A Study on Gender Stereotypes and Their Activation Effects in Academic Performance

Han Wu *

Chengdu Foreign Language School, Sichuan, 610031, China

* Corresponding Author Email: 1910821118@mail.sit.edu.cn

Abstract. Gender stereotypes in educational settings have emerged as an important issue. While many scholars have recognized the stereotypical beliefs regarding academic performance based on gender and the potential harm they may cause, there is still a lack of sufficient research exploring how these stereotypes specifically impact people's score estimations of students from different genders, as well as the triggering and reinforcement mechanisms of such stereotypes. Therefore, this study examined the experimental dataset of Daphne Lenders and Toon Calders, utilizing methods such as validity analysis, stratified chi-square tests, and visualization techniques. The findings revealed that in the absence of any stereotype activation, individuals tend to rate the academic performance of females higher than males. However, by manipulating stereotype activation, such as Case-based and Statistics, this paper observed variations in the intensity of bias: the difference in score estimations between genders significantly increased, particularly for the Statistics condition. This suggests the existence of inherent gender-based stereotypes in everyday life, which can be reinforced by the daily information received by people. The result helps to make the public aware of the intrinsic existence of performance gender stereotypes, increasing the possibility of mitigating such phenomenon, and point to the way for further research afterward.

Keywords: Gender stereotype; stereotype activation; academic performance; stratified chi-square tests.

1. Introduction

Gender biases in social settings have been a subject of significant interest and concern. Gender stereotypes are especially pervasive in educational field with children demonstrated serious gender biases in some subjects, as early as second grade [1, 2].

Currently, there is a prevailing stereotype in academic performance that girls tend to outperform boys and exhibit better study attitude. For instance, in 2011, Robinson and Lubienski's survey indicated that in junior high school, teachers consistently gave higher ratings to females in math and reading, despite cognitive assessments suggesting that males have an advantage [3]. Similarly, another study presented that although not remained uncontested, belief that language learning is predominantly associated with females is widespread. Generalist assumptions, such as "girls excel more in language learning compared to boys," are commonly held [4]. Moreover, gender stereotypes surrounding personality traits often perpetuate the notion that boys tend to be more mischievous, active, and playful, while girls are commonly believed to possess stronger self-control, focused attention, and a more diligent and serious attitude towards learning. It is widely believed that girls exert more effort, seriousness, and care in their academic pursuits [5].

In addition, research suggests that the formation of academic stereotypes may have multiple original sources. The study conducted by Jacobs examined the hypothesis that parental gender stereotypes in relation to mathematical ability have an interactive effect with child's gender, resulting in a direct influence on parents' beliefs about the child's mathematical proficiency and their expectations for future success in mathematics. Furthermore, it was postulated that these stereotypes have an indirect impact on the child's self-perceptions and overall performance in this field [6]. Meanwhile, parents' and educators' beliefs about children's grades frequently exhibit gender biases, which can exert a significant influence on children's attitudes towards and performance in study. These biases, stemming from societal expectations and preconceived notions, have the potential to shape children's perceptions and abilities in the study [7].

This ubiquitous stereotype surrounding academic performance exerts a profoundly deleterious influence on the growth and maturation of students, potentially imprinting lasting effects on individuals' scholastic achievements and others' scholarly anticipations [8]. Males who are perceived to have lower grades in such stereotypes face gender stereotype threat which is a phenomenon in which the activation of a self-relevant stereotype leads people to show stereotype-consistent behavior, thereby perpetuating the stereotypes [8]. The pernicious impacts of gender stereotype threat have been empirically demonstrated in numerous domains, such as performance in business negotiations, caregiving abilities, and driving skills. In essence, individuals who experience the threat of gender stereotypes tend to exhibit poorer performance compared to those unaffected by such threats [9]. Likewise, male students who find themselves under the influence of this stereotype regarding academic achievement tend to unconsciously undermine their own educational pursuits. Simultaneously, this ingrained gender stereotype regarding academic achievement has the potential to impact the support provided by teachers and parents towards students' learning, consequently indirectly influencing students' level of active engagement in their studies [5]. On the other hand, female students who face such gender stereotypes will bear comparatively higher academic expectations from teachers and parents, leading to greater pressure to meet the expectations of others than their male counterparts. If this stereotype persists, it can result in distorted learning conditions for both male and female students.

Currently, despite the extensive research on gender stereotypes in educational psychology, there are still some gaps that need to be filled. The studies on the triggering and moderating factors of stereotype effects on academic achievement are not yet comprehensive and profound enough. For instance, while Penner and Paret examined gender differences in mathematics achievement, there is still a need for further empirical research on how to effectively elicit and control the impact of stereotypes, as well as how these stereotypes specifically influence our predictions of others' performance on study [10].

In this paper, data analysis was performed on the experimental dataset of Daphne Lenders and Toon Calders, called "performance VS predicted performance" to investigate whether people have inherent biases against boys when evaluating their academic performance and whether these biases can be triggered or amplified through stereotype activation. In the main task, participants were presented with eight student profiles, each containing basic information about a student, and were asked to make grade predictions. Crucial manipulations involved swapping the gender labels of some profiles and stereotype activation. In this section, additional details about the study design will be performed. The results of the proof-of-concept study will also be presented, which will form the basis for data analysis and conclusions regarding the existence of gender biases in academic performance evaluations and the factors that may influence these biases, such as stereotype activation [11].

Not only could this paper let individuals aware of their subconscious gender stereotypes regarding academic achievement, but the findings from this study will also contribute to a better understanding of the factors influencing individuals' perceptions of students' performances and shed light on potential interventions to address gender biases in education.

2. Methodology

2.1. Data Sources

The data for this literature was collected from the Kaggle website, which was compiled by Daphne Lenders and Toon Calders from an experiment they conducted using an existing dataset provided by UCL Machine Learning on Kaggle. The dataset was published and last updated in 2023 with a total of 107 participants.

2.2. Variable Selection

In the experiment, Daphne Lenders and Toon Calders collected data through questionnaires. A total of 107 participants were assigned to evaluate 8 students each, with an equal distribution of 4

males and 4 females. Participants were randomly exposed to one of three stereotype activation (None, Casebased, Statistics). Following this, participants were presented with the basic information of eight students, including Name, Sex, Studytime, Freetime, Romantic, Walc, Goout, Parents_ed, absences, and reason. Based on this information, participants estimated the students' exam scores. The present study analyzes these varibales to validate the bias in judging academic performances based on gender, and further explores how this bias is strethened or triggered by different stereotype activation conditions. Table one provides the names, meanings, and value ranges of all independent and dependent variables, which will facilitate readers' understanding of the experiment and the subsequent discussion, as Table 1 shows.

Table 1. Variable description

Variable name	Meaning	Data range
Name	Name of each student	Each of the 8 student profiles, that participants had to grade & rank was randomly matched to one of four boy/girl's names (depending on the sex of the student)
Sex	The gender of each student	Female (F) or male (M)
Studytime	How long a student studied for their exam	1(less than 2 h), 2(2-5h), 3(more than 5h)
Freetime	How long a student doesn't study	1(very low), 2(moderate), 3(very high)
Romantic	whther a student was in a romantic relationship	Yes or No
Walc	How much alcohol each student consumes in the weekend	1(very low), 2(moderate), 3(very high)
goout	How often a student goes out in a week	1(very low), 2(moderate), 3(very high)
Parents_edu	The higher education of both parents	1-4 (rising trend)
absences	The number of absences per student	0-6(reprents the acurrate number), 7(more than 7 times)
reason	Why a student choose to go to the school	home, school's requtation, school's curricular, and other
G3	The actual grade	0-20
Pass	whether G3 >=10	True (>=10), False (less than 10)
Stereotype activation	Prior to making grade predictions, participants were exposed to one of three forms of stereotype activation, where two of them were meant to trigger stereotypes against boys.	None, CaseBased, Statistics
Predicted_Pass_PassFailStrategy	whether "Predicted Grade" was >=10 or not	True (>=10), False (less than 10)

2.3. Research Method

This study selected a dataset called "Student Alcohol Consumption" as the source of student basic information. While the original dataset contained more than 30 attributes to describe each student, only eight attributes were presented per profile to avoid overwhelming participants with information. The chosen attributes had high variability between students and could be legitimately or stereotypically associated with school performance. Validity tests were conducted to assess the correlation between the selected attributes and school performance, ensuring the effectiveness of the study design. All variables were validated to ensure the extraction of meaningful research data. Subsequently, line chart, pie chart and horizontal chart are drawn respectively to clearly show the age, gender and nationality distribution of participants. Furthermore, a stratified chi-square test was employed to verify whether there were significant differences between sex and estimated grades, and whether these differences varied under the influence of stereotype activation in the stratified variable. Classification summaries were used to calculate the average G3 scores and predicted grades for male and female students in three stereotype activation groups. To better illustrate and compare the results of grade estimation under different conditions, bar charts were created to show the average differences in G3 scores and predicted grades between males and females in different stereotype activation conditions. Through these methods, a comprehensive evaluation of the differences in grade estimation

between males and females under different stereotype activation conditions was conducted, and the biases in gender-based grade judgments were examined. These analysis methods and visualization tools provided strong evidence to support the research findings.

3. Results and Discussion

The study initially examined the validity of all independent and dependent variables to ensure the effectiveness and rationality of the questionnaire design, and to derive valid conclusions from the final data. To assess the construct validity of the data, factor analysis was conducted as a data analysis method. As Table 2 shows, this involved comprehensive analysis of indicators such as the KMO value, communalities, variance explained ratio, and factor loading coefficients. The KMO value is 0.658, which is used to assess the adequacy of information extraction, communalities were used to exclude items with insufficient validity, variance explained ratio was used to indicate the level of information extraction, and factor loading coefficients were used to measure the relationship between factors and items.

From Table 2, it can be observed that all researched items had communalities values higher than 0.4, indicating effective extraction of information related to the research items. Additionally, the KMO value was 0.658, which is greater than the threshold of 0.6, indicating effective data extraction. Moreover, the variance explained ratios for the five factors were 19.366%, 11.981%, 11.153%, 10.174%, and 8.695%, respectively, with a cumulative variance explained ratio of 61.368% after rotation, surpassing the 50% threshold. Factor loading coefficients exceeding an absolute value of 0.4 indicate a correspondence between options and factors.

Table 2. Validity analysis results

Name	factor load coefficient					Common degree
	factor1	factor2	factor3	factor4	factor5	
Sex	-0.189	0.130	-0.722	0.099	0.165	0.611
studytime	0.416	0.013	0.499	-0.373	-0.039	0.562
freetime	0.099	0.714	-0.115	-0.102	-0.124	0.588
romantic	-0.179	0.018	0.591	-0.039	0.323	0.487
walc	-0.442	0.490	-0.228	0.049	0.172	0.520
goout	-0.167	0.810	0.098	0.042	0.045	0.697
Parents_edu	0.119	0.005	-0.219	0.717	0.249	0.637
reason	-0.096	-0.060	0.111	0.675	-0.297	0.569
asences	-0.637	0.060	0.224	0.242	0.044	0.521
stereotype activation	0.070	-0.033	0.025	-0.025	0.837	0.708
Predicted grade	0.866	-0.072	0.147	0.147	0.074	0.803
Predicted_Pass_PassFailStrategy	0.816	-0.036	0.109	0.080	0.074	0.691
Feature root (before rotation)	2.646	1.303	1.246	1.136	1.033	-
Variance explanation rate % (before rotation)	22.049%	10.86%	10.380%	9.470%	8.608%	-
Cumulative variance explanation rate % (before rotation)	22.049%	32.909%	43.290%	52.760%	61.368%	-
Feature root value	2.324	1.438	1.338	1.221	1.043	-
Variance interpretation rate % (after rotation)	19.366%	11.981%	11.153%	10.174%	8.695%	-
Cumulative variance explanation rate %(after rotation)	19.366%	31.347%	42.500%	52.673%	61.368%	-

The age, rough nationality, and gender distribution of the participants have been presented using line graphs, bar charts, and pie charts, respectively. Figure 1 illustrates that the majority of participants fall within the age range of 18-34, indicating that they are in a stage of psychological maturity, possess rational thinking abilities, and have developed self-awareness and stable values. Hence, the results obtained from this population are likely to be representative and applicable to a broader context. Additionally, Figure 2 depicts the distribution of participants across different continents such as Europe, North America, and Asia. By collecting responses from participants from various continents,

the experimental data results have a higher degree of universality. Finally, Figure 3 showcases the distribution of participants based on their biological gender, without an extreme uneven distribution between males and females. The dissemination of relevant demographic information about the participants ensures the scientific rigor and authenticity of the final experimental results.

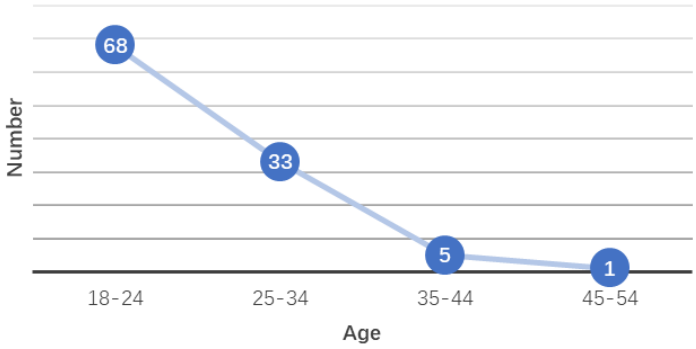

Fig. 1 Number of participants by Age

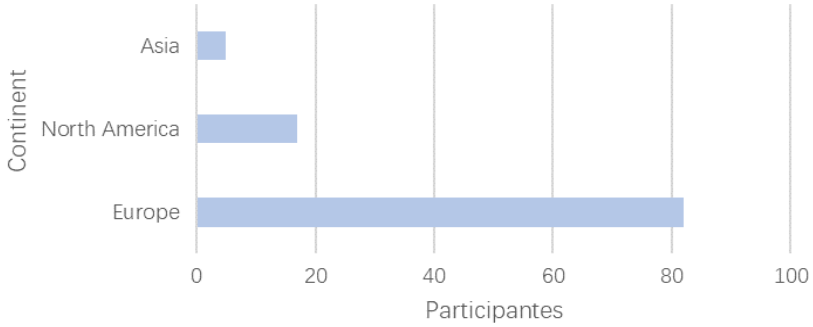

Fig. 2 Number of participants by continent

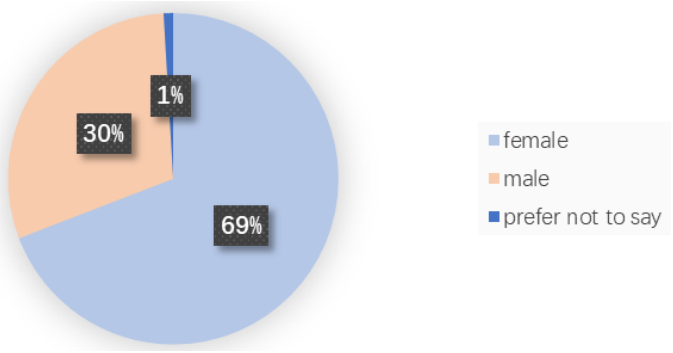

Fig. 3 Number of participants by gender

The findings presented above indicate the results of a stratified chi-square test, where sex serves as the independent variable (X), Predicted_Pass_PassFailStrategy as the dependent variable (Y), and stereotype activation as the stratifying factor. The summary provided in Table 3 shows a significant relationship between sex and Predicted_Pass_PassFailStrategy, suggesting that gender has a notable impact on the predicted pass rate (p=0.000). Notably, the predicted pass rate for females exceeds that of males, suggesting the presence of gender bias in grade prediction.

Table 4, which displays the Breslow-Day test for homogeneity, demonstrates significance (chi=9.615, p=0.008 < 0.05), indicating heterogeneity among the strata and the presence of confounding effects due to stratifying factors. Similarly, Table 5 exhibits the Cochran-Mantel-Haenszel test for conditional independence, which also shows significance (chi=29.170, p=0.000 < 0.05). This outcome suggests a substantial difference between sex and Predicted_Pass_PassFail Strategy after accounting for confounding factors. These two tables, taken together, reveal variations in score prediction for males and females under different stereotype activations. Thus, the findings imply that bias in predicting gender-based grades may be intensified or triggered by stereotype

activation. In the subsequent discussion, this paper will delve into the specific effects of different stereotype activations and perform intuitive comparisons and validations.

Table 3. Hierarchical Chi-square Inspection Summary Form

stereotype	Predicted_Pass_PassFail	FALSE		TRUE		$\chi 2$	p	OR 95% CI
	Sex	F	M	F	M			
Casebased		21	44	115	92	10.649	0.002	0.212～0.687
None		21	26	111	106	0.647	0.520	0.409～1.454
Statistics		11	47	149	113	27.291	0.000	0.088～0.358
Summary		53	117	375	311	30.065	0.000	0.263～0.537

Table 4. Breslow-Day-Tarone

$\chi 2$	df	p value
9.6148	2	0.0082

Table 5. Cochran–Mantel–Haenszel

ORvalue(Mantel-Haenszel Common)	95%CI	$\chi 2$	df	p value
0.3777	0.2643~0.539	29.1698	1	0.0000

According to Table 3, it is evident that women surpass men in every stereotype activation group in terms of pass rates. However, when the average actual scores of men and women in each group are computed and graphically represented in Figure 4, it becomes visually apparent that women consistently achieve higher actual scores than men across all groups. Consequently, in order to further validate the hypothesis conclusion with greater rigor, this study proceeds to conduct an additional stratified chi-square test.

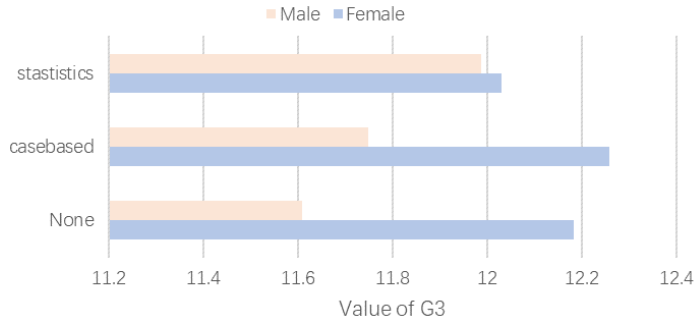

Fig. 4 Average value of G3 in different groups

Table 6. Breslow-Day-Tarone2

$\chi 2$	df	p value
null	16	null

Table 7. Cochran–Mantel–Haenszel2

ORvalue(Mantel-Haenszel Common)	95%CI	$\chi 2$	df	p value
0.3955	0.2754~0.568	null	1	null

The chi-square test was conducted to analyze the relationship between the independent variable "Sex", the dependent variable "Predicted_Pass_PassFailStrategy", and the stratification variable "G3". The homogeneity of the data was evaluated using the Breslow-Day test (Table 7), which revealed that there was no significant difference (chi=nan, p=nan >0.05) among the different layers. This indicates that the relationship between each layer is homogeneous and there is no confounding effect from the

stratification factors. Similarly, Table 8 (chi=nan, p=nan >0.05) also showed no significant difference between "Sex" and "Predicted_Pass_PassFailStrategy" after taking confounding factors into account. Therefore, it can be inferred that the interference of "G3" does not affect the relationship between gender and the predicted pass rate. In most cases, when the G3 of males and females is the same, the predicted pass rate for females is higher than that of males.

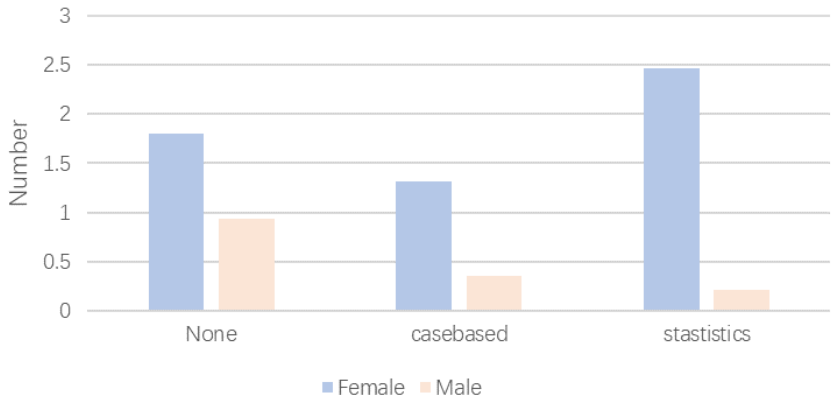

Fig. 5 The difference between the average of G3 and predicted grade in different conditions

In all categories, students obtained higher average predicted scores compared to the average actual scores. Analyzing the differences between the average predicted scores and the average actual score in different categories allows for valid conclusions to be drawn. First, let's examine the column with the stereotype activation set to None in Figure 5. It is evident that, without any external influences, the difference between the predicted average scores and the actual score for female students is higher than that for male students (with a mean difference of 0.871). This suggests an inherent bias in people's predictions of academic performance based on gender, indicating a belief that female students will generally outperform male students. Next, let's consider the columns with stereotype activation set to case-based and statistics. In the case-based group, the difference between different value of the predicted average scores and the actual scores for male and female students slightly increases (with a mean difference of 0.967). In the statistics group, the difference significantly increases (with a mean difference of 2.257). This implies that exposure to cases or data that reinforce biases about academic performance based on gender strengthens participants' stereotypes. Moreover, the influence of concrete data on gender stereotypes related to academic performance is more pronounced. Therefore, in our daily lives, we need to be mindful of potential gender-based biases regarding academic performance. As we encounter a multitude of information every day, it is essential to critically assess academic data or information and avoid generalizing specific cases to entire gender groups. Failure to do so will only fortify gender-based stereotypes concerning academic performance and have adverse effects on both male and female students.

4. Conclusion

Conducting validity analysis on the questionnaire content provided by Daphne Lenders and Toon Calders, performing a stratified chi-square test on the data results from the questionnaire, the result of our study indicates that gender has a substantial effect on pass rate of the grade, which provides compelling evidence to support the existence of fundamental biases in individuals' perceptions of academic performance based on gender. These biased perceptions indicate that a gender-based performance bias exists within our research population which could represent larger group of people in society.

Furthermore, adopting another stratified chi-square test and creating visualizations to represent certain questionnaire outcomes, the study found a great difference between sex and Predicted_Pass_PassFailStrategy after inventions of different stereotype activation: the differences between difference value of average predicted grade and G3 in two gender group increase, suggesting

that stereotype activation has the potential to reinforce these biases. Particularly noteworthy is the stronger impact of statistics-based activation compared to case-based activation in amplifying gender-based biases.

Therefore, this paper must remain vigilant against gender biases in academic performance. When confronted with academic information, it is crucial to evaluate it critically and avoid attributing the good or bad grades of students in individual cases to their gender group. These results help people realize their potential gender stereotype in perceptions of academic performance. By illustrating the malleability of bias through stereotype activation, our findings imply the existence of potential interventions to mitigate and counteract gender-based biases in academic contexts.

The implications of this research are significant for educational settings and policies. By fostering awareness and promoting critical thinking about biased perceptions, this paper can strive towards a fair and equal educational system that rewards merit and eliminates the influence of gender stereotypes on performance evaluations to avoid the suffering of stereotype threat by students. It is important to note that this study had certain limitations. It is difficult to say whether differences of predicted grade can completely be attributed to the various stereotype activation conditions, since participants in each condition got presented with different profiles and direct comparison for predictions on the same profiles is impossible. Future research can fill the gaps of this study by further investigating the effects of much diverse stereotype activations on individuals' stereotypes. Additionally, an analysis can be conducted to understand the reasons behind the varying intensities of effects produced by different activation conditions.

References

[1] Fiske S T. Taylor S E. Social cognition: From brains to culture. Sage, 2013.

[2] Cvencek D, Meltzoff A N, Greenwald A G. Math-gender stereotype in elementary school children. Child development, 2011, 82(3): 766-779.

[3] Robinson J P, Lubienski S T. The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratinfs. American Educational Research Journal, 2011, 48(2): 268-302

[4] Schmenk B. Language learning: A feminine domain? The role of stereotyping in constructing gendered learner identities. TESOL quarterly, 2004, 38(3): 514-524.

[5] Dpng Bo. A study on the impact of Gender stereotype identification on Mathemactics Achievement of junior Middle School students. East China Normal University, 2019.

[6] Tiedeman J. Parents' gender stereotypes and teachers' beliefs as predictor of children's concept of their mathematical ability in elementary school. Journal of Educational psychology, 2000, 92(1): 144.

[7] Bleeker M M, Jacobs J E. Achievement in math and science: Do mothers' beliefs matter 12 years later. Journal of Educational Psychology, 2004, 96(1): 97.

[8] Dar-Nimrod I, Heine S J. Exposure to scientific theories affects women's math performance. Science, 2006, 314(5798): 435-435

[9] Shi Linxi. The impact of gender stereo type threat on male and female Mathematics Achievement and its countermeasures. China Science and Education Innovation Guide, 2012(14): 147.

[10] Penner A M, Paret M. Gender differences in mathematics achievement: Exploring the early grades and the extremes. Social Science Research, 2008, 37(1): 239-253

[11] Lenders D, Calder T. Real-life Performance of Fairness Interverntions-Introducing A New Benchmarking Dataset for Fair ML. Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, 2023, 350-357.