International Journal of Mathematics and Statistics
ISSN (Online): 2208-2212
Volume 09 Issue 04 December 2023
DOI: https://doi.org/10.69980/sbp2cr78

STATISTICAL MODELING OF ACADEMIC ACHIEVEMENT: ASSOCIATIONS BETWEEN STUDENT CHARACTERISTICS AND EXAMINATION PERFORMANCE

Dr. Sophia L. Bennett[1]*, Prof. Daniel A. Ferreira[2]

[1]Department of Educational Statistics and Quantitative Assessment, Westmoor University, Bristol, United Kingdom*
[2]Institute of Educational Research and Applied Data Science, Atlantic University of Social Sciences, Lisbon, Portugal

Abstract:
Academic achievement is a result of the interplay of demographic, socioeconomic, and educational variables, and may differ from domain to domain. Publicly available examination records of 1000 students were used to perform a quantitative cross-sectional analysis to examine relationships between student characteristics and examination performance in mathematics, reading, and writing. The following statistical methods were used: frequencies, percentages, means, standard deviations, independent-samples t-tests, one-way analysis of variance, Pearson correlation and multiple linear regression. Mean scores were 66.09 ± 15.16 for mathematics, 69.17 ± 14.60 for reading, and 68.05 ± 15.20 for writing. The females had a higher overall mean score than the males (69.57 vs 65.84) while the males scored more on mathematics. The overall score for students who received the standard lunch meal and students who finished test preparation was significantly higher than for comparison groups. There were also substantial differences in academic achievement by race/ethnicity and parental education level. Positive correlations of all the three subjects were found with the highest correlation between reading and writing ($r=0.955$, $p<0.001$). Male gender, completion of the test preparation and membership in groups D and E were positively associated with mathematics scores and the categories of free/reduced lunch and high school and some high school parental-education were negatively associated under the adjusted mathematics model. Model accounted for 26.3% of the variation in score. In general, multiple student-related factors were correlated with academic achievement and the high inter-correlations of subject scores indicated a strongly inter-related nature of the learning outcomes. The results reinforce the value of using learning patterns through a multi-variate statistical analysis to guide targeted academic intervention.

Keywords: *academic achievement; examination performance; student characteristics; mathematics scores; socioeconomic factors; multiple linear regression*

1. Introduction:

Academic performance is generally accepted as a measure of students' learning growth and educational progress, especially in terms of subject marks achieved in standardized tests. However, examination results are influenced by more than this. Student performance in various subjects can be affected by the school context, family context or access to educational support. There is evidence that the quality of teacher–student relationships and the socioeconomic environment of the school can make significant contributions to academic achievement differences [1]. This is also supported by wider international studies which have shown that socioeconomic status works on multiple dimensions and is a significant predictor of educational achievement in various societies [2].

Socio-economic influences can be seen at the student, school and national level. The levels have been shown to be interrelated, through comparative evidence derived from reading and mathematics achievement, to influence outcomes of learning [3]. It was also found that a socioeconomic status is a consistent predictor of learning in an umbrella review of the literature conducted [4]. Gender has also been associated with reading and numeracy achievement differences within each school year besides the socioeconomic variables [5]. However, this does not mean that the gender gap in achievement is consistent: broader socioeconomic and cultural contexts may lead to a different picture of the relationship between gender and achievement [6].

The mathematics, reading, and writing scores offer complementary perspectives on students' academic performance. These skills are connected, but each also has a small component of different aspects of learning. Studies have revealed that cognitive processes which underlie reading abilities are also related to mathematical abilities [7]. Other mechanisms that have been associated with the relationship between reading comprehension and mathematics include inferencing and attentional control [8]. Early education has also revealed that reading and mathematical abilities occur as a result of within domain and across domain relationships [9]. Findings indicate that the scores obtained by the subjects should not be measured individually without taking into account their overall performance.

Additionally, student background characteristics can also affect performance in ways that are not measured in exam scores. International assessment research has indicated that part of the link between socioeconomic status and academic achievement is due to cognitive ability [10]. Another key variable that might matter could be educational preparation. There have been found performance differences in high-stakes examinations that have been linked to test-preparation activities, although the relationship can be different depending on context and nature of test preparation [11]. State test results have also revealed that preparation efforts can impact achievement on tests and should be taken into account when evaluating test results [12].

Although many studies have focused on the effects of single factors on achievement, relatively few simple statistical studies have looked at the interrelationships among achievement in mathematics, reading, and writing, as well as the interrelationships among demographic, socioeconomic, and preparation-related factors. The statistical approach can differentiate simple group differences from association that are not explained by other student characteristics. The objectives of the present study were to: describe mathematics, reading, and writing scores, examine differences between mathematics, reading, and writing scores by gender, race/ethnicity, parental education, lunch category, and test preparation status, assess correlation between mathematics, reading, and writing scores, and determine the independent association between student characteristics and mathematics scores with multiple linear regression.

2. Materials and Methods

2.1 Study Design and Source

The study employed quantitative cross-sectional method to investigate the factors associated with academic achievement. The analysis was based on a publicly available collection of students observations, with 1000 observations from students. Every record contained details of one student, including reading, maths and writing scores and educational and demographic details. There were no direct personal identifiers, and there was no interaction with students or collection of new data during the study.

2.2 Study Variables and Outcome Measures

Mathematics, reading, and writing exam scores were used to measure academic achievement on a scale of 0–100. For each student an overall score was made by averaging the scores from the three subjects. Explanatory characteristics consisted of gender, race/ethnicity group, parental level of education, lunch category, and test preparation course completion. The variables were chosen in order to investigate the available demographic, socioeconomic, and educational variables that were reasonably thought to examine the link with examination performance.

2.3 Data Preparation and Descriptive Analysis

Prior to analyzing all records were checked for missing, incorrect, and inconsistent scores. Categorical data were summarized by frequencies and percentages, and examination scores were summarized by means and standard deviations, medians, minimum and maximum scores, and interquartile ranges. All necessary values were present, and then the three subject scores were added together to obtain the overall academic performance score. Simple graphical summaries were also used to explore the shape, spread, and extreme observations in score distributions.

2.4 Inferential Statistical Analysis

Standard inferential procedures were used to examine differences in academic performance by student characteristic. Characteristics that had two categories such as gender, lunch category, and test-preparation completion were analyzed using independent-samples t-tests. The factor of race/ethnicity had more than two groups, thus one-way analysis of variance was used for this factor, as were parental education. Strength and direction of relationship between Mathematics, Reading and Writing scores were measured using Pearson correlation coefficients. All statistical tests were performed in a two-tailed way and $p<0.05$ was regarded as statistically significant.

2.5 Multiple Linear Regression Analysis

Multiple linear regression was used to examine the independent associations between student characteristics and mathematics achievement. Gender, race/ethnicity, parental education, lunch category and test-preparation completion were all included together as explanatory variables. The tests of linearity and residual patterns were done and variance consistency was checked using residual plots and the Breusch-Pagan test. Multicollinearity was checked using the adjusted generalized variance inflation factors (GVIF). Unusual or influential observations were examined using standardized residuals and Cook's distance. The standardized residuals from seven observations were found to be below the value of -2.5 and were omitted prior to fitting the final model. The regression coefficients, standard errors, significance values, R^2, and adjusted R^2 were reported.

3. Results

3.1 Characteristics of the Student Sample

All 1000 student records were analysed, there were no missing data for the eight variables. Females accounted for 518 students (51.8%) and males for 482 (48.2%). Race/ethnicity comprised group A, 89 (8.9%); B, 190 (19.0%); C, 319 (31.9%); D, 262 (26.2%); and E, 140 (14.0%). Some college (22.6%) and an associate's degree (22.2%) were the most common levels of parental education. Sixty-four and a half percent (645) of students reported receiving standard lunch, and 35.5 percent (355) of students reported receiving free/reduced lunch. There were 358 students (35.8%) who finished test preparation (Table 1).

Table 1. Demographic, Socioeconomic, and Educational Characteristics of the Student Sample

Characteristic	Category	n	%
Gender	Female	518	51.8
	Male	482	48.2
Race/ethnicity	Group A	89	8.9
	Group B	190	19.0
	Group C	319	31.9
	Group D	262	26.2
	Group E	140	14.0
Parental education	Some college	226	22.6
	Associate’s degree	222	22.2
	High school	196	19.6
	Some high school	179	17.9
	Bachelor’s degree	118	11.8
	Master’s degree	59	5.9
Lunch category	Standard	645	64.5
	Free/reduced	355	35.5
Test preparation	Completed	358	35.8
	None	642	64.2

Note: n=1,000; no missing values were identified across the eight variables.

3.2 Distribution of Examination Performance

Mean examination scores were 66.09 ± 15.16 for mathematics, 69.17 ± 14.60 for reading, and 68.05 ± 15.20 for writing. The three achievement domains were similar across the medians, which were 66, 70 and 69. The mathematics score was 0-100, reading 17-100 and writing 10-100. The overall academic performance was 67.77 ± 14.26, median 68.33 and interquartile range 58.33–77.67. The average performance for Reading was the highest, followed by Writing and Mathematics (Table 2, Figure 1).

Table 2. Descriptive Statistics of Mathematics, Reading, Writing, and Overall Academic Performance Scores

Measure	Mean ± SD	Median	IQR	Minimum-Maximum
Mathematics	66.09 ± 15.16	66.00	57.00-77.00	0-100
Reading	69.17 ± 14.60	70.00	59.00-79.00	17-100
Writing	68.05 ± 15.20	69.00	57.75-79.00	10-100
Overall academic performance	67.77 ± 14.26	68.33	58.33-77.67	9-100

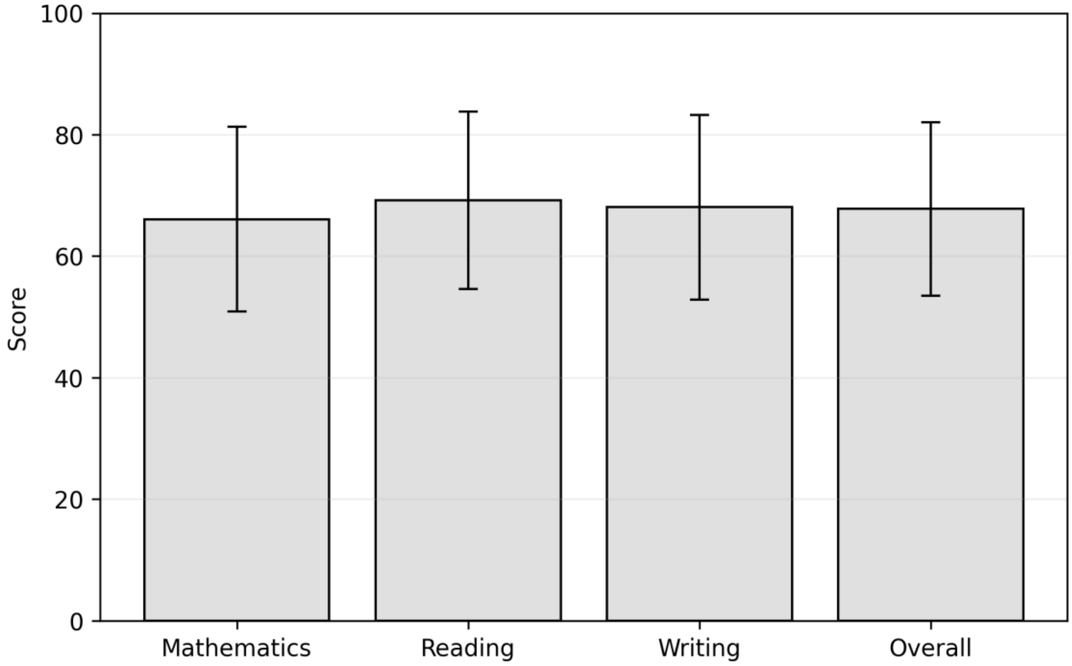

Figure 1. Mean Examination Scores with Standard Deviations.

3.3 Differences in Academic Achievement Across Student Characteristics

There were substantial variations between the student characteristics in overall achievement. Females had a higher composite mean than males (69.57 vs 65.84; $t(998)=4.17$, $p<0.001$), although males scored higher in mathematics (68.73 vs 63.63). Standard-lunch students outperformed free/reduced-lunch students (70.84 vs 62.20; $t(998)=9.58$, $p<0.001$), while test-preparation completers outperformed non-completers (72.67 vs 65.04; $t(998)=8.39$, $p<0.001$). Composite performance also differed by race/ethnicity, $F(4,995)=9.10$, $p<0.001$, and parental education, $F(5,994)=10.75$, $p<0.001$. Table 3 shows that the higher means were found in group E and master's degree households (Figure 2).

Table 3. Comparison of Overall Academic Performance Across Student Characteristics

Factor	Group means (overall score)	Test statistic	p-value
Gender	Female 69.57; Male 65.84	t(998)=4.17	<0.001
Lunch category	Standard 70.84; Free/reduced 62.20	t(998)=9.58	<0.001
Test preparation	Completed 72.67; None 65.04	t(998)=8.39	<0.001
Race/ethnicity	A 62.99; B 65.47; C 67.13; D 69.18; E 72.75	F(4,995)=9.10	<0.001
Parental education	Some college 68.48; Associate’s 69.57; High school 63.10; Some high school 65.11; Bachelor’s 71.92; Master’s 73.60	F(5,994)=10.75	<0.001

Note: Overall academic performance was calculated as the arithmetic mean of mathematics, reading, and writing scores.

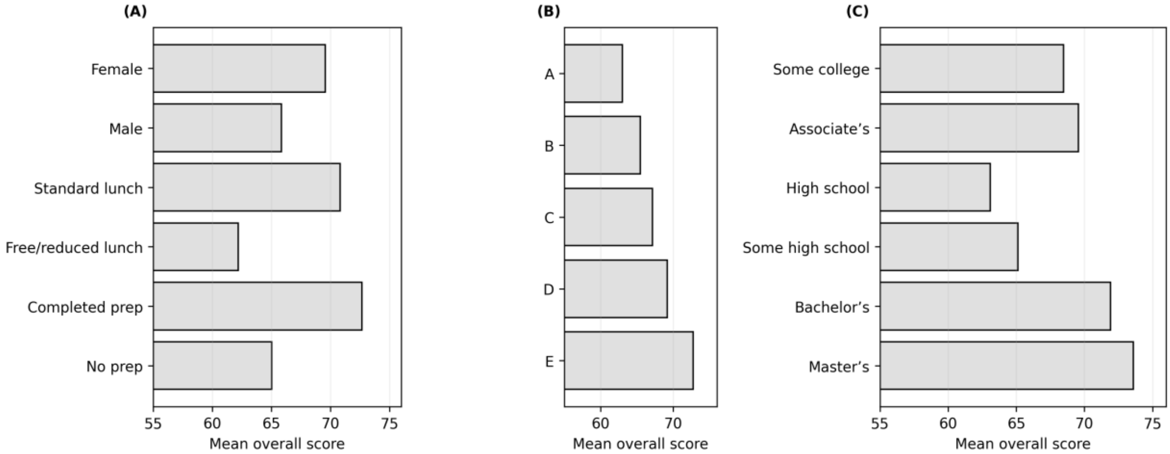

Figure 2. Mean Overall Academic Performance Across Student Characteristics.

3.4 Relationships Among Mathematics, Reading, and Writing Performance

The Pearson analysis revealed high positive relationships between all three examination domains. Mathematics correlated strongly with reading (r=0.818, 95% CI 0.796–0.837, p<0.001) and writing (r=0.803, 95% CI 0.779–0.824, p<0.001). The highest correlation was between reading and writing (r=0.955, 95% CI 0.949–0.960, p<0.001) with 91.1% of the variance shared. Generally, higher scores on one examination domain were related to higher scores on the other examination domains, although mathematics had slightly weaker relationships with literacy-based domains of scores than literacy had with the writing domain (Figure 3).

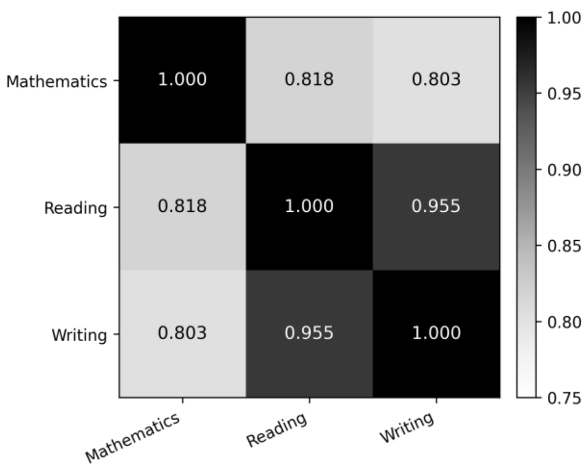

Figure 2. Pearson Correlations Among Mathematics, Reading, and Writing Scores.

3.5 Multivariable Statistical Modeling of Academic Achievement

Diagnostic screening indicated 7 observations were standardized residuals below -2.5 and were excluded, and the mathematics model refitted with 993 students. The model was significant, $F(12,980)=29.13$, $p<0.001$, explaining 26.3% of score variance (adjusted $R^2=0.254$). There was no heteroscedasticity as suggested by the Breusch-Pagan test ($p=0.297$) and the adjusted GVIF values were around 1.00. Math scores were positively related to male gender (B=4.75); groups D (B=5.50) and E (B=10.84); and completion of the test-preparations (B=5.13). The parent education in high school was negatively associated with free/reduced lunch (B=-10.89) and some high school parental education (B=-3.09) (Table 4).

Table 4. Multiple Linear Regression Model for Mathematics Performance

Predictor	B	SE	t	p-value
Intercept	62.55	1.69	37.04	<0.001
Male (vs female)	4.75	0.81	5.84	<0.001
Group B (vs A)	2.71	1.65	1.64	0.100
Group C (vs A)	2.68	1.54	1.74	0.082
Group D (vs A)	5.50	1.57	3.50	<0.001
Group E (vs A)	10.84	1.75	6.20	<0.001
Associate’s degree (vs some college)	0.51	1.21	0.42	0.673
High school (vs some college)	-4.24	1.25	-3.38	<0.001
Some high school (vs some college)	-3.09	1.29	-2.39	0.017
Bachelor’s degree (vs some college)	2.69	1.45	1.85	0.065
Master’s degree (vs some college)	3.28	1.87	1.76	0.079
Free/reduced lunch (vs standard)	-10.89	0.85	-12.87	<0.001
Completed test preparation (vs none)	5.13	0.85	6.05	<0.001

Note: B=unstandardized coefficient; SE=standard error. Reference categories were female, group A, some college, standard lunch, and no test-preparation course. Model n=993; $R^2=0.263$; adjusted $R^2=0.254$; $F(12,980)=29.13$; $p<0.001$.

4. Discussion

The results indicated significant differences in academic achievement for a variety of demographic, socioeconomic and preparation factors. Students' scores in the standard lunch category were higher than those in the free and reduced lunch category, indicating that available socioeconomic indicators (lunch category) influenced students' educational performance. This trend is in line with the evidence that socioeconomic differences may affect academic performance because of differences in learning opportunities, expectations and educational resources [13]. Overall performance was also related to parental education, with students whose parents were more highly educated generally having higher scores. Educational expectations of parents have been found to influence outcomes in lifelong learning and instil academic involvement [14]. But parental education benefits are not necessarily the same in all social and geographical settings [15]. Racial differences can also influence the educational benefits of parental attainment, as shown in similar evidence, so family background cannot be understood without a consideration of the larger structural context [16].

Gender differences were also evident. Male students had better mathematics scores, while female students had better overall academic scores. This variation lends itself to the theory that gender differences might occur within a specific academic area, but not across all areas of academic achievement. Sex-related differences have been found in cognitive processes that underlie mathematical performance, such as visual-spatial and verbal working memory [17]. There can also be social influences as mathematics-related gender stereotypes can affect students' confidence, engagement, and achievement [18]. However, writing performance has been shown to vary somewhat in standardized tests, which suggests the same could happen within a writing task [19]. Gender differences are also sensitive to how achievement is measured, as evidenced by more general comparisons of grades and standardized tests [20].

Lunch category had a significant impact on the multivariable mathematics model, and free or reduced lunch was negatively associated with the mathematics scores after the other characteristics were controlled. This finding corroborates the interpretation that there is a socioeconomic disadvantage independent of the demographic and educational factors measured in this study and that socioeconomic disadvantage may continue to be associated with achievement even when these factors are accounted for concurrently. Evidence from large student samples has also revealed that socioeconomic conditions act through multiple pathways in interactions with cognitive characteristics and school performance [21].

There were generally high positive correlations between all three measures, the highest being between reading and writing. This is a pattern suggesting that skills in academic areas are interrelated, especially in the literacy area. There is previous research that demonstrates the reinforcement that reading and writing can provide to each other through language, understanding, and ways of learning [22]. The somewhat weaker, but still significant, links with the literacy scores indicate that mathematics might share some common elements of the curriculum, but with more specific mathematics demands.

The mathematics model was further adjusted for male gender and group membership (D and E), with positive associations found for these factors, but noting that this is a statistical relationship and not a group characteristic. Evidence shows that mathematical gender gaps can be narrowed or widened under certain individual and social factors [23]. After adjustment, high school, and some high school parental education negatively correlated with mathematics scores compared to some college. Maths performance and structured preparation was also positively correlated with test-preparation completion, further supporting the importance of structured preparation. It has been found that students' motivation and expectations affect their attitude towards the test preparation process [24]. Evidence of the effects of test preparation on performance on large-scale educational assessments is also found in recent meta-analytic studies, with intervention effects varying depending on both the type of intervention and the context of the assessment [25]. Overall, the findings indicate that the observed differences in academic achievement are generated from a combination of socio-economic background, demographic, preparation, and the interrelated subject skills.

5. Conclusion

There were significant differences in academic achievement by demographic, socioeconomic, and educational-preparation factors; and mathematics and reading/writing were closely related. There were higher scores for students getting lunch in the standard lunch group and for those who had done academic preparation, suggesting that there was a relationship between examination outcomes and lunch category and academic preparation. Gender differences reported were of a subject specific nature: males were better in maths, and females had higher overall academic performance. There were also relationships between achievement and parental education and race/ethnicity, but these were to be viewed as association and not cause and effect. The high correlations among the three domains of the examination showed that reading and writing skills are not completely independent and that competencies are closely related. The results of the multiple linear regression also showed that male gender, group D and group E, free/reduced lunch, and the selected categories of parental education were significant predictors of mathematics performance after controlling for one another. The authors conclude that it is important to use descriptive, comparative, correlational, and multivariable statistical methods together when studying academic outcomes. Implementing strategies to improve outcomes for disadvantaged students must take into account both academic readiness and socioeconomic disadvantage. Further studies may reinforce these results by adding more data on previous levels of achievement, attendance, school atmosphere, learning motivation, and instructional quality to offer a fuller account of student performance.

References

1. Xuan X, Xue Y, Zhang C, Luo Y, Jiang W, Qi M, et al. Relationship among school socioeconomic status, teacher-student relationship, and middle school students’ academic achievement in China: using the multilevel mediation model. PLoS One. 2019;14(3):e0213783. doi:10.1371/journal.pone.0213783.
2. Eriksson K, Lindvall J, Helenius O, Ryve A. Socioeconomic status as a multidimensional predictor of student achievement in 77 societies. Front Educ. 2021;6:731634. doi:10.3389/feduc.2021.731634.
3. Finch H, Finch MEH. The relationship of national, school, and student socioeconomic status with academic achievement: a model for Programme for International Student Assessment reading and mathematics scores. Front Educ. 2022;7:857451. doi:10.3389/feduc.2022.857451.
4. Tan CY. Socioeconomic status and student learning: insights from an umbrella review. Educ Psychol Rev. 2024;36:100. doi:10.1007/s10648-024-09929-3.
5. Thomas DP, Hopwood B, Hatisaru V, Hicks D. Gender differences in reading and numeracy achievement across the school years. Aust Educ Res. 2024;51:41-66. doi:10.1007/s13384-022-00583-8.
6. Eriksson K, Lindvall J. Cultural variation in the SES-gender interaction in student achievement. Front Psychol. 2023;14:1120211. doi:10.3389/fpsyg.2023.1120211.
7. Bernabini L, Bonifacci P, de Jong PF. The relationship of reading abilities with the underlying cognitive skills of math: a dimensional approach. Front Psychol. 2021;12:577488. doi:10.3389/fpsyg.2021.577488.
8. Kim YSG. Inferencing skill and attentional control account for the connection between reading comprehension and mathematics. Front Psychol. 2021;12:709944. doi:10.3389/fpsyg.2021.709944.
9. Kwok FY, Bull R, Muñez D. Cross- and within-domain associations of early reading and mathematical skills: changes across the preschool years. Front Psychol. 2021;12:710470. doi:10.3389/fpsyg.2021.710470.

10. Boman B. Is the SES and academic achievement relationship mediated by cognitive ability? Evidence from PISA 2018 using data from 77 countries. Front Psychol. 2023;14:1045568. doi:10.3389/fpsyg.2023.1045568.
11. Gómez Soler SC, Bernal Nisperuza GL, Herrera Idárraga P. Test preparation and students' performance: the case of the Colombian high school exit exam. Cuad Econ. 2020;39(79):31-72. doi:10.15446/cuad.econ.v39n79.77106.
12. Li H, Xiong Y. The relationship between test preparation and state test performance: evidence from the Measure of Effective Teaching (MET) project. Educ Policy Anal Arch. 2018;26:64. doi:10.14507/epaa.26.3530.
13. Destin M, Hanselman P, Buontempo J, Tipton E, Yeager DS. Do student mindsets differ by socioeconomic status and explain disparities in academic achievement in the United States? AERA Open. 2019;5(3):1-12. doi:10.1177/2332858419857706.
14. Lai T, Liu F, Huang Y. The influence of parental educational expectations on children's higher education attainment: re-estimation based on instrumental variables. Front Psychol. 2022;13:899348. doi:10.3389/fpsyg.2022.899348.
15. Assari S, Boyce S, Bazargan M, Caldwell CH, Zimmerman MA. Place-based diminished returns of parental educational attainment on school performance of non-Hispanic White youth. Front Educ. 2020;5:30. doi:10.3389/feduc.2020.00030.
16. Assari S, Boyce S, Caldwell CH, Bazargan M. Parental educational attainment and Black-White adolescents' achievement gap: Blacks' diminished returns. Open J Soc Sci. 2020;8(3):282-297. doi:10.4236/jss.2020.83026.
17. van de Weijer-Bergsma E, Van Luit JEH, Moeller K. Sex differences in the association of math achievement with visual-spatial and verbal working memory: does the type of math test matter? Br J Psychol. 2022;113(3):798-819. doi:10.1111/bjop.12562.
18. Luo Y, Chen X. The impact of math-gender stereotypes on students' academic performance: evidence from China. J Intell. 2024;12(8):75. doi:10.3390/jintelligence12080075.
19. Li Z, Chen MY, Banerjee J. Using corpus analyses to help address the DIF interpretation: gender differences in standardized writing assessment. Front Psychol. 2020;11:1088. doi:10.3389/fpsyg.2020.01088.
20. Lehti H, Laaninen M. The gender achievement gap in grades and standardised tests—what accounts for gender inequality? Front Sociol. 2024;9:1448488. doi:10.3389/fsoc.2024.1448488.
21. Langensee L, Rumetshofer T, Mårtensson J. Interplay of socioeconomic status, cognition, and school performance in the ABCD sample. NPJ Sci Learn. 2024;9(1):17. doi:10.1038/s41539-024-00233-x.
22. Zhang Y, Peng H, Bian Y. Interplay between reading and writing under different teaching models: a study based on Chinese learning by China's ethnic minorities. Front Psychol. 2020;11:2150. doi:10.3389/fpsyg.2020.02150.
23. Paz-Baruch N, Spektor-Levy O. How the relationship between individual and social factors informs the narrowing of gender gaps in elementary mathematical achievements. Front Educ. 2024;9:1339040. doi:10.3389/feduc.2024.1339040.
24. Xu S. Construct-oriented or goal-motivated? Interpreting test preparation of a high-stakes writing test from the perspective of expectancy-value theory. Front Psychol. 2022;13:846413. doi:10.3389/fpsyg.2022.846413.
25. Hao Z, Baird JA, El Masri Y, Double K. The impact of test preparation on performance of large-scale educational tests: a meta-analysis of experimental studies. Rev Educ Res. 2025. doi:10.3102/00346543251360775.