

# 국가 수준의 과학 지식 평가 체제 개발<sup>1)</sup>

권재술 · 최병순 · 김찬종<sup>1</sup>  
(한국교원대학교) · <sup>1</sup>(청주교육대학교)

## Developing National Science Assessment System: Scientific Knowledge Domain<sup>1)</sup>

Kwon, Jae-Sool · Choi, Byung-Soon · Kim, Chan-Jong<sup>1</sup>  
(Korea National University of Education) ·  
<sup>1</sup>(Chongju National University of Education)

### ABSTRACT

Establishing and evaluating science education policies and revising and monitoring the effectiveness of science curriculum should be based upon the results of systematic and scientific research studies. Advanced nations have already been administering and developing national level science assessments for these purposes. The science assessments administered in Korea have been reported having many limitations and problems, and not succeeded in providing data for science education policy making and curriculum reform. The major purpose of the study is developing national level science knowledge assessment system in order to identify longitudinal trends of elementary and secondary school students science knowledge achievements.

The research team consisted of science education experts and teachers from various school levels, decided the directions and major elements of national level science knowledge assessment with the consultation of educational evaluation experts. Item developing ability of the researchers was improved by seminars and workshops on national assessment in advanced nations and developing skills of writing science items. Nearly 500 items were developed and revised. Pilot test was administered with 958 students at various school levels. 380 items were selected and tested with 8766 students, and the characteristics were analyzed in terms of item response theory.

The target populations for national level science knowledge assessment are 5th-grade of elementary school, 2nd-grade of middle school, 1st and 2nd-grade of high school students. The proper period for the assessment is February every year. Multi-stage clustered sampling method is desirable and rotated forms are recommendable for the test format. Bridge items should be introduced to compare the results of multiple tests, and various grades. Anchor items should also be used for longitudinal interpretations of the results. The items for elementary school require low to medium abilities, for middle school and first grade of high school require medium to high abilities and for 2nd-grade of high school high abilities. The discrimination ability of the items developed is high.

\*1998년 7월 16일 받음

<sup>1)</sup> 이 연구는 1994년도 학술진흥재단의 대학부설 중점연구소 지원사업으로 수행되었습.

**Key words** : science assessment, science knowledge, national level assessment, elementary school, middle school, high school.

## I. 서 론

우리는 거의 20,000여개에 이르는 각급 학교를 갖고 있으며, 약 33만 명 이상의 교원과 약 1,200만 여명의 학생을 22만개의 교실에 수용하고 있으며, 일년에 몇 조원에 달하는 예산을 투입하고 있다(교육부, 1997). 우리나라의 어떤 기업이나 조직도 그 규모와 중요성에 있어서 이렇게 큰 사업을 하는 곳이 없다. 특히, 국가 경쟁력 강화와 국제화 시대의 필수 불가결 요소인 과학 과목은 그 중요성 때문에 과학입국, 과학교육의 진흥 등과 같은 많은 구호가 제시되어 왔으며, 과학교육 진흥법과 학교 실험 실습에 필요한 실험실과 기자재의 확보 기준까지 제정되기도 하였다.

그러나 그 동안 투입된 많은 노력과 예산이 얼마나 효율적으로 사용되었으며, 수많은 과학교육과정 개편과 정책들이 우리 나라 학생들의 과학적 소양을 높이고 우수한 과학기술 인력의 양성에 얼마나 기여하였는지에 대한 실증적인 자료는 많지 않다. 학교 과학교육에 대한 문제점 제기와 질적 개선을 요구하는 많은 의견이 있었지만, 체계적이고 과학적인 연구와 조사가 바탕이 된 경우는 거의 없었다. 이는 우리 나라에서 시행되고 있는 과학 교육 평가가 선발(selection)과 학습(learning) 기능에 치우쳐 있으며, 다른 중요한 기능인 점검(monitoring)이 잘 이루어지지 않고 있음을 의미한다(Kim, 1996).

선진국의 경우 많은 나라가 국가 수준에서 학교 과학 교육을 관리한다는 견지에서 국가 교육의 질을 점검하려는 노력을 진행시켜 왔다(Somerset, 1996; 신세호, 1987). 미국의 NAEP(National Assessment of Educational Progress, 1984)나 영국의 APU(Assessment of Performance Unit)와 NCA(National Curriculum Assessment, 1984a, 1984b)가 그 대표적인 예라고 할 수 있다. 예컨대, 미국 연방정부는 1983년 "국가의 위기(A Nation at Risk)"라는 보고서를 국민에게 내놓고 교육 개혁의 필요성과 방향을 천명했는데, 그 보고서의 근거는 바로 미국의 국가수준평가 즉, NAEP의 연구 결과였다. 즉, 당시의 교육개혁 필요성과 방향은 1969년 이후 10여년 이상 미국 교육체제를 면밀히 분석하고 조

사한 평가 결과에 바탕을 두고 설정된 것이었다. 미국은 종종 이러한 평가의 결과를 바탕으로 하여 교육개혁의 방향을 정해 왔다. 영국의 국가 수준의 교육평가 연구 즉, APU는 1989년 영국의 교육과정 정책을 바꾸는데 중요한 자료를 제공하였다. OECD국가들 중심으로 많은 선진국들은 자국의 교육 체제의 효율성이나 효용성을 체계적이고 과학적으로 이해하려는 노력을 게을리하지 않고 있다는 점에 우리는 주목해야 한다. 이러한 노력의 일환으로 국가 수준 학력 평가 연구 사업을 국가의 주요 사업으로 정례화하고 그 규모를 확대하고 있는 것은 좋은 타산지석이 된다(Broadfoot & Gipps, 1996; Madaus & Raczek, 1996).

이처럼 국가수준에서 한 국가의 과학교육 성과를 가늠해보려고 하는 작업은 그 국가의 과학교육의 질을 체계적이고 과학적으로 관리한다는 차원에서 매우 중요하다. 즉, 과학교육 전반에 걸쳐서 교육의 질을 확보하고 관리한다는 견지에서 과학교육 지식 수준 및 성취 수준의 장기적 추이를 밝히고 교육정책자들이 교육 정책에 관련된 의사결정을 하는데 쓰일 수 있는 필수 자료를 제공할 수 있는 체계적이고 장기적인 과학 지식의 조사 연구는 매우 필요하다고 할 수 있다.

우리 나라에서도 학교 교육의 성과를 평가하려는 작업은 제한적이고 산발적이거나 꾸준히 이루어져 왔다(고종렬, 이해선, 박종삼, 1969; 국립교육평가원, 1993a, 1993b, 1993c, 1995, 1997; 권낙원, 이재분, 구자역, 1984, 1985; 김명숙, 강태중, 양미경, 백순근, 1994; 김형립 외, 1980; 박종삼 외, 1972; 이종재 외, 1978; 전용신, 1963; 중앙교육평가원, 1986, 1988; 한국행동과학연구소, 1980a, 1980b; Kim, 1996). 이들 연구들은 당시의 연구 요건이나 수행 여건에 비추어 볼 때, 나름대로 의미 있는 연구 결과들을 내놓았고, 학교교육의 이해에 기여한 바도 적지 않았다. 그러나 이들 기존의 평가 혹은 조사 연구들은 몇 가지 큰 제약점을 안고 있다. 첫째, 과학학력평가를 시행해온 기관들이 자주 바뀌었으며, 따라서 평가의 계획성과 일관성이 결여되고 평가 목적이 뚜렷하지 못하였다. 둘째, 이들 조사 또는 평가 연구들의 대부분은 과학 지식의 장기적 추이를 점검하지 못하고 셋째, 조사 방법이나 도구가 다양하지 못하였으며 과학 능력 중 일부 영역의 평가에 국한되었으며, 넷

제, 관련 변인이나 조건 등에 대한 관심 부족과 결과 처리 및 분석의 심도가 낮아서 과학 교육 개선에 구체적인 시사점을 도출하는데 상대적으로 소홀히 하였다(장석우, 1987). 일부 과학 학력평가에서는 평가의 점검 기능보다는 학교나 지역간 성취도 비교나 학교 과학 학습 평가 수준의 제고를 목적으로 한 경우까지 있었다. 이러한 제한점은 주로 이들 연구가 여러 연구 기관에서 제각기 일회적이고 산발적으로 조사 또는 평가를 수행하는 양태로서 진행되어 왔음에 기인한다. 또한 개개의 연구들이 각 연구의 초점을 달리하고 제각기 다른 연구 설계와 방법을 가지고 학교 과학교육을 진단하고 평가하였기 때문에 그 결과를 서로 비교할 수 있는 가능성을 약화시켰다. 이러한 문제점들은 과학 성취도의 장기적 추이를 파악하는 것을 어렵게 하였다. 우리 나라 학생들의 과학 성취도 수준에 대한 장기적 추이를 관찰하고 과학교육 과정의 올바른 길잡이를 해 줄 수 있는 지표를 구하는 작업이 매우 시급한 과제이며, 선결 과제라고 할 수 있다.

이상의 논의에서 드러나듯이 국가 수준에서의 과학 학력 평가 체제의 구축은 더 이상 미룰 수 없는 중요한 과제가 아닐 수 없다. 즉, 학교 과학교육에서 지향하고 있는 과학적 지식의 성취, 과학적 탐구 능력의 습득, 과학에 관련된 정의적 영역의 함양이라는 세 가지 측면에서 학생들에 대한 평가가 이루어져, 우리 나라 학생들의 과학적 소양에 대한 주기적 변화를 관찰하고 과학교육 과정의 올바른 길잡이를 해 줄 수 있는 지표를 구하는 작업이 매우 시급한 과제이며, 선결 과제라고 할 수 있다.

이 연구의 주요 목적은 우리 나라 초·중·고 학생들의 과학 지식 성취의 경향성을 파악하고 향후 주기적으로 진행되어질 국가적 차원의 평가 지표를 만들기 위한 국가 수준의 과학 지식 평가 체제를 개발하는 것이다. 이를 위해서 다음과 같은 하위 연구 질문에 대한 답을 모색하였다.

- (1) 과학 지식 평가 체제를 위한 적절한 방향과 주요 요소는 무엇인가?
- (2) 과학 지식 평가 체제를 위한 학교급별, 영역별 평가 문항의 수와 내용 및 형태는 무엇인가?
- (3) 과학 지식 평가 체제를 위해 개발한 학교급별, 영역별 문항과 검사의 특성은 어떠한가?

## II. 연구 절차 및 방법

이 연구는 먼저 국가 수준 과학지식 평가체제의 방향을 설정하고, 이어서 이를 위한 평가 도구를 개발하며,

개발한 도구의 특성과 적합성을 검토 분석하는 과정으로 실시되었다.

### 1. 국가 수준 과학지식 평가 체제의 주요 요소 결정

국가 수준의 과학 지식 평가 체제의 방향을 설정하기 위해서 교육평가통계 분야의 전문가들을 초빙하여 연구진과 과학교육 전문가 및 참여 교사들이 참여하여 1995년 2월부터 2개월 동안에 걸쳐 4차례의 세미나를 실시하였다. 세미나의 주제는 국가수준 과학탐구능력 평가 체제, 영국과 미국의 국가 수준 과학교육 평가체제, 과학학력 평가체제 실천방안, 미국 ACT의 개발과정과 활용 등이었다.

세미나에 초빙된 교육평가 전문가와 연구진, 그리고 과학교육 전문가들의 토의를 통해서 주요 선진국의 과학 학력평가체제에 대한 예시와 우리 나라 상황에 대한 논의 등을 바탕으로 국가 과학교육평가 체제의 대상학년, 평가 주기, 평가를 등이 결정되었다. 과학지식평가 체제를 위한 기본적인 요소를 결정한 후 여기에 핵심적인 양질의 평가 도구 개발 과정으로 들어갔다.

### 2. 과학지식평가체제 점검을 위한 과학 지식 평가 문항 개발

과학 지식 영역의 평가 도구를 개발하기 위해서는 본 연구에서는 평가 문항이 어느 한 영역에 편중됨을 방지하고, 체계화하기 위해서 먼저 지식 영역의 평가틀을 만들고, 이 평가틀에 입각하여 교육목표를 상세화하였다. 상세화된 목표에 따라서, 평가 문항을 개발하고, 이를 현장에 투입하여 난이도와 변별도를 조사하여 다시 문항을 수정, 보완하였다.

개발된 과학지식 평가틀을 기초로 연구진은 우리 나라 초·중·고등학교의 6차 교육과정을 분석한 후, 평가목표 상세화 작업을 하였다. 선형 연구와 국가 수준의 평가라는 본 연구의 성격을 고려하여 마련된 평가 목표 상세화를 위한 원칙을 요약하면 다음과 같다.

- 가. 평가 목표는 각 학년별로 제시하되, 해당 학년뿐 아니라 해당 학교급의 지식 성취를 대표할 수 있는 내용으로 한다.
- 나. 평가를 각 차원의 특정 하위 영역에 문항들이 편중되지 않도록 한다.
- 다. 과학교육의 본질상 꼭 포함시켜야 할 내용 혹은

국내 교육과정에는 없거나 중요하지 않게 다루더라도 평가 목표로 바람직하다고 판단되는 내용이 있으면 포함시킨다.

라. 장기간의 과학 지식 성취 수준의 추이를 분석하기 위해서는 교육과정이 바뀌더라도 이의 영향을 최소화할 수 있도록 핵심 내용(core content)을 중심으로 한다.

1) 문항개발팀 구성

본 연구에서 개발한 평가 도구는 초 중 고 학교급별에서 사용되어야 하므로, 각 학교 급별에 따라 공동연구원 및 연구보조원을 구성하고 아울러 현직 교사 및 교육 전문직 등을 대상으로 문항 개발팀을 구성하였다. 문항 개발팀은 동시에 진행된 과학탐구평가체제개발연구와 과학태도평가체제개발연구진을 포함하여 대학교수 11명과 현직교사 34명의 협력연구원으로 구성되었다.

2) 문항 제작을 위한 워크숍

양질의 문항을 작성하기 위해서 연구에 참여하는 현직 교사들을 대상으로 1995년 6월과 7월 동안 문항 개발에 대한 4차례의 워크숍이 있었다. 2차례의 워크숍은 3시간, 나머지 2차례는 1주일에 걸쳐서 집중적으로 진행되었다. 연구진을 포함한 10명의 과학교육 전공 교수들이 워크숍의 진행을 주도하였으며, 참여한 협력연구원들은 문항을 실제로 개발하고 공동으로 검토, 수정하는 기회를 가졌다.

3) 과학 지식 평가 문항 영역별 가중치 결정

과학 지식의 점검을 위해서는 우수한 문항을 많이 확보하는 것이 중요하다. 본 연구는 초기 단계이며, 평가체제의 개발이라는 관점에서 최종 평가 문항 수를 초등 80문항, 중학교, 고등학교 1, 2학년 각각 100문항씩 전체 380문항을 개발하는 것을 목표로 하였다.

평가 문항은 교육목표별로 고르게 개발하여, 평가를 실시할 때마다 동질의 평가가 이루어질 수 있도록 하기 위하여, 학교급별로 영역별 가중치를 미리 결정하였다. Table 1은 각 학교급별의 각 내용 영역(물리, 화학, 생물, 지구과학)별로 평가들에 의해 할당된 문항 수이다.

4) 개발 문항의 수정 및 보완

집중 작업을 통해 개발된 문항의 수정 및 보완을 위하여, 3회의 수정 및 보완 작업을 실시하였다. 1차 수정 및 보완 작업은 각 연구원들의 책임 하에 예비문항을 각 내용 영역(물리, 화학, 생물, 지구과학)에 따라 나누어

**Table 1** Intended distribution of items by content, ability and school level

	Knowledge	Understanding	Application
<b>Elementary school level</b>			
Facts	4	·	·
Concepts	6	7	1
Principles	2	6	1
Theory	·	2	1
Total	12	15	3
<b>Middle school level</b>			
Facts	1	·	·
Concepts	4	7	1
Principles	3	7	3
Theory	·	3	1
Total	8	17	5
<b>High school (1st-grade)</b>			
Facts	1	·	·
Concepts	4	3	3
Principles	5	4	4
Theory	·	3	3
Total	10	10	10
<b>High school (science track)</b>			
Facts	1	·	·
Concepts	2	3	6
Principles	2	4	6
Theory	·	3	3
Total	5	10	12

수정 및 검토 작업을 하였다. 2차 수정 및 보완 작업은 예비문항의 개발을 위한 집중 작업 및 검토 작업에 참여하지 않은 과학교육전공 박사과정 4명 및 현직 초·중·고 교사들에게 수정 및 보완을 의뢰하여 개발된 예비문항의 타당도를 높이도록 하였다. 3차 수정 및 보완 작업은 예비 검사를 위해 삽화 작업 및 편집 작업이 완료되어 검사지 형태로 구성된 후, 3차 수정 및 검토 작업이 이루어졌다.

3. 예비 검사

개발된 과학 지식 평가 문항의 적절성과 타당성을 검증하기 위하여 소규모 집단을 대상으로 예비 검사를 실

**Table 2** Content by ability distribution of pilot items developed

Contents		Ability				Subtotal	Total
		Facts	Concepts	Principles	Theory		
Elementary school level							
Knowledge	Physics.	3	5	2		10	53
	Chemistry	6	7	1		14	
	Life Sci.	6	6			12	
	Earth Sci.	8	5	4		17	
Understanding	Physics		11	8		19	64
	Chemistry		4	9		13	
	Life Sci.		15			15	
	Earth Sci.		14	3		17	
Application	Physics			3		3	14
	Chemistry		2	3		5	
	Life Sci.		3			3	
	Earth Sci.		2	1		3	
Middle school level							
Knowledge	Physics	5	2	1		8	25
	Chemistry	1	3	3		7	
	Life Sci.	4				4	
	Earth Sci.		4	2		6	
Understanding	Physics		10	12	1	23	84
	Chemistry	3	9	6	1	19	
	Life Sci.	1	14	1	8	24	
	Earth Sci.		7	9	2	18	
Application	Physics			3		3	
	Chemistry		3	3		6	
	Life Sci.			2	2	16	
	Earth Sci.		2	2	1	5	
High school 1st-grade level							
Knowledge	Physics	2	2			4	31
	Chemistry	6	4	1		11	
	Life Sci.	4	1			5	
	Earth Sci.	1	7	3	11		
Understanding	Physics		12	9	2	23	66
	Chemistry	1	2	9		12	
	Life Sci.		9	2	6	17	
	Earth Sci.		4	7	3	14	
Application	Physics		1	1	1	3	22
	Chemistry			7	1	8	
	Life Sci.		1		5	6	
	Earth Sci.			3	2	5	

**Table 2** Continued

Contents		Facts	Concepts	Principles	Theory	Subtotal	Total
Science track, high school (2nd-grade) level							
Knowledge	Phycis	·	3	3	·	6	22
	Chemistry	4	1	1	·	6	
	Life Sci.	3	·	·	·	3	
	Earth Sci.	1	3	3	·	7	
Understanding	Physics	·	8	10	·	18	56
	Chemistry	1	4	3	·	8	
	Life Sci.	·	9	9	3	21	
	Earth Sci.	·	3	4	2	9	
Application	Phycis	·	3	3	·	6	45
	Chemistry	·	6	10	·	16	
	Life Sci.	·	1	2	5	8	
	Earth Sci.	·	5	4	6	15	
Total		60	217	170	51	498	498

시하였다. 예비 검사에 사용된 학교별, 영역별 문항 수는 Table 2와 같다.

예비검사를 위하여 표집한 학생 수는 각 학교 급별로 Table 3과 같다. 예비 검사를 실시하여 얻은 자료를 모두 전산 입력하였다. 입력된 전산 자료를 바탕으로 각 문항의 양호도를 검토하여 필요에 따라 문항을 삭제하거나 수정·보완하였다. 문항의 양호도 검토에서는 개발된 예비 문항이 얼마나 교육평가·측정학적으로 바람직한가를 알아보는 것으로 문항의 타당도, 신뢰도, 변별도, 난이도 등을 검토하였다.

3회의 예비 검사 문항의 수정·보완과 예비 현장 검증, 전문가 협의회를 거쳐 최종 선정된 본 검사용 문항을 확정하였다. 학교급별 본 검사용 평가 문항은 40~50 분 동안 검사할 수 있는 분량으로 나누어 검사지를 구성하였다. 초등학교용은 20문항으로 한 검사지가 구성되어 있으며, 중·고등학교용은 25문항으로 한 검사지를 구성하였다. 본 검사용 검사지의 구성은 Table 4와 같다.

#### 4. 본 검사 및 결과 분석

최종 완성된 16종 380문항에 대한 본 검사를 1997년 2월 중에 실시하였다. 본 검사 실시를 위해 지역을 대도시, 중소도시, 읍·면으로 나누고 해당 지역 학생들의 모집단 수를 고려하여 연구 협력 학교를 선정하였다. 초등학교 61개교, 중학교 59개교, 고등학교 77개교에 본 연구의 취지와 조사 내용에 대한 안내 등이 담긴 협조 공문을 발송한 후, 검사지를 우송하였다.

검사를 실시한 후, 회수된 자료 중에서 학습능력이 현저히 떨어지거나(특수아동), 불성실하게 응답한 자료는 제외하고 최종 본 검사 자료처리에 포함된 지역별, 성별 표본 현황은 Table 5와 같다.

본 검사 결과는 전산 입력되었으며, 문항반응이론(item response theory; 성태제, 1991)에 의하여 Assessment Systems Corporation에서 개발한 문항 분석 package(Assessment Systems Corporation, 1996)를

**Table 3** Number of subjects for pilot test

School	Elementary	Middle	High		Total
Grade	5th	2nd	1st	2nd	
Subjects	226	274	210	248	958

**Table 4** Booklets for main test

Booklet name	Number of items	Time required (min)
Elementary school		
NKE1	20	40
NKE2	20	40
NKE3	20	40
NKE4	20	40
Total	80	
Middle school		
NKM1	25	45
NKM2	25	45
NKM3	25	45
NKM4	25	45
Total	100	
High school 1st-grade		
NKC1	25	50
NKC2	25	50
NKC3	25	50
NKC4	25	50
Total	100	
High school science track		
NKH1	25	50
NKH2	25	50
NKH3	25	50
NKH4	25	50
Total	100	

이용하여 분석을 실시하였다. 문항반응이론을 이용하여 분석을 한 이유는 이 방법이 표집 학생의 수준에 관계없이 일관성 있는 문항 특성을 알려주기 때문이다. 이 과정을 통해서 얻은 자료는 문제은행을 운영하여 학력평가를 실시하고자 할 때 적절한 수준의 검사지를 구

성하기 위한 훌륭한 자료가 될 수 있다.

### Ⅲ. 연구결과

#### 1. 과학 지식 평가 체제의 기본 요소

국가수준 과학지식 평가체제의 기본 요소로는 대상 학년, 평가 주기, 평가틀, 평가 시기, 표집 방법 및 비율, 검사지 형식, 평가 소요 시간, 문항 유형 등을 결정하였다.

##### 1) 평가 대상 학년

국가수준 과학지식 평가의 대상은 초등학교 5학년, 중학교 2학년, 고등학교 1학년, 2학년 학생들로 국한하였다. 이는 국가 수준에서 점검해야 할 대상이 각 학년 별 교육 성취가 아니라 주요 학교급별 교육성취임을 의미한다. 또한 우리 나라에서 교육 방법과 학교의 운영 방법이 학년보다는 학교급에 따라 크게 달라지기 때문에 각 학교급별 한 학년을 선정해도 대체적인 실태의 파악은 가능하다고 보기 때문이다. 각 학교급별로 한 학년 씩을 평가 대상으로 잡았으나, 그 학생들이 저학년에서 배운 내용도 평가 항목에 포함되기 때문에 여러 학년에 걸쳐서 배운 결과를 평가할 수 있다. 각 학교급에서 최종학년을 제외한 것은 졸업을 앞두고 진학 준비에 바빠 뿐만 아니라, 입시를 위한 집중적인 공부를 하기 때문에 정상적인 학력을 나타낸다고 보기 어렵기 때문이다. 또, 고등학교에서 1학년과 2학년을 선정한 것은 우리 나라에서는 고등학교 1학년까지는 대학 진학과 관련된 선택 과정을 이수하지 않고, 모든 학생이 같은 과학 내용을 학습하기 때문이다. 고등학교 자연계 학생들을 평가 대상으로 표집한 것은 이 학생들이 이공계 대학으로 진학할 가능성이 높기 때문이다. 이들의 과학학력은 우리 나라 과학기술 인력의 바탕이 얼마나 튼튼한지를 보여주는 자료가 된다. 국제수학과과학력평가연구(TIMSS)

**Table 5** Sampling for main test

School level	Region			Sex		Total
	Metropolitan	City	Rural	M	F	
Elementary	1,139	816	481	1,269	1,167	2436
Middle	1,193	790	313	1,273	1,023	2296
High	1,672	1,447	1,015	2,281	1,753	4034
Total	4,004	3,053	1809	4,823	3,943	8766

에서도 초등학교와 중학교 학생은 물론이고, 고등학교 최종학년의 자연계 또는 과학 선택 학생들의 학력을 평가한다.

### 2) 평가 주기

국가 수준에서 과학 학력 평가를 실시하는 주기는 국가에 따라 매년(APU)에서부터 3년 간격(NAEP) 등 다양하다. 문제는 의미 있는 자료의 수집을 위해서 어느 정도의 주기가 바람직한가를 결정하는 것이다. 우리나라는 국가 교육과정이 제정되어 있고, 학습기회가 비교적 균질한 점으로 미루어 볼 때에는 상당한 간격으로 국가 수준 평가를 실시해도 무방할 수 있다. 그러나 근래에 들어서서 교육과정이 자주 바뀌고, 새로운 교수학습 평가 방법의 도입이 활발하며, 표집과 자료 수집이 비교적 용이한 점으로 볼 때 자주 실시하는 것이 가능하다. 이러한 점을 고려하여 평가 주기를 1년으로 결정하였다.

### 3) 평가틀

과학 학력평가를 보다 의미있게 실시하기 위해서는 과학 교육 평가틀(framework)이 구성되어야 한다. 국가 수준의 대규모 과학 학력평가에서 활용되는 평가틀은 각국의 관심과 목적에 따라 탐구 중심이나 내용 중심, 또는 내용과 과정의 통합 중심 등으로 다양하다. 우리나라의 경우에는 국가 교육과정이 과학 교육의 내용과 방법을 비교적 명확하게 제시하고 있으므로, 평가 영역을 국가 교육과정의 목적에 준하여 구성하는 것이 바람직한 것으로 판단된다. 자연과와 과학과 교과 목표에는 과학의 탐구와, 지식, 태도 및 과학과 기술 및 사회와의 관련성을 강조하고 있다. 이에 준하여 국가 수준 과학학력평가틀을 과학 지식, 과학 탐구, 과학 태도로 크게 구분하였다. STS와 관련된 내용은 과학태도 영역과 과학 지식의 사회적 상황에 포함시켰다.

과학 지식 평가틀은 지식수준, 행동수준, 상황의 3차원으로 구성하였다(Fig. 1). 평가틀을 이와 같이 3차원으로 설정한 것은 과학교육의 목표를 되도록 포괄적으로 반영하기 위함이다. 지식차원은 사실, 개념, 법칙·원리, 개념체계로 하위 요소를 세분화하였고, 행동수준은 지식의 기억, 이해, 적용 수준으로 상세화하였다. 과학 지식의 상황은 순수과학적 상황, 일상적 상황, 기술산업적 상황, 사회적 상황 요소로 구성되어 있다. 지식 영역은 과학의 학문적 구조의 위계를 반영하여 가장 초보적인 사실에서 개념체계까지 위계적으로 나열하였

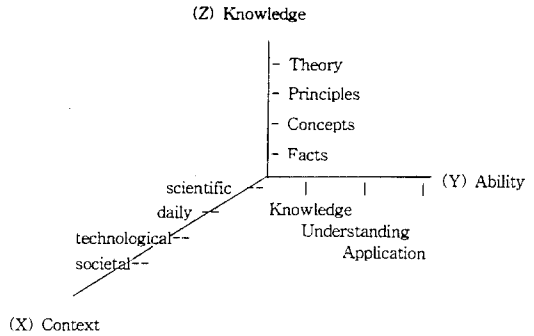


Fig. 1 Science knowledge assessment framework

다. 그러나, 평가 문항이 각 위계별로 골고루 분포해야 하는 것은 아니라고 본다. 특히, 개념체계는 한 개의 문항으로 평가하기는 어려운 수준이라고 생각한다. 상황 영역은 순수과학적 상황에 너무 치우치지 않도록 일상적 상황과 기술, 사회적인 상황을 설정하였다. 이것은 최근 과학교육에서 중요한 목표로 추구하고 있는 STS적인 측면을 반영하기 위해서 설정하였다.

### 4) 기 타

국가 수준 과학 지식 평가는 대상 학년의 교육과정이 마무리되는 2월 중에 실시하기로 하였다. 이는 그 학년까지 학습한 내용을 평가할 수 있다는 점과 그 때가 비교적 학습 부담이 적기 때문이다. 표집 방법은 전체 모집단의 구성을 가장 잘 반영할 수 있도록 다단계 유층화 표집 방법을 사용하고, 유층 구분의 요인으로는 지역과 성별을 고려하였다. 지역과 성별 단위 내에서 학교를 무선 표집하고, 학교 내에서 학급을 무선 표집하며, 표집된 학급은 학급 전체를 평가에 참여시킨다. 표집 비율은 전체 모집단의 1% 내외로 하였다.

각 영역에 대한 의미있는 결과를 얻기 위하여 각 모집단별로 4가지의 검사지를 구성하도록 하였다. 각 검사지는 미리 준비된 문제 은행에서 난이도와 변별도 등과 같은 자료를 참조하여 구성한다. 각 검사지마다 가교 문항을 포함시켜 전체 문항에 대한 성취도를 비교하는 근거로 삼는다. 또한 초등학교에서 고등학교 자연계까지 거의 동일한 가교 문항을 포함시켜 횡단적 비교를 시도한다. 이러한 가교 문항을 5개씩으로 할 때 전체 문항 수는 초등학교 65문항이 되고 중·고등학교는 80문항이 된다. 이 가교 문항들은 이후에 실시될 국가 수



준 학력평가에서 다시 포함되어 중단적인 경향을 판단할 수 있는 기초 자료로 삼는다. 한 검사지별 평가 소요 시간은 초등학교 5학년 40분, 중학교 2학년 45분, 고등학교 50분으로 한다. 이는 1차시 수업시간과 동일한 것으로 시험 실시 조건을 최대한 평소와 동일하게 하려는 시도이다. 평가 문항은 객관식 5지선다형을 원칙으로 한다. 이는 대규모 국가수준 평가에서 흔히 요구되는 채점의 신뢰성과 용이성을 반영하면서 추측가능성을 줄이려는 것이다.

## 2. 문항의 영역별 분포 및 문항의 특성

개발된 최종 문항들의 하위 영역별 분포와 특성은 다음과 같다.

### 1) 문항의 하위 영역별 분포

예비검사 결과를 바탕으로 최종적으로 문제 은행에 수록할 검사용 문항은 초등학생용 80문항, 중학생용 100문항, 고등학교 공통과정용 100문항, 고등학교 자연계열용 100문항이며, 이들 문항의 하위 영역 분포 문항 수는 Table 6과 같다.

초등학생용 문항 분포를 보면, 지식 수준에서는 개념 문항이 50%로 가장 많고, 사실 문항과 법칙 문항이 각각 19%, 20%이다. 행동 수준에서는 기억 문항이 53.8%로 가장 많고, 이해 문항이 37.5%이다. 상황에서는 과학적 상황 문항이 57%로 가장 많고, 일상 상황 문항

이 28.8%이다.

중학생용 문항 분포를 보면, 지식 수준에서는 개념 문항이 44%로 가장 많고, 법칙 원리 문항이 32%이었고, 사실 문항이 10%로 초등학생용에 비해 법칙 원리 문항 비율이 더 높고 사실 문항 비율이 낮다. 행동 수준에서는 이해 문항이 66%로 가장 많고, 기억 문항이 22%이다. 상황에서는 과학적 상황 문항이 79%로 가장 많고, 일상 상황 문항이 21%이다.

고등학생 공통과정용 문항 분포를 보면, 지식 수준에서는 법칙 원리 문항이 37%로 가장 많고, 개념 문항이 33%이었고, 개념 체계 문항이 20%로 초등학생용과 중학생용에 비해 법칙 원리 문항과 개념 체계 문항 비율이 높다. 행동 수준에서는 이해 문항이 63%로 가장 많고, 적용 문항이 16%이다. 초등학생과 중학생용에 비해 적용 문항 비율이 높다. 상황에서는 과학적 상황 문항이 73%로 가장 많고, 일상 상황 문항이 14%, 기술산업 상황 문항이 6%, 사회 상황 문항이 7%로, 상황이 다양하게 구성되었다.

고등학생 자연계열용 문항 분포를 보면, 지식 수준에서는 법칙 원리 문항이 42%로 가장 많고, 개념 문항이 34%이었고, 사실 문항과 개념 체계 문항이 각각 16%, 14%이다. 타 문항지와 비교해 볼 때, 법칙 원리 문항 비율이 가장 높다. 행동 수준에서는 이해 문항이 58%로 가장 많고, 적용 문항이 33%이고, 기억 문항이 15%이다. 타 문항지에 비해 적용 문항 비율이 가장 높고, 기억 문항 비율이 가장 낮다. 상황에서는 과학적 상황 문항이

**Table 6** Content by ability by context distribution of items developed

Knowledge	Ability	Elementary	Middle	High(1st)	High(Science)
Fact	Knowledge	14/5/0/0*	5/0/0/0	6/0/1/1	8/0/0/0
	Understanding	0/1/0/0	4/1/0/0	1/0/1/0	2/0/0/0
	Application	0/0/0/0	0/0/0/0	0/0/0/0	0/0/0/0
Concept	Knowledge	15/3/0/0	8/2/0/0	10/0/0/0	4/0/0/0
	Understanding	11/6/0/0	25/4/0/0	16/5/1/0	20/1/0/0
	Application	2/3/0/0	2/3/0/0	0/0/1/0	8/1/0/0
Principle	Knowledge	5/1/0/0	7/0/0/0	3/0/0/0	3/0/0/0
	Understanding	7/5/0/0	14/7/0/0	18/5/0/2	23/2/0/0
	Application	0/2/0/0	3/1/0/0	5/2/2/0	13/1/0/0
Theory	Knowledge	1/0/0/0	0/0/0/0	0/0/0/0	0/0/0/0
	Understanding	0/0/0/0	10/1/0/0	12/1/0/1	4/0/0/0
	Application	0/0/0/0	1/2/0/0	2/1/0/3	6/4/0/0

\*context (scientific /daily /technological /societal)

91%로 가장 많고, 일상 상황 문항이 9%이다.

2) 문항의 하위 영역별 정답률

개발한 문항의 하위 영역별 정답률은 Table 7에 나타나 있다. '지식 수준'의 하위 영역별 정답률을 비교해 보면, 초등학생용의 경우, '법칙 원리' 영역의 정답률이 63.19%로 가장 높고, '사실' 영역과 '개념' 영역은 각각 55.28%와 54.87%이다. 그리고 '개념체계' 영역의 정답률이 28.85%로 가장 낮게 나왔다. 다른 학교 급간에 비해 초등학생용 문항의 '개념체계' 영역의 정답률이 낮다. 중학생용의 정답률을 보면, '사실' 영역의 정답률이 가장 높고, 그 다음이 '개념' 영역이다. 고등학교 공통과정용의 '지식' 수준의 하위 영역의 정답률은 40%에서 46% 사이로 거의 비슷하게 나왔다. 고등학교 자연계열용의 '지식' 수준의 하위 영역별 정답률을 비교해 보면, '사실'과 '개념체계' 영역의 정답률이 52%~53%로 높게 나왔고, '개념' 영역은 40.91% '법칙 원리' 영역은 33.3%의 정답률을 보인다.

'행동' 수준의 하위 영역별로 정답률을 비교해 보면, 초등학교용의 경우, '기억'과 '이해' 영역은 정답률이 58% 정도로 높게 나왔으나, '적용' 영역의 정답률은 그보다 낮은 46% 정도 나왔다. 중학생용의 경우도 초등학교용의 정답률의 분포와 거의 비슷하였다. 고등학교 공통과정용의 경우는 '행동'수준의 하위 영역별 정답률의 큰 차이는 나지 않았다. 고등학교 자연계열용의 경우 '기억'과 '적용' 영역에 비해 '이해' 영역의 정답률이 낮게 나왔다

'상황' 수준의 하위 영역별 정답률은 학교급 내에서 차이가 많이 나지 않았다. 고등학교 공통과정용 외에는 '기술산업' 영역과 '사회' 영역의 문항을 개발하지 않았기 때문에 학교급간 차이를 비교할 수 없었다. 고등학교 공통과정용의 경우 '일상' 영역의 정답률이 다른 요소에 비해 비교적 높았다.

3. 문항 및 검사지 특성 분석

본 연구에서 개발된 평가 문항은 절대평가의 의미를 지니고 있다. 즉, 국가 수준에서 우리 나라 초·중·고 학생들의 성취도가 어느 정도인지를 알아보기 위한 평가 도구로 개발된 것이다. 문항 반응이론에 입각한 문항의 난이도와 변별도는 문항의 양호도를 판단하는데 도움을 줄뿐만 아니라 특정 문항을 해결하는 데 필요한 능력이 어느 정도인지를 추정할 수 있게 해준다.

각 문항의 난이도, 변별도, 그리고 추측도를 문항반응이론에 입각하여 구하였다. 문항 분석의 결과는 추후 본 평가 문항을 이용할 때, 각 문항의 고유 특성을 파악하고 필요한 요인(과학지식 수준, 행동수준, 평가목표, 난이도, 변별도 등)에 따라 조합에 의해 다양한 검사지를 구성할 수 있도록 하는데 훌륭한 자료로 활용될 것이다. 또한 이러한 자료는 추후 국가 수준의 과학 학력 평가 체제 안에서 구축하게 될 문제은행(item bank)에 데이터베이스화할 자료가 된다.

Table 7 Average percent correct by school level and domain

		School level			
		Elementary	Middle	High(1st)	High(Sci)
Content	Fact	55.28	52.62	40.28	52.96
	Concept	54.87	48.14	44.84	40.91
	Principle	63.19	39.10	46.31	33.30
	Theory	28.85	43.76	43.30	52.34
Ability	Knowledge	58.49	49.29	46.40	46.71
	Understanding	58.74	46.74	43.76	38.51
	Application	45.86	35.27	49.00	44.23
Context	Scientific	56.28	46.09	45.20	40.03
	Daily	57.75	42.94	50.89	44.69
	Technological	.	.	40.40	.
	Societal	.	.	40.39	.

1) 문항의 난이도

개발된 과학 지식 평가 문항의 난이도 분포는 Table 8에 제시되어 있다. 초등학교용 80문항 중에서 해결에 낮은 능력을 요구하는 문항(24)과 중간 정도의 능력을 요구하는 문항(21)이 가장 많았으나, 높은 능력을 요구하는 문항(17)과 매우 높은 능력을 요구하는 문항(15)도 적지 않았다. 중학교용 100 문항 중에서 높은 능력을 요구하는 문항이 44문항으로 절반에 가까웠으며, 그 다음으로는 보통 수준을 요구하는 것이 26문항, 매우 높은 능력을 요구하는 것이 23 문항의 순이었다. 고등학교 공통 과정 문항 100개 중에서 높은 능력을 요구하는 문항이 46개, 매우 높은 능력을 요구하는 문항이 28개, 보통 수준인 문항이 16개였다. 고등학교 자연계 문항들은 매우 높은 수준의 문항이 44개로 가장 많았고, 높은 능력을 요구하는 문항이 33개, 보통인 문항이 15개 순이었다. 대체로 볼 때 학년이 올라갈수록 문항의 난이도가 높아지는 경향을 보였다. 같은 학년을 대상으로 하는 여러 검사지 사이의 난이도는 비교적 유사하게 나타났다.

2) 문항의 변별도

개발된 문항의 변별도에 대한 자료가 Table 9에 제시되어 있다. 개발한 문항들의 변별도는 0.39~0.79에 집중되어 있으며, 이는 변별도가 적절한 수준임을 보여준다. 그 다음으로는 0.80~0.99의 변별도를 가지는 문항들이 많았으며, 이 범위는 변별도가 높음을 의미하는 것이다. 따라서 개발된 문항들은 학생들의 성취도를 구별해 내는데 매우 뛰어난 기능을 가지고 있음을 알 수 있다. 이러한 경향은 학교급별로 큰 차이를 보이지 않았다. 또한 동일한 학년을 대상으로 개발된 여러 평가지 사이에도 거의 차이가 없는 것으로 나타났다.

3) 검사의 신뢰도

검사의 신뢰도는 K-R 21 값을 구하여 조사하였으며, 0.484에서 0.730의 분포를 보인다 (Table 10). 16개 검사지 중에서 14개 검사지는 신뢰도가 0.600보다 높으며 2개 검사지만 그 보다 낮은 값을 보였다. 신뢰도가 낮은 두 검사지는 모두 고등학교 자연계용으로 난이도가 매

**Table 8** The distribution of item difficulties of science knowledge items developed

Item difficulties	NKE					NKM					NKC					NKH				
	1	2	3	4	tt	1	2	3	4	tt	1	2	3	4	tt	1	2	3	4	tt
-2.0<b	·	3	·	·	3	·	·	·	·	·	·	·	·	·	·	·	·	1	·	1
-2.0~-0.5	5	7	5	7	24	1	3	1	2	7	3	3	3	1	10	·	1	5	1	7
-0.5~+0.5	6	3	7	5	21	7	5	6	8	26	4	4	5	3	16	·	5	6	4	15
+0.5~2.0	4	4	5	4	17	11	8	14	11	44	13	12	11	10	46	6	12	9	6	33
+2.0≤		5	3	3	4	15	6	9	4	4	23	5	6	6	11	28	19	7	4	44
Total	20	20	20	20	80	25	25	25	25	100	25	25	25	25	100	25	25	25	25	100

**Table 9** The distribution of discrimination index of science knowledge items developed

Item discrimination index	NKE				NKM				NKC				NKH			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
0.0	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
0.01~0.19	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
0.20~0.38	·	2	·	·	·	·	·	·	·	·	·	·	·	·	·	·
0.39~0.79	14	17	19	16	12	10	19	21	11	24	23	11	18	18	20	23
0.80~0.99	6	1	1	4	10	14	6	4	13	1	2	12	7	7	5	2
1.00	·	·	·	·	3	1	·	·	1	·	·	2	·	·	·	·
∞	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
Total	20	20	20	20	25	25	25	25	25	25	25	25	25	25	25	25

**Table 10** The reliability of science knowledge achievement tests

Reliability	NKE				NKM				NKC				NKH			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
K-R 21	.665	.670	.605	.666	.676	.665	.730	.702	.715	.643	.684	.616	.484	.690	.704	.515

**Table 11** Estimated measurable abilities from test information curve

Test		$\theta$
NKE	1	-1.2~ -0.2
	2	-1.0~ 0.5
	3	-1.2~ 0.2
	4	-1.0~ 0.0
NKM	1	0.2~ 0.7
	2	1.0~ 2.0
	3	0.0~ 1.0
	4	1.0~ 1.0
NKC	1	0.0~ 1.0
	2	0.0~ 1.3
	3	0.2~ 2.0
	4	1.0~ 2.0
NKH	1	1.2~ 1.8
	2	0.0~ 1.3
	3	-0.2~ 1.5
	4	1.5~ 2.0

우 높은 어려운 문항이 많이 포함되어 있다는 공통점을 보인다. 앞으로 실시할 국가수준 과학지식 평가에서는 이러한 문항들을 바탕으로 새로운 검사지를 구성할 예정이며, 이 때에는 난이도의 분포를 조정하게 되므로 신뢰도 지수는 상향 조정될 것으로 기대된다.

문항반응이론에서는 검사정보함수(test information function)를 이용해서 검사의 신뢰도를 나타내기도 한다. 검사정보함수는 특정 검사가 어떤 능력을 가진 학생들에게서 가장 많은 정보를 얻어낼 수 있는지를 알려준다. 최대 정보를 얻어낼 수 있는 학생들의 능력을 추정한 값이 Table 11에 제시되어 있다. 국가 수준 지식 평가를 위해서는 중간 정도의 학생들을 포함하여 비교적 넓은 범위의 능력을 가진 학생들에 대한 정보를 얻어 낼 수 있어야 할 것이다. 초등학교 검사들은 비교적 낮은 능력에서부터 중간 능력 정도의 학생들에게 적합한 것으로 나타났으며, 중학교와 고등학교 공통과정은

중간 능력에서부터 높은 능력의 학생들에게 적합한 것으로 나타났다. 자연계 학생용 검사지는 능력이 매우 우수한 학생들에게 적합한 경우가 4개 중에서 2개에 달하였다. 앞으로 실시할 국가수준 지식 평가를 위한 검사지 구성에서는 문항의 난이도 지수를 조절하여 검사정보함수의 범위를 넓혀서 다양한 능력을 가진 학생들을 평가할 수 있도록 구성하여야 할 것이다.

#### IV. 결론 및 제언

본 연구는 우리 나라 초, 중, 고등학교 학생들의 장기적 과학학력 점검을 위한 국가 수준의 평가 체제 개발을 위한 연구의 일환으로 과학 지식 평가 체제를 개발하는데 그 목적이 있었다. 이를 위해서 국가 수준 과학지식 평가 체제의 방향과 주요 요소를 결정하고, 검사지 구성을 위한 문제 은행용 문항들을 개발하였으며, 이 문항들의 특성을 조사 분석하였다.

##### 1. 결 론

본 연구의 결론은 다음과 같다.

첫째, 우리 나라 국가 수준 과학 지식 평가는 학교급 별 주요 성취의 평가에 초점을 맞추어서 매년 주요 학교급 한 학년을 대상으로 학년말에 실시하는 것이 바람직하다. 다만 고등학교 수준에서는 공통과정과 자연계를 구분하여, 과학적 소양을 중심으로 한 과학 학력의 점검과 과학기술인력의 준비도를 평가할 수 있도록 한다. 평가를 위한 표집은 단단계 유층화 표집 방법을 사용하며, 주요 고려 변인은 지역과 성별로 한다. 검사지는 복수형으로 개발하며 가교문항을 활용하여 검사시간 학교급간, 종단적 비교의 기초로 삼는다.

둘째, 본 연구에서는 국가 수준 과학지식 평가 체제를 위하여 각 모집단 별로 80문항 내지 100문항씩 총 380문항을 개발하였다. 문항은 지식, 행동, 상황으로 구분되고 지식은 다시 사실, 개념, 법칙, 이론으로 구별하였으며, 행동은 블록의 분류를 원용하여 기억, 이해, 적용으로 구분하였으며, 상황은 과학적 상황, 일상적 상황, 기

술산업적 상황, 사회적 상황으로 구분하였다. 문항은 이들 항목에 균형있게 배분되도록 하였다.

셋째, 본 연구에서 개발된 문항과 검사의 특성을 분석하여 본 결과 초등학교용 문항들은 중간 또는 중간 이하의 능력을 가진 학생들에게 적합하였으며, 중학교용과 고등학교 1학년 용 문항들은 중간 또는 중간 이상의 능력을 가진 학생들에게 적합하였다. 고등학교 자연계용 문항들은 비교적 높은 능력을 가진 학생들에게 적합하였다. 이는 문항 자체의 난이도 차에 기인할 수도 있지만, 우리 나라 과학 교육과정의 내용 특성이 반영된 것일 수도 있다. 개발한 문항들의 변별도는 우수하거나 매우 우수한 것으로 나타나서, 학생들의 과학 지식 성취를 정확하게 평가하는 데 적합한 것으로 보인다.

## 2. 제 언

본 연구는 짧은 기간 동안이지만, 국내에서는 지금까지의 어떤 연구와 개발보다 더 많은 인력과 시간이 투입되었으며, 세미나와 워크숍 등을 포함한 체계적인 계획과 교육을 바탕으로 개발되었다. 따라서, 본 평가 도구가 완벽한 것이라고는 할 수 없지만 지금까지 개발된 어떤 도구보다도 그 타당성이 높다는 것은 의심의 여지가 없다고 본다. 이러한 연구의 결과가 결코 사장되어서는 안되며, 장기적인 운영을 통하여 우리 나라 과학교육의 질적 향상에 기여해야 할 것으로 사료된다. 이를 위해서 다음과 같은 몇 가지 제언을 하고자 한다.

첫째, 국가적인 인식의 전환이 필요하다. 앞서서도 지적한 것과 같이 본 연구 결과의 중요성을 행정 당국이 인식하여 본 연구 결과를 적극적으로 활용해야 한다. 국가 수준 과학교육 평가체제는 특성상 장기적인 계획과 지속적인 연구가 필요하며, 시행과정에서 행정적인 뒷받침이 필요하다. 따라서 본 평가 체제를 효율적으로 일관성있게 운영할 기관의 설립이나 지정이 반드시 필요하다.

둘째, 국가수준 과학 지식 평가를 위한 지속적인 연구와 개선이 이루어져야 하며, 새로운 양질의 문항을 개발하여 문제은행을 구성하고 확대하여 나가야 한다. 본 연구가 상당한 노력의 산물이라는 하나, 짧은 연구 기간을 감안할 때, 연구자들이 발견하지 못한 문제가 추후에 발견될 가능성을 배제할 수는 없다. 따라서, 본 평가 체계에 대한 계속적인 분석과 개선이 이루어지는 것이 바람직할 것이다.

셋째, 과학 학습에 영향을 미치는 요인들을 정확하게

파악할 수 있는 방법이 보완되어야 한다. 이를 위해서는 우리 나라 과학교육체제 모형을 개발하고, 이에 대한 정보를 수집할 수 있는 성취도 검사지, 국가, 지역, 학교, 교사, 교실, 학생, 가정 수준의 배경과 학습기회를 파악할 수 있는 검사지를 개발하여 입체적인 정보를 수집하여야 한다. 또한 이러한 결과를 의미 있게 분석할 수 있는 통계적 모형과 분석기법이 도입되어야 한다.

넷째, 장기적인 평가 결과는 교육 정책 입안자와 과학 교육 전문가, 그리고 전국민에게 널리 알려야 한다. 본 연구로 개발된 평가 체제를 적용하여 평가한 결과는 국가적으로 매우 중요한 의미를 갖는다. 따라서, 본 연구 결과를 공개하고, 교육계는 물론 일반 국민들에게 적극적으로 홍보함으로써 국민들이 과학 교육에 관심을 갖고, 잘못된 점을 시정하고 개선하기 위해서 지원을 할 수 있는 풍토를 조성해야 할 것으로 본다. 이를 위해서는 국가수준지식 평가 결과 보고서와 기술보고서(technical report) 이외에도 결과를 알아보기 쉽게 요약한 소책자의 발행이 반드시 필요하다.

## 적 요

과학교육의 정책 수립과 평가 및 과학교육과정의 개편과 효율성 점검 등은 체계적이고 과학적인 연구 결과를 바탕으로 해야 한다. 이를 위해서 주요 선진국에서는 국가 수준 과학 학력평가를 지속적으로 실시하고 발전시켜 왔다. 우리 나라에서 그 동안 실시된 국가 수준 과학 학력평가는 많은 문제점을 가지고 있는 것으로 지적되어 왔으며, 과학 교육 정책 수립 및 평가 그리고 교육 과정 개편을 위한 자료 제공에 성공하지 못하였다. 이 연구의 주요 목적은 우리 나라 초·중·고 학생들의 과학 지식 성취의 경향성을 파악하고 향후 주기적으로 진행되어질 국가적 차원의 평가 지표를 만들기 위한 국가 수준의 과학 지식 평가 체제를 개발하는 것이다.

과학교육전문가와 초·중·고 교사들로 연구팀을 구성하고, 교육평가 전문가들과의 협조와 자문을 통해서 국가 수준 과학지식 평가 체계의 방향과 주요 요소를 결정하였다. 또한 문항 개발을 위한 세미나와 워크숍을 통해서 평가문항개발 능력을 향상시켰으며, 계획한 문항을 개발하고, 수정하였으며, 초·중·고교생 958명을 대상으로 예비검사를 실시하고 그 결과를 바탕으로 최종 문항을 380개를 선정하고 수정하였다. 최종 선정한 문항은 다시 초·중·고 학생 8766명을 대상으로 본 검사를 실시하여 문항반응이론(IRT)을 이용하여 문항의

특성을 분석하였다.

연구 결과 국가 수준 과학 지식 평가는 매년 초등학교 5학년, 중학교 2학년, 고등학교 1학년과 2학년을 대상으로 학년말에 실시하는 것이 바람직하다. 표집은 다단계 유층화 표집 방법을 사용하며, 검사지는 복수형으로 개발하며 가교문항을 활용하여 검사시간 학교급간, 종단적 비교의 기초로 삼는다. 본 연구에서는 국가 수준 과학 지식 평가 체제를 위하여 각 모집단 별로 문제 은행에 포함시킬 80 내지 100문항씩 총 380문항을 개발하였다. 개발된 문항과 검사의 특성을 분석하여 본 결과 초등학교용 문항들은 중간 또는 중간 이하의 능력을 가진 학생들에게, 중학교용과 고등학교 1학년 용 문항들은 중간 또는 중간 이상의 능력을 가진 학생들에게, 고등학교 자연계용 문항들은 비교적 높은 능력을 가진 학생들에게 적합하였다. 개발한 문항들의 변별도는 우수하거나 매우 우수한 것으로 나타나서, 학생들의 과학 지식 성취를 정확하게 평가하는 데 적합한 것으로 보인다.

### 참고 문헌

고종렬, 이혜선, 박종삼 (1969). 국민학교 기초학력조사. 교육과학, 제36호. 중앙교육연구소.

교육부 (1997). 교육통계 연보. 서울: 교육부

국립교육평가원 (1992a). 국민학교 3학년 학업성취도 평가 연구. 연구보고서 93-6.

국립교육평가원 (1992b). 중학교 2학년 학업성취도 평가 연구. 연구보고서 93-8.

국립교육평가원 (1992c). 고등학교 1학년 학업성취도 평가 연구. 연구보고서 93-9.

국립교육평가원 (1993). 중학교 2학년 학업성취도 평가 연구. 서울: 국립교육평가원

국립교육평가원 (1995). 전국 국민학교 학업성취도 평가 연구. 연구보고서 95-11. 서울: 국립교육평가원

국립교육평가원 (1997). 전국 초등학교 학업성취도 평가 연구. 연구보고서 97-16. 서울: 국립교육평가원.

권낙원, 이재분, 구자역 (1984). 국민학교 교육과정 평가 연구(II). - '83, '84년 간의 학생들의 학업성취도 비교- 한국교육개발원 연구보고 RR84-14. 서울: 한국교육개발원.

권낙원, 이재분, 구자역 (1985). 국민학교 교육과정 평가 연구(IV). - '83, '84, '85년 간의 학생들의 학업성취도 비교- 한국교육개발원 연구보고 RR85-7. 서울: 한국교육개발원.

김명숙, 강태중, 양미경, 백순근 (1994). 초·중등학교 교육 성취 조사 연구(II)-조사 설계의 구체화 및 과학 영역의 예비 문항 개발, 한국교육개발원 연구보고 RR94-4. 서울: 한국교육개발원.

김형립 (1980). 새교육체제 제5차 종합시험 연구보고서. 서울: 한국교육개발원.

박종삼 (1972). 초등교육의 질을 향상시키기 위한 종합 학력조사. 중앙교육연구소 연구보고서.

성태제 (1991). 문항반응이론 입문. 서울: 양서원.

신세호 (1987). 교육의 질 관리와 학력평가의 중요성. 제4회 전국교육평가세미나, 교육의 질 관리와 전국 학력평가의 의의 기초 강연, 1987. 10. 16, 중앙교육평가원.

이종재 (1978). 한국 초·중학교 학생의 특성과 학업 성취 수준. 한국교육개발원 연구보고 제 67집. 서울: 한국교육개발원.

장석우 (1987). 국내 학력평가의 반성과 활용 방안. 제4회 전국교육평가세미나, 교육의 질 관리와 전국 학력평가의 의의 기초 강연, 1987. 10. 16, 중앙교육평가원.

전용신 (1963). 국민학교 5~6학년 학생의 학력검사 결과의 비교(1959년도 대 1963년도). 중앙교육연구소 조사 연구 제 27집. 서울: 중앙교육연구소.

중앙교육평가원 (1986). 일반계 고등학교 전국학력평가 문항개발. 서울: 중앙교육평가원.

중앙교육평가원 (1988). 중학교 학력평가 연구. 서울: 중앙교육평가원.

한국행동과학연구소 (1980a). 한국 국민학교 교육의 평가(I): 학력 평가. 한국행동과학연구소.

한국행동과학연구소 (1980b). 한국 국민학교 교육의 평가(II): 아동과 학습 환경. 한국행동과학연구소.

APU (1984a). *Science Assessment Framework, age 11: Science Report for Teachers*, No. 4, DES, DENI and WOED.

APU (1984b). *Science Assessment Framework, age 13 & 15 : Science Report for Teachers*, No. 2 DES, DENI and WOED.

Assessment Systems Corporation (1996). *User's Manual for the ASC Item and Test Analysis Package*. The Author.

Broadfoot, P., & Gipps, C. (1996). *Assessment developments in England and Wales: the triumph of tradition*. A. Little, & A. Wolf. (Eds.) *Assessment in*

- transition: Learning, monitoring and selection in international perspective.* Oxford: Pergamon, pp 134-153.
- Kim, C.-J. (1996). External assessment in Korea. In A. Little, & A. Wolf. (Eds.) *Assessment in Transition: Learning, Monitoring and Selection in International Perspective.* Oxford: Pergamon, pp. 204-218.
- Madaus, G. F., & Raczek, A. E. (1996). A turning point for assessment: reform movements in the United States. In A. Little, & A. Wolf. (Eds.) *Assessment in transition: Learning, monitoring and selection in international perspective,* Oxford: Pergamon, pp. 101-117.
- National Assessment of Educational Progress (1984). *Science objectives,* Fifth National Assessment, 3rd Draft.
- Somerset, A. (1996). Examinations and educational quality. A. Little, & A. Wolf. (Eds.) *Assessment in transition: Learning, monitoring and selection in international perspective,* Oxford: Pergamon, pp. 263-284.