

科學 探究能力 評價 問項 類型 變化에 關한 縱斷的 研究

우종옥 · 이항로 · 구창현*
(한국교원대학교) (국립평가원)*

(1996년 6월 20일 받음)

I. 서론

1950년대에 들어서면서 기하 급수적으로 팽창하는 과학 지식을 과학 교육 현장에서 모두 수용하고 교수·학습 과정을 통해 학생들에게 모두 전달하기란 불가능하게 되었다. 이는 곧바로 과학 교육 방법에 대한 현실적인 개혁을 요구하게 되어, 1950년대까지 미국의 교육 철학을 주도해 오던 존 듀이의 생활 중심 교육 사상을 바탕으로 하던 교육 방법을 개선하기 위하여 브루너는 “교육의 과정(the process of education)”이라는 보고서에서 “주입식 교육, 형식 교육과는 달리 과학의 학문적 본성과 본질을 교육해야 한다”는 학문 중심(discipline centered) 교육 사상을 바탕으로 한 교육을 주장하게 되었다.

그는 구체적인 교육 방법으로 지식을 구조화하고 탐구를 증시한 발견 학습을 주장하였다(Bruner, 1960). 그리고 과학 분야의 탐구론과 피아제의 인지 발달론을 인용하여 나선형 교육과정을 내세웠다(한중하, 1984). 이와 같은 이론을 바탕으로 하여 미주쪽에서는 대학 교수들에 의해, 서유럽 국가들에서는 현장 교사들에 의해 과학 교육 개혁 운동이 주도되었다(이화국, 1989).

이러한 과학 교육 개혁 운동은 우리나라의 과학교육계에도 영향을 미쳤다. 그 결과 제 3차 교육과정부터 도입된 이후 4차 교육과정(1982), 5차 교육과정(1987)에서도 과학 탐구능력의 신장을 계속 강조해 왔고 6차 교육과정(1992)에서는 지식 체계에서 탐구 영역을 분리 독립시켜 이를 더욱 강조하고 있어, 과학 교육에서의 “과학 탐구능력 신장”이라는 목표가 차지하는 중요성을 알 수 있다.

이와 같은 강조에도 불구하고 현장에서 탐구적인 과학교육의 실천과 평가가 거의 이루어지지 않고 있는 근본적인

이유 중 대학입학시험에서 과학 탐구능력이 평가되지 않기 때문이라고 한다면 1994학년도부터 시행된 대학수학능력시험 제도에서의 과학 탐구능력 평가는 현장 과학 교육을 탐구 중심으로 획기적이고 효율적으로 전환시킬 수 있는 절호의 기회라고 판단된다.

그러나 대학수학능력시험이 실시된지 최근 3년이 지난 현시점에서, 과학 탐구능력 측정을 위한 행동 요소의 추출과 평가 목표의 상세화 연구 I, II(우종옥 등, 1991, 1992), 대학수학능력시험 실험 평가 문항 분석 및 교사들의 새입시 제도에 대한 인식도에 관한 연구(김은진 등, 1992; 김상철 등, 1994), 탐구 사고력 문제 해결 과정에 나타난 유형과 특징(김익균 등, 1993), 탐구 문제 유형에 관한 연구(김찬중, 1994)를 제외하고는 보다 효율적인 탐구능력 평가에 대한 연구는 미흡한 실정이다. 이런 대학수학능력시험 제도의 개발도 중요하지만 현장 과학교과 평가 방법의 변화를 수반한다고 보장할 수는 없다(병전옥 등, 1995). 과거 혁신 실패의 이유를 여러가지 들 수 있겠으나 혁신적 제도나 정책만 마련했을 뿐 그 이후의 과정에 관심과 지원을 게을리 했기 때문이라고 볼 수 있다(Hall, 1992).

현장 교육의 개선을 위해서는 과학 교사들의 열람이 가장 중요한 변수이다. 교사 개개인의 변화 없이는 교육적 변화를 기대할 수 없기 때문이다. 특히 교사의 역할이나 행동의 변화가 요구되는 교육 혁신의 경우는 변화에 상당한 시간과 노력이 요구된다(Brown & McIntyre, 1986).

따라서 본 연구는 대학수학능력시험의 실시 전·후의 시간 경과에 따른 현장 과학교사들의 과학 성취도 평가 문항 유형 변화에 대한 종단적 분석을 통해, 과학 교육이 강조하고 있는 탐구 학습 평가에 대한 이론적 연구가 현장에서 얼마나 효율적으로 침투되고 있는가를 밝힘으로써 앞으로 보다

효율적이며 타당도 높은 과학 탐구능력 평가 방법에 관한 단서를 찾고자하였다.

본 연구는 국립교육평가원에서 그 동안 시험 모형을 정립하기 위하여 7차례(1990.12.19~1992.11.10)에 걸쳐 실시했던 실험 평가 문항, '94학년도 1·2차, '95학년도 3 차, '96학년도 4차 대학수학능력 시험 문제 중 과학탐구 영역 문제, 대학수학능력시험 실시 전·후 최근 4년간에 걸쳐 10개 학교에서 표집한 현장 과학학습 성취도 평가 문항을 3차원 과학학습 평가들의 각 축, 평가 문항 유형, 문항 표현 양식을 중심으로 시간의 경과에 따른 종단적 비교를 위하여 다음과 같은 연구 문제를 설정하였다.

- 1) 수학능력시험 실험평가 문항 중 3 차원 평가들에 의한 과학탐구 문항의 평가목표, 문항 유형 및 문항 표현 양식은 어떠한가 ?
- 2) 1, 2, 3, 4차 대학수학능력시험 중 3 차원 평가들에 의한 과학탐구 문항의 평가목표, 문항 유형 및 문항 표현 양식은 어떠한가 ?
- 3) 대학수학능력시험 실시 이전 현장 과학 성취도 문항의 평가 목표 분포, 문항 유형 및 문항 표현 양식 분포는 어떠한가 ?
- 4) 대학수학능력시험 실시 이후 3 년간 현장 과학 성취도 문항의 평가 목표, 문항 유형 및 문항 표현 양식 분포는 어떠한가 ?
- 5) 대학수학능력시험 실시전·후에 평가 목표, 문항 유형, 문항 표현 양식 분포 변화는 어떠한가 ?

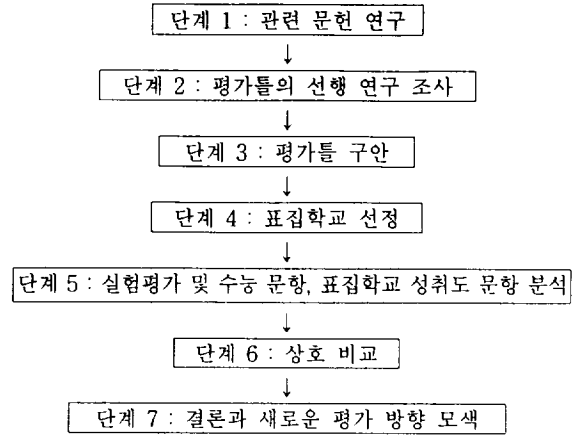
II. 연구 절차

본 연구는 7차례의 실험 평가를 거쳐 이미 4 차례의 수학능력시험이 실시된 현 시점에서, 교육의 현장인 학교에서의 과학 학습 성취도 문항의 평가 목표, 문항 유형과 문항 표현 양식의 분포가 어떻게 변화하고 있는지를 알아보고, 현장 과학 학습 성취도 문항이 추구하고 있는 현상태와 수학능력시험의 과학탐구 영역이 추구하고 있는 상태를 상호 비교·분석하여 보다 바람직한 과학 학습 성취도 평가 방향에 대한 단서를 제시하고자 하는 시도를 하였다. 이를 위하여 수행된 연구 절차를 제시하면 다음 <그림 1>과 같다.

1. 관련 문헌 연구

과학 탐구능력 평가틀, 과학 탐구능력 평가에 관련된 국내외 선행 연구들을 중심으로 문헌을 수집한 후 과학 탐구능력 평가틀, 평가 도구의 개발과 문항 유형으로 세분하여

분석을 하였다.



<그림 1> 본 연구의 흐름도

2. 과학 학습 평가틀 고찰

현대의 과학 교육은 과학자들이 이루어 놓은 지식의 습득에 못지 않게 이런 지식을 창출하는 방법 즉, 과학 탐구능력의 습득도 강조하고 있다(Lawson, 1982). 과학교육에서 이양자를 모두 강조한다면, 탐구학습을 구현하기 위한 교수·학습 및 평가 목표의 설정과 상세화가 우선적으로 필요하다. 과학 탐구능력을 평가하기 위한 평가목표는 과학교육에 관련된 변인들(교사, 학생, 교과, 인지 수준 등)에 따라 너무나 다양하기 때문에 과학 탐구능력 평가목표를 설정하기 이전에 이들의 교수·학습 및 평가목표들을 범주화 해 줄 수 있는 목표 분류 체계가 필요하다. 이 목표분류틀이 평가 목표의 작성과 선정에 이용될 때, 이를 과학 평가 목표 분류틀 또는 탐구능력 평가틀이라고 부른다(김창식 등, 1992). 과학 학습의 결과에 대한 평가는 학교에서 배운 과학 뿐만 아니라 학교밖의 다양한 상황에서 일어나는 실제적인 평가가 되기 위해서는 '내용'과 '행동'의 2차원적인 분류 체계에 '상황' 차원이 포함된 3차원 분류 체계가 바람직하다. 이것은 과학 평가가 수업 시간에 배우는 과학 뿐만 아니라 학교 내외의 일상적인 상황에 걸쳐서 이루어져야 한다는 것을 의미하며, 현실의 과학이 '내용', '행동', '상황'의 3차원에서 밀접하게 상호 작용하고 있다(Kirkham, 1989)는 주장과 맥을 같이한다고 볼 수 있다. 과학과 교수·학습 및 평가 목표들의 범주화를 시도한 연구는 매우 다양하지만 국내외에서 주로 원용되거나 변안되어 사용되는 것을 제시하면 다음<표 1>과 같다.

<표 1> 선행 연구 보고된 과학 학습 평가들

평가들	구분	년도	차원 수	차원명	비고
Bloom의 목표 분류학		1956	2	행동, 내용	일반적 교육 목표 분류들
Klofer의 목표 분류 체계		1973	2	행동, 내용	과학 교육을 위한 목표 분류들
NAEP의 과학 학습 평가들					
제 1 차		1970	1	과학적 소양	국가 수준의 미국의 과학 학습 성취도 평가
제 2 차		1973	2	행동, 과학의 기본적 양상	
제 3 차		1977	2	행동, 인지 영역	
제 4 차		1982	3	내용, 인식, 상황	
제 5 차		1986	3	내용, 인식, 상황	
제 6 차		1990	2	내용, 사고 능력	
APU의 과학 학습 평가들		1984	3	개념, 과정, 상황	영국의 과학 학습 평가 목표 분류들 * 실험 중요시
Swain의 과학 학습 평가들		1989	3	과학적 지식이나 내용, 과학 탐구 기능, 창의적이고 열린 문제 해결 기능	
우종욱 등의 과학 탐구능력 평가들		1993	3	내용, 탐구능력, 탐구 상황	대학수학능력시험의 과학 탐구 영역 평가들

시대 변화에 따른 과학의 응용은 우리들이 살아가고 있는 실생활과 밀접하게 관련되어 있음에도 불구하고 학교에서 가르쳐 지는 대부분의 과학은 우리의 현실과는 거리가 먼 순수과학이다. 그러나 과학교육은 실험실과 교실뿐만이 아니라 가정, 지역사회를 포함한 공동체 안에서 조화롭게 이루어져야 한다(Kirkham, 1989). 이와 같은 과학교육 사조에 부응하여 NAEP와 APU 과학 학습 평가들에서는 '상황' 차원을 첨가한 3차원 평가들을 개발해 국가 수준에서 학생들의 과학 성취도를 평가하는 데 활용하였으며, 우리나라에서도 대학수학능력시험의 과학 탐구 영역에서 과학 탐구능력 평가를 위한 3차원 평가들을 개발해 제시한 바 있다(우종욱 등, 1991, 1992).

3. 문항 분석을 위한 과학 학습 평가들 구안

위에서 살펴본 바와 같이 과학 교육에서 추구하고자 하는 목표는 과학적 문제가 발생되는 상황 속에서의 과학 교과에 대한 기본적인 개념과 이를 창출하는 과정인 과학 탐구능력이 함께 작용하는 종합적인 능력의 습득에 대한 강조가 과학 교육에 대한 현대 인식론자들의 견해라고 볼 수 있다.

대학수학능력시험의 과학 탐구 영역 평가를 위해 변안 제시된 과학 학습 평가들로는 우종욱등(1991, 1992)과 박승재(1991)의 연구가 있다. 박승재(1991)는 상황(context) 차원을 상세화 한 반면, 우종욱 등(1991, 1992)은 과학 탐구능력

차원을 상세화 한 점이 다르고 내용 차원(content)은 대동소이하다. 이 둘 모두는 본질적으로 과학교육에 대한 현대 인식론자들이 공유하고 있는 입장을 충분히 고려한 3차원 평가들을 제시하고 있다.

Cronbach(1970)는 Lindquist의 "교육측정"과 Bloom의 "목표분류학"까지도 학문적 분석이라기 보다는 경험의 추출물이라는 견해를 밝히고 있고 이종기(1988)는 탐구 과정 요소의 지나친 세분화는 평가 문항 타당도의 결여를 야기시킨다는 견해를 밝히고 있다.

이를 근거로 볼 때 보다 종합적인 분석들을 마련하기 위해 박승재(1991)가 제시한 과학 학습 3차원 평가들 중 과학 탐구능력 차원보다는 우종욱등(1991, 1992)의 연구에서 제시한 과학 탐구능력 차원이 본 연구의 분석에 타당하다고 판단되어 상호 교체하고 과학 개념 측과 탐구 상황측은 그대로 사용하여 <표 2>와 같은 평가들을 구안하였다.

4. 연구 대상 및 분석 대상 문항 표집

문항 분석을 위하여 7 차례의 대학수학능력 시험 평가 문항과, 4 차례의 대학수학능력시험 중 과학탐구 영역 평가 문항을 선정하고, 현장 표집 학교는 최근 4년간의 과학 성취도 문항을 무작위(random)로 20 개 학교에서 표집하였다. 현장 과학 성취도 평가 문항은 매 년도 학기당 중간 고사, 기말 고사 문항을 4년간 20 개 학교에서 표집한 후 이 중에

<표 2> 본 연구의 문항 분석을 위한 과학 학습 평가틀

축	구분	대 범 주	소 범 주
X	내용 (과학기본개념)	A. 생 물 B. 지구과학 C. 물 리 D. 화 학 E. 통합과학	A1. 생명의 특성 A2. 생물의 영양 A3. 생물의 항상성 A4. 생명의 연속성 A5. 생물과 환경 B1. 우리의 지구 B2. 지각의 물질 B3. 대기와 해양 B4. 지구의 역사 B5. 지구밖의 환경 C1. 운동과 에너지 C2. 전자기 C3. 빛과 파동 D1. 물질 세계의 규칙성 D2. 화학 반응 D3. 화합물
Y	탐구능력 (탐구사고력)	I. 문제 인식 및 가설 설정 II. 탐구의 설계 III. 탐구의 수행 IV. 자료의 해석 V. 결론 도출 및 평가	I 1. 문제 인식 I 2. 가설 설정 II 1. 실험, 조사, 연구의 계획 II 2. 변인 통제 II 3. 실험 장치의 고안 및 배치 III 1. 관찰, 측정, 분류, 실험 III 2. 실험 절차 III 3. 실험 안전 III 4. 자료의 수집, 전개 IV 1. 정량적 분석 IV 2. 정성적 분석 IV 3. 자료의 변환 V 1. 일반화 V 2. 결론에 대한 비판 V 3. 결론의 의미 V 4. 결론이 미치는 사회적 영향
Z	탐구상황	a. 식식주 b. 일 상 c. 학 교 d. 지 역 e. 지 구	a1. 식품, 부엌 a2. 의류, 침구류 a3. 주택, 정원 a4. 인체, 건강 b1. 용품, 공구 b2. 교통, 여행 b3. 스포츠 b4. 음악, 미술 c1. 교실 c2. 실험실 c3. 특별실 c4. 교실밖 d1. 환경, 오염 d2. 농경, 목장 d3. 공장 d4. 공공기관 e1. 지질, 토양 e2. 바다, 호수 e3. 일기, 기후 e4. 천체

서 4년간 동일 교사가 개발하여 계열성을 유지한 10개교만을 연구 대상 표집으로 하였다. 연구 대상의 표집과 분석 대상 문항을 나타내면 다음 <표 3>, <표 4>와 같다.

<표 3> 연구 대상의 표집

구분	실험평가	수학능력시험	대상 학교
표집	7 차 실험 평가의 과학 탐구 문항	1, 2, 3, 4 차의 과학 탐구 문항	10 개 학교의 4년간 성취도 문항

<표 4> 분석 대상 문항

구분	실험평가	수학능력시험		표집학교	
		1, 2차	66	92년도	883
문항수	210	3 차	36	93년도	727
		4 차	36	94년도	619
				95년도	644
계	210	138	2873		

5. 실험평가 문항, 수학능력시험 중 과학 탐구 영역 평가 문항, 표집학교 성취도 문항 분석

본 연구에서 구안한 과학학습 평가틀을 준거로 실험평가 문항, 수학능력시험 중 과학 탐구 영역 평가 문항, 표집학교 성취도 문항을 3차원 평가틀의 각 차원, 문항 표현 양식, 문

항 유형에 따라 분석하였다.

문항 분석에는 물리, 화학, 생물, 지구과학 각 1인으로 구성된 분석팀에서 공동으로 작성한 체크리스트로 분석했으며, 모든 문항에서 2인 이상의 의견과 본 연구자의 의견이 합치를 보인 경우를 동일한 목표를 측정하는 것으로, 동일 문항 유형으로, 동일 문항 표현 양식으로 확정하는 방법을 사용하였다.

1) X축에 따른 문항 분석 결과 및 논의

대범주 5개의 분야 중 출제된 총 문항수에 대한 백분율을 살펴보면<표 5> 생물 교과 영역은 실험평가 22.9%, 4회의 수학능력 평가에서 20.3%, 지구과학 교과 영역은 실험평가의 B1 0%, B2 6.7%, B3 10.5%, B4 41.4%, B5 6.7%(평균 24.8%), 4회의 수학능력 평가의 B1 0.7%, B2 6.5%, B3 7.2%, B4 42.2%, B5 6.5%(평균 23.2%), 표집학교 성취도 평가의 B1 40%, B2 20.8%, B3 17.1%, B4 12.5%, B5 9.2%로, 물리 교과 영역은 실험평가 22.4%, 4회의 수학능력 평가 22.5%, 화학 교과 영역은 실험평가 21.4%, 4회의 수학능력 평가 23.2%로 비교적 고르게 출제되었으나 통합교과 영역의 출제는 실험평가 8.1%, 4회의 수학능력 평가 10.9%, 표집학교 성취도 평가 0.5%로 통합교과적 소재를 활용한 평가 문항 비율이 매우 낮아 통합 교과를 소재로 중핵으로 하는 평가 방법에 대한 연구가 필요함을 알 수 있다

2) Y 축에 따른 문항 분석 결과 및 논의

과학 탐구과정 5 단계 중 문제인식 및 가설설정 능력 평

<표 5> 3차원 과학학습 평가들의 X축에 따른 문항수 분포

()는 전체 문항에 대한 백분율(%)

대 범 주	A					B					C			D			E	계
	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	D1	D2	D3	E	
1-7 차 실험 평가	4 (1.9)	26 (12.5)		5 (2.4)	13 (6.3)		14 (6.7)	21 (10.5)	3 (1.4)	14 (6.7)	27 (12.0)	20 (9.6)		14 (6.7)	19 (9.1)	13 (6.3)	17 (8.2)	210 (100)
평 균	48 (22.9)					52 (24.8)					47 (22.4)			46 (21.4)			17 (8.1)	
93 1 차 수능 시험			1 (3.0)	2 (6.1)	2 (6.1)		2 (6.1)	2 (6.1)	1 (3.0)	2 (6.1)	5 (15.2)	3 (9.1)		7 (21.2)	1 (3.0)		5 (15.2)	33 (100)
93 2 차 수능 시험		3 (9.1)	1 (3.0)	2 (6.1)	1 (3.0)		2 (6.1)	3 (9.1)	1 (3.0)	3 (9.1)	3 (9.1)	2 (6.1)	2 (6.1)	3 (9.1)	4 (12.1)	1 (3.0)	2 (6.1)	33 (100)
94 3 차 수능 시험		2 (8.3)	3 (8.3)	2 (5.6)	2 (5.6)		2 (5.6)	2 (5.6)	1 (2.8)	3 (8.3)	4 (11.1)	3 (8.3)	2 (5.6)	2 (5.6)	4 (11.1)	2 (5.6)	3 (8.3)	36 (100)
95 4 차 수능 시험	1 (2.8)	1 (2.8)	3 (8.3)	1 (2.8)	2 (5.6)	1 (2.8)	3 (8.3)	3 (8.3)		1 (2.8)	4 (11.1)	2 (5.6)	1 (2.8)	2 (5.6)	5 (13.9)	1 (2.8)	5 (13.9)	36 (100)
평 균	1 (0.7)	5 (3.6)	8 (5.8)	7 (5.1)	7 (5.1)	1 (0.7)	9 (6.5)	10 (7.2)	3 (2.2)	9 (6.5)	16 (11.6)	10 (7.2)	5 (3.6)	14 (10.1)	14 (10.1)	4 (2.9)	15 (10.9)	138 (100)
영역별 평균	28 (20.3)					32 (23.2)					31 (22.5)			32 (23.2)			15 (10.9)	
표집 학교 92 년도						302 (34.2)	188 (21.3)	170 (19.3)	137 (15.5)	86 (9.7)								883 (100)
표집 학교 93 년도						298 (41.0)	174 (23.9)	125 (17.2)	70 (9.6)	52 (7.2)							8 (1.1)	727 (100)
표집 학교 94 년도						209 (33.8)	97 (15.7)	142 (22.9)	98 (15.8)	71 (11.5)							2 (0.3)	619 (100)
표집 학교 95 년도						339 (52.6)	138 (21.4)	55 (8.5)	55 (8.5)	54 (8.4)							3 (0.5)	644 (100)
평 균						1148 (40.0)	597 (20.8)	492 (17.1)	360 (12.5)	263 (9.2)							13 (0.5)	2873 (100)

가는 실험평가에서 10%, 4회의 수학능력 평가에서 13%, 표집학교 성취도 평가에서 0.6%, 탐구의 설계 능력 평가는 실험평가에서 10%, 4회의 수학능력 평가에서 8%, 표집학교 성취도 평가에서 1%, 탐구의 수행능력은 실험평가에서 6.7%, 4회의 수학능력 평가에서 21%, 표집학교 성취도 평가에서 1.7%, 자료의 해석 능력은 실험평가에서 50.5%, 4회의 수학능력 평가에서 40%, 표집학교 성취도 평가에서 38.8%, 결론 도출 및 평가능력은 실험평가에서 22.9%, 4회의 수학능력 평가에서 18.1%, 표집학교 성취도 평가에서 2.8%였다. 실험평가와 수학능력 평가의 모든 문항이 탐구 능력을 측정하는 것으로 나타났으나 표집학교의 경우 전체 문항의 44.9%로 국가 수준의 평가와 학교 수준의 평가에는 2배 이상의 차이가 있는 것으로 나타나 <표 6> 과학 탐구능력 평가 이론이 과학 교육 현장에 유효하게 전이되고 있지 않음을 알 수 있다.

따라서 과학 탐구능력 평가 이론의 습득과 평가 문항 개발에 관한 예비 교사 및 현직 교사에 대한 지속적인 교육이 필요할 것으로 판단된다

3) Z축에 따른 문항 분석 결과 및 논의

탐구 상황 5개 범주 중 의식주 범주는 실험평가에서 11.9%, 4회의 수학능력 평가에서 8.0%, 표집학교 성취도 평가에서 0.1%, 일상 범주는 실험평가에서 6.2%, 4회의 수학능력 평가에서 3.6%, 표집학교 성취도 평가에서 0.5%, 학교 범주는 실험평가에서 49.5%, 4회의 수학능력 평가에서 73.2%, 표집학교 성취도 평가에서 98.4%, 지역 범주는 실험평가에서 5.2%, 4회의 수학능력 평가에서 5.1%, 표집학교 성취도 평가에서 0.1%, 지구 범주는 실험평가에서 25.7%, 4회의 수학능력 평가에서 10.1%, 표집학교 성취도 평가에서 0.9%가 상황과 관련된 능력을 측정하는 것으로 나타났다.

전체적으로 학교, 지구, 의식주, 지역, 일상 상황 순이었으며 학교 상황이 각각 49.5%, 73.2%, 98.4%로 과학 교과에서 쉽게 접할 수 있는 문제 상황인 것으로 나타났다.

특히 실생활 속에서 과학적 탐구 문제를 해결할 수 있게 한다는 과학 교육 목표에 적합한 의식주, 일상 생활, 지역, 지구 상황 속에서의 평가 문항 비율은 실험평가 50.5%, 4회의 수학능력 평가 26.8%, 표집학교 성취도 평가 1.6%로 나타났다<표 7>.

학교 범주에 속하는 문항은 대부분 교과서 내의 상황이기

<표 6> 3차원 과학학습 평가들의 Y축에 따른 문항수 분포

()는 전체 문항에 대한 백분율(%)

대 범 주	I		II			III				IV			V				계
	I1	I2	II1	II2	II3	III1	III2	III3	III4	IV1	IV2	IV3	V1	V2	V3	V4	
1 차 실험 평가	3 (12.5)	3 (12.5)			2 (8.3)	2 (8.3)	1 (4.2)		1 (4.2)	5 (20.8)	3 (12.5)				4 (16.7)		24 (100)
2 차 실험 평가	2 (7.4)		2 (7.4)	1 (3.7)	2 (7.4)	1 (3.7)	2 (7.4)		1 (3.7)	2 (7.4)	5 (18.5)	2 (7.4)			7 (26.0)		27 (100)
3 차 실험 평가			3 (11.1)	3 (11.1)			2 (7.4)			5 (18.5)	1 (3.7)	3 (11.1)		3 (11.1)	7 (26.0)		27 (100)
4 차 실험 평가	3 (9.1)				4 (12.1)		1 (3.0)			10 (30.3)	9 (27.3)	1 (3.0)	1 (3.0)	1 (3.0)	3 (9.1)	1 (3.0)	33 (100)
5 차 실험 평가	2 (6.1)	2 (6.1)		1 (3.0)	1 (3.0)	1 (3.0)		1 (3.0)		9 (27.3)	6 (18.2)	4 (12.1)		1 (3.0)	5 (15.2)		33 (100)
6 차 실험 평가	2 (6.1)			1 (3.0)	1 (3.0)	1 (3.0)				5 (15.2)	7 (21.2)	2 (6.1)		3 (9.1)	6 (18.2)		33 (100)
7 차 실험 평가	3 (9.1)	1 (3.0)								5 (15.2)	14 (42.4)	3 (9.1)		7 (21.2)			33 (100)
전 체 평 균	15 (7.1)	6 (2.9)	5 (2.4)	6 (2.9)	10 (4.8)	5 (2.4)	6 (2.9)	1 (0.5)	2 (1.0)	41 (19.5)	45 (21.4)	20 (9.5)	1 (0.5)	15 (7.1)	32 (15.2)		210 (100)
전 체 단 계 별 평 균	21 (10)		21 (10)			14 (6.7)				106 (50.5)			48 (22.9)				210 (100)
93 년 1 차 수 능 시 험	2 (6.1)	1 (3.0)	1 (3.0)	2 (6.1)	1 (3.0)	3 (9.1)			2 (6.1)	7 (21.2)	9 (27.3)			2 (6.1)	3 (9.1)		33 (100)
93 년 2 차 수 능 시 험	3 (9.1)	3 (9.1)			2 (6.1)		6 (18.2)		3 (9.1)	4 (12.1)	8 (24.2)	2 (6.1)		2 (6.1)			33 (100)
93 년 단 계 별 평 균	9 (13.6)		6 (9.1)			14 (21.2)				30 (45.5)			7 (10.6)				66 (100)
94 년 3 차 수 능 시 험	1 (2.3)	1 (2.3)	1 (2.3)		1 (2.3)	4 (11.1)	3 (8.3)	1 (2.3)	1 (2.3)	5 (13.9)	9 (25)			4 (11.1)	5 (13.9)		36 (100)
95 년 4 차 수 능 시 험	6 (16.7)	1 (2.3)	2 (5.6)		1 (2.3)	1 (2.3)	1 (2.3)		4 (11.1)	2 (5.6)	6 (16.7)	3 (8.3)		1 (2.3)	8 (22.2)		36 (100)
전 체 평 균	12 (8.8)	6 (4.3)	4 (2.9)	2 (1.4)	5 (3.6)	8 (5.8)	10 (7.2)	1 (0.7)	10 (7.2)	18 (13.0)	32 (23.2)	5 (3.6)		9 (6.5)	16 (11.6)		138 (100)
전 체 단 계 별 평 균	18 (13.0)		11 (8.0)			29 (21.0)				55 (40.0)			25 (18.1)				138 (100)
표 집 학 교 92년도	1 (0.1)			1 (0.1)		5 (0.6)	1 (0.1)			14 (1.6)	92 (10.4)	156 (17.7)	5 (0.6)	5 (0.6)	1 (0.1)		281 (31.8)
단 계 별 평 균	1 (0.1)		1 (0.1)			6 (0.7)				262 (29.7)			11 (1.2)				
표 집 학 교 93년도	2 (0.3)	8 (1.1)		14 (1.9)	4 (0.6)	11 (1.5)	6 (0.8)		5 (0.7)	58 (8.0)	108 (14.9)	158 (21.7)	10 (1.4)		25 (3.4)		409 (56.3)
단 계 별 평 균	10 (1.4)		18 (2.5)			22 (3.0)				324 (44.6)			35 (4.8)				
표 집 학 교 94년도	1 (0.2)	2 (0.3)	1 (0.2)	3 (0.5)	2 (0.3)	6 (1.0)	5 (0.8)		2 (0.3)	43 (6.9)	88 (14.2)	168 (27.1)	2 (0.3)	1 (0.2)	11 (1.8)		335 (54.1)
단 계 별 평 균	3 (0.5)		6 (1.0)			13 (2.1)				299 (44.3)			14 (2.3)				
표 집 학 교 95년도	1 (0.2)	3 (0.5)		3 (0.5)	1 (0.2)	5 (0.8)	2 (0.3)		1 (0.2)	17 (2.6)	83 (12.9)	129 (20.0)	11 (1.7)	1 (0.2)	9 (1.4)		386 (41.3)
단 계 별 평 균	4 (0.6)		4 (0.6)			8 (1.2)				229 (35.6)			21 (3.3)				
평 균	18 (0.6)		29 (1.0)			49 (1.7)				1114 (38.8)			81 (2.8)				1291 (44.9)

<표 7> 3차원 과학학습 평가들의 Z축에 따른 문항수 분포

()는 전체 문항에 대한 백분율(%)

대 범 주	a				b				c				d				e				계
	a1	a2	a3	a4	b1	b2	b3	b4	c1	c2	c3	c4	d1	d2	d3	d4	e1	e2	e3	e4	
1-7 차 실험 평가	7 (3.3)	3 (1.4)	4 (1.9)	11 (5.2)		10 (4.8)	3 (1.4)		22 (10.5)	66 (31.4)		16 (7.6)		5 (2.4)	6 (2.9)		16 (7.6)	8 (3.8)	15 (7.1)	15 (7.6)	207 (98.6)
영역별 평균	25 (11.9)				13 (6.2)				104 (49.5)				11 (5.2)				54 (25.7)				
93 1 차 수능 시험				1 (3.0)			2 (6.1)		15 (45.1)	12 (36.1)			1 (3.0)				1 (3.0)		1 (3.0)		.33 (100)
93 2 차 수능 시험			1 (3.0)	2 (6.1)		1 (3.0)			13 (39.4)	12 (36.1)			1 (3.0)	1 (3.0)					1 (3.0)	1 (3.0)	.33 (100)
93 년 영역별 평균	4 (6.1)				3 (4.5)				52 (78.8)				3 (4.5)				4 (6.1)				.33 (100)
94 3 차 수능 시험		1 (2.8)	3 (8.3)			1 (2.8)			9 (25)	14 (38.9)	1 (2.8)	2 (5.6)	1 (2.8)	1 (2.8)					1 (2.8)	2 (5.6)	.36 (100)
94 년 영역별 평균	4 (11.1)				1 (2.8)				26 (72.2)				2 (5.6)				3 (8.3)				.36 (100)
95 4 차 수능 시험			3 (8.3)			1 (2.8)			16 (44.4)	5 (13.9)		2 (5.6)		1 (2.8)	1 (2.8)		1 (2.8)	2 (5.6)	1 (2.8)	3 (8.3)	.36 (100)
95 년 영역별 평균	3 (8.3)				1 (2.8)				23 (63.9)				2 (5.6)				7 (19.4)				.36 (100)
전체 영역별 평균	11 (8.0)				5 (3.6)				101 (73.2)				7 (5.1)				14 (10.1)				138 (100)
표집 학교 92 년도						1 (0.1)			879 (99.5)								1 (0.1)	1 (0.1)	1 (0.1)		883 (100)
영역별 평균					1 (0.1)				879 (99.5)								3 (0.4)				
표집 학교 93 년도					1 (0.1)	5 (0.7)	1 (0.1)		696 (95.7)	9 (1.2)	1 (0.1)	1 (0.1)						7 (1.0)	5 (0.7)	1 (0.1)	727 (100)
영역별 평균					7 (1.0)				707 (97.2)								13 (1.8)				
표집 학교 94 년도	1 (0.2)		1 (0.2)			5 (0.8)			596 (96.3)	6 (1.0)				1 (0.1)	1 (0.1)		3 (0.5)	4 (0.6)	1 (0.1)		619 (100)
영역별 평균	2 (0.2)				5 (0.8)				602 (97.3)				2 (0.3)				8 (1.3)				
표집 학교 95 년도	1 (0.2)		1 (0.2)			2 (0.3)			638 (99.1)	1 (0.2)										1 (0.2)	644 (100)
영역별 평균	2 (0.2)				2 (0.3)				639 (99.2)								1 (0.2)				
전체 영역별 평균	4 (0.1)				15 (0.5)				2827 (98.4)				2 (0.1)				25 (0.9)				2873 (100)

때문에 이를 제외한 나머지 문항비로 볼 때 국가 수준의 평가에서는 어느 정도 반영하고 있으나 표집학교 성취도 평가에서는 현대 인식론자들이 주장하는 실생활속에서 과학 문제를 해결하는 탐구능력의 신장이라는 과학 교육 목표 및 평가에 대한 견해와 많은 차이가 있음을 알 수 있다.

4) 문항 표현 양식 분석 결과 및 논의

문항과 관련된 그림, 그래프나 도표와 같은 다양한 문항 표현 양식을 중심으로 진술된 문항의 비율은 실험평가에서 93.8%, 4회의 수학능력 평가에서 62.3%, 표집학교 성취도 평가에서 35.7%로 나타났다. 이는 국가 수준 평가에서는 대부분 문항 표현 양식이 다양한 반면 표집학교 과학 성취도 평가 문항은 단순 서술식 문항으로 구성되어 있어 다양한

자료를 이용한 문항 개발에 소극적인 것으로 나타났다<표 8>. 평가도 과학 교육 활동의 연장선 상에 있는 하나의 과정으로 볼 수 있기 때문에 학습목표는 평가목표를 구성할 뿐 아니라, 역으로 평가목표가 곧 학습목표를 구성하기도 한다. 과학 개념의 표현 양식에 따라 학습 지속 효과(momentum effects)가 달라진다는 선행 연구(김준태, 1994)에 비추어 볼 때 평가 문항의 표현 양식에 따라 과학 성취도의 차이는 물론 평가의 지속 효과도 다를 것으로 판단된다. 따라서 다양한 문항 표현 양식에 따른 문항 개발에 관한 지속적인 연구가 필요한 것으로 판단된다.

5) 문항 유형 분석 결과 및 논의

본 연구를 위해 분석 자료를 통해 설정한 4가지 문항 유

<표 8> 문항 표현 양식 분포

구분 \ 자 료	그림	그래프	도표	보기	그림+그래프	그림+도표	그림+보기	그래프+도표	그래프+보기	도표+보기	계
1 차 실험 평가	3 (12.5)	4 (16.7)	7 (29.2)	10 (41.7)							24 (100)
2 차 실험 평가	1 (3.7)	6 (22.2)	11 (40.7)			5 (18.5)	1 (3.7)				24 (88.9)
3 차 실험 평가	2 (7.4)	10 (37.0)	4 (14.8)	2 (7.4)	5 (18.5)		3 (11.1)				26 (96.1)
4 차 실험 평가	7 (21.2)	4 (12.1)	13 (39.4)		2 (6.1)	3 (9.1)	1 (3.0)				30 (90.1)
5 차 실험 평가	5 (15.2)	8 (24.2)	5 (15.2)	1 (3.0)	3 (9.1)		5 (15.2)	5 (15.2)		1 (3.0)	33 (100)
6 차 실험 평가	7 (21.2)	5 (15.2)	5 (15.2)		4 (12.1)		3 (9.1)		2 (6.1)	2 (6.1)	28 (84.8)
7 차 실험 평가	4 (12.1)	6 (18.2)	6 (18.2)	4 (12.1)	5 (15.2)	2 (6.1)		1 (6.1)	1 (6.1)	2 (12.1)	31 (93.9)
평 균	29 (13.8)	44 (21.0)	51 (24.3)	17 (8.1)	19 (9.0)	10 (4.8)	13 (6.2)	6 (2.9)	3 (1.4)	5 (2.4)	197 (93.8)
93 1 차 수능 시험	2 (6.1)	1 (3.0)			2 (6.1)	1 (3.0)	1 (3.0)				7 (21.2)
93 2 차 수능 시험		1 (3.0)	1 (3.0)		1 (3.0)		2 (6.1)		2 (6.1)		7 (21.2)
93 년 평균	2 (3.0)	2 (3.0)	1 (1.5)		3 (4.5)	1 (1.5)	3 (4.5)		2 (3.0)		14 (21.2)
94 3 차 수능 시험	4 (11.1)	5 (13.9)	5 (13.9)	8 (22.2)	1 (2.8)		8 (22.2)		3 (8.3)	2 (5.6)	36 (100)
95 4 차 수능 시험	11 (30.6)	2 (5.6)	1 (2.8)	4 (11.1)	5 (13.9)	1 (2.8)	8 (22.2)		3 (8.3)	1 (2.8)	36 (100)
평 균	17 (12.3)	9 (6.5)	7 (5.1)	12 (8.7)	9 (6.5)	2 (1.4)	19 (13.8)		8 (5.8)	3 (2.2)	86 (62.3)
표집 학교 92 년도	116 (13.4)	56 (6.3)	15 (1.7)	31 (3.5)			1 (0.1)				219 (24.8)
표집 학교 93 년도	171 (23.5)	72 (9.9)	64 (8.8)	12 (1.7)	4 (0.6)	2 (0.3)				1 (0.1)	326 (44.8)
표집 학교 94 년도	157 (25.4)	75 (12.1)	35 (5.7)	9 (1.5)	4 (0.6)	1 (0.2)	3 (0.5)	1 (0.2)			285 (46)
표집 학교 95 년도	99 (15.4)	31 (3.1)	20 (3.1)	25 (3.9)	5 (0.8)	4 (0.6)	7 (1.1)		3 (0.5)	1 (0.2)	195 (30.3)
평 균	543 (18.9)	234 (8.1)	134 (4.7)	77 (2.7)	13 (0.5)	7 (0.2)	11 (0.4)	1 (0.03)	3 (0.1)	2 (0.1)	1025 (35.7)

형 중 5지 선다형은 실험평가에서 99.04%, 4회의 수학능력 평가에서 100%, 표집학교 성취도 평가에서 55.6%였다. 이는 실험 평가에서는 우연(chance)에 의한 정답의 확률을 줄이고 국가 수준의 대집단을 평가하여야 하는 제한성 때문에 5지 선다형 형태를 취했던 것으로 해석할 수 있다. 또한 4지 선다형은 실험평가와 수학능력 평가에서는 한 문항도 없었으며, 표집학교 성취도 평가에서만 11.4%를, 논술형 중 단답형은 실험평가와 대수능 평가에서는 없었으며 표집학교 성취도 평가에서만 30.8%를, 이유진술형은 실험평가에서 2.1%, 대수능 평가에서는 한 문항도 없었으며 표집학교 성취도 평가에서는 2.0%, 기타 유형이 0.2%로 나타났다<표 9>. 이는 국가수준의 평가보다 일선 학교의 성취도 평가가 다양한 유형의 문항을 이용하는 것으로 나타나 바람직한 것으로 판단된다. 또한 평가 문항 유형에 따라 측정학적 특성

이 다르기 때문에 각 탐구능력 요소에 보다 타당도가 높은 문항 유형에 대한 이론적 연구와 이에 대한 적용이 필요할 것으로 판단된다.

6) 평가들의 3축 평가 목표를 공유한 문항수의 분석 및 논의

전체 문항에 대한 3축 평가 목표를 공유 문항수비는 실험 평가에서 98.6%, 4회의 수학능력 평가에서 100%, 표집학교 성취도 평가에서 44.9%로 2배 이상의 차이를 보이고 있다 <표 10>.

이는 현장 과학 교육의 평가면에서 3차원 평가들과 이를 근거로한 평가 방법 및 문항 개발에 대한 이해 부족에 연유한 것으로 해석할 수 있다. 과학 교육이 추구하는 목표 중의 중요한 일부로서 3차원적인 목표들을 설정한다면 이에 대

<표 9> 문항 유형(item form) 분포

()는 전체 문항에 대한 백분율(%)

유형 구분	선다형										논술형						기타			계		
	4 지 선다형				5 지 선다형						단답형				이유 진술형		연결형	주객 혼형	퍼즐형			
	단독	2세트	3세트	4세트	단독	2세트	3세트	4세트	5세트	단독	2세트	3세트	4세트	단독	2세트	3세트						
1 차 실험 평가					1 (4.2)	3 (25.0)	2 (25.0)	1 (16.7)	1 (20.8)	1.5 (6.3)					0.5 (2.1)					24 (100)		
2 차 실험 평가					3 (11.1)	7 (51.9)	2 (22.2)	1 (16.7)													27 (100)	
3 차 실험 평가					4 (14.8)	10 (74.1)	1 (11.1)														27 (100)	
4 차 실험 평가					15 (45.5)	9 (54.5)															33 (100)	
5 차 실험 평가					15 (45.5)	9 (54.5)															33 (100)	
6 차 실험 평가					8 (24.2)	11 (66.7)	1 (9.1)														33 (100)	
7 차 실험 평가					7 (21.2)	13 (78.8)															33 (100)	
평균					53 (25.2)	62 (59.0)	6 (8.6)	2 (3.8)	1 (2.4)	1.5 (6.3)					0.5 (2.1)					33 (100)		
문항유형별 평균					208 (99.04)				2 (0.95)												210 (100)	
93 1 차 수능 시험					31 (93.9)	1 (6.1)															33 (100)	
93 2 차 수능 시험					29 (87.9)	2 (12.1)															33 (100)	
93 년 평균					66 (100)																66 (100)	
94 3 차 수능 시험					34 (94.4)	1 (5.6)															36 (100)	
94 년 평균					36 (100)																	
95 4 차 수능 시험					34 (94.4)	1 (5.6)															36 (100)	
95 년 평균					36 (100)																	
전체 평균					138 (100)																	
표집 학교 92 년도	266 (30.1)	9 (2.0)	4 (1.4)	1 (0.5)	279 (31.6)	5 (1.1)	1 (0.3)			189 (21.4)	36 (8.2)	3 (1.0)	1 (0.5)	14 (1.6)	1 (0.2)				1 (0.1)		883 (100)	
문항 유형별 평균	300 (33.9)				292 (33.1)				274 (31.0)				16 (1.8)		1 (0.1)							
표집 학교 93 년도	22 (3.0)		1 (0.4)		435 (59.8)	25 (6.9)	5 (2.1)			164 (22.6)	11 (3.0)	1 (0.4)		11 (1.5)				1 (0.1)	1 (0.1)		727 (100)	
문항 유형별 평균	25 (3.4)				500 (68.8)				189 (26.0)				11 (1.5)		2 (0.5)							
표집 학교 94 년도	3 (0.5)				354 (57.2)	14 (4.5)	1 (0.5)			195 (31.5)	9 (2.9)	1 (0.5)		13 (2.1)				2 (0.3)			619 (100)	
문항 유형별 평균	3 (0.5)				385 (62.2)				216 (34.9)				13 (2.1)		2 (0.3)							
표집 학교 95 년도					380 (59)	19 (5.9)	1 (0.5)			172 (26.7)	15 (4.7)	1 (0.5)		15 (2.3)		1 (0.2)					644 (100)	
문항 유형별 평균					421 (65.4)				205 (31.8)				18 (3.0)									
평균	328 (11.4)				1598 (55.6)				884 (30.8)				58 (2.0)		5 (0.2)							2873 (100)

<표 10> 3 차원 과학학습 평가들의 3 축이 공유한 문항 수 분포

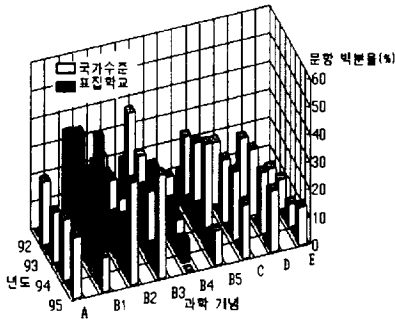
()는 전체 문항에 대한 백분율(%)

시험구분	실 험 평 가							수 학 능 력 시 험				표 집 학 교			
	1 차	2 차	3 차	4 차	5 차	6 차	7 차	93 1 차	93 2 차	94 3 차	95 4 차	92 년도	93 년도	94 년도	95 년도
문항 수	22 (10.5)	27 (12.9)	27 (12.9)	33 (15.7)	33 (15.7)	31 (14.8)	34 (16.2)	33 (100)	33 (100)	36 (100)	36 (100)	281 (31.8)	439 (56.3)	335 (54.1)	266 (41.3)
계	207 (98.6)							138 (100)				1291 (44.9)			

한 교사의 사전·사후(pre or in-service) 교육이 절실히 필요한 것으로 판단할 수 있다. 6차 과학과 교육과정의 목표를 근거로 볼 때 이에 대한 강조는 앞으로도 계속될 것으로 판단되기 때문에 더욱 필요하며, 과학 교육의 본성을 추구한다는 측면에서도 또한 그렇다.

6. 상호 비교 분석 결과와 논의

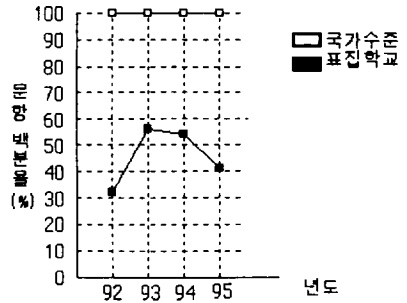
대학수학능력시험 실험평가, 대학수학능력시험 과학 탐구영역 평가 문항, 표집학교 과학 성취도 평가 문항을 대학수학능력 시험 실시전·후 4년간의 시간 경과에 따른 과학 개념 문항비, 탐구능력 문항비, 탐구단계별 문항비, 탐구 상황별 문항비, 문항 표현 양식별 문항비, 평가목표가 3축을 공유한 문항비의 변화를 나타내면 다음과 같다.



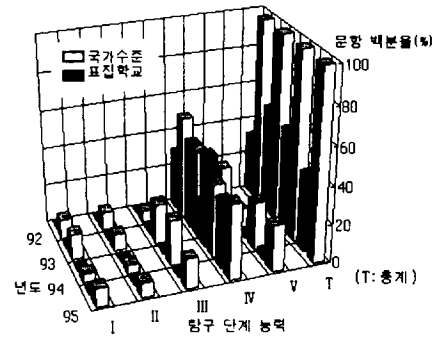
<그림 2> 과학 개념 측정 문항수비 변화

과학 개념측의 문항비를 보면 국가 수준에서는 4개 과학 교과 영역이 균형있게 평가되고 있으나 표집학교 수준에서는 한 교과 중 마지막 단원의 평가 비중은 낮은 것으로 나타났다. 또한 통합교과적 소재를 활용한 평가도 국가 수준에서 9.5% 내외, 표집학교 수준에서 0.5%내외로 매우 낮은 것으로 나타났다<그림 2>. 표집학교의 탐구능력 평가 문항비는 수학능력 시험 실시전 31.8%이던 것이 실시되던 해에 56.3%로 급상승한 반면 그 이후도 47%를 상회하는 것으로 나타나 대학입학시험제도의 평가 목표가 현장 과학 교육에 지대한 영향을 주는 것으로 나타났다.

국가 수준의 모든 문항이 탐구능력을 측정하는 반면 표집학교에서는 평균 44.9% 만이 탐구능력을 측정하고 있었으며, 실험평가, 수학능력 시험, 표집학교 과학 성취도 평가 모두 자료해석과 결론도출 및 평가 능력 측정이 대부분이었으며 문제인식 및 가설설정, 탐구의 설계, 탐구의 수행 능력 평가에 소홀히 하고 있는 것으로 나타났다<그림 3, 4>.

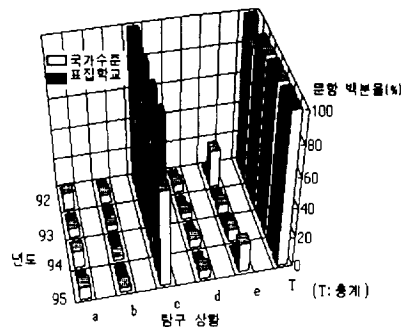


<그림 3> 탐구능력 측정 문항수비 변화



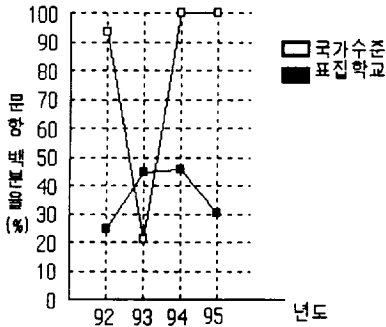
<그림 4> 탐구단계별 측정 문항수비 변화

상황측의 문항비는 대부분 학교와 지구 문제 상황이었으며, 의식주, 일상, 지역의 문제 상황은 매우 낮은 비율로 시간이 지나도 이들의 비율은 안정적인 것으로 나타났다<그림 5>.



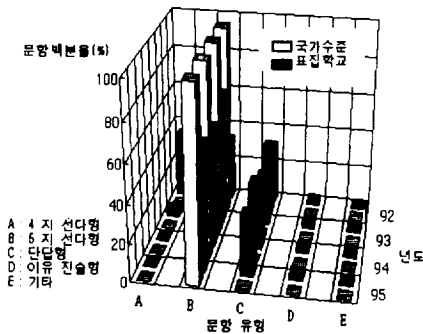
<그림 5> 탐구 상황별 문항비 변화

문항 표현 양식별 문항비를 보면 다양한 자료를 이용한 문항이 국가수준의 93년도 수능시험에서만 21.2%로 낮았을 뿐 모두 90% 이상이었고, 표집학교에서는 수학능력시험 실시전보다 그 이후부터 상승한 것으로 나타났다<그림 6>. 평가 문항 유형별로는 국가수준에서는 실험평가에서의 논술형 0.95%를 제외하고 모두 5지 선다형이었다.



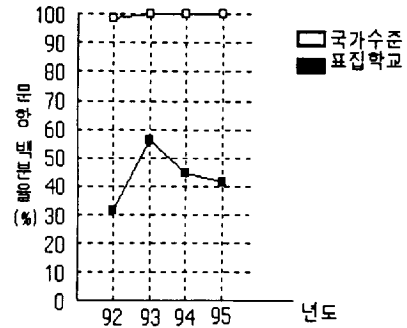
<그림 6> 문항 표현 양식 문항수비 변화

표집학교에서는 수학능력시험 실시 전·후해 4지 선다형은 33.9%에서 1.3%로 급격하게 줄어든 반면, 5지 선다형은 33.1%에서 65.5%로 급상승했다. 또한 단답형은 31%에서 30.9%로, 이유 진술형은 1.8%에서 2.2%로, 기타도 0.1%에서 0.3%로 비교적 안정적인 것으로 나타났다<그림 7>.



<그림 7> 문항 유형별 문항수비 변화

3축이 공유한 문항수의 비도 국가 수준에서는 98.6% ~ 100%인 반면, 표집학교에서는 대학수학능력 시험 실시전 31.8%에서 실시 후 평균 50.8%로 급상승한 것으로 나타났다<그림 8>.



<그림 8> 3축 공유 문항수비 변화

IV. 결론 및 앞으로의 연구 과제

1. 결론

우리 나라에서는 3차 교육 과정때 부터 과학과에서는 학생 중심의 탐구 활동을 핵으로 하는 교육과 그의 평가를 계속 강조해 오고 있다. 그 동안 그호로만 식상하던 탐구능력 중심으로의 혁신적인 전환은 대학수학능력시험의 과학 탐구 영역에서 과학 탐구능력을 측정하게 되어 과학의 본성 교육에 대한 희망을 갖게 되었다. 또한 6차 과학과 교육과정에서는 탐구 영역을 독립시켜 탐구능력 신장의 과학교육 목표를 한층 강조하고 있으며, 과학과 기술과 사회의 관계 속에서 과학적 문제 해결력 배양을 중심으로 하는 공통과학 교과를 신설하여 앞으로도 이에 대한 강조가 계속될 전망이다. 이와 같은 과학교육 사조에 힘입어 국내외에서 그 동안 탐구 학습 방법의 효과와 그를 측정하기 위한 평가도구 개발에 관한 연구가 주종을 이루어왔으며, 우리나라에서도 선행 연구들을 바탕으로 새로운 대학입시제도인 대학수학능력시험의 과학탐구 영역에서 과학 탐구능력을 측정하는 일대 전환을 하게되었다.

이런 대학수학능력시험 제도의 개발도 중요하지만 현장 과학교과 평가 방법의 변화를 수반한다고 보장할 수는 없으며, 과거 혁신 실패의 여러가지 이유 중 그 이후의 과정에 관심과 지원을 게을리 했기 때문이라고 볼 수 있다. 현장 교육의 개선을 위해서는 과학 교사들의 열혈이 가장 중요한 변수이며, 교사 개개인의 변화 없이는 교육적 변화를 기대할 수 없기 때문이다.

따라서 본 연구에서는 탐구 과학교육 활동의 결과를 보다 다양하고 효율적으로 평가하기 위한 방법에 대한 단서를 찾

아보기 위해 대학수학능력시험 실시전·후 현장의 과학 성취도 평가 방법이 어떻게 변하고 있는지를 국가수준의 평가와 상호 비교·분석하는 연구를 수행하였다.

과학 탐구능력 평가에 관한 문헌 연구를 통해 3차원 평가틀을 구안한 후 7차례의 수학능력 실험평가의 과학 탐구능력 측정 210문항, 4차례 수학능력시험의 과학 탐구능력 측정 138 문항, 수학능력 시험 실시 전·후 4년간 10개의 표집 학교에서 과학 성취도 평가 2873문항을 표집하여 본 연구에서 구안한 3차원 평가틀을 근거로 상호 비교·분석한 결과는 다음과 같다.

국가수준의 평가에서는 현행 고등학교 과학과 교육과정에서 제시된 4개 교과 영역의 개념이 균형있게 평가되고 있으나 표집학교 수준의 성취도 평가에서는 교과의 단원 중 마지막 단원의 평가에는 소홀히 다루어지고 있는 것으로 나타났다. 국가수준의 평가에서는 모든 문항이 탐구능력을 측정하고 있으나 표집학교 수준의 성취도 평가에서는 전체 문항 중 평균 약 44.9%만 탐구능력을 측정하는 것으로 나타났다. 5 단계의 탐구 단계별 능력 측정 문항비는 국가수준과 표집학교 수준 모두 자료 해석 능력이 각각 약 45.3%, 35.6%로, 결론 도출 능력이 20.5%, 2.8%로 나타났으며, 문제의 인식 및 가설 설정 능력은 국가 수준에서 평균 약 11.5%, 표집학교 수준에서 평균 약 0.6%, 탐구의 설계 능력은 국가수준에서 평균 약 9%, 표집학교 수준에서 평균 약 1%, 탐구의 수행능력은 국가수준에서 평균 약 13.9%, 표집학교 수준에서 평균 약 1.7%로 나타났다. 대학수학능력 시험 실시전의 표집학교 과학 성취도 총 문항의 31.8%가 과학 탐구능력을 측정하던 것이 대학수학능력 시험 실시 이후 평균적으로 50.5%로 급상승한 것은 현장의 과학 학습 평가가 탐구능력 평가 중심으로 정착되어 가고 있는 것으로 해석할 수 있다. 다만 자료 해석 능력과 결론 도출 및 평가 능력에 치중되고 있는 것으로 나타나 그 이외 3단계 능력의 평가에 관심을 가져야 할 것으로 보인다. 이 점은 국가수준의 평가보다 표집학교 수준에서 더욱 그런 것으로 나타났다.

문제가 제시되는 상황의 문항비는 국가수준과 표집학교 수준에서 각각 99.3%, 100%로 나타났으며 학교 상황이 각각 67.2%, 98.4%, 지구 상황이 각각 17.9%, 0.9%이고 나머지는 그 이외의 상황인 것으로 나타났다. 이는 대부분의 문항 제시 상황이 교과서적인 상황이라는 것을 의미하는 것으로 최근 과학 교육이 강조하는 생활할 속에서의 문제 해결력 신장을 강조하는 목표와는 거리가 있는 것으로 해석할 수 있어 다양한 상황속에서의 문항 개발이 시급한 것으로 나타났다.

문항 표현 양식도 국가수준에서는 78.1%가, 표집학교 수

준에서는 35.7%가 다양한 자료를 이용하여 제시하는 것으로 나타나 2배 이상의 차이가 있었다. 이는 현장 학교 수준의 성취도 평가 문항이 매우 단순한 진술문 형태로 제시되고 있는 것으로 해석이 가능하나 수학능력시험 실시 이전 24.8%이던 것이 이후에 42.2%로 향상되어 현장에서도 많은 노력을 하고 있음을 알 수 있다.

문항 유형은 국가수준의 평가에서는 0.95%를 제외한 모든 문항이 5지 선다형을 취한 것으로 나타났으나, 표집학교 수준에서는 4지 선다형 11.4%, 5지 선다형 55.6%, 단답형 30.8%, 이유진술형 2.0%, 기타가 0.2%의 비율을 보였고 대학수학능력 시험 실시 이후 5지 선다형이 차지하는 비율은 33.1%에서 65.5%로 급격히 상승하여 추측에 의한 오차의 한계를 극복하려고 노력하려는 경향을 보이고 있는 것으로 해석할 수 있다. 이를 근거로 볼 때 국가수준의 평가보다 현장의 학교 과학 성취도 평가가 다양한 문항 형태를 이용하고 있어 국가수준의 평가 문항 유형에 관한 개선점을 시사하는 것으로 볼 수 있었다.

3축이 공유한 문항비는 국가수준의 평가에서는 99.3%, 표집학교 수준에서는 44.9%를 보여 학교 수준의 평가에서 3축 공유의 문항 개발에 어려움을 겪고 있는 것으로 볼 수 있으나 대학수학능력 시험 실시 이전 31.8%이던 것이 대학수학능력 시험 실시 이후 50.6%로 현격한 상승을 보여 현장에서 최근 강조하는 과학교육 목표에 적합한 문항 개발에 부단히 노력하고 있는 것으로 나타났다.

이상과 같은 본 연구의 결과는 대학입시제도의 평가 목표에 의해 현장 과학교육의 평가 목표도 강하게 영향을 받고 있음을 시사하고 있으며, 현행 과학교육이 강조하고 있는 과학과 기술 및 사회 문제 속에서의 문제 해결 능력인 과학 탐구능력 신장의 학습과 평가가 점진적으로 정착되어 가고 있음을 반증하는 것으로 볼 수 있다.

2. 앞으로의 연구 과제

본 연구에서는 대학수학능력시험 실시전·후의 시간 경과에 따른 현장의 과학 학습 성취도 평가 문항이 추구하고 있는 목표의 변화, 문항 유형의 변화, 문항 표현 양식의 변화를 국가수준의 평가와 종단적으로 비교·분석해 보았다. 본 연구를 수행하는 과정에서 밝혀진 문제점을 바탕으로 좀더 바람직하고 타당도 높은 과학 탐구능력 평가를 위한 향후의 연구 문제를 제시하면 다음과 같다.

1) 과학 학습 평가틀의 국내외 문헌 연구를 하는 과정에서 Cronbach의 견해와 같이 평가틀의 작성과 설정 과정에 대한 학문적 전개가 약하고 학자들마다 다르게 제시되고 있어

이 분야에 대한 연구자들에게 혼란을 야기시키고 있는 것으로 보아 좀더 체계적이며 통합하는 연구가 필요하다.

- 2) 문항별로 측정하는 과학 탐구능력이 다르기 때문에 5단계 중 자료해석과 결론도출 및 평가와 같은 특정 능력에 치중하고 있는 것으로 나타나 균형있는 평가가 어렵다고 판단된다. 따라서 하나의 문제 상황속에서 5단계를 종합적으로 평가할 수 있는 문항 개발법에 관한 연구가 필요하다.
- 3) 단시간에 다수를 대상으로 하는 국가수준에서는 모두 5지 선다형을 택하고 있으나 측정학적 측정하고자 하는 능력에 따라 문항 유형이 다르다. 따라서 측정하고자 하는 과학 탐구능력과 문항 유형과의 상관 관계나 인과 관계에 관한 연구도 필요하다
- 4) 학생들이 문항 개발자가 의도하는 방향으로 탐구 과정을 거쳐 문제를 해결하는지 아니면 다른 방법으로 해결하는지 등의 일반화된 방법을 찾아 이를 바탕으로 하는 보다 타당도 높은 평가 방법에 관한 연구도 필요하다
- 5) 국가수준은 물론이고 현장 학교 수준에서도 과학에 대한 정의적 영역 평가는 전혀 실시하지 않는 것으로 나타났다. 과학적 태도나 과학에 대한 태도도 이 두 수준의 평가에 병행할 수 있는 구체적인 이론과 실천 연구가 필요하다

감사의 글 : 본 연구는 20개 학교의 선생님들이 4년간 자료를 수집·분석하는 과정을 거쳐 이루어졌습니다. 본 연구를 위해 협조해 주신 선생님들의 노고에 지면을 통해 감사드립니다.

참 고 문 헌

교육부(1992). 대학수학능력시험 해설, 서울 : 극동문화사
 구창현(1993). 과학적 탐구 학습 지도와 평가에 관한 세미나 및 워크숍(내용), 강원대학교부설 과학교육연구소
 김상철,권재술(1994). 대학수학능력시험에서 탐구영역의 분석, 한국과학교육학회지, 14(2), pp.214~224.
 김은진,김영수(1992). 대학수학능력시험 실험 평가 문제의 분석, 한국과학교육학회지, 12(1), pp.30~46.
 김익균, 황유정(1993). 고등학생의 탐구 사고력 문제 해결 과정에 나타난 유형과 특성, 한국과학교육학회지, 13(2), pp.30 ~ 46.
 김준태, 권재술(1994). 과학 개념의 표현 양식별 학습 지속 효과, 한국과학교육학회지, 14(2), pp.111~122.
 김찬중(1993). 과학 탐구 학습 문제의 유형 및 특성 분석 연구, 한국과학교육학회지, 13(2), pp.359~365.
 김창식 등(1992). 과학학습 평가, 서울 : 교육과학사.

명전옥,박승재(1995). 대학수학능력시험 도입에 따른 과학적 탐구사고력 평가에 대한 과학 교사들의 관심과 필요 사항, 한국과학교육학회지, 15(4), pp.417~428.
 박도순(1989). 문항작성방법론, 서울:중앙교육진흥연구소. pp.1 ~ 3.
 박승재(1991). 과학적 탐구사고력 평가, 서울대학교사범대학 물리교육과 물리학습연구실
 박학규(1993). 학생들의 물리 문제해결 과정과 문제공간의 유형 분석. 한국교원대학교 박사학위논문.
 우종옥,이항로,이경훈(1991). 대학수학능력시험의 수리·탐구 영역 중 지구과학 교과에 관련된 탐구능력 측정을 위한 행동 요소의 추출과 평가. 목표의 상세화 연구 I, 한국과학교육학회지, 11(1), pp.83~96.
 우종옥,이항로,이경훈(1992). 대학수학능력시험의 자연과학 탐구능력 평가를 위한 행동 요소의 추출과 평가목표의 상세화 연구 II, 한국과학교육학회지, 12(2), pp.81~95.
 우종옥,이경훈,이항로(1994). 과학교육목표에 관한 연구. 한국과학교육학회지. 14(2). pp.159~169.
 이종기(1988). 고등학생의 과학 탐구능력 측정을 위한 평가 도구 개발. 한국교원대학교 석사 학위논문
 이항로(1992). 고등학생의 과학 탐구능력 측정을 위한 평가 도구 개발-지구과학소재를 중심으로-, 한국교원대학교 석사 학위 논문.
 이화국(1989). 외국의 과학교육의 동향. 과학교육 박승재 편지, 서울 : 교육과학사, p.52.
 이 무(1992). 과학적 사고력 평가 문항 형태에 관한 연구. 서울대학교 박사학위 논문.
 한종하(1984). 미국 과학교육의 동향과 변화-1960 ~ 1980-, 한국교육개발원 기타 연구, OP 84-3.
 APU(1984). Science Assessment Framework, Age 13 & 15 : *Science Report for Teachers*, No.2, DES, DEM and WOED.
 Nielsen, B. S. (1995). South Carolina Science Framework-Field Review Draft, South Carolina : *South Carolina Science Curriculum Framework Writing Team*.
 Bell, B. & Osborne, R.(1981). Interviewing Children - *A Checklist for the I.A.I.Interviewer. Learning in Science Project, Working Paper, Document Resume*. ED 230 594.
 Brown, S & McIntyre, D.C.(1986). Influence upon teacher's attitudes to difference types of innovation : A study of Scottish integrated Science. In Brown

- et al(Eds). *Science in School. Philadelphia, PA : The Open University Press*
- Bruner, J.S.,(1960). *The Process of Education*, Cambridge Mass : Havard University Press, pp.19 ~ 22.
- Liu, Chin-Tang, Lieu, J. and Yager, R. E.(1995). 1996 Iowa Assesment Handbook. Iowa : The Assessment Office Science Education Center The University of Iowa.
- Cronbach, L.J.(1970). Review of On the theory of achievement test items by J. R. Bormuth. *Psychometrika*, 35, pp.509 ~ 511.
- Hall, G.E.(1992). The local educational change process and policy implementation, *Journal of Science Education*, 29(8), pp.877~904.
- Kirkham, J,(1989). Balanced Science : equilibrium between context, process, and content. In W.Jerry(Ed.), *Skills and process in Science Education : A Critical Analysis*, London & New York : Routledge, pp.130~150.
- Lawson, A.E.(1982). The reality of general cognitive operations. *Science Education*, 66(2), pp.229 ~ 241.
- NAEP(1984), Science Objectives. fifth National Assessment 3rd Draft.
- NAEP(1984), Science Objectives. 1990 Assessment, The Nation's Report Card.
- Robertta H. Barba & Peter A. Rubba(1993). Expert and novice, earth and space: teacher's declarative, procedural and structural knowledge, *International Journal of Science Education*, 15(3), pp.273 ~ 282.
- Yager, R. E.(1993). The Science, Technology, Society Movement. Volume 7. Washington DC : NSTA.
- Doran, R.L., Boorman J., Chan, F. & Hejaily, N.(1993). Alternative Assessment of High School Laboratory Skills, *Journal of Research Science Teaching*, 30(9), pp.1121 ~ 1131.
- Shaw, E. L., Jr et al(1988). Assessing the skill of Variable Identification with Multiple Choice Items, Document Resume, ED 292 663.
- Tamir, P(1990). Justifying the selection of answers in multiple choice items, *International Journal of Science Education*, 12(5), pp.251 ~ 259.
- Swain, J.R.L,(1988). The development of a framework for the assessment of process skills in a Graded Assessment in Science Projects, *International Journal of Science Education*, 11(3), pp.251 ~ 259.

(ABSTRACT)

Longitudinal Study about Science Process Skills Item Forms Transition before and after Scholastic Ability Test for College

Woo, Jong-Ok · Lee, Hang-Ro
(Korea National University of Education)

Goo, Chang-Hyun
(Korea National Evaluation Institute)

This study investigated the literature about science process skills' evaluation to analyse transition of evaluation objectives before and after a Scholastic Ability Test for College Entrance. In the literature survey the researcher established a 3 dimensional science assessment framework with X axis as science concept, Y axis as science process skills and Z axis as problem context.

In order to analysis and compare each items the researcher selected 210 items from the 1st to the 7th trials and 138 items from the 1st to 4th Scholastic Ability Test for College Entrance and sampled 2873 science achievement test items from 10 high schools. In accordance with this taxonomy the researcher analysed and compared science process skills item forms.

The following results were drawn :

The items were evenly distributed in all the four areas(Earth Science, Biology, Physics and Chemistry) of the science concept domain, but they were heavily concentrated on data analysis and drawing a conclusion in science process domain. In the domain of problem context school context was the majority. In spite of distribution like this the ratio on science process skills measurement items and science achievement test items was increased after the Scholastic Ability Test for College Entrance was given. Also the ratio on item expression type was increased. Item form was almost 5 options selection type in the national level test. Although there were 4 options selection type, 5 options selection type, short answer type, essay type in school level test, rising from 33.1% to 65.5% on 5 options selection type is exhibited. This study showed that the school level item form was better various than the nation level. This point like this is the evidence for the improvement toward the science process skills test and influenced by Scholastic Ability Test for College Entrance. The ratio on the item which joined with the 3 axes had a mean of 99.3% in nation level test and mean 44.9% in school achievement test level. But the ratio in the school achievement test level increased after the Scholastic Ability Test for College Entrance was given.

In view of this study we must furthermore study the item types which can evaluate validately science process skill's five stage each and evaluation method by the high school students' problem solving patterns and features in scientific inquiry on all science process skills elements.