札幌大学総合論叢　第5号（1998年3月）

〈研究ノート〉

Critique of Stumpf and Stanley's Gender-Related Differences on the College Board's Advanced Placement and Achievement Test, 1982-1992

Kazuo Akasaka

Introduction

The independent and dependent variables could be determined from the title of the article. The IV was gender (male and female student's) and the DVs were scores on the College Board's Advanced Placement tests (AP) and the Achievement Tests (ATs) from 1982-1992.

The Introduction provided an extensive historical background on gender differences on cognitive ability tests. Interest in the area was stimulated by research in the early 1970's. Since that time much research has been done in the area. In general, the authors found previous studies reporting an advantage for male students in the natural (mathematics, physics, chemistry) and computer sciences. Female students had a slight advantage in the languages (French, German, Spanish). As well, they found large differences in the enrollment in classes, with a higher percentage of males in the natural and computer sciences and the opposite for languages.

Of particular interest to the authors were the inconsistent findings regarding the decreasing trend in gender-related differences on cognitive ability tests and changes in enrollment. In order to examine the trend over time, the authors studied the College Board AP tests from 1984-1992 (1984 was the first year that the Computer Science test was included) and the Achievement Tests (ATs) from 1982-1992. The two tests differ in 3 main respects: a) the AP tests are college level, whereas the ATs are not, b) the AP test consists of some free response items whereas the the ATs are strictly multiple choice, c) special classes are taught to prepare students for the AP test. The authors believed these factors may be important in gender-related differences, as female students tend to score better than males on free response tests and generally get better grades in high school courses. Therefore, they expected gender-related differences to be smaller on the AP tests. However, theoretical explanations were not emphasized in this article. Instead, the authors state that their purpose was descriptive. Their aim was to study "the 'what' of gender-related differences on these tests and offer only some hypotheses about the 'why'". The authors do not state a hypothesis to be tested in the introduction. In fact, the authors stressed the limitations of the study. Specifically, that the populations were strongly self-selected and could not be used to generalize to a larger population. Their objective was to determine the trend (if any) in gender-related differences over time and stimulate further research.

Methods

In order to examine the temporal trends in gender-related differences the authors examined the AP tests and the ATs over time. 4 spaced apart years of each test between 1982-1992 were considered. The subjects, therefore, were the students who took these tests during the intervals studied. The total number of examines was 236 314 (1984) and 566 036 (1992) for the AP tests, and 196 991 (1982) and 199 538 (1992) for the ATs.

In addition to the descriptive statistics compiled by the testing company (score means, standard deviations and percentage enrollment by gender), the authors also calculated the effect size (d) and Upper (UTR) and Lower (LTR) Tail Ratios. The UTR and LTR examined the ratio of male to female students in the highest and lowest possible score ranges. For example, a high UTR showed an advantage for that gender (3:1 for males indicated 3 males scored in the top score range for each female) and a high LTR showed a disadvantage for that gender (3:1 for males indicated that 3 males scored in the lowest range for each female).

Results

The results section was divided into 5 aspects of interest to the authors.

Gender imbalance in Participation.

The authors displayed two tables, on for the AP tests and one for the ATs. Each table showed, for each subject area, the number of examines (N), the percentage of females enrolled (%), effect size indexes for the mean differences between males and females (d), and the UTR and LTR ratios. These statistics were shown for 1982 and 1992 for the AT and 1984-1992 for the AP. The results of this section showed that there were large differences in gender participation in various subjects. Females predominated in languages and Psychology, whereas males predominated in Computer Science, Physics, Chemistry and Mathematics.

Changes in Participation

The authors compared the average percentage of male examines in 1982 (AT) and 1984 (AP) with the average percentage in 1992 (AT & AP). The comparisons revealed a slightly higher female attendance and more balanced distribution across subject areas. However, even in 1992, large differences in enrollment remained. Notably, far more males than females took the Computer Science test (79% and 87% male enrollment for the two tests) and Physics tests (69%, 78%, and 80% male enrollment for the three tests).

Gender-Related Differences in AP examination Scores

The two statistics used examine differences were effect size and UTR and LTRs. The authors found that several natural science subject areas had effect sizes over .20. Computer Science was the highest with an effect size of .59. Males also showed an advantage in the UTR for Computer Science (2.77:1) and in each of the Physics tests (2.15:1). Females showed a slight advantage in the languages (no effect size was over .40). However the overall advantage of male students was not dramatic (males had an advantage of an unweighted 0.13 standard deviation). The UTR of scores across AP examinations was only 1.31:1 in favor of male students. There was no noteworthy difference in LTR overall (1.04:1 for males).

Gender-Related Differences in AT Examination Scores

The effect size differences were higher for various subjects than the AP tests, as expected by the authors. For example, European History had an effect size of d=.63. The UTRs and LTRs also showed a large advantage for male students in the natural sciences and American History. For the ATs, gender-related differences across all subjects were larger than the AP examinations. Males outscored females by an unweighted 0.27 standard deviation. The UTR averaged across subject areas was 2.43:1, indicating a much higher ratio of males in the top score range. However, more males scored in the lowest range than on the AP tests, with an average LTR of 1.43:1 in favor of males. This means that more males scored in the top and bottom ranges.

Changes in d over time

The authors conclude that the changes in effect sized over time was smaller than they expected. Female students gained in Computer Science, General Art, Music Theory and Spanish Literature, but lost ground in Latin and French Literature.

The authors focused on Computer Science and American History in this section as they were both high enrollment classes that showed some change in effect size. They provided a graph for both subjects displaying effect size, UTR and LTR over time. The Computer Science graph clearly indicates that female students are increasingly scoring closer to male students. The difference trend in effect size, UTR and LTR was narrowing as female students scores improved relative to the males. The American History graph displayed a steady but not dramatic narrowing of the difference trend as females began to score slightly better. The trend was not as strong as the Computer Science trend, but a change could be perceived.

The authors emphasized, however, that most of the tests showed little change over time and that Computer Science and American History were the only noteworthy changes.

Discussion

The Discussion section was relatively small. The authors made three observations form the data: a) the percentage of males and females taking the tests was still unbalanced, b) there were small to moderate gender-related differences in the scores for a number of subjects (Computer and Natural Sciences), and c) though 2 subjects

(Computer Science and American History) showed some long term changes in score differences favoring females, the overall pattern of differences remained stable over time. Male students were favored in Natural and Computer Sciences, and female students were a little ahead in some languages.

The authors carefully discussed the limitations in interpreting their results. Most importantly was the descriptive nature of their study. The population was not randomly selected but rather strongly self-selected and therefore the results are not gereralizable. As well, they distinguish between academic achievement (ability to perform well on tests) and ability (the ability to perform within the subject area). They believe that abilities may not be accurately represented in achievement and placement tests.

They conclude that some of the subject examinations attract a specific gender. For instance, Psychology, French and Spanish attract more female students, whereas Computer Science, Physics and Mathematics attract more male students. They believe there is some indication that there is a trend toward more balanced registration. However, they state that with the current data, no firm conclusion can yet be drawn in this respect.

The authors offer several possible reasons why female students scores have improved relative to their male counterparts in Computer Science and American History: a) female students have become more successful in academic training in these areas (change in perceptions and attitudes), b) the tests may vary from year to year, sometimes favoring males and other times females, and c) the abilities of students choosing to take certain tests may change from year to year. Note, however, that these reasons were offered after the study results had been examined, and were not a *priori* hypotheses that were tested.

The authors offered 2 future areas for research. The first is to study why the difference in gender enrollment is still high in some subject areas. They note that there is a strong correlation between the percentage of males taking a test and the performance by males on that test. The higher the percentage of males enrolled, the higher the scores for males compared to females. The second area offered was to examine the reason for the observed narrowing of gender-related differences in Computer Science and American History, but not in other disciplines.

If I were to design a study related to this article, I would attempt to conduct an experiment to test whether gender-related differences on cognitive tests are physiological. The key to the study would be to randomly select both males and females. The randomly selected students would then take a short course in Computer Science (computer language or related material). After the course, the students would be tested. If the results showed that one gender outperformed the other it would be some evidence that physiological differences may play a role in gender-related differences on cognitive tests.

Personal Reactions

I had many personal reactions to this study. I was most curious about the observation that male students were more likely to score in the top range as well as the bottom range. It seems that male students can perform very well and very badly, whereas female students are more consistent. In Japan, this may result from the different pressures and expectations placed on boys and girls. More pressure is placed on male students to perform well in school, and higher expectations are also placed on the male students. This may result in some males responding well to the pressure and scoring high, and others rebelling against the pressure and scoring low. In Japan, less pressure and expectations are placed on females, perhaps allowing them to perform more consistently. However, even in Japan, this may be changing a little by little.

I believe that female students are scoring better in Computer Science because of the prevalence of computers in today's society. Both females and males now regularly use computers, so females may perceive this subject area as more important, thus performing better on the academic tests.

The findings on enrollment differences in this study parallel my own experience here at the Sapporo University. I belong to the Foreign Language department, and have noticed that enrollment in this department is almost entirely female students. There are some male applicants who want to enroll, but are unsuccessful. I am curious as to why.

Finally, I have noticed, that in Japan, almost all top professionals are male. This is also true in female related areas like medicine, academics, fashion design, and politics. In fact, in Japan, whenever a female becomes a top professional the newspapers often report it as big news. For example recently a woman professor was appointed President of a public university. This event was reported in every Japanese newspaper as big news. Why does this occur. I find it strange that this should be considered big news, but I suppose it reflects our culture's attitudes towards women in top positions. I am curious as to the difference between Japan and Western cultures in this respect.

References

Benbow, C. P. (1990). Sex differences in mathematical reasoning ability among the intelectually talented: Further thoughts. Behavioral and Brain Sciences, 13.

Bridgeman, B., & Lewis, C. (1994). The relationship of essay and multiple-choice scores with grades in college courses. Journal of Educational Measurements, 31.

Clearly, T. A. (1992). Gender differences in aptitude and achievement test scores. In J. Pfleiderer (Ed.), Sex equity in educational opportunity, achievement, and testing. Princeton, NJ: Educational Testing Service.

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

Halpern, D. F. (1992). Sex differences in cognitive abilities (2nd ed.). Hilsdale, NJ: Erlbaum.

Kimball, M. M. (1989). Anew perspective on women's math achievement. Psychological Bulletin, 105.

Stanley, J. C. (1993a). Boys and girls who reason well mathematically. In G. R. Bock & K. Ackrill (Eds.), The origins and developement of high ability. New York: Wiley.

Williams, S. W., Ogletree, S. M., Woodburn, W., & Raffeld, P. (1993). Gender Roles, computer attitudes, and dynamic computer interaction performance in college students. Sex Roles, 29.