The Impact of Family Size and Siblings on Chinese Students' Academic Performance

Meiling Guo

Nanyang Technological University - School of Social Science, 639798, Singapore

guomeiling01@qq.com

Keywords: family size, sibling composition, educational economics

Abstract: This paper aims to examine the impact of family size and siblings on the academic performance of 19487 Chinese middle school students. The researcher believes that the siblings' effect is nonnegligible in children's intellectual development. Thus, on the one hand, the number and sex of siblings are included when measuring the family size effect. On the other hand, the sibling composition effect will be tested under each family size. The findings indicate that academic performance is first positive but later negative correlated with family size, and the optimal family size is 2. In addition, siblings can benefit children's academic learning in general, but having younger sisters makes a boy get a relatively poor test score.

1. Introduction

The effect of family size on educational input has long been a popular research field. One of the most famous models is the quantity and quality tradeoff introduced by Becker, and numerous empirical researches on it draw different conclusions. There are some reasons for the differentiates: first, the premise of quantity and quality tradeoff is that the resources are fixed, but some parents having more children will work harder and reduce entertainment expenditure, so more resources can be used for children's education. Second, interactions among siblings are ignored in this model. Older siblings always serve as the teachers to younger siblings, which not only stimulates the intellectual development of elder siblings but also saves the parents' time to take care of the young. Third, whether the quantity and quality tradeoff are prominent depends on the macro circumstances of the country. Lee (2008) believed that if a country's public education system and social welfare system are functioning well, the burden of raising children will be released, so the quantity and quality tradeoff is not apparent[1].

To date, most of the family size effect studies have been conducted in the West. There are very few published studies that report on the family size effect in the China context due to One-Child Policy, which was implemented for decades. However, this restriction on family size has been released since 2016, and increasing Chinese families have more than one child.

One task of this research is to explore the relationship between family size and academic scores among Chinese students aged 13-15. besides controlling individual, family, and school characteristics, the impact of siblings should also be considered. Another task is to examine the effect of siblings on academic achievement under each family size.

This paper unfolds as follows. Section 2 presents the existing empirical evidence. Section 3, 4, and 5 describe our method, data, and regression analysis, respectively. Section 6 concludes.

2. Literature review

Previous research findings into family size have been inconsistent and contradictory. Taking private tutoring expenditure in Korea as a measurement, Kang (2011) has argued that family size has a strong negative impact on girls' educational investment but has little effect on boys[2]. In contrast to Kang, Qian (2004) has suggested that the arrival of the second child increased the enrollment rate of the first girl by 8%-17% in China[3]. He used the idea that rural families that

  DOI: 10.25236/ssehr.2021.027

firstborn girls can enjoy the relaxations in China's One-Child Policy and have another baby. Additionally, in Norway, Black et. Al. (2004) found that family size does not affect education after controlling for birth order[4].

To eliminate the endogeneity of the family size, researchers have come up with various instruments. Several studies used twins as instrument (e.g. Rosenzweig, 1980; Black, 2010; MAKINO, 2017). Conley (2004) assumed that parents with two children of the same sex are more likely to have a third child than equivalent parents with two children of the opposite sex.

The development of children's cognitive not only relies on the resources invested by parents but also the interaction between siblings. Parish et, AI. (1993) found that elder siblings, especially elder sisters, might replace the mother's role to some extent[5]. Ha and Tam (2011) believed that elder siblings always serve as intellectual resources to younger siblings, which stimulates the intellectual development of elder siblings. While the youngest and only children never have the chance to become the "teacher", so they do not perform as well as firstborn in academics[6]. Hence, the impact on academic scores brought by siblings' interaction should also be controlled when measuring the family size effect. Momoe Makino (2016) used the samples from India to detect whether there exists gender bias, especially to later-born girls. She found that once the girls survive to primary school age, they would be treated equally. Instead, girls seem to be at an advantage if they have elder sisters[7].

3. Method

The estimation equation of family size is presented as follows:

$$y_{ij} = \beta_0 + \beta_1 FamilySize_{ij}^2 + \beta_2 FamilySize_{ij} + \beta_3 X_{ij} + \beta_4 C_j + \beta_5 Sibling_{ij} + \varepsilon_{ij}$$

The y_{ij} is the standard test score. As for the family size effect, this research adopts two forms: nonlinear and binaries. X_{ij} is a vector of child i's individual and family background characteristics: father's educational background, whether the child lives with parents, family's wealth, sex and health condition. C_j describes the characteristics of class j: the ratio of students to teachers and the proportion of the students having rural hukou. $Siblings_{ij}$ describes the number of elder brothers, elder sisters, younger sisters. Children from larger families always have more siblings, so the number of siblings should be controlled. Given birth order is colinear with the sum of elder brothers and sisters, this paper doesn't control it. In addition, the impact of siblings on test scores is probably determined by the child's sex, so this paper sets the interactive term for sex and the number of siblings.

The estimation equation of sibling effect is presented as follows:

$$y_{ij} = \beta_0 + \beta_3 X_{ij} + \beta_4 C_j + \beta_5 Sibling_{ij} + \varepsilon_{ij} \quad \text{for each family size}$$

In order obtain a more precise estimation of the sibling effect, the researcher regresses the exact number of siblings under each family size. This idea origins from Black's paper, in which they regress birth order binaries under each family size[4]. This paper will divide birth order into the number of older brothers and sisters. Additionally, the number of younger siblings is involved because educating and taking care of the young develop one's psychological ability[8].

Given the endogeneity of the family size, this research adopts the sex of the first child and whether the first two children are of the same sex as instruments. The descriptions of instruments are presented in table 3-1.

Table 3-1 the description about instruments

Instrument	Validity	Description
Sex of the eldest child	Relevance	Parents whose first child is a girl are more likely to have a second child than parents whose first child is a boy.

	Exogeneity	Parents' preference does not determine the sex of the fetus, and Chinese parents cannot know the sex of the fetus before birth.
Whether the first and second child are of the same sex	Relevance	Parents having two children of the same sex are more likely to have a third child than equivalent parents having two children of the opposite sex
	Exogeneity	The same as "sex of the eldest child"

4. Data

The data was released by the China Education Panel Survey, a nationally representative survey project designed and implemented by the National Survey Research Center at Renmin University of China. This survey uses the average educational level of the population and the proportion of the floating population as stratified variables. 112 schools and 438 classes are randomly selected from 28 county-level units across the country. All students in the selected classes are sampled. It aims to reveal the impact of family, school, community, and macro social environment on personal educational output.

The characteristics of dependent and independent variables are demonstrated in table 4-1. The test score has been standardized to a normal distribution with a mean of 70 and a standard deviation of 10.

Table 4-1 Characteristics of variables

variable	N	mean	std	max	min
standard Chinese score	19001	70	10	98	-2
family size	19487	2	0.86	13	1

Table 4-2 indicates that the sample size decreases as family size increases. In order to get a representative regression result, the samples from family size greater than 5 will be dropped due to their small sample sizes.

Table 4-2 Frequency distribution of family size

Family size	frequency	percentage
1	8755	44.93
2	8043	41.27
3	2068	10.61
4	436	2.24
5	113	0.58
>5	72	0.37
total	19487	100

Several individual and family characteristics – sex, health condition, family income, father's education, the parental company – are added as control variables. The school characteristics – the ratio of students to tutor and the proportion of the students having rural hukou – are also involved. Additionally, the number of siblings should be controlled.

The father's education is reported on a 0-8 scale where 0 indicates that the father is illiterate, and 8 indicates that the father has a master degree or higher. The parental company is stratified as a 0-2 scale where 0 indicates that the child is left behind, and 2 indicates that both parents accompany the child. The family income is classified as a 0-4 scale from very poor to very rich. The health condition is divided into a 0-4 scale from very poor to very good.

Table 4-3 Descriptive statistics for control variables

	observations	size1	size2	size3	size4	size5

father's education	19487	4.02	2.73	2.44	2.36	2.24
		(2.20)	(1.77)	(1.63)	(1.63)	(1.58)
parental company	19487	1.76	1.62	1.50	1.48	1.49
		(0.55)	(0.71)	(0.78)	(0.78)	(0.78)
family income	19487	2.00	1.86	1.72	1.70	1.62
		(0.66)	(0.66)	(0.68)	(0.69)	(0.76)
health	19487	3.08	3.02	2.88	2.88	2.92
		(0.93)	(0.92)	(0.97)	(0.95)	(1.02)
sex	19487	0.55	0.50	0.42	0.43	0.45
		(0.50)	(0.50)	(0.49)	(0.50)	(0.50)
students - tutor ratio	19487	47.35	48.66	50.78	51.49	48.66
		(12.83)	(12.46)	(13.40)	(13.59)	(11.77)
proportion of rural hukou	19487	0.41	0.65	0.69	0.68	0.60
		(0.28)	(0.25)	(0.23)	(0.24)	(0.25)
elder bro	8672		0.20	0.40	0.52	0.74
			(0.40)	(0.61)	(0.77)	(0.87)
elder sis	9071		0.35	0.81	1.41	1.71
			(0.48)	(0.74)	(1.00)	(1.02)
younger sis	8772		0.26	0.48	0.69	0.93
			(0.44)	(0.59)	(0.84)	(1.03)
younger bro	9155		0.40	0.64	0.74	0.91
			(0.49)	(0.59)	(0.66)	(0.75)

The father's education, parental company, and family income drop monotonously according to family size, and the number of siblings is positively correlated to the family size. Therefore, this research will add them as control variables to avoid interference with family size effects.

The proportion of boys decreases when the family size is less than 4, but the rule does not work due to the small sample size, which indicates that girls are more likely to come from a large family than boys, which is consistent with the conclusion drawn by Jensen and Clark - in a society with a preference for sons, parents continue to give birth until they have a boy.

5. Regression analysis

5.1 Family size effect

The regression results about family size are presented in the table 5-1(a) and 5-1(b).

Table 5-1(a) Use nonlinear form representing family size

	(1) without controls	(2) control individual and family	(3) control class	(4) control siblings	(5) use instruments
Family size	1.248***	1.382***	0.970**	1.156**	0.326
	(0.370)	(0.360)	(0.374)	(0.385)	(0.960)
Family size^2	-0.284***	-0.319***	-0.253**	-0.286***	-0.315*
	(0.0850)	(0.0806)	(0.0824)	(0.0847)	(0.154)
Sex		-5.961***	-5.986***	0.285***	0.267***
		(0.137)	(0.137)	(0.0404)	(0.0442)
Live with parents		0.208	0.276*	0.272*	0.237*
		(0.108)	(0.109)	(0.109)	(0.117)

father education		0.228***	0.283***	0.637***	0.616***
		(0.0372)	(0.0404)	(0.119)	(0.121)
Wealth		0.587***	0.642***	0.233**	0.228**
		(0.118)	(0.118)	(0.0779)	(0.0781)
Health		0.223**	0.236**	-5.616***	-5.771***
		(0.0778)	(0.0778)	(0.170)	(0.232)
students - tutor ratio			0.0150**	0.0151**	0.0169**
			(0.00543)	(0.00543)	(0.00574)
rural hukou			1.218***	1.189***	1.454***
			(0.299)	(0.299)	(0.400)
elder bro				0.0183	0.697
				(0.278)	(0.774)
elder sis				0.308	1.266
				(0.275)	(1.085)
younger sis				0.567*	1.503
				(0.264)	(1.043)
elder bro * sex				-0.551	-0.350
				(0.407)	(0.466)
Elder sis * sex				-0.615*	-0.737*
				(0.288)	(0.335)
younger sis * sex				-1.088**	-1.166**
				(0.374)	(0.392)
Constant	68.89***	68.96***	67.61***	67.28***	68.39***
	(0.359)	(0.545)	(0.622)	(0.631)	(1.308)
Observations	18930	18930	18930	18930	18930
adj. R-sq	0.001	0.096	0.098	0.098	0.096

Standard errors in parentheses

="* p<0.05 ** p<0.01 *** p<0.001"

Table 5-1(b) Use binaries representing family size

	(1)	(2)	(3)	(4)
	without controls	control individual and family	control class	control siblings
size1	2.792*	2.740*	2.846**	2.796*
	(1.144)	(1.067)	(1.066)	(1.243)
size2	3.116**	3.157**	3.039**	3.089**
	(1.144)	(1.065)	(1.065)	(1.186)
size3	3.103**	2.892**	2.732*	2.783*
	(1.161)	(1.081)	(1.081)	(1.146)
size4	2.591*	2.485*	2.340*	2.356*
	(1.238)	(1.150)	(1.149)	(1.162)
Sex		-5.962***	-5.986***	-5.616***
		(0.137)	(0.138)	(0.170)
Live with parents		0.208	0.276*	0.272*

		(0.108)	(0.109)	(0.109)
father education		0.228***	0.283***	0.284***
		(0.0372)	(0.0404)	(0.0404)
Wealth		0.587***	0.642***	0.637***
		(0.118)	(0.118)	(0.119)
Health		0.223**	0.237**	0.233**
		(0.0779)	(0.0779)	(0.0780)
students-tutor ratio			0.0149**	0.0151**
			(0.00543)	(0.00543)
rural hukou			1.222***	1.192***
			(0.299)	(0.300)
elder bro				0.0146
				(0.277)
elder sis				0.304
				(0.274)
younger sis				0.558*
				(0.263)
elder bro * sex				-0.543
				(0.407)
Elder sis * sex				-0.621*
				(0.288)
younger sis * sex				-1.075**
				(0.374)
Constant	67.09***	67.29***	65.49***	65.36***
	(1.139)	(1.112)	(1.174)	(1.336)
Observations	18930	18930	18930	18930
adj. R-sq	0.001	0.096	0.097	0.098

t statistics in parentheses

="* p<0.05 ** p<0.01 *** p<0.001"

Model(3) and (4) in table 5-1(a), the model(4) in table 5-1(b) can better represent the family size effect than others.

In table 5-1(a), the estimator of family size is always positive, and its square is negative, which indicates a parabola opening down. This parabola is shown in Fig 1 based on models (3) and (4). Model(4) adds siblings and the interaction term of siblings and sex based on model(3). It is apparent that the test score gets down in each family size after controlling the sibling effect, which means having siblings helps students perform better in learning. More precisely, having a younger sister increases the test score by 0.567 (α=0.05), a boy having an elder sister and a younger sister can decrease his test score by 0.615 (α=0.05) and 1.088 (α=0.01) respectively.

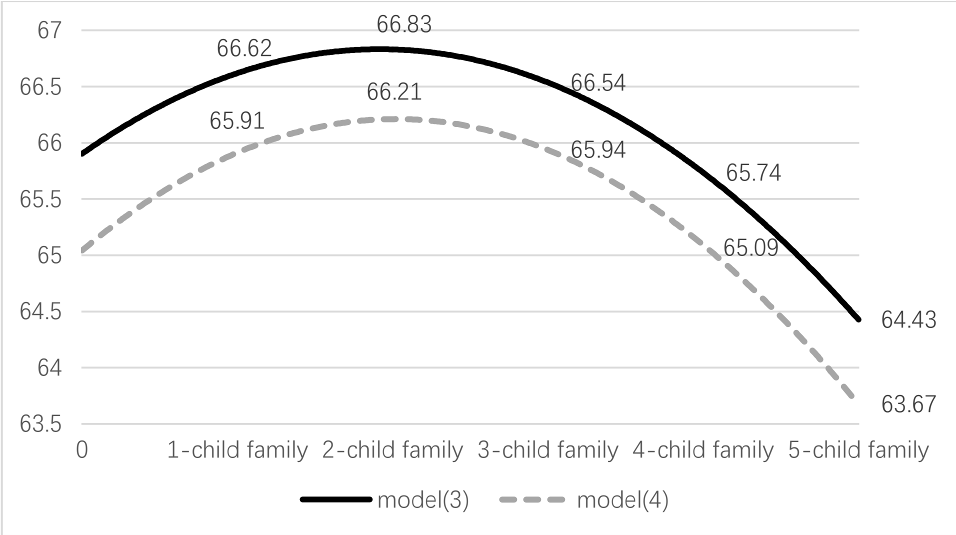

Fig 1. Nonlinear family size effect

Nevertheless, the impact of siblings is not that apparent when using binaries. As shown in Fig 2, curves of the model(3) and model(4) are almost the same, but the estimators of siblings are still significant: having a younger sister increases the test score by 0.558 (α=0.05), a boy having an elder sister and a younger sister can decrease his test score by 0.621 (α=0.05) and 1.075 (α=0.01) respectively.

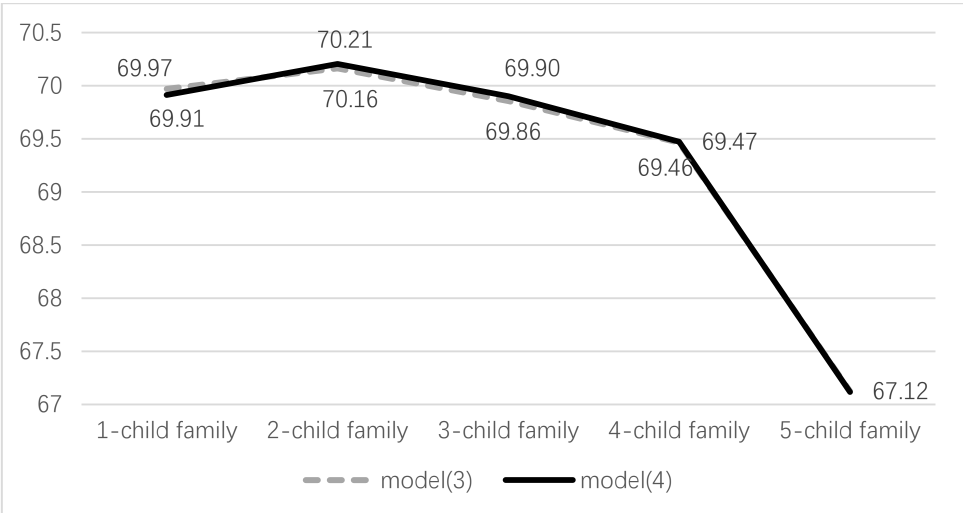

Fig 2. Family size effect using binaries

Fig 1 and Fig 2 reflect the same rule about family size: the optimal family size is 2 in China. This conclusion is interesting because only children should perform better than children in the family size of 2 according to quantity and quality tradeoff. However, it is the children from 2-child families that get the highest score on average, which indicates that even though only children can occupy the whole family's resources, they lose the opportunities to educate and take care of young siblings, and the benefit cannot cover the loss. Additionally, the educational investment is not strictly constrained because parents of 2 children will find ways like working harder and reducing entertainment expenditure. When family size expands to 3, the burden of rearing many children is so heavy that family sources limit the children's intellectual development, then quantity and quality tradeoff works.

Given that family size is endogenously determined by parents' preference, this paper presents the family size curve adding instruments in Fig 3. Disappointedly, the optimal family size is 0.52, far away from 2. This is because only the estimator of family size is not significant in model(5) and leads to the inaccuracy of the symmetry axis, and this is why the adjusted R square of the model(5) is not as high as that of the model(3) and (4).

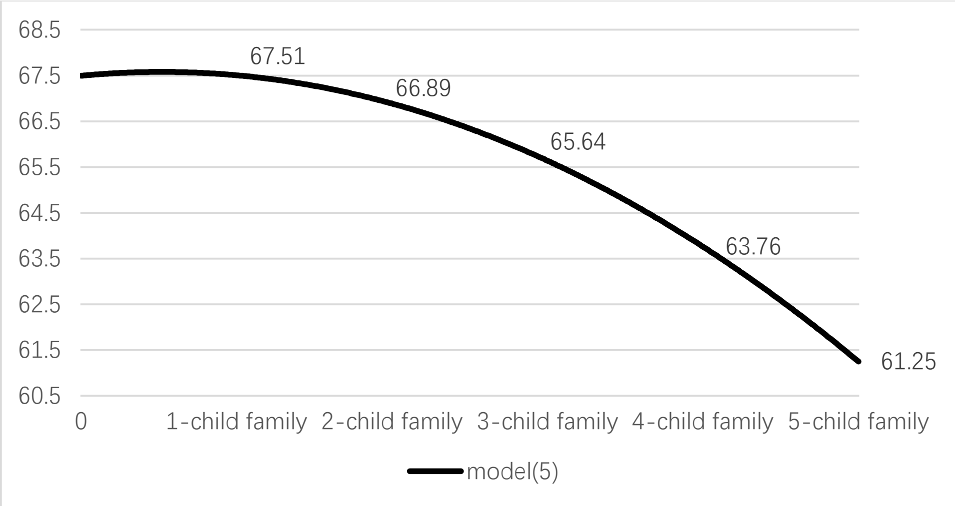

Fig 3. Family size effect using instruments

5.2 Sibling composition effect

The impact of siblings is nonnegligible in children's intellectual development, and the number of siblings is positively correlated with family size based on table 4-3. To extract the sibling composition effect from family size, the regression of sibling composition effect is under each family size, and the results in table 5-2 are presented in family size order

Table 5-2 Sex composition effect under each family size

	(1) 2-child family	(2) 3-child family	(3) 4-child family	(4) 5-child family
elder bro	0.0456	-0.386	1.552*	-0.132
	(0.12)	(-0.76)	(1.77)	(-0.06)
elder sis	0.299	-0.633	2.092*	2.065
	(0.81)	(-1.10)	(1.95)	(0.79)
younger sis	0.674*	-0.271	2.437**	1.962
	(1.87)	(-0.49)	(2.33)	(0.86)
sex	-5.561***	-4.724***	-3.247	-12.64
	(-14.49)	(-3.45)	(-0.81)	(-1.24)
elder bro * sex	-0.674	-1.361	-1.007	2.421
	(-1.01)	(-1.36)	(-0.59)	(0.69)
elder sis * sex	-0.413	-0.739	-2.126	1.081
	(-0.74)	(-0.82)	(-1.29)	(0.33)
younger sis * sex	-1.176**	-0.242	-4.948**	0.861
	(-1.96)	(-0.21)	(-2.26)	(0.23)
father's education	0.117*	0.00283	-0.734**	-0.852
	(1.73)	(0.02)	(-2.48)	(-1.24)
Live with parents	0.233	-0.109	1.169*	2.814*
	(1.50)	(-0.39)	(1.96)	(1.67)
Wealth	0.351*	0.655*	0.286	1.161
	(1.93)	(1.91)	(0.51)	(1.05)

Health	0.183	0.105	0.253	-0.203
	(1.50)	(0.46)	(0.50)	(-0.18)
students-tutor ratio	0.00373	-0.00118	-0.0123	0.0677
	(0.42)	(-0.08)	(-0.38)	(0.73)
rural hukou	1.632***	0.0796	0.990	0.237
	(3.45)	(0.08)	(0.49)	(0.05)
Constant	69.84***	72.16***	67.50***	58.67***
	(80.43)	(43.66)	(19.09)	(6.24)
Observations	7824	2002	419	108
adj. R-sq	0.094	0.102	0.205	0.143

t statistics in parentheses

="* p<0.1 ** p<0.05 *** p<0.01"

In a 2-child family, having a younger sister is conducive for girls' academic learning but bad for boys'. In a 4-child family, having an elder brother, an elder sister, a younger sister increases test scores by 1.552, 2.092, 2.437 points, respectively (α=0.1, 0.1, 0.05). However, a boy in a family of 4 children having an elder younger sister will drop his grades by 4.948 points (α=0.05).

Generally, the estimators of siblings who are significant in the regression are always positive, which indicates that having siblings benefits children's academic performance. However, having a younger sister is harmful to a boy's academic performance.

6. Conclusion

The optimal family size is two children, which is the turning point of the family size effect from positive to negative. This is not just because the only children cannot serve as "teachers" to educate the young, but also because the educational resources are not strictly constrained. Thus, children from the family size of 2 perform better than others. Nevertheless, when family size continues to expand, there is a decline in children's quality due to limited resources.

As for the sibling composition effect, having a younger sister is always harmful to boys, but in general, having siblings is conducive to children's academic performance because the significant estimators are always positive.

Reference

[1] Li, H., Zhang, J., & Zhu, Y. The quantity-quality trade-off of children in a developing country: Identification using Chinese twins. Demography, vol.45, no.01, pp.223-243, 2008.

[2] Kang, C. Family size and educational investments in children: Evidence from private tutoring expenditures in South Korea. Oxford Bulletin of Economics and Statistics, vol.73, no.01, pp.59-78, 2011.

[3] Qian, N. Quantity-quality and the one child policy: The only-child disadvantage in school enrollment in rural China (No. w14973). National Bureau of Economic Research. , 2009.

[4] Black, S. E., Devereux, P. J., & Salvanes, K. G. The more the merrier? The effect of family size and birth order on children's education. The Quarterly Journal of Economics, vol.120, no.02, pp.669-700, 2005.

[5] Parish, W. L., & Willis, R. J. Daughters, education, and family budgets Taiwan experiences. Journal of Human Resources, no.05, pp.863-898, 1993.

[6] Ha, T. S., & Tam, C. L. A study of birth order, academic performance, and personality. In International Conference on Social Science and Humanity, IPEDR, vol.5, no.01, pp.28-32, 2011.

[7] Makino, M. Birth order and sibling sex composition effects among surviving children in India: enrollment status and test scores. The Developing Economies, vol.56, no.03, pp.157-196, 2018.

[8] Zhao Dongmei, Liao Hongling.Mechanism of action of child birth order and parental parenting mode on the development of psychological theory. Journal of South China Normal University (Social Science Edition), no.01, pp.110-122, 2020.