.327***
Guizhou Province	-0.115	0.066	-0.409**
Yunnan Province	-0.359***	0.211***	-0.457*
Xinijang Uygur Autonomous Region	-0 737***	0 260**	-1 522***

	Table 10.	Heterogeneity	Analysis	by Province.
--	-----------	---------------	----------	--------------

Notes: The report in brackets is the clustering standard error adjusted at the major level; The control variables are the same as those in columns (3), (6) and (9) of Table 3. *** p<0.01, ** p<0.05, * p<0.1.

5.4. Regional heterogeneity

In this paper, the samples are divided into three regions according to the native provinces of students. The three regions were regressed by samples, and the corresponding regression results were reported in Table 11. The results show that the difference between the scores of ethnic minority students and Han students from the western

region is significantly higher than the benchmark results. The possible reason is that China's ethnic minorities are characterized by small settlements and large mixed settlements in geographical space, and they are mainly distributed in the western region. Differences in language, culture and social networks make it more difficult for ethnic minorities in these regions to integrate into society than in the eastern and central regions (Wang&Zhao, 2021). The regression results of the east central sample are not significant, indicating that there is no significant difference between the academic performance of ethnic minority students and Han students in the east-central region. The reason may be that the ethnic minorities in the eastern and central regions are highly sinicized, and ethnic minority families pay more attention to their children's education, which leads to a small difference in academic performance between ethnic minority students and Han students in these regions. To sum up, the difference in academic performance between ethnic minority students and Han students in the western region is the main reason for the benchmark results.

	GPA	Rank	CET-4
	(1)	(2)	(3)
Eastern region	-0.014	-0.004	0.006
	(0.054)	(0.032)	(0.168)
Central region	-0.097	0.069	-0.052
	(0.063)	(0.041)	(0.133)
Western region	-0.150***	0.084***	-0.345***
	(0.024)	(0.012)	(0.047)

Table 11. Heterogeneity Analysis by Region.

Notes: The report in brackets is the clustering standard error adjusted at the major level; The control variables are the same as those in columns (3), (6) and (9) of Table 3. *** p<0.01, ** p<0.05, * p<0.1.

6. Mechanism analysis

This section further explains the benchmark results from the perspective of mechanism inspection. National identity may affect academic performance through "peer effect". For example, national identity may affect academic performance through roommate level peer effect and professional level peer effect. In order to verify the above mechanism, the following model is adopted for mechanism verification:

 $Y_{imsgp} = \alpha + \delta minority_i + \beta X_{imsgp} + \rho M_{imsgp} + \eta minority_i * M_{imsgp} + \gamma_m + \tau_s + \pi_g + \varphi_p + \varphi_p * \pi_g + \varepsilon_{imsgp}(2)$

Among which, the definitions of *Y*, minority and *X* are consistent with model (1). *M* are variables, which are measured by the average score of roommates and the average cultural distance between roommates. Coefficient of interaction term η Is the parameter of interest in this paper, if η If the estimated value of is statistically significant, it indicates that the mechanism is established.

6.1. Average score of roommates

The better the performance of individual or class level students, the better the academic performance of minority students (Min et al., 2019). The ethnic minority students with better roommates may be affected by the learning atmosphere in their dormitories and work harder at their own studies, thereby improving their academic performance and narrowing the difference with the Han students. In order to verify whether the mechanism is established, this paper calculates the average score of roommates (excluding themselves) at the bedroom level, and re-estimates the model (2). The results are reported in Table 12. The results in columns (1) and (3) show that the coefficient of the average score of ethnic minorities * roommates in the interactive item is positive and significant at the 5% level. This shows that the increase of roommate's average score is one of the effective ways to narrow the difference between minority students and Han students.

Variables	GPA	Rank	CET-4
	(1)	(2)	(3)
Ethnic minorities	-1.175**	0.077	-1.983**
	(0.464)	(0.228)	(0.813)
Average score of roommates	0.021***	-0.011***	0.003
	(0.002)	(0.001)	(0.004)
Ethnic minorities * Average	0.013**	-0.000	0.021**
score of roommates	(0.006)	(0.003)	(0.010)
Controls	YES	YES	YES
Ν	9006	9006	8508
R2	0.377	0.290	0.296

Fable 12. Mechanism Test - Average Score of Roommates.

Notes: The report in brackets is the clustering standard error adjusted at the major level; The control variables are the same as those in columns (3), (6) and (9) of Table 3. *** p<0.01, ** p<0.05, * p<0.1.

6.2. Cultural distance of roommates

The cultural distance between minorities and major groups will increase the difficulty of minority cultural adaptation, which will have a negative impact on their academic performance (Bat 1 r & Okur, 2016) . Narrowing the cultural distance between ethnic minority students and their roommates will help reduce the difficulty of cultural integration, thereby improving their academic performance and narrowing the difference with Han students. To test this mechanism, referring to the research methods of Fan and Huang (2017), this paper uses the distance between provincial capitals to measure cultural distance. We calculate the average cultural distance between ethnic minority students and their Han roommates according to the following steps: (1) At the dormitory level, calculate the spherical distance between any two students' provincial capitals according to the longitude and latitude of their provincial capitals;⁸ (2) Calculate the average distance from Han roommates and add one to get the logarithm. Then re-estimate the model (2), and the corresponding regression results are reported in Table 13. The results in columns (1) and (2) show that when the dependent variable is GPA and major, the coefficient of logarithm of ethnic minority * cultural distance in the interactive item is positive and negative respectively. This means that the increase of cultural distance will expand the GPA and ranking differences between ethnic minority students and Han students; that is, cultural distance is one of the mechanisms of the benchmark results.

Variables	GPA	Rank	CET-4
	(1)	(2)	(3)
Ethnic minorities	0.401***	-0.150*	0.889
	(0.127)	(0.077)	(0.706)
Logarithm of cultural distance	0.013	-0.009*	-0.001
	(0.008)	(0.005)	(0.018)
Ethnic minorities * Logarithm	-0.076***	0.032***	-0.170
of cultural distance	(0.018)	(0.011)	(0.102)
Controls	YES	YES	YES
Ν	9006	9006	8508
R2	0.361	0.280	0.296

Notes: The report in brackets is the clustering standard error adjusted at the major level; The control variables are the same as those in columns (3), (6) and (9) of Table 3. *** p<0.01, ** p<0.05, * p<0.1.

⁸ The result of calculating cultural distance based on the straight-line distance between provincial capitals is consistent with this, and this part of the result is not reported due to space limitations.

7. Further analysis

The above analysis shows that the academic performance of minority students is significantly lower than that of Han students. However, this is a static relationship based on the analysis of mixed-section data. The author pays more attention to the dynamic relationship between ethnic identity and academic performance, that is, the time-varying trend of the differences in academic performance between ethnic minority students and Han students. Therefore, the author defines time trends as two types: (1) semester time trends. That is, under the same sample conditions, the difference between the scores of ethnic minority students and Han students changes with the semester; (2) Year time trend. That is, under different sample conditions, the change trend of the difference between the scores of ethnic minority students and Han students minority students and Han students with years.

7.1. Semester Trends

In order to investigate the change rule of academic performance difference between ethnic minority students and Han students over the term, the empirical model is set as follows:

$$Y_{imsgp,t} = \alpha + \delta minority_i + \rho Y_{imsgp,t-1} + \beta X_{imsgpt} + \gamma_m + \tau_s + \pi_g + \varphi_p + + \varphi_p * \pi_g + \varepsilon_{imsgpt}$$
(3)

Where, $Y_{imsgp,t}$ means the GPA or major ranking of students i from grade g major m school s of city p in semester t, and the definitions of other variables are consistent with model (1).⁹ δ is the focusing parameter of this paper, which means the difference of GPA (or major ranking) between ethnic minority students and Han students in semester t and t - 1 under the condition that other factors remain unchanged, that is, the speed of change of differences in scores of ethnic minority students in a unit semester.¹⁰ Table A3 and A4 report the specific regression results of model (3).

Based on Attached Tables A3 and A4, the regression coefficient and significance level of ethnic minorities in each semester when the dependent variable is GPA and major ranking are visualized. The results are summarized in Figure 1. We find that: first, ethnic differences in academic performance persist. The differences in GPA and major between ethnic minority students and Han students persist in different semesters, and show a gradually expanding trend; Second, the first-order difference of national differences in student performance converges gradually over time. The increase rate of this difference is the largest in the first two semesters, and it gradually decreases with the semester. The difference between the fifth semester and the sixth semester is not statistically significant. This may be related to the urban location of the school. The school is located in a new first tier city dominated by Han people. Minority students may not only face different language environments in their study and life, but also be inferior to Han students in terms of customs and family education resources. The rate of change of this difference caused by the "Matthew effect" was the most significant in the first two semesters (Kermanshachi&Safapour, 2017), and continued to exist in the subsequent semesters, and converged to the sixth semester.

 $^{^{9}}$ When t = 1 (the first semester), the scores of the last semester are replaced by the scores of the college entrance examination.

¹⁰ This paper only uses the GPA scores and professional rankings of the first six semesters, not the GPA scores and professional rankings of the seventh and eighth semesters. The reason is that in the last two semesters, there were few courses (only one or two courses per semester), and these courses were basically non professional compulsory courses. The dependent variable of CET-4 is not considered here. Because students will not take the CET-4 every semester, and the school only provides students with the highest score of CET-4 during their school years.



Figure 1. The first order difference in GPA and ranking of students from different nationalities changes with the semester.

Notes: * * * p<0.01, * * p<0.05, * p<0.1.

7.2. Annual trends

In order to investigate the change trend of academic achievement difference of students of different nationalities with years, this paper uses econometric model (1) to regress the annual samples respectively, and the corresponding estimation results are reported in Schedule A5. Similarly, the regression coefficient and significance level of ethnic minorities in each grade when the dependent variable is GPA and major ranking are visualized. The results are summarized in Figure 2. The basic law of the results is: the difference between the scores of ethnic minority students and Han students has gradually expanded since 2010 and narrowed again after 2012.

How to explain the result? This is related to the time node when the author collected data and the lack of data. First of all, the author obtained all the data of undergraduate administration in August 2016. At that time, students in 2013, 2014 and 2015 had just completed their junior, sophomore and freshman courses respectively, and only students in 2010-2012 had completed all the four-year college courses. Therefore, the regression coefficients of the 2013, 2014 and 2015 samples correspond to the GPA and major differences between minority students and Han students in the first three academic years, the first two academic years and the first academic year, respectively. Therefore, the results of these three grades cannot represent the entire college life. The difference in scores of ethnic students in 2013 narrowed sharply and was not statistically significant. This may be due to a large number of missing values in the national identity indicators of the sample in 2013. In 2013, the sample size was 514, accounting for only 5.7% of the analysis samples. If the regression results of this year are not taken into account, the GPA difference between ethnic minority students and Han students tends to expand gradually with the year, which is basically consistent with the conclusion in Figure 2. However, the difference in major remains roughly the same over the years.



Figure 2. Trends of GPA and Major Ranking Differences of Students from Different Nationalities over the Years. *Notes:* * * * *p*<0.01, * * *p*<0.05, * *p*<0.1.

8. Summary

This paper examines the relationship between students' national identity and academic performance, the mechanism of action and the time trend using the administrative data of a liberal arts finance and economics college from 2010 to 2015. The results showed that the GPA, major and CET-4 scores of ethnic minority students were significantly lower than those of Han students. The results are still stable after replacing key variable indicators, empirical models and regression samples. The differences in academic performance between ethnic minority students are mainly driven by the fact that the gender is female, the ethnic group is Kazakh or Uyghur, the native place is Xinjiang Uygur Autonomous Region, Yunnan Province or Hubei Province, and the samples from the western region. The mechanism analysis shows that the increase of average scores of roommates and the decrease of cultural distance between roommates are the main ways to narrow the differences in academic performance are always significant. The vertical trend analysis shows that although the difference tends to expand with the year and semester, its speed tends to converge.

This research involves the two themes of high-quality education system construction and ethnic equality proposed by the Chinese government, which will provide empirical basis for improving the quality of education in ethnic areas, promoting the unity and common prosperity of all ethnic groups and other political goals. First, the common requirement of building a high-quality education system and ethnic equality is to narrow the ethnic differences in students' academic performance. Peer effect and cultural distance are the main reasons for this difference; that is, the improvement of roommate performance and the reduction of cultural distance can effectively reduce the difference in ethnic education output. Secondly, to narrow the differences in ethnic education policies. For ethnic minority girls, Kazakhs or Uighurs, Xinjiang Uygur Autonomous Region, Yunnan Province or Hubei Province and other provinces, as well as ethnic minority students from the western region, we should formulate appropriate ethnic education policies and increase investment in education, especially the relevant systems related to the accommodation arrangements for ethnic minority students at the school level. We should not only try to create a better peer environment for them, but also take full account of the cultural distance between roommates, so as to

achieve the goal of narrowing the differences in the scores of ethnic students. Only in this way can we further improve the equalization level of ethnic education on the basis of improving China's high-quality education system.

Funding Statement

This research received no external funding.

Conflict of interest

All the authors claim that the manuscript is completely original. The authors also declare no conflict of interest.

Appendix

Grade	Number of minority students	Number of Han students	Proportion of minority students (%)
	(1)	(2)	(3)
2010	355	3616	8.94
2011	365	3344	9.84
2012	427	3544	10.8
2013	111	1203	8.45
2014	468	3263	12.5
2015	430	3320	11.5

A1. Proportion of Minority Students in Different Years.

Notes: The data are from the full sample data of undergraduate administration.

A2. Robustness Test - Replacement Capacity Index and the Three Provinces with the Largest Proportion of Nationalities Removed.

The minimum score for arts and s		n score for admis arts and sciences	sion of liberal	Remove Guangxi, Yunnan and Guizhou		
Variables	GPA (1)	Rank (2)	CET-4 (3)	GPA (4)	Rank (5)	CET-6 (6)
Ethnic	-0.130**	0.074***	-0.288***	-0.164***	0.087***	-0.337***
minorities	(0.021)	(0.013)	(0.045)	(0.016)	(0.009)	(0.055)
Controls	YES	YES	YES	YES	YES	YES
Ν	8988	8988	8491	8357	8357	7936
R2	0.357	0.277	0.294	0.246	0.168	0.176

Notes: The report in brackets is the clustering standard error adjusted at the major level; The control variables are the same as those in columns (3), (6) and (9) of Table 3. *** p<0.01, ** p<0.05, * p<0.1.

A3. Trend of GPA Differences of Ethnic Students	Changing with Semester.
-------------------------------------------------	-------------------------

		Average GPA of each semester				
Variables	GPA1	GPA2	GPA3	GPA4	GPA5	GPA6
	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic minorities	-0.105***	-0.053***	-0.031*	-0.031**	-0.029*	-0.027
	(0.016)	(0.018)	(0.017)	(0.014)	(0.015)	(0.021)
Controls	YES	YES	YES	YES	YES	YES
Ν	7332	7330	7327	7323	7246	7188
R2	0.373	0.661	0.658	0.696	0.601	0.594

Notes: The report in brackets is the clustering standard error adjusted at the major level; The control variables are the same as those in columns (3), (6) and (9) of Table 3. *** p<0.01, ** p<0.05, * p<0.1.

	Ranking of each semester					
Variables	Rank1	Rank2	Rank3	Rank4	Rank5	Rank6
	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic minorities	0.066***	0.037***	0.013	0.010	0.019*	0.013
	(0.012)	(0.011)	(0.009)	(0.008)	(0.010)	(0.011)
Controls	YES	YES	YES	YES	YES	YES
Ν	7332	7330	7327	7323	7246	7188
R2	0.265	0.579	0.622	0.660	0.583	0.580

A4. Trend of Major Ranking Differences of Ethnic Students Changing with Semester.

Notes: The report in brackets is the clustering standard error adjusted at the major level; The control variables are the same as those in columns (3), (6) and (9) of Table 3. *** p<0.01, ** p<0.05, * p<0.1.

A5. Trends of National Students	s' Performance	Differences v	with Years.
---------------------------------	----------------	---------------	-------------

	GPA (1)	Rank (2)
2010	-0.066**	0.071***
2011	-0.102**	0.064**
2012	-0.144***	0.089***
2013	-0.008	0.032
2014	-0.125**	0.064**
2015	-0.184**	0.067**

Notes: The report in brackets is the clustering standard error adjusted at the major level; The control variables are the same as those in columns (3), (6) and (9) of Table 3. *** p<0.01, ** p<0.05, * p<0.1.

References

- Barkindo, H. M., Shehu, I. Y., & Aliyu, M. M. (2018). Relationship between Parents' Socio-economic Background, Parental Involvement and Students' Academic Performance in Higher Institutions of Learning in Adamawa State, Nigeria. Proceedings of ICGET, 25-29.
- Blatt, L., & Votruba-Drzal, E. (2021). District-Level School Choice and Racial/Ethnic Test Score Gaps. American Educational Research Journal, 0002831221999405.
- Boen, C., Keister, L., & Aronson, B. (2020). Beyond Net Worth: Racial Differences in Wealth Portfolios and Black– White Health Inequality across the Life Course. Journal of health and social behavior, 61 (2), 153-169.
- Cassidy, A. (2020). Ethnicity, Parenting Styles, and Adolescent Health Behaviours Doctoral dissertation, University of Glasgow.
- Condon, B. P., Worley, P. S., Condon, J. R., & Prideaux, D. J. (2017). Student Academic Performance in Rural Clinical Schools: The impact of Cohort Size and Competition. Medical teacher, 39 (3), 262-268.
- Fahle, E. M., Reardon, S. F., Kalogrides, D., Weathers, E. S., & Jang, H. (2020). Racial Segregation and School Poverty in the United States, 1999–2016. Race and Social Problems, 12 (1), 42-56.
- Fang, Y., Wang, Q. (2019). A Study of Higher Education Equality and Its Realization Paths for Ethnic Minorities in China: from the Perspective of Equal Access to Education, Journal of Research on Education for Ethnic Minorities 30 (02): 42-49. (in Chinese).
- Henry, D. A., Betancur Cortés, L., & Votruba-Drzal, E. (2020). Black–White Achievement Gaps Differ by Family Socioeconomic Status from Early Childhood through Early Adolescence. Journal of Educational Psychology, 112 (8), 1471.
- Hirata, G., & Soares, R. R. (2020). Competition and the Racial Wage Gap: Evidence from Brazil. Journal of Development Economics, 146, 102519.
- Kermanshachi, S., & Safapour, E. (2017). Assessing Students' Higher Education Performance in Minority and Nonminority Serving Universities. In 2017 IEEE Frontiers in Education Conference (FIE) (pp. 1-6). IEEE.
- Kuhfeld, M., Condron, D. J., & Downey, D. B. (2021). When Does Inequality Grow? A Seasonal Analysis of Racial/Ethnic Disparities in Learning from Kindergarten through Eighth Grade. Educational Researcher, 50 (4), 225-238.

- Ma, L., Bu, S. (2019). The Effects of the Independent Addmission Policy of Talents Selection in Chinese Elite Universities, Peking University Education Review 17 (02): 109-126+190. (in Chinese).
- Nguyen, H. T., Connelly, L. B., Le, H. T., Mitrou, F., Taylor, C. L., & Zubrick, S. R. (2020). Ethnicity Differentials in Academic Achievements: The Role of Time Investments. Journal of Population Economics, 33, 1381-1418.
- Niemesh, G. T., & Shester, K. L. (2020). Racial Residential Segregation and Black Low Birth Weight, 1970–2010. Regional Science and Urban Economics, 83, 103542.
- Pilkington, A. (2018). The Rise and Fall in the Salience of Race Equality in Higher Education. In Dismantling Race in Higher Education (pp. 27-45). Palgrave Macmillan, Cham.
- Pratiwi, I. A., Zulianti, H., & Rohimah, S. (2020). The Influence of Read Cover Remember Retell (RCRR) towards Students' Reading Ability at the First Semester of Eleventh Grade of Mas Nurul Islam Jati Agung in 2020/2021. Journal of English Education Students, 1 (1), 1-19.
- Schindler, D., and Westcott, M. (2021). Shocking Racial Attitudes: Black GIs in Europe, The Review of Economic Studies, 88 (1), 489-520.
- Schneider, B. (2018). Parents, Their Children, and Schools: An Introduction. In Parents, their children, and schools (pp. 1-12). Routledge.
- Sutton, M. Y., Anachebe, N. F., Lee, R., & Skanes, H. (2021). Racial and Ethnic Disparities in Reproductive Health Services and Outcomes, 2020. Obstetrics and Gynecology, 137 (2), 225.
- Veldkamp, S. A. M., Zondervan-Zwijnenburg, M. A. J., van Bergen, E., Barzeva, S. A., Tamayo-Martinez, N., Becht, A. I., ... & Hartman, C. (2020). Parental Age in Relation to Offspring's Neurodevelopment. Journal of Clinical Child & Adolescent Psychology, 1-13.
- Wang, C., Xiao, A. (2019). Social Networking and Academic Performance Within Classroom: An Experimental Study of Random Seating, China Economic Quarterly, 18 (03): 1123-1152. (in Chinese).
- Wang, J., Zhang, H., Li, Z., Yin, H., Zhang, Z., (2011). Comparative Study of Teacher-student Interaction under Multinationality Cultural Background: Taking the Classroom Teaching in Han, Tibetan, Hui, and Uighur Regions as Examples, Educationa research, 32 (09): 68-75. (in Chinese).
- Wang, Z., and Zhao, G. (2021). Cultural Integration among Non-Han Migrants in Urban China: To What Extent does Intermarriage Matter, Growth and Change, 00, 1–23.
- Wu, J., Wu, G., Li, B. (2021). Can Early-life Health Input Boost Children's Long-life Cognitive Ability, China Economic Quarterly, 21 (01):157-180. (in Chinese).
- Wu, Y., Zhang, F. (2020). The Health Cost of Attending Higher-achievement Schools: Peer Effects on Adolescents' Academic Performance and Mental Health, Educationa research, 41 (07): 123-142. (in Chinese).
- Xu, Y., A, H., Yan, H., Yu, B., Chen, X., Li, S. (2017). Influential factors of acculturative stress among Xinjiang Kazak students studying in inland cities of China, Chinese Journal of Public Health, 33 (04): 642-646. (in Chinese).