

# 국가 수준의 과학탐구능력 평가체제 개발<sup>1)</sup>

우종옥 · 김범기 · 한안진<sup>1</sup> · 허 명<sup>2</sup>

(한국교원대학교) · <sup>1</sup>(인천교육대학교) · <sup>2</sup>(이화여자대학교)

## Development of National Assessment System: Scientific Inquiry Domain

Jong-Ok Woo · Beom-Ki Kim · Ann-Chin Hann<sup>1</sup> · Myung Hur<sup>2</sup>

(Korea National University of Education) · <sup>1</sup>(Incheon National  
University of Education) · <sup>2</sup>(Ewha Womans University)

### ABSTRACT

Inquiry approach in teaching science has been widespread in Korea since the Third National Curriculum in 1970's. But the effect of the approach has never been evaluated systematically in Korea, so the science educators do not know if the inquiry approach really works and if any critical problems inhibit the effect of the approach. In this context, the NAEP of the United States and the APU of the United Kingdom were model programs for assessing the effect of the inquiry approach in teaching science.

The purpose of this study is to develop an assessment system for evaluating the effect of inquiry teaching in elementary and secondary schools. For this purpose, an assessment framework and 240 test items were developed and tried with a sample of 8,906 students. The results say that the developed tests are reliable.

The average Cronbach  $\alpha$  reliability coefficient of the tests was 0.69.

The discrimination index(point-biserial correlation coefficient) ranged from 0.39 to 0.54 with a mean of 0.49, which indicate they are excellent in discriminating students in terms of their inquiry achievement. The test items were also analyzed by "item response theory." The results also say that the items are successfully developed.

**Key words** : inquiry, assessment, science education, elementary, secondary.

### I. 연구의 필요성 및 목적

우리 나라의 경우 학생들의 탐구 능력을 배양하기 위한 탐구 학습의 도입은 이미 70년대 초 제3차 교육과정 때 받아들여졌다. 즉, 기계적 암기 학습(rota learning)

으로 부터의 탈피를 강조하면서, "자연을 과학적으로 탐구하는 능력을 신장시키고, 문제 해결에 이를 활용하게"하는 것을 핵심적인 교육목표로 제시해 오고 있다. 제 4차 및 5차 교육과정에서도 탐구능력의 배양을 과학 교육의 주요 목표로 설정하였고, 특히 1996학년도부터

\*1998년 8월 19일 받음

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

적용하고 있는 6차 과학과 교육과정에서는 내용 체계를 지식 영역과 탐구 영역으로 구분하여 탐구 능력을 더욱 강조하고 있다. 한편, 1994 학년도 대학입시에서부터 실시되기 시작한 대학수학능력 시험에서도 학생들의 탐구 능력을 특별히 강조하였고, 2000년대 초에 적용되는 제 7차 과학교육과정에서도 탐구 능력은 여전히 과학교육의 주요한 학습 목표로 제시되고 있다.

그동안 교육사조의 변화와 교육과정의 변화, 대학입시 제도의 변화 등 과학교육의 질적 개선노력이 끊임없이 요구되어 왔다. 그러나, 우리는 과학교육의 질적 개선에 따라 학생들의 과학 탐구능력이 과거에 비하여 향상되었는지, 과학에 대한 관심과 호기심이 향상되었는지, 학교 과학교육과정은 의도한 방향으로 운영되고 있는지 등 우리 나라 학교 과학교육에 대한 많은 의문에 대하여 답변할 자료를 축적하지 못하고 있다. 이는 그간 우리 과학교육에 대한 깊은 반성과 함께 과학교육의 질적 개선을 위한 국가 수준의 평가 연구가 시작되어야 함을 말해주고 있다. 이러한 점에서 미국의 NAEP (National Assessment of Educational Progress)나 영국의 APU (Assessment of Performance Unit)와 NCA(National Curriculum Assessment) 가 우리에게 주는 시사점은 매우 크다고 할 수 있다. 미국의 경우, 1983년 “국가의 위기(A Nation at Risk)”라는 보고서에서 교육 개혁의 필요성과 방향을 천명했는데, 그 보고서의 근거는 바로 미국의 국가수준평가 즉, NAEP의 연구 결과이었다. 즉, 당시의 교육개혁 필요성과 방향은 1969년 이후 10여년 이상 미국 교육체제를 면밀히 분석하고 조사한 평가 결과에 바탕을 두고 설정된 것이었다. 이에 비하여 우리 나라의 교육개혁은 어떠한 평가 결과나 정보에 근거를 둔 것이 아니라 소수의 정책 입안자나 위탁 연구를 수행한 몇몇 전문가들에 의하여 개혁의 필요성과 방향이 결정되었다고 해도 과언이 아니다.

국내에서도 학교 과학교육의 성과를 평가하려는 연구가 전혀 이루어지지 않은 것은 아니지만(전용신, 1963; 고종렬 외, 1969; 김형립 외, 1980; 권낙원 외, 1985; 한국행동과학연구소, 1980; 국립교육평가원, 1992a, 1992b, 1992c), 이들 연구는 대부분 일회적이고 산발성 조사 연구로 끝났거나, 정치적 이유로 연구 결과가 공개되지 못했던 것이 사실이다. 따라서 객관적이고 타당한 설계와 방법으로 학교 과학교육의 성취 수준을 진단하고 평가하기 위한 장기적 연구의 필요성이 제기되고 있다. 이러한 연구는 해당 전문가에 의해서 장기적이고 주기적으로 수행되고, 연구 결과의 공개와 함께 조사 자료를

공유할 수 있도록 진행되어야 한다.

본 연구는 이러한 필요성에 따라 1994년 12월부터 한국학술진흥재단의 지원을 받아 3년 동안 수행한 『초·중·고 학생들의 장기적 과학학력 점검을 위한 국가 수준의 평가 체계 개발』의 세부과제로써 우리 나라 초·중·고 학생들의 과학 탐구능력 성취의 경향성을 파악하고 향후 주기적으로 진행되어질 국가적 수준에서의 평가 지표를 만들기 위한 국가 수준의 과학 탐구능력 평가체계를 개발하는데 목적이 있다.

## II. 연구 방법 및 내용

### 1. 탐구능력 평가들의 개발

과학탐구능력의 평가들은 문항 개발의 방향과 내용을 결정하는데 중요한 역할을 한다. 본 연구에서는 그 동안 국내외에서 발표된 탐구능력 평가들을 참고하였으며, 특히 미국의 NAEP, 영국의 APU 및 우리나라의 대학수학능력시험 과학탐구 평가들을 주요 참고자료로 활용하여 새로운 평가들을 개발하였다. 새로운 평가들의 개발은 이러한 문헌 연구와 함께 8차에 걸친 세미나를 통하여 여러 사람의 의견을 수렴하고 여러 차례에 걸친 수정보완을 통하여 점진적으로 이루어졌다. 이러한 연구 과정을 통하여 개발된 과학탐구 평가들은 탐구 과정의 단계를 5단계, 즉 문제 인식 및 가설 설정, 탐구 설계, 탐구 수행, 자료 해석, 결론도출 및 평가 단계 등으로 나누고, 각 단계별로 초등 및 중등 수준에 알맞은 탐구요소를 아래의 표와 같이 설정하였다.

#### 1) 문제 인식 및 가설설정

##### (1) 문제 인식

주어진 상황에서의 연구 문제 도출 능력으로 문제발상 능력이 포함된다.

##### (2) 가설 설정

검증 가능한 가설을 찾아내어 검증할 수 있는 실험 방법을 제시하거나 주어진 실험 상황에서 가설을 세울 수 있는 능력을 말한다.

#### 2) 탐구 설계

##### (1) 변인 통제

연구에 대한 가설이 주어지면, 종속변인과 독립변인을 찾아내고 실험에서 통제되어야 할 변인을 구별하는

탐구의 단계	초 등	중 등
1단계: 문제 인식 및 가설 설정	1-1 문제 인식 1-2 가설 설정	1-1 문제 인식 1-2 가설 설정
2단계: 탐구의 설계	2-1 변인 통제 2-2 실험 설계	2-1 변인 통제 2-2 실험 설계 2-3 실험과정 개발
3단계: 탐구의 수행	3-1 기구 조작 3-2 관찰 3-3 측정 3-4 분류	3-1 기구 조작 3-2 관찰 3-3 측정 3-4 분류 3-5 자료변형
4단계: 자료의 해석	4-1 추리 4-2 예상	4-1 추리 4-2 예상(내삽/외연) 4-3 상관관계 4-4 인과관계
5단계: 결론 도출 및 평가	5-1 결론	5-1 결론 5-2 일반화 5-3 평가

것이다.

### (2) 실험 설계

주어진 가설 속에 내재해 있는 독립변인과 종속변인 사이의 관계를 추출하기 위하여 여러가지 실험방법 및 과정을 생각해 내는 능력이며, 종합적이고 전체적인 실험 계획을 수립하는 것을 의미한다.

### (3) 실험 과정 개발

전체적인 실험의 수행을 위하여 부분적인 실험 절차를 고안하는 것을 의미한다.

## 3) 탐구 수행

### (1) 기구 조작

설정된 가설 속에 내재해 있는 독립변인과 종속변인 사이의 관계를 추출하기 위해 논리적으로 서열화된 실험 절차에 따라 실험 기구를 조작하는 것을 말한다.

### (2) 관 찰

사물의 형태나 사건의 현장으로부터 오감을 이용하여 올바른 정보나 자료를 찾아내고 기술하는 것이다.

### (3) 측 정

주어진 물체의 특성에 따라 적절한 측정 도구를 이용하여 관찰한 것을 정량화하는 것이다.

### (4) 분 류

수집된 사물이나 사건으로부터 여러 가지 특성을 찾아내고 그 특성에 따라 나누는 것이다.

### (5) 자료 변형

관찰, 측정 등의 결과가 수량적으로 표시된 자료를 모아 정리하고 진술하는 능력을 의미한다.

## 4) 자료 해석

### (1) 추 리

주어진 사건 혹은 사실로부터 이미 일어난 사건을 돌이켜 생각하는 것이다. 즉 관찰이나 실험에서 얻은 자료를 수집하여 직접 관찰이 안된 새로운 사실을 이끌어 내는 과정이다.

### (2) 예 상

현재의 관찰 결과를 토대로 앞으로의 결과를 예측하는 활동이다. 즉 주어진 사건 혹은 사실로부터 아직 일어나지 않은 사건을 미리 생각하는 것이다.

### (3) 상관 관계

관찰된 사실을 토대로 2개 혹은 그 이상의 변인들 간의 관계를 알아내는 능력으로 관찰된 사실내에 들어 있지 않은 부분을 추리하는 것이 아니라, 단지 주어진 사실내에서 관계를 해석하는 것이다.

### (4) 인과 관계

인과 관계는 이미 관찰된 사실이나 과거의 경험적 지식을 근거로 실험, 관찰, 관측 결과나 자연적 결과와 효과에 대한 원인을 확인하고 설명하는 것이다.

## 5) 결론 도출 및 평가

### (1) 결 론

수집된 자료로부터 모든 중요한 정보들을 추출하여, 실험 사실을 포괄적으로 설명할 수 있는 종합적인 아이디어를 이끌어내는 과정이다.

(2) 일반화

주어진 자료의 경향성과 규칙성을 파악하여 보다 넓은 범위로의 해석이나, 그러한 특성을 폭 넓게 설명할 수 있는 모델을 설정하는 것이다.

(3) 평가

탐구 활동 결과의 가치 특성에 관한 판단 과정으로서 비판적 검토 및 탐구 과정을 거쳐 내린 결론이 그 자체에 내재해 있는 뜻이나 탐구 문제 해결에 어떤 역할을 하고 있는지를 파악하고 해석하는 능력이다.

탐구 상황은 탐구과정이 일어나는 맥락을 의미하며, 순수과학적 상황, 실험실적 상황, 일상적 상황, 기술·사회적 상황, 자연환경적 상황 등 모두 5가지 상황으로 구분하였다.

① 순수과학적 상황 : 기본 과학 개념의 체계적 이해와 이들 개념의 형성에 상호작용하는 과학 교과내적 상황을 의미한다.

② 실험실 상황 : 기본 과학 개념의 형성 정도나 개념들간의 관계에 대한 이해 및 적용 능력을 평가하기 위하여, 교실 또는 실험실에서 제기되는 문제 상황을 의미한다.

③ 일상적 상황 : 학생들이 일상 생활에서 직면하는 문제의 탐구와 해결에 과학적 사실이나 원리를 활용하고 일상적 상황 내에서 기본 과학 개념이나 탐구능력을 적용할 수 있는지를 평가하는 문제 상황을 의미한다.

④ 기술·사회적 상황 : 과학과 기술의 발달이 인간과 사회에 미치는 영향을 과학적 자료에 근거를 두고 의사 결정하는 능력이나, 과학 지식과 방법이 산업적 혹은 실용적 목적으로 응용되는 상황을 의미한다.

⑤ 자연환경적 상황 : 학습한 기본 과학 개념과 탐구 능력을 활용하여 해결할 수 있는 과학교과외의 외적 자연환경 상황을 의미한다.

2. 탐구능력 평가 문항의 개발

탐구능력 평가 문항의 개발은 R & D(Research & Development) 과정에 의하여 진행되었다. R & D 과정은 개발된 예비 문항을 현장 적용해 본 후 수정 보완을 거치기 때문에, Borg(1989)는 이러한 방법이 연구와 실제 현장과의 괴리를 극복할 수 있는 좋은 방법이라고 하였다.

1) 평가 목표 진술

평가 목표와 내용이 어떤 방식으로 진술되어야 하는가는 중요한 문제다. 이 평가 목표와 내용은 평가 문항 개발의 지침으로 이용되기 때문에 평가 목표의 진술 형태는 과학 탐구 평가에 큰 영향을 미치게 된다. 따라서 평가 목표는 평가 기준이 명확히 제시된 형태로 진술되어야 한다. 예를 들어 평가 목표는 다음과 같이 나타낼 수 있다.

실험실에서 거리와 시차와의 관계를 알아보기 위해

(a) (b)

스탠드, 막대자, 막대, 실, 단추를 이용한 실험 계획을 세울 수 있다.

(c)

즉, a는 실험실 '상황'을, b는 거리와 시차와의 관계에 관한 '내용' 영역을 나타내며, c는 '과정' 영역으로 탐구의 단계 중 실험 설계에 해당된다. 이와 같이 '상황', '내용', '과정'의 3차원적 요소가 한 평가 목표에 모두 진술되어 있다. 이러한 평가 목표의 상세화 절차에 따라 초등학교 3학년에서 고등학교 3학년까지의 교육과정에 대하여 평가목표 상세화 작업을 하였다.

개발된 평가들의 개발과 이를 바탕으로 한 평가 목표의 상세화 작업을 한 후, 과학탐구 평가 문항을 개발하기 위한 설계를 다음과 같이 구체화하였다.

(1) 평가 대상

초등학교 5학년  
중학교 2학년  
고등학교 1학년(공통과정)  
고등학교 2·3학년(자연계)

(2) 평가 시기

평가 대상 학년의 해당 교육과정이 끝나는 2월 중

(3) 표집 수

전체 모집단의 1%

(4) 개발 문항 수

해당 학년 별 100문항 정도

(5) 평가 시간

초등학교 5학년 40분  
중학교 2학년 45분  
고등학교 50분

(6) 평가 문항 유형

객관식 5지 선다형

2) 예비 문항 개발

문항 개발을 위하여 연구원 및 협력위원, 연구보조원이 사전에 2회에 걸친 문항 제작 워크숍; 실시하여 문항 제작의 숙련도를 높였다. 예비문항은 연구원 및 협력위원, 연구보조원이 함께 집중 작업을 통해 개발하였다. 집중작업에 참여한 연구인력은 과학교육 전공 교수 11명, 연구보조원(과학교육전공 대학원생) 33명이다. 예비문항은 각기 낱장의 문항 카드에 기록하도록 하였다. 문항 카드는 문항 코드와 함께 평가들에 의한 분류, 평가 목표, 문항의 정답, 예상 난이도 등을 체크할 수 있도록 구성되어 있다.

공동작업을 통해 개발된 예비문항은 3회의 수정 및 보완 작업을 거쳤다. 1차 수정 및 보완 작업에서는 각 연구원들의 책임하에 문항 카드를 각 내용 영역(물리, 화학, 생물, 지구과학)별로 나누어 문항의 내용상 오류, 문제부와 답지부의 적절성, 어휘력의 수준 등을 검토하였다. 2차 수정 및 보완 작업은 문항 개발을 위한 공동 작업 및 검토 작업에 참여하지 않은 현직 초·중·고 교사 및 과학교육전공 대학원생들에게 수정 및 보완을 의뢰하여 예비문항의 적절성을 높이도록 하였다. 2차 수정 작업까지 마친 문항을 토대로 삽화 제작을 의뢰하고 각 문항을 전산 입력하는 작업을 실시하였다. 삽화 작업 및 편집 작업이 완료되어 검사지 형태로 구성된 후, 3차 수정 및 검토 작업이 이루어졌다. 개발된 문항들에 대한 3차에 걸친 수정 및 보완 작업을 실시한 후, 예비 검사를 위해 최종 선별된 문항수는 총 325문항이었다. 이를 학교급별, 내용 영역별로 나누어 보면 다음과 같다 (Table 1).

3) 예비 검사

개발된 예비 검사 문항의 적절성과 타당성을 검증하

**Table 1** Distribution of pilot test items for evaluating inquiry process skills

School	Content	Number of items
Elementary	Physics	27
	Chemistry	26
	Biology	25
	Earth science	26
	Subtotal	104
Middle	Physics	25
	Chemistry	29
	Biology	30
	Earth science	25
	Subtotal	109
High	Physics	25
	Chemistry	28
	Biology	30
	Earth science	29
	Subtotal	112
Total		325

기 위하여 소규모 집단을 대상으로 예비 검사를 실시하였다 (Table 2).

예비 검사를 실시하여 얻은 자료는 모두 전산 입력되었으며, 입력된 전산 자료를 바탕으로 각 문항의 양호도를 검토하여 필요에 따라 문항을 삭제하거나 수정·보완하였다. 문항의 양호도 검토는 개발된 예비 문항이 얼마나 바람직한가를 알아보는 것으로 문항의 타당도, 신뢰도, 변별도, 난이도 등이 검토되었다.

4) 본검사 문항 확정

3회에 걸친 예비 검사 문항의 수정·보완과 예비 검사 및 전문가 협의회를 거쳐 최종 선정된 본 검사용 문항은 총 240문항으로 각 학교급별로 80문항씩이다.

본 검사용 문항은 한 검사지에 20문항씩 배치하여 각

**Table 2** Number of sampled students for the pilot test

	High school			
	5th grade	8th grade	10th grade	11th and 12th grade (science track)
Number of students	226	274	210	248

학교급별로 4종의 검사지로 만들어 총 12종의 검사지를 본 검사에서 사용하였다. 검사지에 문항을 나누어 재배치할 때는 탐구내용, 탐구단계가 각 검사지에 고르게 배치되도록 하였으며, 예비 검사를 통해 조사된 문항의 난이도 또한 문항의 배치에 고려하였다.

5) 본 검사 실시

최종 완성된 240문항에 대한 본검사를 1997년 2월 중에 실시하였다. 본검사 실시를 위해 대도시, 중소도시, 읍·면 지역으로 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. 우편으로 회수된 응답자의 지역별, 성별 현황은 다음과 같다(Table 3).

본 연구는 탐구능력을 측정하기 위한 평가 도구의 개발에 주 목적을 두고 있으므로 여기에서는 변인(지역, 성, 학년)에 따른 탐구능력의 성취 수준 차이는 별도로 분석하지 않고 문항 분석에 초점을 맞추었다.

III. 연구 결과

1. 고전검사이론에 의한 문항 분석

현재 국내에서 대부분의 문항 분석에 적용되고 있는 이론은 고전검사이론이다. 고전검사이론은 최근의 문항반응이론과 대비되는 이론으로서 전통적 방법에 의하여 난이도, 변별도, 신뢰도 등이 산출될 수 있다. 본 연구에서 문항 난이도(difficulty index)는 문항에 대한 정답율로, 변별도는 양분상관계수(biserial correlation coefficient)로, 신뢰도는 Cronbach  $\alpha$  계수로 산출되었다. 이러한 계산과 분석을 위해 컴퓨터 프로그램인 ITEMAN (Version 3.50)이 사용되었다. 고전검사이론

에 의한 문항 분석 결과를 정리하면 Table 4와 같다.

초등학생용 검사지 A, B, C, D에 대한 학생들의 평균 점수는 11.0~12.2로서 난이도 지수로 환산할 경우 0.55~0.61이다. 5지 선다형 문항의 경우 변별도가 극대화될 수 있는 난이도가 0.6인 것을 감안하면, 검사지의 난이도는 매우 적절한 것으로 판단된다. 이는 실제로 산출된 문항 변별도의 평균이 검사지별로 0.51~0.54라는 사실로 뒷받침된다. Doran(1980)은 변별도가 0.3이상이면 적절하고, 0.4이상이면 매우 좋은 문항이라고 하였는데 이러한 기준을 적용할 경우 본 검사지의 문항은 매우 우수한 문항이라고 할 수 있다. 신뢰도(Crabach  $\alpha$ )는 검사지별로 0.69~0.73으로써 보통의 수준이다.

중학생용 검사지의 경우 평균 점수는 9.3~10.3으로서 초등학생용보다 다소 낮게 나왔으나, 변별도는 0.47~0.53으로서 초등학생용과 별 차이가 없었다. 고등학생용 검사지의 경우 평균 점수는 8.0~10.7로 검사지에 따라 비교적 큰 차이를 보였다. 이는 각 검사지별로 시험을 본 학생들이 서로 다른 집단인 것도 원인이 될 수 있으나, 검사지별로 난이도 면에서 문항 배분이 균등하게 이루어지지 않았을 가능성도 시사한다. 문항 배분을 위한 참고 자료로 예비 검사 결과를 이용하였는데, 예비 검사에서는 표집의 크기가 작아 문항 난이도 산출시 비교적 오차가 컸을 것으로 생각된다. 앞으로의 평가에서는 예비 검사 결과보다는 본 검사 결과를 이용하여 문항을 배분할 수 있기 때문에 이러한 문제가 해결될 수 있을 것이다. 변별도는 0.39~0.49로서 초등학생용 및 중학생용 검사지보다 다소 낮으나, 이러한 수치 역시 매우 우수한 변별도를 나타내고 있다. 신뢰도는 0.50~0.70으로서 높지 않은 수치이다.

본 연구에서 개발된 초등학생용, 중학생용, 고등학생용 문항은 난이도가 적절하고 변별도는 매우 우수한 반면, 신뢰도는 기대한 것만큼 높게 나오지 않았다. 신뢰도가 문제될 정도는 아니나 충분히 높게 나오지 않은 가

Table 3 Sampled students for the main test

School	Residence			Sex		Total	
	Large city	Small city	Rural	Male	Female		
Elementary	1,302	843	430	1,347	1,228	2,575	
Middle	1,097	789	336	1,200	1,022	2,222	
High	10th grade	683	689	522	1,138	756	1,894
	11th, 12th grade	1,146	733	336	1,375	840	2,215
Total	3,082	2,321	1,288	5,060	3,846	8,906	

**Table 4** Results of item analysis by traditional method

School	Test	Number of items	Number of students	Mean	SD	Discrimination index(Average)	Reliability (Cronbach $\alpha$ )
Elementary	A	20	623	11.7	3.6	0.54	0.73
	B	20	537	11.4	3.5	0.51	0.69
	C	20	694	11.0	3.6	0.51	0.71
	D	20	721	12.2	3.5	0.51	0.69
	Total	80	2575	11.6	3.6	0.52	0.71
Middle	A	20	478	10.3	3.8	0.53	0.74
	B	20	617	10.1	3.8	0.52	0.75
	C	20	389	9.3	3.5	0.47	0.67
	D	20	738	10.0	3.6	0.49	0.72
	Total	80	2222	9.9	3.7	0.50	0.72
High	A	20	1002	8.0	2.9	0.39	0.50
	B	20	722	9.7	3.6	0.49	0.70
	C	20	1258	10.7	3.4	0.49	0.67
	D	20	1127	9.1	3.3	0.45	0.64
	Total	80	4109	9.4	3.3	0.46	0.63

장 큰 이유는 본 연구에서 개발된 검사지가 특정 단원에 대한 동질적인 문항들로 구성된 것이 아니라, 전 학년에 걸친 내용을 이용하여 개발한 다양한 문항을 한 검사지에 포함시켜서 문항의 동질성이 낮기 때문인 것으로 판단된다.

## 2. 문항반응이론에 의한 문항 분석

문항반응이론(Item Response Theory)은 문항 분석에 관한 새로운 이론으로서 문항의 특성을 고전검사이론에 의한 분석보다 더 정확히 분석할 수 있다는 장점이 있다. 고전검사이론에 의하여 문항을 분석할 때, 능력이 높은 집단에서 검사가 실시되었다면 그 문항은 쉬운 문항으로 분석되고, 능력이 낮은 집단에서 검사가 실시되었다면 똑같은 문항이 어려운 문항으로 분석된다. 그러나 문항반응이론에 의하면 난이도, 변별도 등의 특성이 피험자 집단의 특성에 따라 변하지 않고 문항은 불변하는 고유한 속성을 지니고 있는 것으로 분석된다.

이것을 문항특성 불변성 개념이라고 한다(성태제, 1998). 문항반응이론에 의한 문항 분석에서는 피험자 자료를 이용하여 문항특성곡선을 그리게 되는데, 이 곡선을 이용하여 난이도와 변별도 등을 계산한다. 본 연구에서는 문항반응이론에 의한 문항 분석 전문 컴퓨터 프

로그래밍 XCALIBRE(Version 1.10)를 사용하여 문항반응이론 난이도(IRT difficulty index)와 문항반응이론 변별도(IRT discrimination index)를 계산하였다(Table 5).

문항반응이론에 의한 난이도가 0일 경우 중간 정도의 난이도를 뜻하며, 큰 음수값을 가질수록 쉬운 문항, 큰 양수값을 가질수록 어려운 문항이다. Table 5를 보면 초등학교용 검사지의 경우 문항반응이론에 의한 난이도가 -2.0~-0.5에 해당하는 문항이 27개, -0.5~0.5에 해당하는 문항이 25개인데, 이는 대부분의 문항이 약간 쉽거나 중간 정도의 난이도를 가지고 있다는 것을 뜻한다. 중학교용 검사지의 경우 중간 정도의 난이도인 -0.5~0.5에 속하는 문항이 34개, 약간 어려운 문항(난이도 0.5~2.0)과 매우 어려운(난이도 2.0이상) 문항이 각각 25개로서 중학교용 검사지보다도 더 어려운 검사지임을 알 수 있다.

문항반응이론에서 변별도가 0.39~0.80이면 보통, 0.80~1.0이면 높은 것으로 해석되는데(성태제, 1998), 본 연구에서 개발된 문항의 경우 변별도가 모두 적절하거나 높은 것으로 나타났다. 이러한 결과는 고전검사이론에 의한 문항 분석 결과와도 대체로 일치한다고 볼 수 있다. 고전검사이론에 의한 변별도 분석 결과 매우 어려운 문항의 경우에는 문항 변별도가 낮게 나온 경우가 간

**Table 5** Number of items according to the results of item analysis by item response theory

School	Test	IRT difficulty index					Total	IRT discrimination index					Total
		below -2.0	-2.0~ -0.5	-0.5 ~0.5	0.5 ~2.0	above 2.0		below -2.0	-2.0~ 0.39	0.39~ 0.80	0.80~ 1.0	above 1.0	
Elementary	A		8	5	5	2	20		15	4	1	20	
	B		5	7	5	3	20		17	3		20	
	C	1	5	6	6	2	20		17	3		20	
	D		9	7	2	2	20		18	2		20	
	Total	1	27	25	18	9	80		67	12	1	80	
Middle	A		2	8	7	3	20		9	11		20	
	B		3	10	3	4	20		6	14		20	
	C			8	6	6	20		13	7		20	
	D		4	8	4	4	20		9	11		20	
	Total		9	34	20	17	80		37	43		80	
High	A			1	8	11	20		20			20	
	B		3	7	6	4	20		16	4		20	
	C		7	3	7	3	20		17	3		20	
	D		2	7	4	7	20		17	3		20	
	Total		12	18	25	25	80		70	10		80	

혹 있었으나, 문항반응이론에 의한 분석에서는 어려운 문항도 변별도가 모두 적절하거나 높게 나온 것이 특징적이다. 그리고 본 연구에서 개발된 문항들은 고전검사 이론 뿐만 아니라 문항반응이론에 의한 분석에서도 매우 양호한 문항 특성을 가진 것으로 밝혀졌다.

#### IV. 결론 및 제언

국가적 수준에서 우리 나라의 과학 교육 성과를 가능해보려는 작업은 우리 나라 과학교육의 질을 체계적으로 관리한다는 차원에서 매우 중요한 일이다.

이러한 노력은 과학교육 체제의 효율성이나 효용성에 대해 체계적이고 과학적으로 이해하려는 작업이다. 본 연구는 우리 나라 초·중·고등학생들의 과학 탐구능력 성취의 경향성을 파악하고 향후 주기적으로 진행되어질 국가적 차원의 평가 지표를 만들기 위한 과학탐구능력 평가 체계를 개발하는데 목적이 있다.

과학탐구능력의 평가들을 개발하기 위하여 기존의 국내의 평가들을 탐색함과 아울러 국내 관련 전문가들의 초청 세미나를 실시하였다. 이러한 과정은 기존 연구들이 시사하는 점을 보다 포괄적이며 체계적으로 수용하며, 아울러 개발될 평가들의 타당성과 객관도를 높이기

위함이었다. 개발된 평가들은 탐구과정, 탐구상황, 탐구내용의 3요소로 구성된 3차원 평가틀이다. 개발된 평가틀에 의해 초등학교 3학년에서 고등학교까지의 전 과정에 대한 목표 분류 및 평가목표 진술을 하였다. 연구 협의를 통해 평가 대상을 초등학교 5학년, 중학교 2학년, 고등학교 1학년과 자연계열로 하고, 평가 시기는 해당학년의 교육과정 이수가 끝나는 학년말(2월)로 하였다.

문항 개발 워크事: 통해 연구원, 연구 협력위원, 연구 보조원(현직교사 또한 과학교육 전공 대학원생)등 44명이 문항 제작 기법을 숙련하였으며, 집중 작업을 통해 예비 문항을 개발하였다. 개발된 문항에 대하여 3차례의 수정·보완 작업을 실시하였고, 1회의 예비 검사를 실시하였다. 325문항의 예비 검사를 통해 문항의 얻은 조사 결과를 바탕으로 문항의 제거, 수정, 보완 등의 작업을 통해 최종 초등학교용 80문항, 중학교용 80문항, 고등학교용 80문항 등 모두 240문항이 선정되었다.

선정된 240문항을 전국에서 지역, 성별을 고려하여 표본을 추출한 후, 모두 8,906명의 본 검사 조사 결과를 분석하였다. 고전검사이론 및 문항반응이론에 의하여 문항 분석을 한 결과, 난이도, 변별도, 신뢰도 등의 문항 특성이 모두 우수하다는 결과를 얻었다.

본 연구는 그동안 우리 나라 과학교육의 개선이 과학



적이고 체계적인 조사 자료에 근거하지 않았으며, 과학교육의 질관리가 합리적이고 타당하게 이루어지지 못하고 있다는 문제점의 인식하에 시작되었다. 본 연구는 우리 나라 국가 수준의 과학 탐구능력의 성취 수준을 주기적으로 측정하여 그 추이를 분석하기 위하여 평가도구를 개발하는 기초 연구의 성격을 띠고 있다. 본 연구를 통해 개발된 평가 체제의 활용 방안을 구체적으로 진술하면 다음과 같다.

- 1) 현재 국내에는 국가 수준의 과학 탐구능력 평가 체제가 마련되어 있지 않으며, 우리 나라 초·중·고 학생들의 학력을 평가할 수 있는 타당하고 신뢰로운 도구가 없다. 따라서 본 연구에서 개발되는 도구는 우리 나라에서 활용할 수 있는 국가 수준의 과학 탐구능력 평가 도구로 활용될 수 있다.
- 2) 본 연구의 결과는 효과적이고 합리적인 과학교육 정책의 방향을 결정하고 과학교육 행정과 학교 과학교육의 추진 계획을 수립하는데 실증적 자료를 제공하는 데 기여할 수 있다. 즉, 국가 수준의 과학 탐구능력 평가의 결과는 학교에서 달성하려는 교육목표를 얼마나 성취하고 있는가? 성공적으로 달성하고 있는 목표는 무엇인가? 특별히 결손이 일어나는 내용과 목표는 무엇인가? 그리고 그 요인은 어디에 있는가? 등에 대한 실증적 자료를 얻는 기초가 될 수 있다.
- 3) 국가 수준의 과학 탐구능력 평가의 정착으로 교수-학습 방법의 개선을 유도할 수 있다. 수학능력시험 제도의 등장은 곧 고등학교는 물론 초·중학교의 교수-학습 방법의 전환을 가져오게 하였다. 이처럼 평가의 방법은 학교 교육의 방향을 유도한다. 따라서 본 연구의 결과는 학교 현장에서 이루어지고 있는 과학평가가 학생들의 탐구능력을 자극하지 못하는 문제를 극복하기 위한 대안으로 양질의 평가 문항 개발과 보급의 역할을 할 수 있다.
- 4) 국가 수준의 과학 탐구능력 평가는 정확한 장학자료를 제공할 수 있으므로 장학력의 제고를 꾀할 수 있다. 장학의 본질은 수업의 질적 개선을 통하여 학생들의 행동변화를 촉진하는데 있다. 따라서 장학의 효율성을 제고하기 위해서는 학생들의 학업 성취 수준 및 그것과 관련된 학습자 요인, 수업자 요인, 교육행정 요인 등에 대한 정확한 정보를 파악해야 하기 때문이다.
- 5) 국가수준의 과학 탐구능력 평가는 학교 과학교육의 질관리를 과학적으로 할 수 있게 해 준다. 학교

과학교육의 질을 과학적으로 관리하기 위해서는 교육과정이 추구하는 교육 목표의 달성이 성공적으로 이루어지고 있는지를 국가수준에서 주기적으로 측정하여 부진의 원인을 진단하고, 개선 방안을 찾아야 정상적인 교육이 이루어지기 때문이다.

- 6) 국가 수준의 과학 탐구능력 평가는 국민으로부터 위임 받은 학교 과학교육의 책무성에 대한 반성의 자료를 제공해 줄 수 있다. 교육의 의무와 권리를 가지고 있는 국민은 그 권리를 학교에 위임하고 있다. 따라서 학교는 국민으로부터 위임받은 교육의 책무가 어떻게 수행되고 있는가를 반성하고 이를 토대로 보다 효과적인 수행 방안을 모색해야 하는데, 국가 수준의 과학 탐구능력 평가 결과는 그 자료로서의 구실을 수행하게 된다.
- 7) 본 연구를 통해 얻은 원자료(raw data)는 자격을 갖춘 여러 연구자에게 공개하여 다양한 분석이 이루어질 수 있도록 한다. 이는 과학교육 연구의 활성화를 기할 수 있으며, 조사 결과에 대한 다양한 분석을 통해 과학교육 발전을 가져올 수 있다.

본 연구 결과로 개발된 평가 도구가 제대로 활용되어 충분한 기대효과를 얻을 수 있게 하기 위해서 다음과 같이 제언하고자 한다.

- 1) 우리 나라 학교 과학교육의 성취 수준을 조사하는 연구를 위해서는 법적, 제도적, 행정적, 재정적 지원을 지속적으로 해야한다. 타당하고 신뢰로운 평가 도구의 개발을 위해서는 많은 전문 인력의 활용되어야 하고 이에 따라 경비의 지출이 필수적이다. 또한 전국적인 규모의 표본을 추출하여 개발된 평가도구를 적용하기 위해서는 법적, 제도적, 행정적 지원이 수반되어야 한다. 그러나 아직 우리나라에서는 이러한 지원이 거의 이루어지지 않고 있으며, 또한 교육 현장에서 갖고 있는 학력 평가에 대한 부정적 견해로 인한 어려움이 많이 있다. 따라서 이러한 연구와 관련된 연구 기구 또는 기관을 설립하거나 지정하고 이에 따른 여러 가지 지원을 해야 할 것이다.
- 2) 본 평가 체제를 운영하기 위해서는 지속적인 연구와 개발이 필수적이다. 즉, 주기적이면서 장기적인 평가 사업을 하려면 평가도구의 신뢰성과 타당성을 확보해야 하며, 또한 주기적인 측정시 사용할 평가도구가 일관성을 유지해야 한다. 이를 위해서

는 지속적인 평가 문항의 교체, 수정, 보완 작업이 이루어져야 한다.

## 적 요

본 연구의 목적은 국가적 수준에서 우리나라 초·중·고등학교 학생의 과학탐구 능력을 점검해볼 수 있는 평가체제를 개발하는 것이다. 이를 위하여 많은 전문인력이 연구에 참여하였고, 과학탐구 평가들과 평가 문항을 개발하였다. 개발된 평가 문항을 여러 차례 수정보완한 후, 325개의 문항을 958명의 학생에게 예비 검사를 실시하여 현장 적용 가능성을 검토하였다. 예비 검사 결과 문항 특성이 만족스럽지 못한 문항은 삭제하여 240개의 문항을 본 검사 문항으로 확정하였다. 본 검사는 총 8,906명의 학생을 대상으로 실시하였으며, 검사결과를 고전검사이론과 문항반응이론에 의하여 분석한 결과 난이도, 변별도, 신뢰도 등이 매우 우수하다는 것이 판명되었다. 앞으로 본 연구에서 개발된 과학탐구 평가들을 이용하여 더 많은 문항이 개발될 수 있으며, 이러한 문항을 이용하여 우리나라 학생들의 과학탐구능력을 주기적으로 점검할 수 있을 것이다.

## 참 고 문 헌

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

- 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.
- 권낙원, 이재분, 구자역(1985). 국민학교 교육과정 평가 연구(IV). - '83, '84, '85년 간의 학생들의 학업성취도 비교-한국교육개발원 연구보고 RR85-7.
- 김형립 외(1980). 새 교육체제 제5차 시범 연구보고서. 한국교육개발원 연구보고 제 112집.
- 성태제(1998). 문항 제작 및 분석의 이론과 실제. 학지사.
- 전용신(1963). 국민학교 5-6학년 학생의 학력검사 결과의 비교(1959년도 대 1963년도). 중앙교육연구소 조사 연구 제 27집.
- 한국행동과학연구소(1980). 한국 국민학교 교육의 평가(I): 학력 평가. 한국행동과학연구소.
- Borg, Walter R.(1989). *Educational Research*. 5th ed. New York: Longman.
- Doran, Rodney L.(1980). *Basic Measurement and Evaluation of Science Instruction*. Washington, D. C. : NSTA.