

평가 방법에 따른 과학 성취도에서의 성 차이

신동희 · 김동영
(단국대학교) · (신림고등학교)

Assessment type and gender differences in science achievement

Shin, Donghee · Kim, Dongyoung
(Dankook University) · (Shillim High School)

ABSTRACT

The purpose of this study is to investigate the differences of science achievement between female- and male students in different assessment types, expecting the reconsideration of their scientific ability. To accomplish this purpose, the school science achievement score of 1,487 female- and male students in 2000 were analyzed from the perspectives of assessment methods and gender differences, achievement level and gender differences, and gender characteristics in correlation between each assessment type and total science score. The science achievement score in this study includes mid-term and final paper-and-pencil test in two semesters and performance assessment. The results are as follows. First, in the total science achievement score, whereas female students achieved significantly better than male counterparts in the 7th grade, there was no significant difference between two genders in the 8th and 9th grades. Second, as for the close-ended items, all students, regardless of their gender and grade, showed no significant difference. As for the open-ended items, whereas female students achieved significantly better than male counterparts in the 7th grade, there was no significant difference between genders in the 8th- and 9th grades. As for the performance assessment, all female students, regardless of their achievement level, achieved significant better than male counterparts, which is very noteworthy. Third, in the uppermost achievement level, there was no significant difference not only in close- and open ended items but also in performance assessment. Fourth, male students' correlation between performance assessment and total science score was relatively higher than female students'. The result of this study, that female students did significantly better than male students in performance assessment, is interesting in that performance assessment is widely accepted as more reliable and valid method in measuring students' various ability than traditional paper-and-pencil test.

Key words: performance assessment, assessment type, gender differences in science achievement, correlation between performance assessment and total science score

I. 서 론

과학 성취도에서 남학생이 여학생보다 높은 성취 수준을 나타낸다고 보고한 연구는 많다(Beaton *et al.*, 1996;

Becker, 1989; Beller & Gafni, 1996; Goldstein, 1993; Hedges & Nowell, 1995; Martin *et al.*, 2000; OECD, 2001; Preece *et al.*, 1999; Ramsbottom-Lucier, 1995; Ryan & Fan, 1996; Schmitt & Dorans, 1990;

*2003.2.13(접수) 2003.5.26(최종 통과) **E-mail: dss25@dankook.ac.kr(신동희)

***이 연구는 2003학년도 단국대학교 대학연구비의 지원으로 연구되었음.

Steinkamp & Maehr, 1983; Valencia, 1995). 이들 연구는 대부분 대규모의 평가 연구로 표준화 검사(standardized test) 형태로 수행되었다. 객관식 문항 위주의 대규모 표준화 검사는 학생들이 가지고 있는 경험의 다양성이 반영되지 않아, 특별한 교육이 필요한 학생들을 잘못 인도할 가능성이 있다(Darling-Hammond, 1991; Oakes, 1985)는 문제점을 들 수 있다. 이 밖에도 표준화 검사를 지나치게 자주 활용하면 정보 활용 능력, 새로운 상황에서의 문제 해결 능력 등 학교 졸업 후 생활에서 필요한 능력이 감소한다(Bailey, 1989; Carnevale *et al.*, 1989)는 주장과 표준화 검사는 정보를 분화시키거나(decompose) 탈맥락적(decontextualize)으로 만든다(Resnick & Resnick, 1992)는 주장 등도 제기되었다. 이렇듯 객관식 문항 위주의 대규모 표준화 검사가 학생들의 능력을 신뢰 롭고 타당하게 측정하는 데 많은 한계를 드러내는 것으로 밝혀지고 있음에도 불구하고, 대규모 검사가 요구하는 채점의 객관성과 경제성, 시간 제약 등의 이유로 표준화 검사가 빈번하게 사용되고 있다.

최근 이러한 객관식 문항 위주의 표준화 검사 한계를 극복하기 위한 수행 평가 방법이 부상하고 있는데, 표준화 검사에서 나타난 성취도에 있어서 성별 차이가 수행 평가를 통해 해소될 것이라는 연구도 보고되고 있다. 그 근거로 여학생과 남학생은 서로 다른 능력을 가지고 있다기보다는 문항 유형의 영향을 받아 서로 다르게 문제를 해결한다(Frederikson, 1984; O'Neil & Brown, 1998; Paris *et al.*, 1991)는 주장, 수행 평가를 통해 학생들에게 그들의 지식과 이해를 표현하는 다양한 기회를 제공함으로써 성 차이를 비롯한 집단 간 차이를 줄일 수 있다(Wolf, 1993)는 주장 등이 있다. 실제로 미국의 과학 교육 개혁에서도 수행 평가를 활용하는 것이 여학생에게 유리할 것으로 판단(Council for Educational Development and Research, 1993)한 바 있다.

최근 수행 평가를 활용해 과학과에서의 성 차이가 줄어들거나 여학생 우위인 경향으로 나타난 결과가 많이 보고되고 있다. Bolger와 Kellaghan(1990)은 문항의 내용을 같게 하고 유형만 달리하여 여학생과 남학생의 성취도를 비교한 결과, 객관식 문항보다 주관식 문항이 여학생에게 유리하게 작용한다는 것을 밝혔다. Klein *et al.*(1997)와 DeMars(1998)의 연구에서도 여학생이 주관식 문항에서 남학생보다 더 높은 성취도를 보이는 결과가 나왔다고 보고했다. 또한, Jovanovic *et al.*(1994)도 과학과 객관식 평

가에서는 남학생과 여학생의 성취도가 의미 있는 차이를 보였으나, 과학과 수행 평가 문항에서는 성별에 따른 의미 있는 성취도 차이가 나타나지 않았다고 보고했다. Elwood(1994)도 수행 평가의 한 형태인 수업 시간에 하는 일(coursework element)을 중심으로 16세 학생들을 평가한 결과, 생물에서는 남학생들이 의미 있게 우세하고, 화학과 통합 과학에서는 성 차이가 거의 없으며, 물리에서는 여학생이 의미 있게 우세한 것으로 나타났다. 이는 기존의 객관식 위주의 평가에서 빈번히 보고되던 물상 과학에서의 남학생 우세, 생명 과학에서의 여학생 우세라는 대부분의 연구(Beaton *et al.*, 1996; Burkam *et al.*, 1996; Lee & Burkam, 1996; Bell, 2001; Johnson & Bell, 1987; Martin *et al.*, 2000)와 상반되는 주목할 만한 것이다.

한편, 교실에서 이루어지는 활동 위주의 수행 평가를 실시해도 과학과 성 차이가 줄어들지 않았다는 연구 결과(Jovanovic & Dreves, 1998)도 보고된 바 있다. 이 연구에서는 수행 평가의 인지적 기능보다 정의적 기능이 강조되었는데, 흥미, 중요성, 유용성, 자아 능력 인식 등의 정의적 측면에서 볼 때, 수행 평가 중심의 학습이 성 차이 해소에 큰 역할을 하지 못한 것으로 나타났다. 이러한 결과의 원인으로 이 연구에서는 남녀 학생이 가진 이전의 경험 차이, 교실 내 활동에서의 남학생의 적극성 등이 작용했기 때문으로 분석하고 있다.

우리 나라의 경우도 과학 성취도에서 성 차이가 매우 심각한 수준인데, 최근 이루어진 과학과 국제 비교 연구에서 우리 나라 학생들의 성 차이가 매우 큰 것으로 나타났다(신동희 등, 2002; 신동희와 박정, 2002; Martin *et al.*, 2000; OECD, 2001). 1999년 실시된 TIMSS-R(Third International Mathematics and Science Study-Repeat)의 경우도 1995년의 TIMSS에 이어 수학과 과학 공통적으로 여학생의 성취도가 남학생의 성취도에 비해 유의미하게 낮게 나타난 국가 중 하나에 우리 나라가 포함되었다(Martin *et al.*, 2000). 성 차이는 PISA(Programme for International Student Assessment) 1주기 연구에서 더욱 심화되어 과학 영역의 경우 우리 나라는 참여국 중 성 차이가 가장 큰 국가인 것으로 드러났다(OECD, 2001). 그러나, PISA에서 우리 나라 여학생들은 특히 상위 성취 수준의 여학생들은 남학생들보다 서술형 문항에 유리하고(신동희 등, 2002), 지식보다는 탐구를 측정하는 문항에 유리한 것으로 나타난 점(신동희와 박정, 2002)은

우리 나라 여학생들의 과학 성취도와 관련하여 다양한 가능성을 모색할 수 있는 계기를 마련해 주었다.

이상과 같은 논의를 바탕으로, 이 연구에서는 평가 방법에 따라 여학생과 남학생의 성취도가 각각 어떤 경향을 보이는지 파악하고자 한다. 특히, 과학 성취도에 있어서 여학생의 능력을 평가 방법별로 분석함으로써 여학생들과 남학생들의 과학 성취 능력에 대해 올바르게 인식하는 데 도움을 주고자 한다. 구체적으로 평가 방법에 따른 성 차이 경향, 성취 수준별 평가 방법에 따른 성 차이 경향, 성별 평가 유형 사이의 상관 관계 등을 알아보하고자 한다.

II. 연구 방법

1. 연구 대상

이 연구에서는 서울특별시에 위치한 남녀 공학 중학교¹⁾ 3개 학년 전 학생이 2000년도 1, 2학기 전체에 걸쳐 실시된 중간 고사, 기말 고사, 수행 평가 등에서 기록된 과학 학업 성취도를 분석했다. 분석 대상이 된 학생의 수는 중학교 1학년 442명(여학생: 198명, 남학생: 244명), 중학교 2학년 574명(여학생: 272명, 남학생: 302명), 중학교 3학년 461명(여학생: 207명, 남학생: 254명) 등 총 1,487명이다.

2. 평가 내용과 방법 및 자료 분석

분석 대상 학교에서 한 학기 동안 지필식으로 이루어지는 중간 고사와 기말 고사는 전체 평가 점수의 70%를 차지하며,

지필 고사 문항 중 최소 30%는 주관식 문항으로 구성되었다. 지필 검사는 크게 객관식 문항과 주관식 문항으로 구성되어 있고, 선다형, 연결형, 완성형, 단답형, 서술형 등의 유형을 갖는다. 수행 평가는 전체 평가 점수의 30%를 차지하며, 인지적 기능, 심동적 기능, 정의적 기능의 평가 내용을 모두 포함한다. 구체적인 수행 평가 방법²⁾으로 학기 중의 수업 태도, 과제물, 발표, 공책 정리, 실험 기구 취급, 실험 기능, 방학 과제물 등을 활용했다 (Table 1). 학생들의 학업 성취도 자료는 SPSS Window Program (Version 10.1)을 사용하여 분석했다. 학생들의 점수는 표준 점수인 T 점수로 환산하여 처리했다.

III. 연구 결과 및 논의

1. 평가 유형별 성 차이

전체적으로 학교 과학 성취도에 있어서 학년에 따라 성 차이 경향이 다르게 나타났는데(Table 2), 7학년의 경우 여학생이 남학생보다 통계적으로 유의미하게 높은 성취도를 보인 반면, 8, 9학년은 의미 있는 성 차이가 나타나지 않았다. 이는 학년이 올라갈수록 남학생들의 과학 성취도가 여학생들보다 높아진다는 기존의 연구 결과(Johnson, 1987; Keeves & Kotte, 1992)와 동일하다.

한편, 객관식 문항, 주관식 문항, 수행 평가 등 평가 유형에 따른 성 차이는 학년에 따라 다소 차이를 보였다. 객관식 문항의 경우, 7, 8, 9학년 모두 의미 있는 성 차이를 나타내지 않았다. 주관식 문항의 경우, 7학년에서 여학생들이 통계적으로 의미 있는 차이를 보이며 남학생들

Table 1. Assessment framework

Area	Assessment method	Time
Cognitive domain	Paper-and-pencil test - Close-ended - Open-ended	Mid- and final exam in the 1st and 2nd semester
Psychomotor skill	- Experiment/Activity report - Notebook organization Performance assessment	All learning processes during the 1st and 2nd semester
Affective domain	- Presentation - Homework - Science event	

1) 이 중학교는 2000년 당시 서울시교육청 수행 평가 시범 학교로 운영 중이었으며, 중산층 거주 지역에 위치하고 있다.

2) 보통 주관식 문항은 수행 평가의 한 유형으로 고려되고 있으나(백순근 편, 1998), 이 연구에서는 지필 검사에서의 주관식 문항을 수행 평가와 구분하기로 한다.

Table 2. Mean(T-score) for close-ended items, open-ended items, and performance assessment by grade

Grade	Gender	N	Close-ended				Open-ended				Performance				Total			
			Mean	SD	f	p	Mean	SD	f	p	Mean	SD	f	p	Mean	SD	f	p
7	female	198	49.65	10.44	.658	.418	51.48	9.20	8.006	.005	54.61	6.82	91.951	.000	51.62	8.86	9.538	.002
	male	244	50.42	9.43			48.79	10.47			46.26	10.60			48.69	10.68		
8	female	272	49.40	9.87	1.848	.175	50.15	9.86	.110	.740	52.69	9.70	40.111	.000	50.07	9.87	.025	.875
	male	302	50.54	10.10			49.87	10.14			47.57	9.65			49.94	10.14		
9	female	207	49.72	10.33	.294	.588	49.81	10.79	.141	.707	51.28	10.30	6.230	.013	49.99	10.24	.001	.977
	male	254	50.23	9.74			50.16	9.33			48.95	9.64			50.01	9.82		

Table 3. Gender differences for 3 assessment types by achievement level in Grade 7 (N=442)

Grade	Gender	N	Close-ended				Open-ended				Performance				Total			
			Mean	SD	f	p	Mean	SD	f	p	Mean	SD	f	p	Mean	SD	f	p
High (>25%)	female	56	60.30	2.55	12.158	.001	59.46	3.03	1.689	.196	59.85	3.36	5.545	.020	60.95	2.08	3.933	.050
	male	54	61.77	1.80			60.18	2.79			58.13	4.28			61.71	1.89		
Medium (25%-75%)	female	113	49.83	5.19	14.244	.000	51.69	5.92	.051	.822	54.09	5.58	52.599	.000	51.21	4.25	.323	.571
	male	109	52.45	5.16			51.86	5.46			47.58	7.67			51.53	4.34		
Low (<25%)	female	29	33.70	4.40	10.126	.002	35.29	6.17	1.505	.223	46.50	7.46	36.941	.000	35.18	4.03	.717	.399
	male	81	37.80	6.42			37.09	6.98			36.57	7.58			36.19	5.91		

보다 높은 성취도를 나타낸 반면, 8, 9학년에서는 의미 있는 성 차이가 나타나지 않았다. 수행 평가의 경우, 7, 8, 9학년 모두 여학생이 남학생보다 의미 있는 차이를 보이며 높은 성취도를 나타낸 점이 주목할 만하다. 이는 과학 성취도에서 주관식 문항이나 수행 평가를 활용하는 것이 여학생들에게 유리하다는 최근의 연구 결과들(신동희 등, 2002; Bolger & Kellaghan, 1990; DeMars, 1998; Elwood, 1994; Jovanovic et al., 1994; Klein et al., 1997)과 일치하는 것으로 우리나라도 외국의 경향과 다르지 않음이 드러났다.

2. 성취 수준별 평가 유형에 따른 성 차이 경향

학생들의 성취 수준에 따라 상위 25% 이내의 집단, 중위 50%의 집단, 하위 25% 이내의 집단 등 세 개 집단으로 나누어 평가 유형에 따른 성 차이 경향을 살펴보았다. 7학년의 경우(Table 3), 전체적으로 상위 집단의 남학생들이 상위 집단의 여학생들보다 유의미하게 높은 성취도를 보였으나, 중위와 하위 집단의 경우 의미 있는 성 차이가

나타나지 않았다. 객관식 문항에서는 상, 중, 하 집단 모두 남학생이 여학생보다 유의미하게 높은 성취도를 나타낸 반면, 주관식 문항에서는 상, 중, 하 집단 모두 유의미한 성 차이를 보이지 않았다. 수행 평가에서는 상, 중, 하 집단 모두 여학생이 남학생보다 유의미하게 높은 성취도를 보였다.

8학년의 경우(Table 4), 전체적으로 상, 중, 하 집단 모두 유의미한 성 차이가 나타나지 않았다. 객관식 문항과 주관식 문항에서도 상, 중, 하 집단 모두 유의미한 성 차이가 나타나지 않았다. 반면, 수행 평가에서는 7학년과 마찬가지로 상, 중, 하 집단 모두 여학생이 남학생보다 유의미하게 높은 성취도를 보였다.

9학년의 경우(Table 5), 8학년과 마찬가지로 전체적으로 상, 중, 하 집단 모두 유의미한 성 차이가 나타나지 않았다. 객관식 문항과 주관식 문항에서도 8학년과 마찬가지로 상, 중, 하 집단 모두 유의미한 성 차이가 나타나지 않았다. 반면, 수행 평가에서는 7학년이나 8학년과는 다소 차이 나는 경향을 보였는데, 중위 집단에서만 여학생이 남학생보다 유의미하게 높은 성취도를 보였고, 7학년과 9학

Table 4. Gender differences for 3 assessment types by achievement level in Grade 8 (N=574)

Grade	Gender	N	Close-ended				Open-ended				Performance				Total			
			Mean	SD	f	p	Mean	SD	f	p	Mean	SD	f	p	Mean	SD	f	p
High (>25%)	female	67	61.23	1.98	1,566	.213	61.54	2.04	.883	.349	60.69	4.55	22,178	.000	61.92	1.80	1,465	.228
	male	76	61.74	2.79			61.13	3.04			56.53	5.83			61.54	1.99		
Medium (25%-75%)	female	133	50.48	5.05	3,223	.074	51.39	5.51	.935	.334	53.37	6.87	59,094	.000	51.21	4.78	.330	.566
	male	155	51.59	5.43			50.75	5.71			47.52	6.03			50.88	4.92		
Low (<25%)	female	72	36.42	3.94	.057	.812	37.24	4.02	3,342	.070	44.00	10.69	11,965	.001	36.93	4.01	3,880	.051
	male	71	36.25	4.48			35.89	4.79			37.93	10.20			35.47	4.87		

Table 5. Gender differences for 3 assessment types by achievement level in Grade 9 (N=461)

Grade	Gender	N	Close-ended				Open-ended				Performance				Total			
			Mean	SD	f	p	Mean	SD	f	p	Mean	SD	f	p	Mean	SD	f	p
High (>25%)	female	55	61.33	3.80	.473	.493	59.54	4.69	.059	.808	57.73	5.37	3,226	.075	61.25	1.95	1,682	.197
	male	60	60.94	2.00			59.70	2.21			56.20	3.66			60.78	1.89		
Medium (25%-75%)	female	99	50.57	4.67	3,478	.063	51.71	5.40	.004	.947	53.61	5.91	12,689	.000	51.51	3.70	.214	.644
	male	132	51.77	2.92			51.67	4.74			50.75	6.14			51.76	4.24		
Low (<25%)	female	53	36.08	5.72	.229	.633	36.15	9.28	1,001	.319	40.04	12.02	.963	.329	35.44	5.63	.155	.695
	male	62	36.58	5.50			37.71	7.46			37.95	10.60			35.87	5.80		

년에는 유의미한 성 차이가 나타나지 않았다.

이상의 학년별 성취 수준에 따른 과학 성취도 성 차이 결과를 평가 유형별로 정리하면 Fig. 1과 같다. 객관식 문항의 경우, 7학년에서 상, 중, 하 집단 모두 유의미한 성 차이가 나타난 반면, 8학년과 9학년에서는 유의미한 성 차이가 나타나지 않았다. 이와 같이, 남학생들이 정확히 몰라도 추측해서 답할 수 있는 객관식 문항에서 유리하다는 기존의 연구들(Ben-Shakhar & Sinai, 1991; Champagne & Newell, 1992; Hanna, 1986; Rowley, 1974)과 가장 일치하는 결과는 7학년에서 나타났고, 8학년과 9학년에서는 뚜렷한 경향을 보이지 않았다. 주관식 문항의 경우, 7, 8, 9학년 모두 성취 수준별로 유의미한 성 차이가 나타나지 않아 주관식 문항에서 여학생들이 유리하거나 유의미한 성 차이가 나타나지 않는다는 기존의 연구들(Bolger & Kellaghan, 1990; DeMars, 1998; Klein 외, 1997)과 비슷하다.

가장 두드러진 특징을 보인 것은 수행 평가의 경우인데, 7, 8학년의 상, 중, 하 성취 집단 모두 여학생이 남학생보다 유의미하게 높은 성취도를 보인 반면, 9학년의 경

우 중위 집단의 여학생만 중위 집단의 남학생보다 유의미하게 높은 성취도를 보였고 상위와 하위 집단에서는 유의미한 성 차이가 나타나지는 않았다. 그러나, 이 경우에 있어서도 여학생의 성취도가 일관성 있게 남학생의 성취도보다 높게 나타났다. 이와 같이 상위, 중위, 하위 성취 수준 및 학년에 상관없이 수행 평가에서는 여학생이 남학생보다 우세한 것으로 나타나 수행 평가가 과학 성취도에서의 성 차이를 해소할 것이라고 주장한 연구 결과(Jovanovic 외, 1994; Wolf, 1993)를 뒷받침한다.

한편, 상위 10% 이내에 들어가는 최상위권 학생들의 평가 유형별 성취도 차이를 파악한 결과, 객관식 문항과 주관식 문항에서는 물론이고, 수행 평가에서도 유의미한 성 차이가 나타나지 않았다(Table 6). 다시 말해, 성취도가 매우 높은 학생 집단에서는 어떤 평가 방법을 사용해도 성취도 결과에서 아무런 성 차이가 나타나지 않은 셈이다.

3. 성별 평가 유형 간 상관 관계 파악

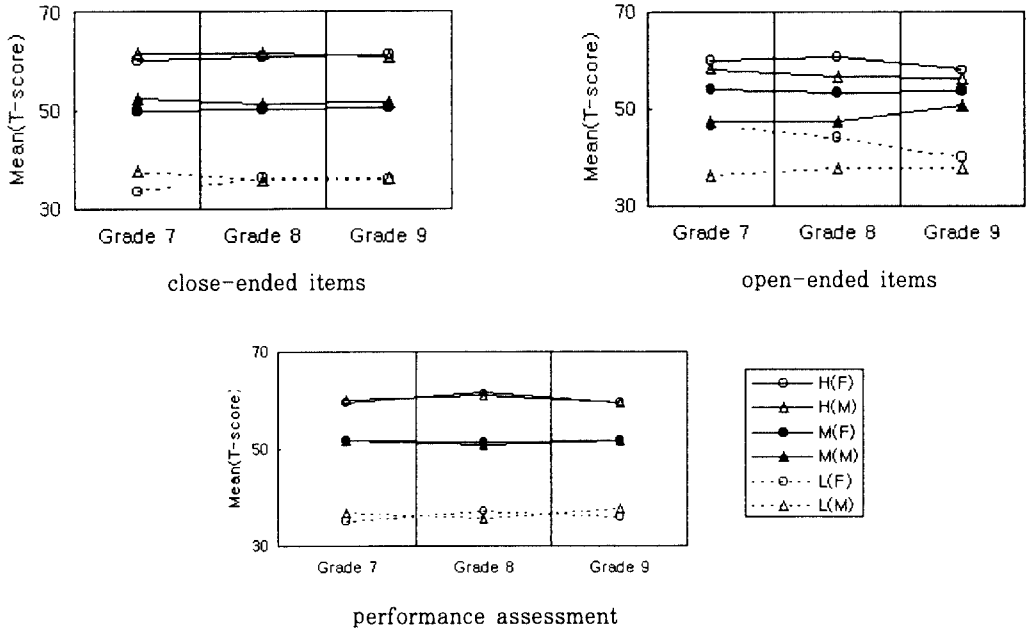


Fig. 1. Mean comparison of assessment types by achievement level

Table 6. Gender differences for 3 assessment types in the uppermost achievement level

Grade	Gender	N	Close-ended				Open-ended				Performance				Total			
			Mean	SD	f	p	Mean	SD	f	p	Mean	SD	f	p	Mean	SD	f	p
7	female	19	136.74	3.48			73.11	1.73			76.63	1.74			286.47	4.66		
	male	25	136.88	3.88	.016	.900	72.96	2.57	.045	.833	76.32	2.41	.227	.636	286.16	5.29	.042	.839
	total	44	136.82	3.67			73.02	2.23			76.45	2.13			286.30	4.97		
8	female	28	175.50	5.18			127.64	2.86			72.93	3.80			376.07	7.90		
	male	29	179.21	11.84	2.316	.134	125.86	8.80	1.041	.312	71.31	4.21	2.316	.134	376.38	8.08	.021	.885
	total	57	177.39	9.30			126.74	6.59			72.11	4.06			376.23	7.92		
9	female	27	196.07	18.85			91.98	13.59			45.76	9.83			162.86	28.68		
	male	19	195.21	5.63	.037	.848	92.95	12.56	.498	.481	41.13	10.46	2.328	.134	164.18	31.70	.022	.883
	total	46	195.72	14.77			92.53	13.00			43.29	10.32			163.56	30.01		

과학 전체 성취도 총점과 평가 유형별 성취도 사이의 상관 관계를 성별로 파악한 결과, 성별, 학년별 상관없이 모두 객관식 문항 성취도(.936<r(.986), 주관식 문항 성취도(.898<r(.974), 수행 평가 성취도(.642<r(.714) 등의 순서로 전체 총점과의 상관 관계가 가장 높은 것으로 나타났다(Table 7). 다시 말해, 학생들의 전체 과학 성취도 총점과 가장 밀접한 관계가 있는 평가 유형은 객관식이고, 상대적으로 가장 관계가 적은 평가 유형은 수행 평가인

것이 드러났다. 이는 지필 평가와는 달리 수행 평가의 경우, 기본 점수가 주어졌던 것에서 일부 원인을 찾을 수 있다.

평가 유형 사이의 상관 관계를 보면, 객관식 문항과 주관식 문항 성취도 사이의 상관 관계(.705<r(.942)는 남녀 학생 모두 매우 높은 반면, 주관식 문항과 수행 평가 성취도 사이의 상관 관계(.486<r(.811) 및 객관식 문항과 수행 평가 성취도 사이의 상관 관계(.455<r(.656)는 남녀 학생

Table 7. Correlation between assessment types

	Grade	Total score		Close-ended		Open-ended		Performance	
		Female	Male	Female	Male	Female	Male	Female	Male
Total score	7 (N = 198)	1.000	1.000	.975	.978	.907	.943	.642	.771
	8 (N = 272)	1.000	1.000	.986	.983	.974	.963	.709	.724
	9 (N = 206)	1.000	1.000	.936	.968	.898	.923	.714	.795
Close-ended	7 (N = 198)			1.000	1.000	.826	.898	.529	.656
	8 (N = 272)			1.000	1.000	.942	.918	.615	.638
	9 (N = 206)			1.000	1.000	.705	.822	.455	.656
Open-ended	7 (N = 198)					1.000	1.000	.486	.642
	8 (N = 272)					1.000	1.000	.644	.644
	9 (N = 206)					1.000	1.000	.811	.763
Performance	7 (N = 198)							1.000	1.000
	8 (N = 272)							1.000	1.000
	9 (N = 206)							1.000	1.000

모두 상대적으로 낮은 것으로 나타났다. 이는 과학과 수행 평가와 주관식 문항 성취도 사이의 상관 관계가 매우 낮게($r=.34$) 나타난 Supovitz(1998)의 연구 결과와 유사하다. 결국 수행 평가가 객관식 문항이나 주관식 문항이 측정하는 것과는 다소 다른 기능을 측정하고 있음이 확인되었다.

한편, 과학 전체 총점과 유형별 상관 관계, 특히 수행 평가와 상관 관계 경향은 성별로 다소 차이를 보였는데, 남학생이 여학생보다 수행 평가와 과학 총점과의 상관 관계(남: $.724 < r < .795$, 여: $.642 < r < .714$) 및 수행 평가와 객관식 문항과의 상관 관계(남: $.638 < r < .656$, 여: $.455 < r < .615$)가 다소 높은 것으로 나타났다. 다시 말해, 남학생들의 수행 평가 성취도가 여학생들의 수행 평가 성취도보다 과학 성취도 총점과 더 밀접한 관계가 있는 셈이다.

IV. 결론 및 제언

일반적으로 여학생들은 남학생들에 비해 과학에 대한 태도(Weinburgh, 1995), 흥미도(Schibeci & Riley, 1986; Simpson & Oliver, 1985), 인식(Handley & Morse, 1984; Kelly, 1987) 등이 낮게 나타나는데, 이렇듯 과학 교육의 정의적 측면에서 여학생의 소극성은 인지적 측면에서 여학생의 낮은 성취도와 연결된다. 지난 수 십 년 동안, 대규모 과학과 표준화 검사에서 남학생들이 꾸준히 우세를 지켜온 결과, 여학생들은 실제로 교육 과정에 근

거한 학교 과학 성취도에서 남학생에 비해 결코 뒤지지 않음(이재천과 김범기, 1996; Kozlow, 1986)에도 불구하고 과학을 배우는 자신들의 능력에 대한 확신이 부족했다(Khale & Rennie, 1993; Licht *et al.*, 1989). 학년이 올라 갈수록 여학생들의 이러한 자신감 상실은 과학 성취도나 과학 관련 직업 선택에도 부정적인 영향을 주었다(Jones *et al.*, 1992; Simpson & Oliver, 1985; 1990).

이 연구 결과가 보여주는 가장 주목할 만한 두 가지는 바로 수행 평가에서의 여학생 우세와 국제 비교 연구 결과와는 달리 학교 과학 성취도에서 성 차이가 나타나지 않은 점이다. 이 밖에도, 수행 평가는 지필 평가에서 측정할 수 없는 다른 능력을 측정함도 밝혀졌는데, 이 능력에 대한 깊이 있는 연구가 필요하다. 수행 평가가 객관식 위주의 전통적 지필 평가의 한계를 넘어서는 대안적 평가로 많은 교육학자들에 의해 지지 받고 있는 현실을 감안할 때, 이 연구에서 나타난 과학과 수행 평가에서 여학생들의 우세는 매우 의미 있는 결과다. 만약, 이 연구에서 분석의 대상이 된 수행 평가 시범 학교에서 수행 평가의 비율이 30%보다 높았고, 수행 평가에서 기본 점수가 주어지지 않았다면, 매우 흥미 있는 성별 성취도 결과가 나타났을 것으로 예상된다. 또한 거의 모든 평가 유형별 학교 과학과 성취도에 있어 성 차이가 나타나지 않은 점도 여학생들의 과학 능력에 대한 인식을 새롭게 하는 데 기여할 수 있을 것이다.

이제 대규모 객관식 문항 위주의 성취도 검사에서 나타

난 남학생들의 우세가 남학생들의 과학적 능력이 우월함으로 직접 연결되어 온 지금까지의 인식에 전환이 필요할 시기다. 객관식 문항을 통해 측정되는 능력이 과학을 하는데 필요한 능력의 전부가 아니라는 것에는 의문의 여지가 없다. 과학적 능력은 매우 다양하며, 이렇듯 다양한 능력은 다양한 평가 방법을 활용함으로써 타당하게 측정될 수 있을 것이다. 또한 다양한 평가 방법의 활용을 통해 과학적 능력에 관한 고정적인 성(性) 인식도 제거될 수 있을 것이다.

다만, 이 연구에서 사용된 수행 평가를 구성하는 내용의 타당성이 검증되지 않은 상태에서 수행 평가와 성별 성취도 문제를 객관화시키는 데 물론 한계가 있을 수 있다. 그러나, 적어도 이 연구에서 활용된 수행 평가가 전통적인 지필 평가와는 전혀 다른 방법으로 이루어진 점에 근거해, 평가 방법의 변화가 지금까지 밝혀진 성별 성취 경향에 큰 영향을 준다는 사실은 우리 나라 여학생들의 과학에 대한 자신감 회복과 이에 따라 여성 인력의 적극적인 과학 활동 참여로 진전할 수 있는 의미 있는 결과다. 이제 이 연구에서 분석한 중학생들의 과학 성취도 결과뿐만 아니라, 초등 학생이나 고등 학생들을 대상으로 연구를 확장해 평가 유형이 성별 과학 성취도에 어떠한 영향을 주는지에 대한 체계적인 연구가 계속 진행되어야 할 것이다.

국문 요약

이 연구에서는 평가 방법에 따라 여학생과 남학생의 과학 성취도가 각각 어떤 차이를 보이는지 파악함으로써 여학생들과 남학생들의 과학 성취 능력에 대해 올바르게 인식하는 데 도움을 주고자 한다. 이를 위해 서울 소재 중학교 전학년 학생 1,487명의 2000년 과학 성적을 분석하여 평가 방법에 따른 성 차이 경향, 성취 수준별 평가 방법에 따른 성 차이 경향, 성별 평가 방법 사이의 상관 관계를 알아보았다. 이 학교의 과학 성적에는 지필 검사 위주의 1, 2학기 중간과 기말 고사 및 수행 평가가 포함되었다. 분석 결과는 다음과 같다. 첫째, 전체 과학 성취도에 있어 7학년은 여학생이 남학생보다 유의미하게 높은 결과를 나타냈으나, 8, 9학년은 의미 있는 성 차이를 보이지 않았다. 둘째, 객관식 문항의 경우 7, 8, 9학년 모두 남녀 학생 간 과학 성취도에 있어 의미 있는 차이를 보이지 않았다. 주관식 문항에서는 7학년의 경우 여학생의 과학 성

취도가 남학생보다 유의미하게 높았으나, 8, 9학년에서는 의미 있는 성 차이가 나타나지 않았다. 반면, 수행 평가의 경우 3개 학년 모두 여학생이 의미 있는 차이를 보이며 남학생보다 높은 성취도를 나타냈다. 셋째, 상, 중, 하 성취 집단 모두 객관식 문항에서는 남학생이 유의미하게 높은 성취도를 보였고, 주관식 문항에서는 유의미한 성 차이가 나타나지 않은 반면, 수행 평가에서는 세 집단 모두 여학생이 유의미하게 높은 성취도를 보였다. 넷째, 상위 10% 이내의 학생 집단의 경우 전체적인 과학 성취도 결과와 그 경향이 전혀 달라, 이들 최상위권 학생들의 경우 모든 학년에서 객관식과 주관식 문항은 물론이고 수행 평가에서도 성별 성취도 차이가 나타나지 않았다. 다섯째, 전체 과학 성취도와 평가 방법들과의 상관 관계를 살펴본 결과, 남녀 학생 모두 모든 학년에 걸쳐 전체 과학 성취도와 가장 큰 상관 관계가 있는 평가 유형은 객관식 문항인 것으로 나타난 반면, 전체 과학 성취도와 가장 작은 상관 관계가 있는 평가 유형은 수행 평가인 것으로 나타났다. 한편, 평가 방법과 전체 과학 성취도 결과와의 상관 관계는 성별로 다소 다른 경향을 보였는데, 여학생보다는 남학생의 경우 수행 평가와 전체 과학 성취도와의 상관 관계가 높은 것으로 나타났다. 수행 평가가 학생들의 능력을 더욱 타당하게 평가할 수 있다는 기존의 연구 결과에 비추어 볼 때, 이 연구에서 여학생들이 남학생들에 비해 수행 평가에서 높은 성취도를 보인 점은 매우 주목할 만하다.

참고 문헌

- 백순근 편 (1998). 수행 평가의 이론과 실제. 서울: 원미사.
- 신동희, 박 정 (2002). 국제 비교 연구에 나타난 우리나라 학생들의 지구 과학 성취도: 성 차이를 중심으로. 한국지구과학회지, 23(3), 207-220.
- 신동희, 박 정, 노국향 (2002). OECD 주관 학생 성취도 국제 비교 연구(PISA 2000) 지구 환경 과학 영역 성취도에서의 성 차이. 한국과학교육학회지, 22(1), 40-53.
- 이재천, 김범기 (1996). 고등 학생들의 과학에 대한 정의적 인식과 과학 탐구 능력 및 과학 학습 성취도의 구조 분석. 한국과학교육학회지, 16(3).
- Bailey, T. (1989). *Changes in the nature and structure*

- of work: Implications for skill requirements and skill formation (Technical paper no. 9). New York, NY: National Center on Education and Employment, Teachers College, Columbia University.
- Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzales, E.J., Smith, T. A., & Kelly, D. L.(1996). *Science achievement in the middle school years: IEA's Third International Mathematics and Science Study*. Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Becker, J. R.(1989). Gender and science achievement: A reanalysis of studies from two meta-analysis. *Journal of Research in Science teaching*, 26, 141-169.
- Bell, J. F.(2001). Investigating gender differences in the science performance of 16-year old pupils in the UK. *Science Education*, 23(5), 469-487.
- Beller, M., & Gafni, N.(1996). The 1991 International Assessment of Educational Progress in mathematics and science: The gender differences perspective. *Journal of Educational Psychology*, 88, 365-377.
- Bolger, N. & Kellaghan, T.(1990). Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement*, 27(2), 165-174.
- Burkam, D. T., Lee, V.E., & Smerdon, B. A.(1997). Gender and science learning early in high school: Subject matter and laboratory experiences. *Educational Research Journal*, 34(2), 297-331.
- Carnevale, A. P., Gainer, L. J., & Meltzer, A. S. (1989). *Workplace basics: The skills employers want*. Alexandria, VA: American Society for Training and Development.
- Council for Educational Development and Research (1993). *What we know about science teaching and learning*. Washington, DC: EDTALK.
- Darling-Hammond, L.(1991). The implications of testing policy for quality and equality. *Phi Delta Kappan*, 11, 220-225.
- DeMars, C. E.(1998). Gender Differences in Mathematics and Science on a High School Proficiency Exam: The Role of Response Format. *Applied Measurement in Education*, 11, 3 279-299.
- Elwood, J.(1994). *Equity issues in performance assessment*. Paper presented at the annual meeting of the American Educational Research Association(New Orleans, LA April 4-8. 1994). 12pp.
- Frederiksen, N.(1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39, 193-202.
- Goldstein, H.(1993). Assessing group differences. *Oxford Review of Education*, 19(2), 141-150.
- Handley, H. M. & Morse, L. W.(1984). Two-year study relating adolescents' self-concept and gender role perceptions to achievement and attitudes toward science. *Journal of Research in Science Teaching*, 21, 599-607.
- Hedges, L. V., & Nowell, A.(1995). Sex differences in mental test scores, variability, and numbers of high scoring individuals. *Science*, 269, 41-45.
- Johnson, S. & Bell, J. F.(1987). Gender differences in science: option choices. *School Science Review*, 69(247), 268-276.
- Jones, L. R., Mullis, I. V., Raizen, S. A., Weiss, I. R. & Weston, E. A.(1992). *The 1990 science report card: NAEP's assessment of fourth, eighth, and twelfth graders*. Princeton, NJ: Educational Testing Services.
- Jovanovic, J. & Dreves, C.(1998). Students' science attitudes in the performance-based classroom: Did we close the gender gap? *Journal of Women and Minorities in Science and Engineering*, 4, 235-248.
- Jovanovic, J., Solano-Flores, G. & Shavelson, R. (1994). Performance-based assessment: Will gender differences in science achievement be eliminated? *Education and Urban Society*, 26(4),

- 352-366.
- Kelly, A.(1987). *Science for girls?* Milton Keynes, England: Open University Press.
- Khale, J. B. & Rennie, L. J.(1993). Ameliorating gender differences in attitudes about science: A cross-national study. *Journal of Research in Science Teaching*, 2, 321-334.
- Klein, S. P., Jovanovic, J., Strecher, B. M., McCaffrey, D., Shavelson, R. J., Haertel, E., Solano-Flores, G. & Comfort, K.(1997). Gender and racial/ethnic differences on performance assessment in science. *Educational Evaluation and Policy Analysis*, 19(2), 83-97.
- Kozlow, M.(1986). Personal correspondence. December 23. In S. M. Haggerty, *et al.*, 1987, Gender and science achievement: A case study, *International Journal of Science Education*, 9(3), 272.
- Lee, V. E. & Burkam, D. T.(1996). Gender differences in middle grade science achievement: subject domain, ability level, and course emphasis. *Science Education*, 80(6), 613-650.
- Licht, B. G., Stader, S. R. & Swenson, C. C.(1989). Children's achievement-related beliefs: Effects of academic area, sex, and achievement level. *Journal of Educational Research*, 82, 253-260.
- 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.
- Oakes, J.(1985). *Keeping track: How schools structure inequality*. New Haven: Yale University Press.
- OECD(2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: OECD Publications.
- O'Neil, H. F. & Brwon, 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. & Roth, J. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher*, 20(5), 12-20.
- Preece, P. F. W., Skinner, N. C., & Riall, R. A. H. (1999). The gender gap and discriminating power in the National Curriculum Key Stage three science assessments in England and Wales. *International Journal of Science Education*, 21(9), 979-987.
- Ramsbottom-Lucier, M.(1995). Age and gender differences in students' preadmission qualifications and medical school performances. *Academic Medicine*, 70(3), 236-239.
- Resnick, L. B. & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. Gifford & M.C. O'Connor (Eds), *Changing assessment: Alternative views of aptitude, achievement and instruction* (pp. 37-75). Boston, MA: Kluwer Academic Publishers.
- Ryan, K. E. & Fan, M.(1996). Examining gender DIF on a multiple choice test of mathematics: A confirmatory approach. *Educational Measurement: Issues and Practice*, 15(4), 15-20, 38.
- Schibeci, R. A. & Riley, J. P.(1986). Influence of students' background and perceptions on science attitudes and achievement. *Journal of Research in Science Teaching*, 23, 177-187.
- Schmitt, A. P. & Dorans, N. J.(1990). Differential item functioning for minority examinees on the SAT. *Journal of Educational Measurement*, 27(1), 67-81.
- Simpson, R. D. & Oliver, J. S.(1985). Attitude toward science and achievement motivation profiles of male and female science students in grades six through ten. *Science Education*, 60, 511-526.
- Simpson, R. D. & Oliver, J. S.(1990). A summary of major influences on attitudes toward and achievement in science among adolescent students. *Science Education*, 74, 1-18.
- Steinkamp, M. W. & Maehr, M. L.(1983). Affect, ability, and science achievement: A quantitative

- synthesis of correlational research. *Review of Educational Research*, 53, 369-396.
- Supovitz, J. A.(1998). Gender and racial/ethnic differences on alternative science assessment. *Journal of Women and Minorities in Science and Engineering*, 4, 129-140.
- Valencia, R.(1995). K-ABC content bias: Comparisons between Mexican-American and White children. *Psychology in the Schools*, 32(3), 153-169.
- Weinburgh, M. H.(1995). Gender differences in student attitudes toward science: A meta analysis of the literature from 1970 to 1991. *Journal of Research in Science Teaching*, 32, 387-398.
- Wolf, D. P.(1993). Assessment as an episode of learning. In R. Bennet & W. Ward (Eds), *Construction Versus Choice in Cognitive Measurement*. NJ: Lawrence Erlbaum Associates.