Journal of the Korean Association for Research in Science Education Vol. 24 No. 1(Special Issue), pp. 17~28 (2004)

Differential Effect of Item Characteristics on Science Achievement Between Genders[1)]

Donghee Shin · Nanmoo Moon
(Dankook University)

ABSTRACT

The purpose of this study is to determine the patterns of differences between genders in science achievement. Eleventh grade, 140 female and male students were sampled from a school in Seoul. According to the analysis results of pilot study, 20 items were finally selected for the main study. To sharpen our interpretations of the factors and provide some confirmation, we supplemented the statistical analysis with a more detailed cognitive study of the items using think-aloud protocols and interviews with student test takers. The analysis of this study took into account the different item formats, contexts, and presentation styles. The findings are as follows: First, there was no significant gender difference between multiple-choice and open-ended items. Second, male students achieved significantly better in the context of everyday life in multiple-choice items. Third, male students favored items presented as written texts. Fourth, in problem-solving process, female students tend to apply their science concepts, whereas male students tend to apply their everyday experiences. The results of this investigation indicate that gender difference in science achievement depends heavily on item characteristics.

Key words: gender differences, science achievement, item formats, contexts, presentation styles, problem-solving process

Ⅰ. Introduction

Female students have made lower scores than their male counterparts in most large-scale science achievement tests (Beller & Gafni, 1996; Hedges & Nowell, 1995; Linn & Hyde, 1989; Shin *et al.*, 2002). Gender difference in science achievement works as one of the various factors that reaffirm the universal belief of 'science as a male-friendly subject'. Such a deep-rooted perception should be considered as a serious problem in that it may weaken female students' motive for continuing their science learning. Whereas the importance of science and technology has been increasing, the number of students intending to major in science-related areas has been decreasing every year. From this perspective, it is inevitable to be concerned about gender differences in science achievement which result in female students' loss of interest in learning science.

Gender differences in science achievement have already been an issue in the science education community since 1980s (Jones & Kirk, 1990; Kahle & Lakes, 1983; Kahle *et al.*, 1985) and

*Received on 27 May 2003 **Donghee Shin(dss25@dankook.ac.kr)
1)The present research was conducted by the research fund of Dankook University in 2003.

several studies of science education in Korea have recently been interested in gender issues (Choi & Kim, 2001; Song *et al.*, 1992; Shin, 2000). In particular, the result that Korea emerged as a country with one of the largest gender gaps in the recent international science comparative studies has accelerated gender-related research (Shin *et al.*, 2002; Shin & Park, 2002).

Most previous studies have used the results of large-scale achievement tests to support their argumentation of male students superiority in science, and it is important to note that most of the tests consisted of multiple-choice items. Therefore, it should be examined whether multiple-choice items may cause gender inequality. Bolger & Kellaghan(1990) argued that item formats affect assessment results between genders.

The hypothesis that multiple-choice items favor male students were suggested on the basis of the fact that most large-scale achievement tests, which result in male superiority, consist of multiple-choice items (Bell & Hay, 1987; Hamilton, 1998; Harding, 1980; Klein *et al.*, 1997; Paris *et al.*, 1991). Several studies agreed that male students achieved better both in multiple-choice and open-ended items, whereas the mean difference was larger in multiple-choice items (Bolger & Kellaghan, 1990; Shin & Noh, 2002). Klein *et al.*(1997) also reported that female students excelled male students in open-ended items. These results are connected to the fact that multiple-choice items favor male students who tend to "guess" and "take risks" in solving problems (Ben-Sakhar & Sinai, 1991). Female students' excellence in open-ended items may attribute to female students' outstanding verbal ability (Stumpf & Stanley, 1996), which is denied by the results that gender difference between two-item formats was larger in mathematics rather than in English (Bolger & Kellaghan, 1990). Although it was clear that mean difference was larger in multiple-choice items, there was no relationship between writing ability and achievement (Bridgeman & Lewis, 1994).

Young *et al.*(1994) denied male- and female students' differential effect on item formats. Cole(1997), Wester & Henriksson(2000) reported that item formats did not affect on gender difference in achievement. The differential effect of item format by science area was significant only in open-ended items of physics, which showed male students' significantly higher grades (Dimitrov, 1999).

Some studies have shown that gender differences are affected by the factor of item context. Male students did better in a novelty context (Haggerty, 1987), in science knowledge acquired from out-of-school (Hamilton, 1998), and in science items from out-of-curriculum (Hanna, 1989). Whereas female students achieved better in in-school test such as a formative test, male students achieved better in out-of-school test such as a large-scale standardized test (Rennie, 1991). Above mentioned studies all agree that female students tend to favor familiar items.

Gender differences have also been found by item presentation styles. As a result of Swedish Scholastic Aptitude Test(SweSAT), male students excelled female counterparts in items presented as charts, tables, maps, and diagrams (Willingham & Cole, 1997). Such results agree with Feingold's study(1988), with SAT in USA, and Shin & Noh's study(2002) with international comparative study in Korea. Wester(1992) presented some characteristics of female-favored items; items which required more accuracy in recognizing it, such as items with several written texts. That is, female students tend to solve problems more carefully and male students tend to guess the answer (Makitalo, 1993).

Wester-Wedman(1992) argued that gender difference in science achievement is caused, not by item formats such as multiple-choice or open-ended, but by item presentation styles given, such as graphs, diagrams, or written texts. It is clear problem-solving strategy presented in a table is one thing, and problem-solving strategy presented in a written text is another (Winn, 1993). Interpreting diagrams usually requires a spatial ability (Maury *et al.*, 1990), and finding hidden information in visualized data requires a kind of inferential ability (Winn & Holiday, 1982). Aberg-Bengtsson(1999) interpreted male students' superiority in items with graphs as their excellence in calculating ability.

This study was initiated to make a more detailed analysis from recent results that, whereas Korean male students outperformed female counterparts in most large scale science achievement studies(Kim *et al.*, 2001; Mullis *et al.*, 2000; Shin & Noh, 2002), male- and female students' science achievement are dependent on item characteristics, such as item format, item context, and item presentation style(Shin & Park, 2002; Shin *et al.*, 2002, Shin & Kim, 2003). In this study, we investigated the gender differences in science achievement according to item formats, item contexts, and item presentation styles. We also investigated the gender differences in problem-solving strategy by interviewing female and male students' problem-solving process. The result of this study may be a suggestion to provide a basis of finding characteristics of female-favored items and reflecting them in a real science assessment in Korea.

Ⅱ. Method

1. Subjects

Eleventh grade, 140 female and male students were sampled from a high school in Seoul. They all had been exposed to the same science educational opportunities at school, in that they learned the same science subjects until 10th grade. To minimize the effect of science learning experience, students who intended to apply to colleges of humanities were selected. Seventy-one students (female: 39, male: 32) were examined in multiple-choice items, 69 students(female: 39, male: 30) were examined in open-ended items. Sixteen volunteer students out of 140 students were interviewed. The gender ratio in interviewees were same. It took approximately 40 minutes to interview a person.

2. Instruments

In the pilot study, 28 earth science items in the 7th grade "Science" subject were developed. All items were developed in two sets, multiple-choice and open-ended formats. All characteristics of two sets of items, including item contents, were the same, except item format. In the case of multiple-choice items, correct answers were coded as 1, and incorrect answers as 0. In the case of open-ended items, partial credits were applied, which was coded 1 as full credit, 0.5 as partial credit, and 0 as no credit. The highest score was 20, and the lowest score was 0. The content of the instrument was validated by an expert in geoscience education, a geologist, and a science teacher. The instruments used in the pilot study were estimated from item difficulty level, discrimination level, and reliability. The instruments in the pilot study had .541 in item difficulty level, .366 in item discrimination level, and .782 in reliability. According to the analysis results of the pilot study, 20 items were finally selected for the main study (Table 1). The item reliability of the main study in Cronbach's α was .706 in multiple-choice items and

.656 in open-ended items. Four out of the final 20 items, showing the meaningful gender differences in various perspectives, were selected for an interview.

3. Data Analysis

The quantitative analysis used SPSS(version 11.0). To sharpen our interpretations of the factors and provide some confirmation, we supplemented the statistical analysis with a more detailed cognitive study of the items, using think-aloud protocols and interviews with student test takers. Students were asked to speak aloud their problem-solving process. They summarized and elaborated their reasoning in 4 items with showing the most meaningful gender differences in achievement. To gain additional insight into the meaning and nature of item context, students were asked where they had learned about the content of two items. Students were encouraged to elaborate on their responses, when appropriate. The protocol also included detailed instructions concerning the procedure. The ability of students' problem-solving strategy was estimated by four stages, understanding, deriving from related concepts, setting up, and reviewing, which were modified from Jeon & Noh(2001)'s study. The full credit in each stage was 3 (Table 4).

Ⅲ. Results and Discussion

1. Gender differences in item formats, contexts, and presentation styles

The total mean difference between genders is presented in Table 1. In the multiple-choice items, the mean of female and male students were 6.98 and 8.40, respectively. In the open-ended items, the mean of female- and male students were 4.51 and 5.27, respectively. This means male students achieved better than female students in both item formats, with no significant difference. This is in contrast with the achievement in large-scale standardized tests, which resulted in male students' outstanding score (Martin *et al.*, 1999; OECD, 2001) and in an agreement with the result of female students' significantly higher achievement in in-school tests (Rennie, 1991). Such mean comparison of two item formats between genders does not support previous studies showing that male students favor multiple-choice items, and female students favor open-ended items (Bolger & Kellaghan, 1990; Bridgeman & Lewis, 1994; Klein *et al.*, 1997). And it supports Wester-Wedman(1992)'s and Wester(1995)'s argument that gender differences in science achievement cannot be explained by item format.

Although there was no significant gender difference by item formats, mean differences between two genders were 1.60 in multiple-choice items and 0.75 in open-ended items, which was larger in multiple-choice items. It is noteworthy that items showing significant gender differences were all multiple-choice items, whereas open-ended items did not reveal any significant gender difference (Table 1).

Figure 1 indicates that the minimum score was almost the same in both genders, whereas the maximum score of male students was much higher than that of female students. Despite the fact there was no significant difference between genders, male students' median was higher than female students', meaning that more male students got scores above mean. And male students' score distribution was larger than female students', which indicates male students

Table 1. A classification of item characteristics

Item	Area	Presentation Style	Description	Multiple-choice					Open-ended				
				Mean			F	p	Mean			F	p
				Female	Male	Total			Female	Male	Total		
1	Fluid	Data	Solar radiation energy on color	.03	.19	.10	5.432	.023*	.12	.07	.09	.658	.420
2	Fluid	Sentence	Definition of high pressure	.41	.66	.52	4.407	.039*	.03	.06	.06	1.710	.196
3	Fluid	Sentence	Ozone of atmospheric sphere	.36	.63	.48	5.210	.026*	.37	.37	.37	.002	.965
4	Fluid	Sentence	The extermination of a dinosaur	.64	.63	.63	.019	.891	.36	.48	.41	1.083	.302
5	Fluid	Graph	Reversed layer of temperature	.51	.59	.55	.455	.502	.41	.47	.44	.524	.472
6	Fluid	Graph	Salinity distribution	.64	.34	.51	6.619	.012*	.55	.68	.61	1.353	.249
7	Fluid	Sentence	Humidity in mountain climbing	.44	.69	.55	4.663	.034*	.31	.22	.27	.716	.401
8	Solid	Graph	Earth's inner structure	.15	.38	.25	4.717	.033*	.15	.15	.15	.003	.960
9	Solid	Sentence	Use of granite and marble	.33	.56	.44	3.851	.054	.49	.53	.51	.361	.550
10	Solid	Table	Igneous rocks	.21	.44	.31	4.601	.035*	.30	.30	.30	.002	.961
11	Solid	Figure	Movements of plates	.44	.50	.47	.283	.596	.03	.10	.06	1.710	.196
12	Solid	Figure	The upper side of strata	.41	.44	.42	.052	.820	.41	.40	.41	.009	.924
13	Solid	Graph	Distribution of coral fossil	.51	.31	.42	2.929	.092	.05	.10	.07	.587	.446
14	Planetary	Sentence	Season changing	.26	.47	.35	3.550	.064	.09	.25	.16	3.511	.065
15	Planetary	Sentence	Relation between tidal range and moon	.15	.25	.20	1.012	.318	.04	.07	.05	.353	.554
16	Planetary	Sentence	Cause of Moon's phase change	.33	.25	.30	.574	.452	.03	.12	.07	2.467	.121
17	Planetary	Figure	The conjunction of sun and moon	.41	.34	.38	.322	.572	.14	.13	.15	.013	.910
18	Planetary	Sentence	Elements of planet atmosphere	.33	.25	.30	.574	.451	.21	.17	.19	.160	.691
19	Planetary	Picture	The relative size of stars	.28	.25	.27	.090	.766	.21	.22	.21	.017	.895
20	Planetary	Sentence	Our galaxy	.15	.22	.18	.484	.438	.24	.28	.26	.265	.609
Total				6.98	8.40	7.63	2.863	.095	4.51	5.27	4.84	.873	.353

$^{*}p < .05$

occupy both the highest and the lowest score in the open-ended items, both minimum and maximum scores of female students were lower than those of male students (Figure 2).

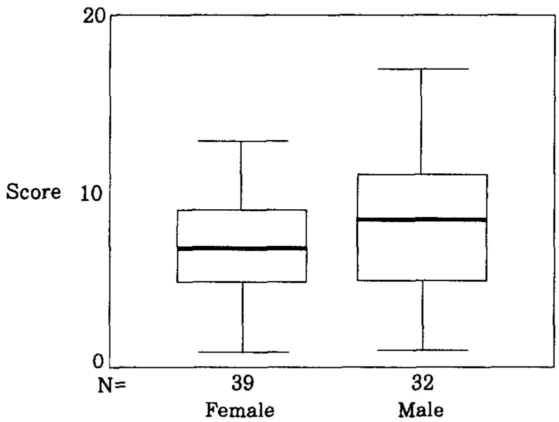

Fig. 1. Comparison of mean, median, and distribution between genders in multiple-choice items

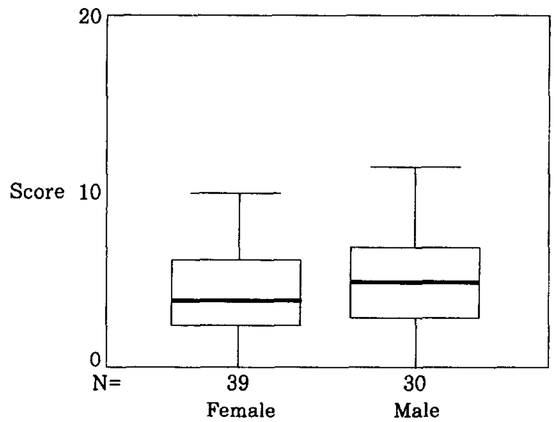

Fig. 2. Comparison of mean, median, and distribution between genders in open-ended items

Multiple-choice item 6 was the item showing the largest gender difference in achievement, as well as the only item showing female students' significant superiority over male counterparts among all items in this study (Table 1). Item 6 asks to select the area of the highest salinity in ocean water, from the graph of 'the distribution of evaporation and precipitation by the latitude'. The correct answer is '30° in latitude' with more evaporation than precipitation. More male students chose the incorrect answer of '40° in latitude', a meeting point of evaporation and precipitation in the given graph. Students who chose '40° in latitude' probably confused this graph in item 6 with another familiar graph of 'absorption and emission of solar energy in the earth' where the pattern is very similar. This indicates that male students chose the answer without reading the question carefully.

In contrast with such results of multiple-choice item 6, it is interesting to note that open-ended item 6, consisting of same contents with multiple-choice item 6, favored male students. In the case of open-ended items, students should read the question carefully because they select a correct answer without a reference to given graphs or examples. Male students' different response rate by item formats indicates that they tend to choose an answer easily without a thorough consideration. This can be connected to Klein *et al.*(1997)'s result, explaining female students' higher achievement in the items which required delicate attention.

Except for multiple-choice item 6, the other six items showing significant gender differences (numbers 1, 2, 3, 7, 8, and 10) were all male-favored, which is noteworthy because all of them were presented in the context of everyday life instead of school science. Table 2 compares the gender difference in achievements by item contexts, showing that male students achieved significantly better than female students in multiple-choice items presented in the context of everyday life. This result is consistent with previous studies that show female students favored items in familiar context such as science curriculum- or textbook-based (Haggerty, 1987; Hamilton, 1998; Hanna, 1989).

Table 2. Mean comparison between genders by item context

Context	Multiple-choice					Open-ended				
	Mean (SD)			F	p	Mean (SD)			F	p
	Female	Male	Total			Female	Male	Total		
School -science	4.28 (2.21)	4.75 (2.77)	4.49 (2.47)	.627	.431	2.90 (2.33)	3.15 (2.22)	3.01 (2.27)	.208	.650
Everyday life	2.72 (1.72)	3.66 (1.85)	3.14 (1.83)	4.873	.031*	1.73 (1.29)	1.77 (1.28)	1.75 (1.28)	.013	.909

*p<.05

Gender differences in item presentation styles were presented in Table 3. Whereas there was no significant gender difference in items presented as tables, graphs, or charts, male students achieved significantly better than female counterparts in items presented as written texts. This is in contrast with the previous results that male students significantly outstood female students in items presented as figure or graph (Hamilton, 1998; Shin & Park, 2002).

Table 3. Comparison of item presentation style

	Mean (SD)			F	p
	Female	Male	Total		
Tables, Graphs, or Charts	2.96 (1.93)	3.21 (2.21)	3.07 (2.05)	.522	.471
Written texts	2.56 (1.89)	3.38 (2.18)	2.92 (2.06)	5.688	.018*

*p<.05

2. Gender differences in problem-solving strategy

As a result of quantitative analysis, multiple-choice items 2 and 8, which revealed the largest gender differences in achievement, were selected for interview items to understand female and male students' significant gender differences in problem-solving strategy (Table 4). Comparison of female and male students' scores in each problem-solving process indicates that the score of male students in items 2 and 8 was higher at the first stage, understanding item. This result consistently appeared, regardless of interviewers' total achievements, which can be interpreted partly as male students spent more time to understand item, and partly as female students tended to read questions more quickly than male students.

At the stage of deriving from related concepts, although there was no large difference between genders, the information sources from which students derive their concepts were quite different between genders. Female students had a tendency to solve problems by introducing science concepts, for example, "Red color has the largest proportion to the solar constant so it absorbs sunlight in the highest amount" (dY). On the contrary, male students had a tendency to solve problems by introducing their life experiences, for example, "I felt hotter in black T-shirts so the black color absorbs solar energy in the highest amount" (aD). In the stages of setting-up

Table 4. Problem-solving comparison in items with the largest gender differences(Items 2 & 8)

		Multiple-choice								Open-ended							
		Female				Male				Female				Male			
Item		aL	bK	cC	dY	aD	BS	cY	dK	aN	bJ	cK	dG	aK	bH	cG	dP
2	Score*	1	1	1	0	0	1	1	0	0	0	0	0	1	0	1	0
	Understanding	2	2	1	1	2	2	1	2	1	1	1	1	1	1	1	1
	Deriving	1	1	1	0	0	1	0	1	1	1	0	0	1	0	1	1
	Setting up	1	1	1	1	1	1	1	1	1	2	3	2	1	0	1	2
	Reviewing	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0
8	Score	1	0	0	1	1	1	0	1	0	0	1	0	0	0	1	1
	Understanding	1	1	1	1	1	3	1	2	2	1	1	2	2	1	2	2
	Deriving	0	1	1	1	2	1	0	1	2	0	0	2	2	1	1	1
	Setting up	1	1	1	1	2	1	0	2	0	2	2	2	1	1	2	1
	Reviewing	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0

* 1 means correct answer and 0 means incorrect answer.

and reviewing problems, there were little differences between genders.

As a result of quantitative analysis, items 15 and 17, which assess similar concept and cognitive process but present items differently, were selected for interview items to understand female- and male students' differences by item presentation styles (Table 5). Item 15 asks moon phase when the tidal range is the smallest, presented in written texts and item 17 asks the position when the moon is not seen in the earth, presented in the figure of position of the sun, earth, and moon in space. Female students spent more time to understand items with figure, whereas male students spent more time to understand items with written texts. This means female students firstly try to understand figures before deriving related concepts. Male students tend not to understand figures carefully, and even a male student solved it without using a figure.

Table 5. Problem-solving comparison by item presentation styles (Items 15 & 17)

		Multiple-choice								Open-ended							
		Female				Male				Female				Male			
Item		aL	bK	cC	dY	aD	bS	cY	dK	aN	bJ	cK	dG	aK	bH	cG	dP
15	Score*	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0
	Understanding	1	1	1	1	3	1	2	2	1	1	1	1	3	2	1	1
	Deriving	0	2	1	1	1	1	0	1	2	1	1	0	2	1	0	1
	Setting up	1	2	1	1	1	2	1	1	3	2	0	3	3	1	2	2
	Reviewing	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0
17	Score	1	1	1	0	0	1	0	1	0	0	0	0	1	0	1	1
	Understanding	2	1	2	2	1	1	3	2	1	2	1	1	1	1	2	1
	Deriving	0	0	1	1	1	0	1	1	1	0	1	0	1	0	1	0
	Setting up	2	2	1	2	2	3	1	1	1	1	2	1	2	2	1	1
	Reviewing	0	0	0	1	0	0	1	1	0	0	0	1	0	0	1	0

* 1 means correct answer and 0 means incorrect answer.

Whereas there was no gender differences in item 17 presented in a figure, male students did significantly better than female counterparts in item 15 presented in written texts. From this is induced that male students focused on understanding what the question asked, by reading it repeatedly, which supports the quantitative result that male students achieved better in items with written texts. There was no gender difference at the stage of deriving from related concepts, setting up, and reviewing. In summary, above results coincide with previous studies arguing problem-solving process is differed by each item and each student (Hamilton *et al.*, 1997; O'Neil & Brown, 1998).

IV . Conclusions

At a national level, females' recovery of self-confidence and active participation in science are essential to promote a nation's competitiveness in science and technology, and further in economic power. This study has showed that female and male students' reaction depended on item characteristics, such as item formats, contexts, and presentation styles. Specifically, this study supports previous results that female students favored items in the context of school science. This result points out another issue in female students' science achievement. That is, female and male students tend to show similar achievement in a unstandardized in-school test setting with many textbook-or teacher-oriented items, whereas female students tend to be inferior in a standardized out-of-school test setting with many unfamiliar items, such as national or international science achievement study.

Another interesting point in this study is female students' favor in items with figures, rather than in written texts, although they spent more time to solve such items. Through this result, it is premature to conclude male students' superiority in verbal ability, even if they achieved better in items presented as written texts in this study. Likewise, we should not jump to conclusion that female students are inferior in science ability, even if they achieved worse in the item of everyday life context in this study.

As a result of this study, we can rightly free ourselves from the stereo-typed judgement on female or male students' science ability to the recognition of female-and male students' differential reaction depending on item characteristics. Unless the objective and valid perception of female students' scientific ability is embedded in our society, the lack of females in science community will continue. Therefore, the science education community in Korea should pay more attention to developing more gender-equitable science assessment tools.

References

Aberg-Bengtsson, L. (1999). Dimension of performance in the interpretation of diagrams, tables, and maps: Some gender differences in the Swedish Scholarstic Aptitude Test. *Journal of Research in Science Teaching*, 36(5), 565-582.

Bell, R.C. & Hay, J.A. (1987). Differences and biases in English language examination formats. *British Journal of Educational Psychology*, 57, 212-220.

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

Ben-Shakhar, G. & Sinai, Y. (1991). Gender differences in multiple-choice tests: The role of differential guessing tendencies. *Journal of Educational Measurement*, 28(1), 23-35.

Bolger, N. & Kellaghan, T. (1990). Method of measurement and gender differential guessing tendencies. *Journal of Educational Measurement*, 28, 23-25.

Bridgeman, B. & Lewis, C. (1994). The relationship of essay and multiple-choice scores with grades in college courses. *Journal of Educational Measurement*. 31, 37-50.

Choi, K. & Kim, K. (2001). The effects of girl-friendly science content and method on middle school girl's attitude toward science. *Journal of the Korean Association for Research in Science Education*, 21(1), 149-159.

Cole, N. S. (1997). *The ETS gender study: How males and females perform in educational settings*. Prinston, NJ: Educational Testing Service.

Dimitrov, D. M. (1999). Gender differences in science achievement: Differential effect of ability, response format, and strands of learning outcomes. *School Science and Mathematics*, 99(8), 445-450.

Feingold, A. (1995). The additive effects of differences in central tendency and variability are important in comparisons between groups. *American Psychologist*, 50, 5-13.

Haggerty, S. M. (1987). Gender and science achievement: A case study. *International Journal of Science Education*, 9, 271-279.

Hamilton, L. S. (1998). Gender differences on high school science achievement tests: Do format and content matter? *Educational Evaluation and Policy Analysis*. 20, 179-195.

Hamilton, L. S., Nussbaum, E.M., & Snow, R.E. (1997). Interview procedures for validating science assessments. *Applied Measurement in Education*, 10(2), 181-200.

Hanna, G. (1989). Mathematics achievement of girls and boys in grade eight: Results from twenty countries. *Educational Studies in Mathematics*, 20, 225-232.

Harding, J. (1980). Sex differences in performance in science examinations. In R. Deem (Eds.), *Schooling for Women's Work*. London: Routledge and Kegan Paul.

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

Jeon, K. & Noh, T. (2001). The effects of problem solving strategy and paired think-aloud problem solving on high school students' chemistry problem solving. *Journal of the Korean Association for Research in Science Education*, 21(2), 289-298.

Jones, A. T. & Kirk, C. M. (1990). Gender differences in students' interests in applications of school physics. *Physical Education*, 25, 308-313.

Kahle, J. B. & Lakes, M. L. (1983). The myth of equality in science classroom. *Journal of Research in Science Teaching*, 20(2), 131-140.

Kahle, J. B., Lakes, M. L., & Cho, H. (1985). An assessment of the impact of science experiences on the career choices of male and female biology student. *Journal of Research in Science Teaching*, 22(5), 385-394.

Kim, J., Lee, M., Yoo, J., & Han, K. (2001). *National assessment of educational achievement in science: Year 2001*. Korea Institute of Curriculum & Evaluation Research Report PRE 2001-5-5.

Klein, S. P., Jovanovic, J., Stecher, B. M., McCaffrey, D., Shavelson, R. J., Haertel, E., Solano-Flores, G. & Comfort, K. (1997). Gender and racial/ethnic differences on performance assessments in science. *Educational Evaluation and Policy analysis*, 19(2), 83-97.

Linn, M. C. & Hyde, J. S. (1989). Gender, mathematics, and science. *Educational Researcher*, 18, 17-19.

Makitalo, A. (1993). *Gender differences in performance on the DTM subtest in the Swedish Scholastic Aptitude Test as a function of item position and cognitive demands.* Unpublished thesis, Goteborg University, Sweden.

Martin, M. O., Mullis, I. V. S., Gonzales, E. J., Gregory, K. D., Smith, T. A., Chrostowski, S. J., Garden, R. A. & O'Connor, K.M. (2000). TIMSS 1999: *International Science Report.* International Study Center Boston College Lynch School of Education.

Maury, S., Janvier, M. & Baille, J. (1990). Diagram processing procedures. *European Journal of Psychology of Education*, 5, 293-307.

Mullis, I. V. S., Martin, M. O., Fierros, E. G., Goldberg, A. L., & Stemler, S. E. (2000). *Gender differences in achievement: IEA's Third International Mathematics and Science Study*(TIMSS). International Study Center. Lynch School of Education, Boston College.

OECD (2001). *Knowledge and skills for life: First results from PISA 2000.* Paris: OECD Publications, 322p.

O'Neil, H. F. & Brown, R. S. (1998). Differential effects of question formats in math assessment on metacognition and affect. *Applied Measurement in Education*, 11, 331-351.

Paris, S. G., Lawton, T. A., Turner, J. C. & Ruth, J. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher*, 20(5), 12-20.

Rennie, L. J. (1991). The relationship between affect and achievement in science. *Journal of Research in Science Teaching*, 28(2), 193-209.

Shin, D. (2000). An analysis of elementary "Chayon" textbooks: From a perspective of gender fairness. *Journal of the Korean Association for Research in Science Education, 20(2), 193-* 199.

Shin, D. & Park, C. (2002). Korean students' achievement of earth science in international comparative studies: From the perspectives of gender differences. *Journal of Korean Earth Science Society*, 23(3), 207-220.

Shin, D., Park, C. & Noh, K. (2002). Gender differences in achievement of earth and environmental area in PISA 2000. *Journal of the Korean Association for Research in Science Education*, 22(1), 40-53.

Shin, D. & Kim, D. (2003). Assessment type and gender differences in science achievement. *Journal of Korean Earth Science Society, 23(3), 265-275.*

Song, J., Pak, S. & Jang, K. (1992). Attitudes of boys and girls in elementary and secondary schools towards science lessons and scientists. *Journal of the Korean Association for Research in Science Education*, 12(3), 109-118.

Stumpf, H. & Stanley, J. C. (1996). Gender-related differences on the college board's advanced placement and achievement tests, 1982-1992. *Journal of Educational Psychology*, 88, 353-364.

Wester, A. (1992). *Item bias with respect to gender interpreted in the light of problem-solving strategies* (EM No. 3). Umea, Sweden: Department of Education, University of Umea.

Wester, A. (1995). The importance of item format with respect to gender differences in test performance: A study of open-format items in the DTM test. *Scandinavian Journal of Educational Research*, 39, 335-346.

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

Wester-Wedman, A. (1992). *Using open-ended questions in the DTM test (PM No. 56).* Umea, Sweden: Department of Education, University of Umea.

Willingham, W. W. & Cole, N. (1997). *Gender and fair assessment.* Mahwah, NJ: Erlbaum.

Winn, W. (1993). An account of how readers search for information in diagrams. *Contemporary Educational Psychology*, 18, 162-185.

Winn, W. & Holiday, W. (1982). Design principles for diagrams and charts. In D.H. Jonassen (Eds.). *Englewood Cliffs*, NJ: Educational Technology Publications.

Young, D. J. & Fraser, B. J. (1994). Gender differences in science achievement: Do school effects make a difference? *Journal of Research in Science Teaching*, 31(8), 857-871.