Science Education International
31(2), 195-202
https://doi.org/10.33828/sei.v31.i2.9

International Council of Association for Science Education

ORIGINAL ARTICLE

# The Gender Differences in Science Achievement, Interest, Habit, and Creativity

**Chaochao Jia[1]*, Tao Yang[1], Yu Qian[2], Xinye Wu[3]**

[1]Collaborative Innovation Center of Assessment for Basic Education Quality, Beijing Normal University, China, [2]College of Physics and Optoelectronics Technology, Baoji University of Arts and Science, China, [3]Department of Physics, Beijing Normal University, Beijing, China

***Corresponding Author:** jiachaochao@mail.bnu.edu.cn

## ABSTRACT

In this study, we explored the gender differences in science achievement, interest, creativity, and so on, using a national representative sample from the National Assessment of Education Quality (NAEQ). NAEQ assessed the Grade-4 (n = 112,314) and Grade-8 (n = 74,808) students' scientific literacy of Mainland China in 2017. The findings indicated that there was no gender difference of science academic achievement in both Grade 4 and Grade 8 in China. However, scientific interests varied in different grades and disciplines. Girls had a higher interest in science in Grade 4 and higher interest in biology in Grade 8. The 8th-grade boys had much higher creativity than girls ($p < 0.001$, Cohen's $d = 0.301$), and outperformed girls in multiple-choice items. Our study also highlighted that boys were likely to outperform in the higher score ranges, as well as lower score ranges. Finally, we also found that the boys had significantly greater variabilities in science achievement and interest (variance ratio [VR] >1.1), while the girls had slight greater variability in creativity ($0.9 < VR < 1.0$). Our results provide a reliable Chinese evidence for international studies on gender differences. However, further research is demanded to study the reasons behind gender differences deeply.

**KEY WORDS:** creativity, gender, science achievement, science interest

## INTRODCUTION

Research on gender differences has a long history (Fennema, 1974), among which the gender gaps in mathematics and science have been the focus of attention (Steinkamp and Maehr, 1983, 1984; Keller, 1985; Kahle, 1986; Whyte, 1986; Ditchfield and Scott, 1987; Kelly, 1987; Becker, 1989). Is boys' superiority in science learning a myth or a reality? On this question, different researchers give different answers. However, it is still an interesting and important question to explore.

We were eager to explore the gender differences between boys and girls in science performance in China. If gender differences do not exist, then educators should discard their belief about boys' superiority. On the contrary, if there are significant gender differences in learning science, then the education profession has a direct responsibility to enable boys and girls to fulfill equally their cognitive potential in science.

In this study, we explored the gender differences between boys and girls in science performance in China. The data we used were collected from the Chinese National Assessment of Education Quality (CNAEQ), which was launched by the Collaborative Innovation Center of Assessment for Basic Education Quality with authorization from the Ministry of Education of the People's Republic of China. The CNAEQ is China's largest continuing and nationally representative education quality assessment which started in 2015 and covers six essential disciplines including language, mathematics, science, art, physical education and health, and civics. The assessment schedule is conducted across a 3-year period with two disciplines every year (Wu et al., 2019). In the science assessment of CNAEQ, students' science achievement, science interest, learning habit, and creativity were investigated, which provided us a high-quality source to explore the gender differences in China.

## LITERATURE REVIEW

There are numerous quantitative studies on the performance of boys and girls in different subjects. The performance compared in studies contains academic achievement, learning attitudes, and emotions, such as learning interests, habits, and creativity, which are all considered as important factors of academic performance nowadays. Different subjects as well as different item forms are also investigated in different researches. The gender gaps over time also attract researchers' interest.

In academic achievement, many studies investigated the academic gender gaps, most found that boys outperform girl in science, and the gaps are larger than mathematics. It has been reported that boys are significantly better than girls in biology, introduction to science, and physics (Becker, 1989; Steinkamp and Maehr, 1983), while girls are superior in language ability (Halpern et al., 2007). Furthermore, gender effect sizes (ESs) for science were found to be somewhat larger than those for

Science Education International
31(2), 195-202
https://doi.org/10.33828/sei.v31.i2.9



mathematics (e.g., Cleary, 1991; Hedges and Nowell, 1995; Linn, 1991). Cleary (1991) found that boys outperformed girls on science tests across all age groups and this advantage increased with age. A meta-analysis revealed a male advantage on science tests ranging from 0.11 to 0.50 standard deviations (SDs) (Hedges and Nowell, 1995). However, less attention has been paid to research on gender differences in science than in mathematics (Beller and Gafni, 1996).

Now, learning attitude and emotion are also considered as important outcomes of academic performance, and positive relationship between emotion outcomes and academic achievement has already been proved (Yu, 2017; Ru and Wu, 2010; Cheng, 2010; Koeller et al., 2001; Fisher et al., 2012). There are studies on the gender gaps in students' learning attitude, but these results are inconsistent. Some indicated girls scored lower in academic self-concept, interest, and motivation (Preckel et al., 2008), others pointed out that girls were significantly more interested in science than boys (Akpınar et al., 2009), and others found no significant differences between boys and girls in their enjoyment of science and scientific experiments (Akpınar et al., 2009).

In science, creativity also plays a crucial role. It is essential for the development and progress of human civilization (Chang et al., 2016) and feeding creative activities when children are young will provide society with imaginative thinkers and leaders of scientific entrepreneurial discovery (Yates and Twigg, 2017). Recently, creativity has roused increased interest, and developing creativity is viewed as an educational imperative worldwide. Education drives the creative potential of students and expands comprehensive and balanced attitudes to transcend their lifestyles (Piguave, 2014). It has been reported that boys' divergent thinking ability, which is related to creativity, is significantly better than girls' and increases with grade (Wu and Xue, 1985). However, Castillo-Vergara et al. (2018) found females exhibited higher results than males in creativity's three dimension, fluency, flexibility, and originality. The results of creativity also vary in different studies.

There are several research studies that have explored the gender gaps in different item forms. These have found that boys were good at multiple-choice (MC) items while girls were good at construct response (CR) items (Wester and Henriksson, 2000; Reardon et al., 2018).

The gender gaps over time have also attracted researchers' interest. There has been a narrowing of the gender gap on tests (Linn, 1991, Hyde and Mertz, 2009). Hyde and Linn (1988) and Linn and Hyde (1989) examined a number of meta-analyses and other studies on gender differences and concluded that the majority of gender differences in cognitive skills were small and have become less pronounced in recent years. Moreover, in the past few decades, growing attention has been paid to gender differences in performance when screening items in the test construction process (Burton and Burton, 1993), which in itself, to some extent, may reduce the magnitude of the observed score differences between the genders.

However, Wilder and Powell (1989) pointed out that convergence of scores is not evident in the higher score ranges, where men are still overrepresented and where the gap may even have increased. Humphreys (1988) stressed the importance of studying score variability and showed the effects of mean and variance differences on ratios of gifted boys to gifted girls. Unequal variabilities of two groups with equal means can produce the same results as mean differences between groups with equal variability, which may explain the gender differences obtained in highly selective samples (Humphreys, 1988), which indicated greater score variability among boys than among girls (Benbow, 1988, 1990; Martin and Hoover, 1987). Hedges and Nowell (1995) found that scores for boys had consistently larger variance. Therefore, the selectivity in samples should be considered carefully when interpreting research results of gender differences.

It is noteworthy that different researchers' results highlight inconsistencies. These inconsistent results are due to several reasons. First of all, the samples vary a lot in different studies, most of which are local, selected, and not representative. Second, the test formats affect the results. Third, some of the studies only report significant p-value, which is not enough to indicate a significant difference in a large sample. Finally, the gender differences also vary over time and the tendency may change, which makes the gender differences worth of long-term researches.

Gender differences in performance on standardized tests have been the focus of many studies, the majority of which were conducted in the United States. There are few standardized tests in China, let alone the studies of gender differences. In our study, we employed a representative sample of China to provide an insight for researchers from Chinese cultural context. Besides, the ES, which is a more valid indicator, is used to testify the gender differences. What is more, we also analyzed the variance ratio (VR) of boys and girls to examine the greater male variability hypothesis.

# METHODOLOGY

## Data and Sample

The data used in our study were collected from CNAEQ 2017. CNAEQ 2017 was carried out on May 25, 2017. Thirty-one Provinces (or municipalities, hereinafter referred to as provinces) and one Corp, Xinjiang Production and Construction Corp which is a provincial unit in China with a separate education and teaching system, participated in the national assessment representing Mainland China. To be representative at the national level, a three-stage stratification cluster sampling design with systematic probability proportional to size (PPS) technique was employed (Zhang and Tang, 2017). First, counties in provinces were selected according to their GDP and educational development levels in the first stage. Second, 12 primary schools were selected from each county based on their location, schooling quality, and school size. Thirdly, 30 students were randomly selected within each school. If the

Science Education International
31(2), 195-202
https://doi.org/10.33828/sei.v31.i2.9



total number of schools in a district, or students in a school, is less than the demanded sample size, then all of them will be picked to take the test (CNAEQ, 2018).

The effective sample used in this study consisted of 111,743 Grade-4 students (52.9% boys), and 74,595 Grade-8 students (52.7% boys), shown in Table 1. 571 Grade-4 students and 213 Grade-8 students are missing due to the blank answer of gender. Details of the participants from the different provinces are listed in the Appendix.

## Measures

### *Science achievement*

Science Achievement was assessed by students' paper-pencil test. Six parallel tests were used to assess Grade 4 and 8 students' scientific literacy, including scientific understanding (SU), scientific inquiry (SI), and scientific thinking (ST). SU refers to students' understanding of scientific concepts and laws, which is the basis of explaining natural phenomena and solving practical problems. SI refers to the ability of carrying out an inquiry activity, which including steps such as raising a question, obtaining evidence, explaining the phenomenon, drawing a conclusion, and reflecting, communicating, and evaluating. ST means that students can understand the essential properties, internal laws, and mutual relations of things, construct models in an abstract and general way, and use reasoning and other methods to question and judge different viewpoints and conclusions based on factual evidence and scientific reasoning (CNAEQ, 2016).

Each Grade 4 parallel test contains 28 MC items and 6 to 7 CR items, and Grade 8 parallel test contains 35 MC items and 5 CR items. The internal consistency was 0.85–0.88. They were all allowed 80 min to finish the test. The score was generated by Rasch Model with a mean of 500 and a SD of 100 (CNAEQ, 2018).

### *Interest, habit, and creativity*

The science interest, habit, and creativity were assessed by students' questionnaires, among which science interests were measured separately in physics, biology, and geography subject, and habit and creativity were measured overall subjects integrated. Students' interest in science learning (physics, biology, and geography in Grade 8) was measured by four items (e.g., "I like science/physics/biology/geography (hereinafter referred to as science)" and "I like science experiment"). Learning habit was measured by seven items (e.g., "I read books related to science" and "I watch science show"). Creativity was measured by eight items (e.g., "I will raise new question on my own" and "When I come across something unknown, I can wait to solve"), composed of personality traits, thinking styles, and creative activities and achievements. All these instruments employed a 4-point Likert-type scale response format ranging from 1 (strongly disagree), 2 (disagree), 3 (agree), to 4 (strongly agree). Responses to the scale indicated the extent of their agreement with each item. The mean score of the items was used ranging from 1 to 4 with higher scores suggested higher interest, better habit, and more creative. The means of the items were calculated and the samples answering more than half amounts of items were taken into account. The scale internal consistency was acceptable (α above 0.70), according to CNAEQ (2018).

**Table 1: The total sample of CNAEQ 2017**

| Total sample | Boys | Girls | Missing | Effective sample | Sum |
|---|---|---|---|---|---|
| Grade 4 | 59,137 | 52,606 | 571 | 111,743 | 112,314 |
| Grade 8 | 39,315 | 35,280 | 213 | 74,595 | 74,808 |

## Analytic Strategy

The main statistics used in our study are ES and VR.

### *ES*

The ES reflects the difference in SD between the mean values of two distributions, which represents the actual difference between two populations despite of the sample size. We use Cohen's d to represent the ES, which was calculated as below:

$$D=(X_m-X_f)/S_W \tag{1}$$

Where $X_m$ indicates the mean score of boys and $X_f$ indicates that of girls. $S_W$ is the pooled SD. Moreover, the positive value of $d$ indicates boys outperform girls, and the negative value indicates girls outperform boys in science. According to the empirical standard, where $d < 0.1$ means no difference between two groups (Hyde, 2005); $d > 0.5$ indicates small effect; and $d > 0.5$ indicates medium to big effect (Cohen, 1992).

We use this indicator to examine the similarity hypothesis, which means boys and girls are more alike than different.

### *VR*

The VR is calculated by the variance of boys divided that of girls and is used to compare the amount of boys and girls at top and bottom groups.

$$VR=\text{Male variance/female variance} \tag{2}$$

It was accepted that VR >1.0 means boys have greater variation, and VR <1.0 means girls have greater variation than boys. If VR >1.1 or VR <0.9, it means the variation make a practical sense (Feingold, 1992).

We use this indicator to examine the greater male variability hypothesis, which means boys have bigger or greater variation.

### *Analysis dimensions*

We analyzed several dimensions in science, as well as different grades and item forms. For example, we analyzed the integrated science in both Grade 4 and 8, and also analyzed the three subjects (physics, biology, and geography) in Grade 8. We also examined the differences in total science achievement, and the three dimensions (SU, SI, and ST). We studied the gender gaps in different item forms (MC items vs. CR items) as well. We examined the gender differences in science interest (physics interest, biology interest, and geography interest in Grade 8), habit, and creativity. In each part, we examined both the ES (to indicates the similarity or difference) and the VR (to indicate the greater male/female variability) values to examine the two hypotheses.

# FINDINGS

## Science Achievement

We investigated the gender gap in science achievement shown in Tables 2 and 3. We found no significant differences in total scientific literacy, as well as in three subdomains (SU, SI, and ST) across both grades. The findings in physics, biology, and geography subjects were the same, which indicated the similarity hypothesis in science achievement. There was no gender difference in either Grade 4 or 8.

While analyzing the variabilities of boys and girls, we found the VRs of scientific literacy (as well as SU, SI, and ST), physics, biology, and geography subjects were all above 1.1 (written in italic in Tables 2 and 3). This indicated boys had greater variability than girls in science achievement in both Grade 4 and 8 nationally (i.e., there were more boys in the top or bottom groups), in accord with greater male variability hypothesis.

Figures 1 and 2 emerged with regard to gender differences in all participating provinces. Although the average ESs of the total sample was smaller than 0.1, some provinces had a significant ESs. The ESs of Grade 4 range from –0.12 to 0.22, and the ESs of Grade 8 range from –0.13 to 0.31. There were 10 provinces which had significant gender differences in science in Grade 4, most show positive SEs (i.e., in favor of boys), while only one of them was negative (i.e., girls significantly outperform boys in province 21). Figure 2 shows that 11 provinces had significant gender differences in Grade 8, most positive SEs, while still one of them was negative (girls significantly outperform boys in province 20).

**Table 2: Gender differences in science cognitive dimensions**

| National | Gap | VR | Significance test | | | |
|---|---|---|---|---|---|---|
| **Grade 4** | | | **ES (Cohen's d)** | **t** | **Sig. (2-tailed)** | |
| Total | 4.40 | **1.154** | 0.044 | 7.418 | 0.000 | *** |
| SU | 2.76 | **1.155** | 0.056 | 9.317 | 0.000 | *** |
| SI | 1.88 | **1.139** | 0.031 | 5.235 | 0.000 | *** |
| ST | 1.86 | **1.157** | 0.039 | 6.501 | 0.000 | *** |
| **National** | **Gap** | **VR** | **Significance test** | | | |
| **Grade 8** | | | **ES (Cohen's d)** | **t** | **Sig. (2-tailed)** | |
| Total | 6.57 | **1.173** | 0.066 | 9.008 | 0.000 | *** |
| SU | 4.10 | **1.161** | 0.081 | 11.119 | 0.000 | *** |
| SI | 2.30 | **1.178** | 0.059 | 8.050 | 0.000 | *** |
| ST | 3.03 | **1.161** | 0.061 | 8.317 | 0.000 | *** |

SU: Scientific understanding, SI: Scientific inquiry, ST: Scientific thinking. Gaps are calculated by boys minus girls. The numbers in bold indicate boys have significant greater variability, or the gender differences are significant. ***indicates statistically differences of p<0.001, VR: Variance ratio, ES: Effect size

**Table 3: Gender differences in different subjects**

| National | Gap | VR | Significance test | | | |
|---|---|---|---|---|---|---|
| **Grade 8** | | | **ES (Cohen's d)** | **t** | **Sig. (2-tailed)** | |
| Total | 6.57 | **1.173** | 0.066 | 9.008 | 0.000 | *** |
| Physics | 2.93 | **1.176** | 0.060 | 8.234 | 0.000 | *** |
| Biology | 2.73 | **1.162** | 0.073 | 9.912 | 0.000 | *** |
| Geography | 3.27 | **1.151** | 0.066 | 8.951 | 0.000 | *** |

The gaps are calculated by boys minus girls. The numbers in bold indicate boys have significant greater variability, or the gender differences are significant. ***indicates statistically differences of p<0.001. VR: Variance ratio, ES: Effect size

## Science Interest, Habit, and Creativity

We investigated the gender gaps in science interest, learning habit, and creativity, Table 4. According to the results, we found that there was no significant difference in physics and geography interest in Grade 8. However, girls had a significantly higher interest in science in Grade 4 and in biology in Grade 8. The VRs of science interest of Grade 4, and physics, biology, and geography interests of Grade 8 were all above 1.1 (written in italic in Table 4). It means boys have greater variability than girls in science interests in both Grade 4 and 8 nationally, the same as that in science achievement.

However, the learning habit and creativity of Grade 4 and 8 were not the same as interest. Boys have big, but not significantly greater variability in both Grade 4 and 8 (1.0 <VR <1.1). Girls have big but not significantly greater variability in two grades (0.9 <VR <1.0). The gender differences in habit and creativity were not significant in Grade 4, but were significantly different in Grade 8, in favor of boys.

## Item Forms

We investigated the different item forms of science in Grade 4 and 8, Table 5. We found that boys were good at MC items, while girls scored higher in CR items. However, the differences were not significant expect in CR items of Grade 8.

The VRs of different item forms were all above 1.0, and three in four were above 1.1, which indicated greater variability in boys.

# DISCUSSIONS

## Gender Differences

We used CNAEQ data to analyze the gender differences across mainland China, a representative sample does not suffer from selection biases. Thus, the results we found in this study can reveal the factual gender differences in China.

According to our study, we found that there was no gender difference of science achievement in both Grade 4 and Grade 8, which supports "The Gender Similarities Hypothesis" (Hyde, 2005). With regard to the different cognitive dimensions, the results were the same. While we examined the different provinces in China, we found that in 24 provinces, boys scored higher than girls in science achievement in both grades, among which nine provinces' boys significantly outperform girls in Grade 4, and ten provinces' boys significantly outperform girls in Grade 8 ($p < 0.05$, Cohen's $d > 0.1$). Meanwhile, there was

Science Education International
31(2), 195-202
https://doi.org/10.33828/sei.v31.i2.9



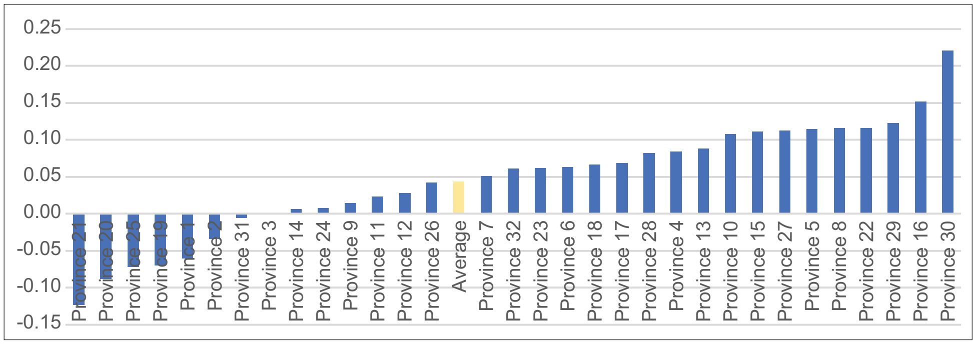


**Figure 1:** Gender effect size of Grade-4 students in science

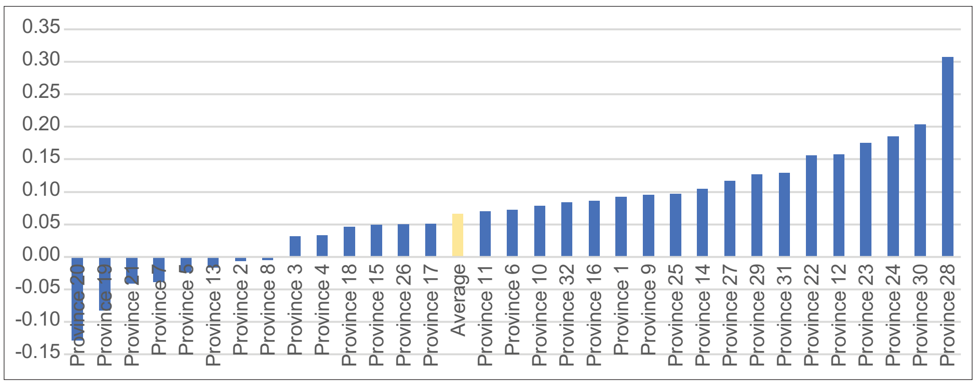


**Figure 2:** Gender effect size of Grade-8 students in science

**Table 4: Gender differences in science interest, habit, and creativity**

| National | Gap | VR | Significance test | | | |
|---|---|---|---|---|---|---|
| **Grade 4** | | | **ES (Cohen's d)** | **t** | **Sig. (2-tailed)** | |
| Science interest | −0.07 | **1.186** | **−0.136** | **−22.527** | **0.000** | *** |
| Learning habit | −0.01 | 1.026 | −0.011 | −1.899 | 0.058 | |
| Creativity | 0.05 | 0.979 | 0.085 | 14.158 | 0.000 | *** |
| **National** | **Gap** | **VR** | **Significance test** | | | |
| **Grade 8** | | | **ES (Cohen's d)** | **t** | **Sig. (2-tailed)** | |
| Physics interest | 0.02 | **1.118** | 0.025 | 3.328 | 0.001 | *** |
| Biology interest | −0.08 | **1.257** | **−0.127** | **−16.918** | **0.000** | *** |
| Geography interest | −0.00 | **1.150** | −0.000 | −0.006 | 0.995 | |
| Learning habit | 0.15 | 1.049 | **0.223** | **30.341** | **0.000** | *** |
| Creativity | 0.17 | 0.968 | **0.301** | **40.998** | **0.000** | *** |

The gaps are calculated by boys minus girls. The numbers in bold indicate boys have significant greater variability, or the gender differences are significant. ***indicates statistically differences of p<0.001. VR: Variance ratio, ES: Effect size

only one province in which the girls significantly outperform boys in Grade 4 and 8 ($p < 0.05$, Cohen's $d > 0.1$).

However, the scientific interests varied in different grades and disciplines. In Grade 4, girls had a higher interest in science than boys ($p < 0.001$, Cohen's $d = -0.136$), which was out of our expectation. Besides, girls had a higher interest in biology than boys ($p < 0.001$, Cohen's $d = -0.127$), while having no significant differences in physics and geography in Grade 8. Although girls may not perform better than boys, they were more interested in science. What is leading the girls to underperform in science may be the question for science researchers.

The 8th-grade boys had much higher creativity than girls ($p < 0.001$, Cohen's $d = 0.301$), while the creativity of Grade 4 was almost the same between boys and girls. What factors may influence students' creativity? Are these factors inherent or environmental? The lack of gender differences in creativity until fourth grade would suggest that these factors may be environmental. On the other hand, it suggests that inherent factors may be more important.

Science Education International
31(2), 195-202
https://doi.org/10.33828/sei.v31.i2.9



**Table 5: Gender differences in different item forms**

| National | Gap | VR | Significance test | | | |
|---|---|---|---|---|---|---|
| Grade 4 | | | ES (Cohen's d) | t | Sig. (2-tailed) | |
| MC | 0.35 | **1.132** | 0.070 | 11.74 | 0.000 | *** |
| CR | −0.06 | 1.083 | −0.017 | −2.77 | 0.006 | ** |
| National | Gap | VR | Significance test | | | |
| Grade 8 | | | ES (Cohen's d) | t | Sig. (2-tailed) | |
| MC | 0.79 | **1.124** | **0.125** | **17.04** | **0.000** | *** |
| CR | −0.24 | **1.156** | −0.072 | −9.90 | 0.000 | *** |

MC: Multiple-choice items; CR: Construct response items. Gaps are calculated by boys minus girls. The numbers in bold indicate boys have significant greater variability, or the gender differences are significant. **indicates statistically differences of p<0.01, ***indicates statistically differences of p<0.001.VR: Variance ratio, ES: Effect size

We found that boys outperformed girls in MC items (Cohen's $d = 0.125$) in Grade 8, while there were no significant differences CR items of both grades. As differential item functioning was tested during the item development process, the differences found in our study were smaller than previous studies.

The findings suggest that the boys had significantly greater variabilities in science achievement and interest (VR >1.1), while the girls had slightly and indistinctively greater variability only in creativity (0.9 <VR <1.0). It may be an evidence of boys outperform in the higher score ranges to Wilder and Powell (1989). However, it may also indicate that boys are more likely to score lower.

### Gender Gaps Over Time

ESs tended to increase with grade (0.04 and 0.07 in favor of boys on the total score for Grade 4 and 8, respectively). In 32 provinces, the gender differences in Grade 8 (ESs range from −0.13 to 0.31) were bigger than that of Grade 4 (ESs range from –0.12 to 0.22). Although we had not tested higher grades, we have reasons to believe that, the higher the grade, the bigger the gender gaps will be.

### A Rebuttal of Stereotype

According to the results, we found that there was no gender difference of science achievement in both Grade 4 and Grade 8, and girls had higher science interest in Grade 4 and higher biology interest in Grade 8. It means that boys and girls were more alike than different in science except in creativity in Grade 8. Hence, educators, parents, and even students themselves should discard their belief about boys' superiority.

Girls have a higher interest in science in early grade (Grade 4), then become the approximate interest in physics and geography in later grade (Grade 8). Moreover, the total interest scores are also declined with grade. Besides, the gender differences are broadened with grade. Whether it has something to do with the development of identification of sex roles? It calls for further researches.

### Limitation

First, we have not compared the science experiment ability in this study due to the CNAEQ program setting.

Second, also due to the setting of CNAEQ, we have only two grades' student samples in this study. If we can examine more grades, the results will be more interesting.

Third, although we compared several important outcomes of education, we have not discussed the factors which may influence gender differences. Moreover, a lot of further studies are still needed in this topic.

## CONCLUSION

Despite all these limitations, our study is important as it adds to the literature of Chinese evidence in the following ways.

First of all, our results provide a reliable Chinese evidence for international studies on gender differences. Second, the findings from this study suggest that boys and girls are more alike than different. Third, girls have higher science interests in Grade 4 and in the biology of Grade 8. Fourth, our study also indicates that boys are more likely to outperform in the higher score ranges (i.e., higher levels), as well as lower score ranges.

### Ethics Statement

The study was consistent with the ethical principles of human subjects. We told the detailed content of the assessment to the participants in advance. Moreover, we got the consent of participants' parents before doing the research.

## REFERENCES

Akpınar, E., Yıldız, E., Tatar, N., & Ergin, Ö. (2009). Students' attitudes toward science and technology: An investigation of gender, grade level, and academic achievement. *Procedia Social and Behavioral Sciences*, 1, 2804-2808.

Becker, J.B. (1989). Gender and science achievements: A reanalysis of studies from meta-analyses. *Journal of Research in Science Teaching*, 26(2), 141-169.

Beller, M., & Gafni, N. (1996). The 1991 international assessment of educational progress in mathematics and sciences: The gender differences perspective. *Journal of Educational Psychology*, 88(2), 365-377.

Benbow, C.P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences*, 11, 169-183.

Burton, E., & Burton, N.W. (1993). The effect of item screening on test scores and test characteristics. In: Holland, P.W., & Wainer, H. (Eds.), *Differential Item Functioning.* New Jersey: Erlbaum. p321-335.

Castillo-Vergara, M., Barrios Galleguillos, N., Jofré Cuello, L., Alvarez-Marin, A., & Acuña-Opazo, C. (2018). Does socioeconomic status influence student creativity? *Thinking Skills and Creativity*, 29, 142-152.

Chang, S., Wang, C., & Lee, J. (2016). Do award-winning experiences benefit students' creative self-efficacy and creativity? The moderated mediation effects of perceived school support for creativity. *Learning and Individual Differences*, 51, 291-298.

Cheng, L. (2010). *Parents' Expectation, Self-expectation and Their Relations with Learning Achievement on Junior Middle School Student*, *Master Dissertation*. China: Henan University.

Chinese National Assessment of Education Quality (CNAEQ). (2016). *National Science Assessment Framework for Compulsory Education Quality*. Restricted. China: Chinese National Assessment of Education Quality.

Chinese National Assessment of Education Quality (CNAEQ). (2018). *CNAEQ 2017 Technical Report*. Restricted. China: Chinese National Assessment of Education Quality.

Cleary, A. (1991). Gender differences in aptitude and achievement test

Science Education International
31(2), 195-202
https://doi.org/10.33828/sei.v31.i2.9



scores. In: *Sex Equity in Educational Opportunity, Achievement, and Testing*, *Proceedings of the 1991 ETS Invitational Conference*. New Jersey: Educational Testing Service.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.

Ditchfield, C., & Scott, L. (1987). *Better Science: For Both Girls and Boys*. USA: Heineman.

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research*, 62, 61-84.

Fennema, E. (1974). Mathematics learning and the sexes: A review. *Journal of Research in Mathematics Education*, 5(3), 126-139.

Fisher, P.H., Dobbs-Oates, J., Doctoroff, G.L., & Arnold, D.H. (2012). Early math interest and the development of math skills. *Journal of Educational Psychology*, 104, 673-681.

Halpern, D.F., Benbow, C.P., Geary, D.C., Gur, R.C., Hyde, J.S., & Gernsbacher, M.A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, 8(1), 1-51.

Hedges, L.V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269, 41-45.

Humphreys, L.G. (1988). Sex differences in variability may be more important than sex differences in means. *Behavioral and Brain Sciences*, 11, 195-196.

Hyde, J.S. (2005). The gender similarities hypothesis. *American Psychologist*, 60, 581-592.

Hyde, J.S., & Linn, M.C. (1988). A meta-analysis of gender differences in verbal abilities. *Psychological Bulletin*, 104, 53-69.

Kahle, J.B. (1986). *Women in Science: A Report From the Field*. London: Falmer Press.

Keller, E.F. (1985). R*eflections on Gender and Science*. Connecticut: Yale University Press.

Kelly, A. (Ed.). (1987). *Science for Girls?* United Kingdom: Open University Press.

Koeller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education*, 32, 448-470.

Linn, M.C. (1991). Gender differences in educational achievement. In: *Sex Equity in Educational Opportunity, Achievement, and Testing, Proceedings of the 1991 ETS Invitational Conference*. New Jersey: Educational Testing Service.

Linn, M.C., & Hyde, J.S. (1989). Gender, mathematics, and science. *Educational Researcher*, 18, 17-27.

Martin, D.J., & Hoover, H.D. (1987). Sex differences in educational achievement: A longitudinal study. *Journal of Early Adolescence*, 7, 65-83.

Piguave, V. (2014). Importance of creativity development for commercial engineering career students from the teaching-learning process (importancia del desarrollo de la creatividad para los estudiantes de la carrera de ingeniería comercial desde el proceso de enseñanza-aprendizaje). *Educación*, 23(44), 29-47.

Preckel, F., Goetz, T., Pekrun, R., & Kleine, M. (2008). Gender differences in gifted and average-ability students. *Gifted Child Quarterly*, 52(2), 146-159.

Reardon, S.F., Kalogrides, D., Fahle, E.M., Podolsky, A., & Zárate, R.C. (2018). The relationship between test item format and gender achievement gaps on math and ELA tests in fourth and eighth grades. *Educational Researcher*, 47(5), 284-294.

Ru, H., & Wu, L. (2010). The research on the correspondence of students' expectation and their achievement. E*ducation and Teaching Research*, 24(7), 40-43.

Steinkamp, M.W., & Maehr, M.L. (1984). Gender differences in motivational orientations toward achievement in school science: A quantitative synthesis. *American Educational Research Journal*, 21, 39-59.

Wester, A., & Henriksson, W. (2000). The interaction between item format and gender differences in mathematics performance based on TIMSS data. *Studies in Educational Evaluation*, 26, 79-90.

Whyte, J. (1986). *Girls Into Science and Technology: The Story of a Project*. London: Routledge & Kegan Paul.

Wilder, G.Z., & Powell, K. (1989). *Sex Differences in Test Performance: A Survey of the Literature*. New Jersey: Educational Testing Service.

Wu, J., & Xue, H. (1985). Test report on divergent thinking ability of primary and secondary school students. *Journal of Psychological Science*, 5, 27-32.

Wu, L., Ma, X., Shi, Y., Tao, S., Yu, Y., Wang, S., Luo, L., Xin, T. & Li, Y. (2019). China national assessment of education quality-physical education and health (CNAEQ-PEH) 2015: An introduction. *Research Quarterly for Exercise and Sport*, 90(2), 105-112.

Yates, E., & Twigg, E. (2017). Developing creativity in early childhood studies students. *Thinking Skills and Creativity*, 23, 42-57.

Yu, H. (2017). *Academic Achievement and Its Relation to Perceived Teacher/Parent's Expectation and Academic Self-Concept among Middle School Students*, *Master Dissertation*. China: Shandong Normal University.

Zhang, D., & Tang, X. (2017). The influence of extracurricular activities on middle school students' science learning in China. *International Journal of Science Education*, 39(10), 1381-1402.

Science Education International
31(2), 195-202
https://doi.org/10.33828/sei.v31.i2.9


# APPENDIX

**Details of participators of different provinces**

| Sub-Samples | Boys | Girls | Missing | Effective sample | Sum | Boys | Girls | Missing | Effective sample | Sum |
|---|---|---|---|---|---|---|---|---|---|---|
| Province 1 | 1,132 | 987 | 2 | 2,119 | 2,121 | 755 | 623 | 5 | 1,378 | 1,383 |
| Province 2 | 1,119 | 990 | 9 | 2,109 | 2,118 | 747 | 621 | 8 | 1,368 | 1,376 |
| Province 3 | 2,617 | 2,318 | 14 | 4,935 | 4,949 | 1,696 | 1,573 | 5 | 3,269 | 3,274 |
| Province 4 | 1,781 | 1,569 | 9 | 3,350 | 3,359 | 1,219 | 1,079 | 10 | 2,298 | 2,308 |
| Province 5 | 1,172 | 1,109 | 10 | 2,281 | 2,291 | 828 | 784 | 2 | 1,612 | 1,614 |
| Province 6 | 1,427 | 1,359 | 12 | 2,786 | 2,798 | 958 | 876 | 7 | 1,834 | 1,841 |
| Province 7 | 1,242 | 1,175 | 9 | 2,417 | 2,426 | 831 | 776 | 7 | 1,607 | 1,614 |
| Province 8 | 1,351 | 1,248 | 16 | 2,599 | 2,615 | 928 | 880 | 2 | 1,808 | 1,810 |
| Province 9 | 1,148 | 967 | 3 | 2,115 | 2,118 | 729 | 668 | 5 | 1,397 | 1,402 |
| Province 10 | 2,282 | 1,941 | 7 | 4,223 | 4,230 | 1,513 | 1,311 | 1 | 2,824 | 2,825 |
| Province 11 | 2,096 | 1,824 | 5 | 3,920 | 3,925 | 1,371 | 1,214 | 4 | 2,585 | 2,589 |
| Province 12 | 2,166 | 1,936 | 14 | 4,102 | 4,116 | 1,526 | 1,294 | 3 | 2,820 | 2,823 |
| Province 13 | 1,867 | 1,654 | 11 | 3,521 | 3,532 | 1,250 | 1,096 | 3 | 2,346 | 2,349 |
| Province 14 | 2,270 | 1,944 | 28 | 4,214 | 4,242 | 1,523 | 1,301 | 7 | 2,824 | 2,831 |
| Province 15 | 2,974 | 2,651 | 8 | 5,625 | 5,633 | 1,968 | 1,804 | 8 | 3,772 | 3,780 |
| Province 16 | 3,766 | 3,239 | 18 | 7,005 | 7,023 | 2,464 | 2,171 | 7 | 4,635 | 4,642 |
| Province 17 | 2,045 | 1,825 | 17 | 3,870 | 3,887 | 1,411 | 1,188 | 8 | 2,599 | 2,607 |
| Province 18 | 2,195 | 2,004 | 30 | 4,199 | 4,229 | 1,516 | 1,300 | 11 | 2,816 | 2,827 |
| Province 19 | 3,388 | 2,869 | 30 | 6,257 | 6,287 | 2,281 | 1,923 | 18 | 4,204 | 4,222 |
| Province 20 | 2,203 | 1,986 | 24 | 4,189 | 4,213 | 1,397 | 1,359 | 10 | 2,756 | 2,766 |
| Province 21 | 1,553 | 1,262 | 39 | 2,815 | 2,854 | 1,005 | 816 | 14 | 1,821 | 1,835 |
| Province 22 | 1,499 | 1,344 | 11 | 2,843 | 2,854 | 999 | 887 | 3 | 1,886 | 1,889 |
| Province 23 | 2,916 | 2,616 | 32 | 5,532 | 5,564 | 1,965 | 1,812 | 12 | 3,777 | 3,789 |
| Province 24 | 2,251 | 1,950 | 42 | 4,201 | 4,243 | 1,442 | 1,340 | 7 | 2,782 | 2,789 |
| Province 25 | 2,466 | 2,343 | 58 | 4,809 | 4,867 | 1,650 | 1,566 | 11 | 3,216 | 3,227 |
| Province 26 | 651 | 630 | 20 | 1,281 | 1,301 | 353 | 321 | 6 | 674 | 680 |
| Province 27 | 1,818 | 1,626 | 7 | 3,444 | 3,451 | 1,255 | 1,080 | 5 | 2,335 | 2,340 |
| Province 28 | 1,803 | 1,583 | 20 | 3,386 | 3,406 | 1,245 | 1,077 | 5 | 2,322 | 2,327 |
| Province 29 | 927 | 841 | 14 | 1,768 | 1,782 | 568 | 532 | 6 | 1,100 | 1,106 |
| Province 30 | 1,113 | 1,015 | 8 | 2,128 | 2,136 | 691 | 726 | 5 | 1,417 | 1,422 |
| Province 31 | 1,190 | 1,130 | 36 | 2,320 | 2,356 | 779 | 805 | 4 | 1,584 | 1,588 |
| Province 32 | 709 | 671 | 8 | 1,380 | 1,388 | 452 | 477 | 4 | 929 | 933 |
| Sum | 59,137 | 52,606 | 571 | 111,743 | 112,314 | 39,315 | 35,280 | 213 | 74,595 | 74,808 |