

國民學生의 科學 探究能力 測定을 위한 評價道具 開發

정완호 허명
(한국고원대학교) (이화여자대학교)

은경용
(한국고원대학교)

(1993년 5월 12일 받음)

I. 서론

1. 연구의 필요성

현대는 눈부신 과학 혁명의 시대이다. 불과 2-3년 전의 첨단 과학 기술이 오늘은 쓸모없는 낡은 지식이 되어 버리는 속도감 있는 현 시점에서 다가오는 21세기 미래 사회에 대응하기 위해서는 과학교육의 방법이 달라져야 한다고 본다.

이러한 시대적 요청에서 1973년 제 3 차 국민학교 교육과정 개정에서 우리나라는 과학 교육과정의 개혁을 단행하였다.

이는 2차 교육과정기 때까지의 생활중심 교육과정에서 과감하게 벗어나 학문중심 사상을 수용하여 자연과 교육의 목표에 '과학적 탐구 방법을 체득시켜 자연의 규칙성을 발견하는 능력과 태도를 기른다.'고 명시함으로써 탐구 과학교육을 강조하였다(문교부, 1973).

그 이후 4차 교육과정 개정이나 현행 5차 교육과정(문교부, 1982, 1987), 그리고 1992년 공포된 6차 교육과정에서도 과학 탐구능력을 계속 강조하고 있다(교육부, 1992).

이와 같이 자연과 교육에서 탐구능력이 중요시되자 학교 현장의 연구나 여러 개인 연구에서 탐구능력 신장 방안 또는 탐구학습에 대한 연구가 많았다. 그러나 탐구학습의 결과를 어떻게 평가할 것인가에 대한 연구는 부분적일 뿐이었다. 김주훈과 이양락(1984)의 조사에 의하면 현행 자연과 평가가 바람직하게 이루어지지 않고 있다고 반응한 교사가 70%나 되고 있다. 이러한 반응은

지금까지의 탐구능력 평가가 진정한 의미의 탐구능력을 잴 수 있는 교과 내용을 중심으로 한 내용 성취도 평가가 자연과 평가의 주류를 이루었다고 할 수 있다. 즉 탐구학습의 결과를 평가하기보다 전형적으로 과학 지식 내용의 습득 정도에 초점을 두는 평가였다. 그러므로 탐구능력을 잴 수 있는 보다 객관적이며 타당성이 있고, 신뢰도가 높은 과학 탐구능력 측정 도구의 개발이 요청된다.

2. 연구 목적과 내용

본 연구의 목적은 과학교육 목표 중의 하나인 과학탐구의 기초적인 방법 습득 정도를 재는 평가도구 개발이다. 따라서 우리 나라 국민학교 고학년생들에게 적용 가능하고, 기초적인 탐구과정 요소를 포함하며 타당도와 신뢰도가 높은 탐구능력 검사용 도구를 개발하는 데 있다.

3. 연구 절차 및 제한점

1) 연구의 절차

평가도구 개발을 위하여 탐구과정 요소 및 평가에 대한 문헌 조사를 하였다. 그리고 다음과 같은 R&D 과정에 따라 도구를 개발하였다.

① 국민학교에 적용 가능한 탐구요소 10개 항목을 선정한다.

② 과학의 내용영역을 물질과 에너지, 생명과 환경, 지구와 우주로 나누고 총 30문항을 개발하였다. 이는 각 내용영역당 10개 문항씩이며, 탐구 과정 요소별로는 각

요소당 3문항씩이다.

③ 개발된 평가문항은 과학 교육 전문가와 현장 교사의 점검 및 1차 현장 검사를 통하여 문항을 분석하고 수정한다.

④ 수정된 평가 문항은 다시 과학교육 전문가에게서 내용 타당도와 정답의 객관도를 점검하고 2차 현장 검사를 통하여 문항 분석을 하고 난이도, 변별도와 평가도구의 신뢰도 등을 산출하여 완성한다.

2) 연구의 제한점

본 연구는 국민학생의 탐구능력을 측정할 목적에서 국민학생의 과학탐구 능력 평가도구를 개발하는 것으로 다음과 같은 제한점을 갖는다.

① 본 연구는 평가도구의 개발 과정에 드는 노력과 시간, 비용 및 국민학생의 인지 발달 단계를 고려하여 적용 대상을 국민학교 고학년(5,6학년)으로 한정하였다.

② 표집은 국민학교의 특성상 사립학교를 제외하고 국·공립학교로 한정하였다.

II. 선행연구

듀이는 "교육적으로 가치 있는 경험이란 사고의 과정이 포함되어야 한다."고 한다. 여기서 말하는 사고는 "반성적 사고(reflective thinking)"를 말하며(Dewey, 1933) 이는 탐구과정이고, 과학적 합리적 사고 방식을 말하기도 한다.

듀이는 사고 과정을 ①암시 ②지성화 ③가설 ④추리 ⑤검증의 5단계로 설명하고 있다(김종건, 1974).

현대 과학의 본질을 날카롭게 통찰하고 주로 이론적인 면에서 탐구학습의 필요성을 주장한 사람은 시카고 대학의 Schwab이었다. 그는 1961년 하버드대학에서 행한 기념강연 "탐구로서의 과학교육(teaching of science as inquiry)"에서 과학론에 중점을 둔 과학교육을 주장하였다. 그후 "탐구학습"(Inquiry learning)이란 말을 Schwab(1964)이 처음 사용하였다(장남기,1990).

Bruner(1956)에 의하면 "개념 획득 과정이란 우리가 획득하고자 하는 개념(부류)의 예물과 비 예물을 구별하는 속성을 발견하는 과정"이라고 했다. 그는 개념 획득 과정을 (1)문제인식 (2)가설설정 (3)가설검증 (4)결론짓기의 4 단계로 설명하면서 사고 과정과 탐구기능을 중요시 하였다.

듀이와 부르너의 인식방법과 교육방법을 간단히 비교

하면(김종건, 1974) 두 사람의 교육방법의 공통점은 암기, 기억에 의존하는 것이 아니고 학생들의 능동적인 지적활동 즉 '탐구'를 강조하고 있다는 것이다.

이러한 이론을 바탕으로 하여 1960년대 이후 과학교육계는 탐구학습이 주류를 이루어 왔다. 이에 맞춰 여러 학자들은 탐구학습을 평가하기 위한 평가틀을 개발하였다. 그리고 이 평가틀에 맞추어 평가 도구도 개발되었다. 이러한 평가도구를 몇 가지 살펴본다.

Walbeser 등은 SAPA 관련 교과과정을 이수한 어린 이들의 탐구의 기초기능 및 통합과정을 측정하기 위한 평가도구를 개발했다.(Doran,1978).

NAEP(1969)에서는 10개의 탐구과정 부분의 요소를 중심으로 문항을 개발하여 평가대상을 4 그룹(9세, 13세, 17세, 성인)으로 구분하여 측정하였다(Welch et al., 1981).

Moliter 와 George(1976)는 TSPS(A Test of Science Process Skills)를 개발했다. 이는 4~6학년을 평가 대상으로 했으며, 추리와 검증의 탐구요소를 측정하기 위한 선다형 각 9문항으로 신뢰도(K-R 20)는 0.55 ~ 0.84, 난이도 지수 0.22~0.66, 변별도 지수 0.44 ~ 1.00이다.

Padilla(1985)등은 SAPA가 제시한 기초적 탐구기능 중에서 관찰, 분류, 추리, 예상, 측정, 의사전달의 6가지 요소를 평가하는 도구인 BAPS(Test of Basic Process Skills)를 개발하였다.

Smith 와 Welliver (1990)는 13가지 탐구능력을 측정하는 문항으로 구성된 평가 도구 SPA (Science Process Assessment)를 개발하였다. 이 평가의 신뢰도(K-R20)는 0.80 ~ 0.82이며 4지 선다형 총 40문항으로 이루어졌다.

이 외에도 Nelson과 Abraham(1973)의 ISM(Inquiry Skill Measures) 등이 있다.

우리 나라에서의 과학 탐구능력 평가 연구는 1973년 이후 제 3 차 교육과정기 부터 관심의 대상이었으나 과학교육에 대한 평가는 주로 내용 성취도 평가 수준에 머물러 왔다. 그러다가 제 4 차 교육과정기부터는 국민학교 교육과정 자연과의 지도 및 평가상의 유의점에서 탐구능력의 평가를 하도록 명문화 하였다(문교부,1981). 이에 따라 탐구능력 평가를 위한 연구가 이루어졌는데 근자의 국내 연구를 살펴보면 다음과 같다.

김주훈과 이양락(1984)은 '국민학교 자연과 평가의 원리와 실제'에서 국민학교 5, 6학년의 자연과 전 단원에 대한 내용 분석을 거쳐 지식·이해영역, 정의적 영역, 탐구과정 영역의 지필평가 문항을 개발하였다.

이종기(1988)는 우리나라 고등학생을 대상으로한

TSIS(Test of Science Inquiry Skills)를 개발하였다. 이 도구는 12개 탐구요소를 선정하고 각 요소당 3문항씩 모두 36문항으로 구성되었다. 평가 시간은 50분간, 적용 대상은 고등학교 1,2,3학년이며 변별도 지수 0.46, 문항 난이도 지수 58.9%, 신뢰도(K-R 20) 0.86, 측정의 표준 오차 2.45이다.

이연우(1989)는 중학교 2학년생의 자료 분석과 해석 능력을 중심으로 한 과학 탐구능력 측정을 위한 표준화 검사지를 개발하였다.

박승재(1989)는 과학학습 평가(국민학교)에서 과학적 탐구 사고력의 평가 범주를 첫째, 관찰·분류 및 측정, 둘째, 문제 발견 및 해결방안, 셋째, 데이터 해석 및 일반화로 정하고 탐구 사고력 평가 예시 문항을 개발하였다.

이 외에도 김명환(1981), 송형호와 채남식(1984), 이항노(1991), 이무(1992) 등의 연구가 있다.

이상에서 살펴본 바와 같이 우리나라에서 개발된 과학 탐구능력 검사는 수 편의 논문을 제외하고는 대부분이 과학 교과서에 있는 내용과 실험을 중심으로 했으며, 평가 도구의 구비 조건인 타당도, 신뢰도, 변별도, 난이도 등에 대한 정보가 고려되지 않은 비형식적 검사지로 과학 학습 성취도 평가의 성격이 짙다. 또 검사 도구로서의 요건을 갖추고 문항 분석이 이루어진 평가 도구도 중등학교의 두 세 편을 제외하고는 국민학생을 대상으로 한 평가도구는 개발되지 않았다.

III. 평가도구 개발

1 과학 탐구과정 요소의 선정

평가도구를 개발하는데 필요한 탐구과정 요소의 선정은 과학 탐구의 평가틀과 과학 탐구능력 평가도구 개발에 관한 광범위한 선행연구를 근거로 하였다. 그리고, 국민학생들에게 적용 가능한 과학 탐구과정 요소 10가지를 선정하였다.

과학 탐구과정 요소는 주로 SAPA 과정안을 근간으로 SIEI(Hur, 1984), Klopfer(1971)의 탐구 과정 평가틀, APU, NAEP, TAPS, 박승재의 과학적 탐구 사고력 평가틀(1991), 우중욱 등의 평가틀(1991) 등을 참고로 하였다.

SAPA 과정안을 근간으로 선택한 이유는 SAPA 과정안이 개발 당시부터 국민학생을 대상으로 개발된 안이기에 국민학생의 탐구능력을 평가하는데 더 적합한 것

으로 판단되기 때문이다.

<표 1> 선정된 10가지 탐구과정 요소

-
- ① 관찰 (observing)
 - ② 분류 (classifying)
 - ③ 문제 인식 (recognizing problems)
 - ④ 측정 (measuring)
 - ⑤ 추리 (infering)
 - ⑥ 가설 설정(formulating Hypothesis)
 - ⑦ 변인 조절 (controlling Variables)
 - ⑧ 실험 (experimenting)
 - ⑨ 결과 해석 (interpreting)
 - ⑩ 결론 (drawing a conclusion)
-

이러한 과정을 거쳐 선정된 과학 탐구과정 요소는 <표 1>과 같다.

2. 문항개발

탐구과정 요소 10개 항이 선정되자 문항 개발을 위하여 각 탐구과정 요소에 대한 평가 목표를 진술한 다음 2원 목적 분류표를 작성하였다. 그리고 나서 문항 개발에 착수하였다.

1) 과학 탐구과정 평가 목표 진술

평가 도구에 대한 내용타당도를 높이기 위해서 NSTA, NAEP, ASPA가 제시한 평가 목표, Dillashaw 등이 평가 문항을 개발할 때 사용한 평가 목표, APU 평가틀과 허명의 탐구 평가표에 제시된 평가 목표 등을 참고로 하여 평가 목표를 다음과 같이 기술하였다.

관찰 - 인간의 감각기관을 최대한 활용하여 사물이나 자연 현상에 대한 1차적 정보를 있는 그대로 수집할 수 있다.

분류 - 관찰에서 수집된 사실을 근거로 하여 사물이나 자연현상의 상호 유사성, 상이성, 관련성 등을 파악할 수 있다.

문제인식 - 자연 현상이나 사물의 관찰에서 규칙성을 찾거나 변화의 원인을 규명하고 해결하는데 필요한 문제를 찾아 낼 수 있다.

측정 - 사물의 현상을 양적으로 관찰하기 위하여 측정 대상에 맞는 도구의 선택과 측정 방법을 알고 측정할 수 있다.

추리 - 실험이나 자연 현상에서 관찰 가능한 사실이

표, 그래프, 그림, 언어 등으로 제시되면, 그 사실로부터 직접 관찰이 안되는 새로운 사실을 이끌어 낼 수 있다.

가설 - 독립 변인과 종속 변인이 구체적으로 제시된 실험이 주어지면, 이로부터 검증 가능한 가설을 제안하거나 확인할 수 있다.

변인 - 가설이나 연구 문제가 주어지면 주어진 문제나 가설을 검증하기 위해서 반드시 통제되어야 할 변인을 확인하거나 통제할 수 있다.

실험 - 가설이나 연구 문제가 주어지면 그 문제나 가설의 타당성을 객관적으로 검증하기 위해 적절한 실험 방법을 설계하거나 실험 설계에 의하여 실험을 수행할 수 있다.

결과해석 - 관찰이나 실험에서 얻어진 정리된 자료(표, 그래프, 그림 등)를 통하여 결과를 해석할 수 있다.

결론 - 실험내용과 결과가 언어나 그래프로 제시되면 실험 결과를 근거로 이들을 종합하여 결론을 도출할 수 있다.

2) 이원 목적 분류표 작성

문항제작에 앞서 과학의 내용영역을 <표 2>와 같이 물질과 에너지, 생명과 환경, 지구와 우주로 나누었다.

그리고 탐구과정 요소를 10개 항으로 하여 각 내용영역이 과학 탐구과정 각각의 요소에 해당되도록 10문항씩을 개발하기로 하였다.

<표 2> 과학 하위 내용 영역 분류표

상 위 영 역	하 위 영 역
물질과 에너지	물질, 힘과 운동, 빛과 열, 전·자기, 공통
생명과 환경	다양성, 형태 특성, 조절 생식과 발생, 생물과 환경
지구와 우주	암석, 기상, 지구 운동, 지구의 역사, 우주

과학의 각 내용영역은 현행 초·중·고등학교의 과학과 교육과정(문교부, 1987 1988)을 참고로 하여 국민학생의 수준에 맞도록 구분하였다. 하위 영역 중 물질과 에너지 영역은 물질, 힘과 운동, 빛과 열, 전자기, 공통으로 나누었다. 생명과 환경은 생명의 다양성, 형태특성, 조절과

<표 3> 물질·에너지 영역 탐구능력 측정 이원 분류표

번호	탐구요소	내 용	물질	힘·운동	빛과열	전·자기	공통	정답
1	관찰	역세의 혼합	√					
2	분류	분류의 기준					√	
3	문제인식	활의 부피계량	√					
4	측정	질량측정	√					
5	추리	자석의 성질				√		
6	가설	빛의 반사			√			
7	변인	열 계량			√			
8	실험	회로 잇기				√		
9	결과	그래프 해석		√				
10	결론	요오드 반응	√					

<표 4> 생명과 환경 영역 탐구능력 측정 이원 분류표

번호	탐구요소	내 용	다양성	형태	조절	생식	환경	정답
1	관찰	현상의 관찰	√					
2	분류	식물의 분류	√					
3	문제인식	환경 보존					√	
4	측정	생체 측정		√				
5	추리	꽃잎의 색깔			√			
6	가설	동물의 특징		√				
7	변인	엽록체의 생성			√			
8	실험	광산작용			√			
9	결과	유기물의 분해					√	
10	결론	생식				√		

<표 5> 지구와 우주 영역 탐구능력 측정 이원 분류표

번호	탐구요소	내 용	암석	기상	지구운동	지구역사	천문	정답
1	관찰	현상의 관찰		√				
2	분류	암석의 분류	√					
3	문제인식	우주 탐색					√	
4	측정	태양 고도 측정			√			
5	추리	기온의 변화		√				
6	가설	태양과 달의 거리					√	
7	변인	효과는 물의 작용			√			
8	실험	회색안물기				√		
9	결과	계절의 변화		√				
10	결론	지구의 역사				√		

항상성, 생식과 발생, 유전, 생물과 환경으로 나누었다. 지구와 우주 영역은 암석, 기상, 지구운동, 지구의 역사, 우주와 천문의 5개 소 범주의 하위 내용 영역으로 분류하여 이원 목적 분류표를 <표 3> <표 4> <표 5> 와 같이 제작하였다.

3) 문항제작

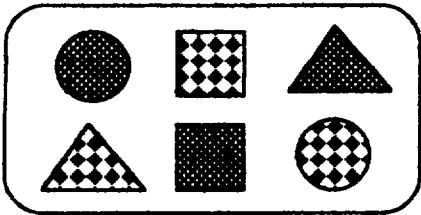
문항의 개발은 이원 목적 분류표에 의하여 다음과 같이 카드를 만들어 제작하였다.

평가 문항 제작지

내용영역	물질·에너지(V) 생물() 지구와 우주()		
탐구과정	관찰()분류()문제()측정()추리()		
요소	가설()변인()실험()결과()결론()		
문항번호	2	정답	3

(문항 예)

영희는 다음 도형을 분류하려고 한다. 분류의 기준으로 적당할 것은 ?



- ① 모양과 크기 ② 무늬와 크기
- ③ 모양과 무늬 ④ 둥근 것과 모난 것

4) 문항점검

내용영역에 따라 과학 탐구과정 요소별로 개발된 평가문항은 1차 현장검사 실시와 동시에 7명의 전문가와 과학교육 경험이 풍부한 현장교사 5명에게서 타당도, 객관도, 문항의 명료성을 점검 받았다.

2차 점검은 1차 점검시 비교적 지적이 많았던 3명의 전문가로부터 점검을 받았다.

그 후 이들 전문가의 의견과 1차 현장 검사의 결과 분석을 통하여 문항을 수정·보완하였다.

가) 내용 타당도 점검

내용 타당도 점검은 문항별로 적절 1, 부적절 0으로 하고 평가 목표에 근거하여 각 평가 문항이 측정하고자 하는 탐구과정 요소를 점검표에 직접 기록하도록 하는 방법으로 처리하였다.

점검 결과 1차는 내용 타당도 지수가 88.8 % 2차는 91.1%로 비교적 만족할만한 지수를 얻었다.

<표 6> 2차 내용타당도 점검표

내용 전문가	가	나	다	계
일치한 문항 수	30	23	29	82
타당도 지수(%)	100	76.7	96.7	91.1

나) 정답의 객관도 점검

1차 점검에서 12명의 의뢰자 중 답안을 작성해 준 8명의 결과를 분석한 결과 총 240문항 중 225문항의 정답이 연구자와 일치하여 93.8%의 객관성을 확보하였다.

2차 점검은 총 90개 문항중 전부가 문항 개발자와 일치율을 보여 정답의 객관도는 100%를 확보하였다. 정답의 객관도 만큼은 만족한 평가도구이다.

다) 문항의 명료성 점검

내용 전문가 및 현장 교사들에게 평가 문항과 문항 명료성의 점검표를 제시하고 다음의 문항 점검표 각 항목에 문제가 있다고 판단되는 점을 점검표에 기록하도록 하였으며 그 결과를 평가 문항 수정에 이용하였다.

평가 문항의 명료성 점검표는 다음과 같다.

- (1) 평가문항이 의도하는 구체적 평가 조건과 요구행동에 알맞게 문항 장면을 선택하여 구성하고 있는가 ?
- (2) 평가문항 장면이 국민학생의 과학 탐구능력 수준을 측정하는데 적절한가 ?
- (3) 문항 장면이 탐구능력을 측정하는데 보다 새로운 내용으로 구성되어 있으며, 피험자의 흥미와 동기 유발에 적절한가 ?
- (4) 문항 장면이 특수한 과학 지식이 아닌 과학의 전 영역에서 선택한 일반적인 내용으로 구성되어 있는가 ?
- (5) 문항 장면이 논리적이고 정확하게 기술되어 있는가 ?
- (6) 각 문항은 문법적으로 정확하며 적절한가 ?
- (7) 애매한 질문이나 지나치게 길고 복잡한 지시문은 없는가 ?
- (8) 중요하고 특별한 학습목표를 측정하기 위한 경우를 제외하고 부정적인 지시문을 지나치게 많이 사용하고 있지 않은가 ?
- (9) 문항의 배열순서와 각 문항내의 답지의 배열순서

는 적절한가 ?

(10) 문항내에서 제시된 표, 도표, 그래프, 그림은 적절한가 ?

(11) 문제지와 답지의 전체적인 구성은 적절한가 ?

(12) 각 평가 문항의 정답은 1개인가 ?

(13) 선택형 보기 중에서 오답지들은 그럴 듯 한가 ?

(14) 질문하는 문장과 선택형 보기 사이에 언어상으로 암시를 주는 것은 없는가 ?

(15) 선택형 보기의 문장 길이나 형태에 정답을 제공하는 단서는 없는가 ?

(16) 선택형 보기내의 정답의 위치는 무선적이며, 그 수는 대체로 비슷한가 ?

(17) 선택형 보기가 숫자일 경우 숫자의 크기 순서로 나열하였는가 ?

위와 같은 점검표를 의한 점검 결과 지적된 내용은 2차 현장검사를 위한 문항 수정시에 반영하였다.

5) 문항의 현장검사 및 수정보완

가) 1차 현장 검사

연구 대상을 표집하기 전에 광주 직할시 봉선국민학교의 5학년생 8명을 무작위 추출하여 예비 검사를 실시하였다. 이 예비 검사에서 소요 시간, 평가 도구에 쓰인 용어, 득점율 등을 간단히 측정하였다. 그 결과 8명 전원이 35분 내에 문제를 해결했으며 모르는 단어가 있으면 질문을 허용했으나 단 1건의 질문도 없었다. 그리고 정답 반응율이 71%이었다. <표 7>에서 보듯이 개발한 평가도구를 현장에 투입해도 별반 무리가 없을 것으로 간주하고 현장 검사를 시행하였다.

<표 7> 현장 검사를 위한 예비 검사 결과

대상	이원	소요시간	정답율	질문 수
5학년	8	29분	71%	0

(1) 연구 대상의 표집

1차 현장 검사를 위한 연구 대상 학교의 표집은 우선 인구 100 만 이상의 대 도시 1개 학교와 인구 5만 이하의 읍 지역 1개 학교를 무선적으로 표집하고 그 다음에 4 5 6학년 각 1 학급씩을 무작위로 표집하는 무선 집단표집 방법을 이용하였다. 국민학교는 각 반에 남녀가

공학을 하므로 성별은 감안하지 않았다.

<표 8> 현장 검증의 연구 대상 표집

지역	학 교 명	표집수 (단위 : 명)			합계
		4학년	5학년	6학년	
광주 읍	봉선국민학교	46	52	52	150
	전남화순국민학교	38	40	41	119
계	2개교	84	92	93	269

1차 현장 검사는 1992년 10월 중(10월 6일~10월 8일)에 실시하였다.

(2) 문항 분석 및 평가 도구의 수정 보완

1차 현장 검사에서 학생들의 반응을 분석한 결과는 <표 9>와 같다.

<표 9> 1차 현장검사의 분석결과

학 년	4	5	6	합계
학생수	84	92	93	269
평 균	14.08	17.36	17.36	16.34
득점율	46.94	57.86	57.85	54.45
표준편차	4.07	4.15	4.78	4.61
신뢰도	0.63	0.72	0.76	0.73
표준오차	2.48	2.20	2.34	2.39

㉠ 신뢰도

본 연구에서는 상관 계수의 단점과 Hoyt(1941)가 제안한 변량 분석의 한계를 극복하고 측정학적 결함이 적은(Nitko, 1983; 황정규,1989) 방법을 선택하였다. 맞는 것을 1, 틀린 것을 0으로 채점하고 Cronbach' α 와 계산방법이 같은 K-R 20을 이용했다.

신뢰도는 <표 9>에 제시된 바와 같다.

㉡ 문항 난이도

본 연구에서는 평가 도구가 4지 선다형 객관식 문항으로 제작되어 있으므로 각 문항에서 추측에 의한 정답의 가능성을 제외한 Davis (1951)의 난이도 계산 공식을 적용하였다.

난이도는 <표 10>과 같다.

<표 10> 문항 난이도 및 변별도 지수

	4학년	5학년	6학년	평균
난이도	29.3	44.14	44.12	39.83
변별도	0.33	0.34	0.40	0.39

이는 전반적으로 어려운 문항이며 전체적으로 허용 범위를 벗어난 문항이 7문항 이었다. 이 중 한 문항은 너무 쉽게 출제되었으며, 나머지 6문항은 어려웠다. 이들 문항 중 쉬웠던 문항은 어렵게, 어려웠던 문항은 쉽게 수정하거나 교체 하였다.

㉞ 문항 변별도

본 연구에서는 검사의 총점이라는 내적 준거 의하여 변별도 지수<표10>를 산출했다.

표에서 보는 바와 같이 전체적인 변별도는 만족하다. 그러나 각 문항별로는 다섯 문항의 변별도가 너무 낮아 이 역시 문항을 수정하였다.

나) 2차 현장 검사

2차 현장 검사는 1992년 11월 1일 부터 11월 6일에 걸쳐 실시하였다.

(1) 연구 대상의 표집

연구 대상 학교의 선정은 인구 1000만 이상과 100만 이상의 대도시에서 각 1개 국민학교씩 2 개 국민학교, 인구 30만 이상과 5만 이상의 중·소 도시에서 각 1개 국민학교씩 2개 국민학교, 그리고 읍·면 지역에서 각 1개 국민학교씩 2개 국민학교를 선정하였다.

<표 11> 2차 현장 검사의 연구 대상 표집

지역	학교명	표집수(당위:명)		합계
		5학년	6학년	
대도시	서울 계남국민학교	56	54	212
	광주 봉선국민학교	49	53	
중·소	원주 명륜국민학교	45	47	184
	공주교대 부속국교	47	45	
읍·면	경북 현풍국민학교	45	36	163
	제주 하귀국민학교	35	37	
계	6개 국민학교	277	272	549

또 지역별로는 수도권 지역에서 1개 국민학교, 영동권 1개 국민학교, 중부권 1개 국민학교, 호남권 1개 국민학교, 영남권 1개 국민학교, 제주권 1개 국민학교 등

전국적으로 6개 국민학교 5·6학년에서 한 학급씩 549명을 무선표집하였다. 연구 대상의 학년은 4·5·6학년을 대상으로 하였으나 분석결과 4학년과 5·6학년간의 평균, 표준편차, 신뢰도, 난이도, 변별도 등에서 격차가 너무 심하므로 통계에서 4학년은 제외하고 5·6학년 549명을 대상으로 통계처리를 하였다.

연구 대상 학교 및 학생 수는 <표 11>과 같다.

(2) 문항분석

1차 현장 검사의 문항 분석과 같은 방법으로 2차 현장 검사를 분석하였다. 그 결과는<표 12>와 같다.

<표 12> 2차 현장 검사의 분석결과

학 년	5	6	합계
학생수	277	272	549
평균	18.80	18.31	18.56
득점율	62.66	61.03	61.85
표준편차	5.29	4.90	5.11
신뢰도	.80	.76	.78
표준오차	2.37	2.42	2.40

㉟ 신뢰도

신뢰도(KR-20)는 5학년이 0.80, 6학년 0.76으로 이 평가 문항은 5학년을 대상으로 할 때 더 신뢰도가 높았다. 전체의 신뢰도는 0.78이다. 이 값은 1차 현장 검사보다는 크게 향상되었다.

㊱ 문항 난이도

문항 난이도 지수 평균은 <표 13>과 같이 5학년 50.29%, 6학년 48.05%, 전체 49.13 % 이다.

<표 13> 2차 현장 검사의 난이도

문항난이도 지수(%)	5학년	6학년	전체
0 ~ 20.0	3	1	
20.0 ~ 40.0	7	10	10
40.0 ~ 60.0	9	10	9
60.0 ~ 80.0	11	9	11
평균	50.3	48.1	49.1

<표 13>에서 보는 바와 같이 전체적으로 난이도 지수의 허용 범위를 벗어난 문항은 없으며 학년별로 범위를 약간 벗어난 문항은 4문항(5학년에서 5, 11, 24번 문항 6학년에서 5번 문항)이 있었다. 그리고 문항 난이도가 20.0%에서 80.0%사이에 고르게 산재하여 바람직한 분포를 보이고 있다.

<표 14> 문항별 난이도 지수 (%)

N=549			
문항번호	5학년	6학년	전체
1	20.1	20.0	20.1
2	57.1	52.5	54.8
3	76.4	72.2	73.8
4	62.5	56.9	59.7
5	*19.6	22.4	21.1
6	58.6	39.2	49.0
7	76.9	74.0	75.5
8	50.0	38.2	43.7
9	58.1	65.2	61.6
10	64.4	70.1	67.2
11	*19.1	21.6	20.3
12	74.5	75.9	75.2
13	75.5	78.9	77.2
14	24.5	34.8	29.1
15	63.9	57.8	60.9
16	63.4	58.3	60.9
17	69.2	72.5	70.9
18	38.9	21.1	30.1
19	62.0	47.5	54.8
20	65.3	67.2	66.2
21	78.3	79.9	79.1
22	46.6	47.5	47.0
23	38.9	23.1	31.0
24	*11.4	42.6	26.9
25	23.9	*19.0	21.3
26	47.0	43.6	45.4
27	48.0	32.8	40.5
28	21.6	21.6	21.5
29	38.8	33.3	35.9
30	54.3	51.9	53.1
평균	50.3	48.1	49.1

*: 허용범위를 벗어난 문항

㊤ 문항 변별도

문항 변별도 지수는 <표 15>과 같이 5학년이 0.46, 6학년 0.40, 전체가 0.42로 비교적 만족할 만한 지수이다. 그러나 전체적으로 변별도 지수의 허용 범위를 0.20 이상 0.80으로 볼 때 허용 범위를 벗어난 문항은 없으나 학년별로는 5학년에서 2 문항(1, 14번 문항), 6학년 1 문항(23번 문항)이 약간 벗어났다.

<표 15> 2차 현장 검사의 변별도

문항수 = 30			
변별도 지수(D.I)	5학년	6학년	전체
0.00 ~ 0.20	2	1	1
0.21 ~ 0.40	10	14	11
0.41 ~ 0.60	15	14	18
0.61 ~ 0.80	3	1	.
평균	0.46	0.40	0.42

<표 16> 문항별 변별도 지수

N=549			
문항번호	5학년	6학년	전체
1	*.19	.30	.22
2	.31	.34	.29
3	.40	.22	.29
4	.61	.56	.58
5	.51	.38	.42
6	.57	.33	.47
7	.51	.43	.45
8	.57	.47	.53
9	.55	.38	.47
10	.40	.40	.37
11	.31	.30	.31
12	.44	.36	.39
13	.39	.26	.30
14	*.11	.30	.20
15	.45	.63	.52
16	.29	.44	.36
17	.37	.51	.42
18	.49	.32	.41
19	.59	.52	.55
20	.56	.47	.51
21	.33	.33	.31
22	.72	.47	.57
23	.45	*.15	.28
24	.65	.52	.55
25	.57	.43	.47
26	.52	.49	.52
27	.76	.45	.60
28	.39	.25	.32
29	.45	.58	.49
30	.40	.51	.43
평균	.46	.40	.42

*: 허용범위를 벗어난 문항

<표 17> 과학 탐구능력 검사의 탐구과정 요소별 반응분포

탐구 요소	5학년			6학년			전체		
	%	P	D.I	%	P	D.I	%	P	D.I
관찰	54.4	39.3	.28	55.4	40.5	.31	54.9	39.8	.28
분류	69.5	59.4	.49	69.0	58.6	.39	69.3	59.0	.42
문제인식	72.7	63.6	.41	68.3	58.0	.21	70.5	60.7	.29
측정	49.3	32.8	.46	58.6	44.8	.46	53.9	38.6	.44
추리	51.9	35.8	.51	49.8	33.1	.48	50.8	34.4	.47
가설	67.3	56.3	.46	60.3	47.0	.42	63.8	51.8	.45
변인통제	73.5	64.7	.55	69.9	59.8	.46	71.7	62.3	.49
실험	52.3	36.8	.48	45.2	27.0	.35	48.9	31.8	.42
결과해석	64.6	53.0	.53	61.5	48.7	.49	63.0	50.8	.50
결론	71.0	61.3	.45	72.3	63.1	.46	71.6	62.2	.44
평균	62.7	50.3	.46	61.0	48.1	.40	61.9	49.1	.42

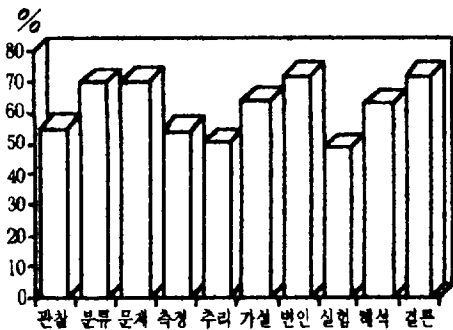
(3) 과학 탐구과정 요소별 반응 결과

2차 현장 검사에서는 본 연구에서 선정한 과학탐구 요소별 반응 결과도 분석하였다.

그 결과는 <표 17>에 제시한 바와 같이 전체적으로 볼때 탐구 과정 요소별 난이도는 관찰 39.8%, 분류 59.0%, 문제 인식 60.7%, 측정 38.6%, 추리 34.4%, 가설 51.8%, 변인 62.3%, 실험 31.8%, 결과 해석 50.8%, 결론 62.2%이다. 이 결과로 보면 변인요소의 문항이 난이도가 높아 우리나라의 학생들이 비교적 변인통제나 조절을 잘 한다고 볼 수 있다. 그럼에도 불구하고 실험에 관한 과정 요소 능력이 낮게 나왔는데 이는 실험 설계 및 실험의 수행에 관한 문항 제작상의 기술 및 피험자의 실험에 관한 탐구 능력의 부족에 기인한 복합적인 원인으로 추측된다. 또 과학의 시작이라 할 수 있는 관찰 능력이 낮게 나온 것은 우리의 자연과 교육이 과정보다는 결과에 치중한 원인으로 분석된다.

(4) 내용 영역별 분석 결과

과학 탐구능력 검사도구를 과학의 내용 영역별로도 활용할 수 있도록 하기 위하여 문제 번호 부여부터 탐구과정 요소별로 내용영역에 따라 일관성있게 하였다.



<그림 1> 탐구 요소별 정답율

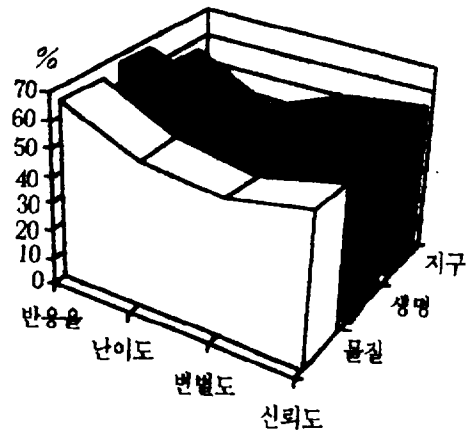
<표 18> 내용 영역별 분석 결과

N=549

영역	A	B	C	D
문항수	10	10	10	30
득점	6.44	6.59	5.51	18.56
득점율	64.48	65.92	55.16	61.85
표준편차	2.05	1.97	2.19	5.11
신뢰도	.56	.53	.57	.78
표준오차	1.36	1.35	1.44	2.40
난이도	52.65	54.56	40.17	49.13
변별도	.49	.47	.53	.42

A: 물질과 에너지 B: 생물과 환경
C: 지구와 우주 D: 전체

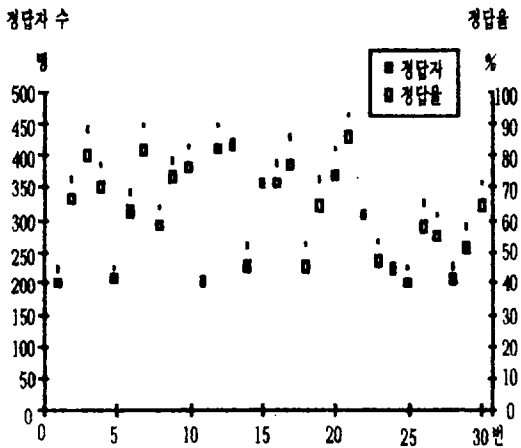
그리고 1번 부터 10번 문항까지는 물질과 에너지, 11번



<그림 2> 내용 영역별 분석 그림

부터 20번 문항까지는 생명과 환경, 21번 부터 30번까지는 지구와 우주 영역의 문항으로 구성하였다. 따라서 본 탐구 능력 평가 도구는 과학 일반에 관한 측정도구이지만 내용 영역별로도 분리하여 활용할 수 있도록 구성되어 있다.

2차 현장 검사의 내용 영역별 분석 결과는 <표 18>에 제시된 바와 같다.



<그림 3> 문항별 정답자 수 및 정답률

내용 영역별 정답률은 생명과 환경이 65.92%로 가장 높게 나왔고, 낮게 나온 지구와 우주 영역이 55.16%로 그 차는 10.76%이다. 이와 같은 결과는 국민학생들은 무생물보다는 살아있는 생물체에 더욱 관심을 갖는 데에 기인한 것으로 생각된다.

IV. 결 론

본 연구의 목적은 우리나라 국민학교 고학년(5,6학년) 학생에게 적용가능하며 타당도와 신뢰도가 높은 과학 탐구능력 평가도구를 개발하는 것이다.

본 연구에서는 이론과 실제의 차이를 가장 극소화 할 수 있어 교육자료를 개발할 때 자주 사용하는 R&D 과정을 택하였다.

개발된 평가도구는 내용타당도, 정답의 객관도, 문항의 명료성 등을 2 차에 걸쳐 과학교육 전문가 7명과 과학교육 전공 현장교사 5명이 점검하였다. 그 결과 내용 타당도면에서 91.1%, 객관도 100%가 일치했다. 또 이들이 문항의 명료성에 대하여 지적한 내용도 수정 보완하였다.

1차 현장 검사는 도시와 읍지역 두 학교에서 269명에게 실시하였다. 1차 현장검사의 분석결과를 근거로 하여 문항을 수정 보완한 후 전국에 걸쳐 대도시, 중·소도시, 읍·면지역의 학교에서 549명을 대상으로 2차 현장검사를 실시하였다. 보완한 평가도구는 내용타당도 지수 91.1%, 정답의 객관도 100%, 2차 현장 검사의 실시 결과는 표준편차 5.11, 신뢰도(K-R 20) 0.78, 문항 변별도 지수 평균 0.42, 전체 정답률 61.85%, 문항 난이도 지수 평균 49.13, 측정의 표준오차 2.40이다.

본 연구에서 개발된 평가도구는 국민학생의 과학탐구 능력을 측정할 뿐만 아니라 물질과 에너지, 생명과 환경, 지구와 우주 내용으로 분리하여 평가를 실시할 수 있다. 그리고 탐구학습을 위한 진단 평가, 또는 총괄 평가로도 활용될 수 있다. 또한 일선 학교에서 과학교육 성취도 평가 문항 제작시에 탐구과정 요소 문항 개발의 지침으로 활용될 수 있을 것이다.

그러나 탐구능력의 측정은 본 연구에서 개발한 지필 검사만으로는 충분하지 않으므로, 다른 방법의 평가(실험 실기 검사, 실험 보고서, 과제, 점검표, 면담 등)와 병행하여 실시되어야 한다.

참 고 문 헌

교육부, 1992, 국민학교 교육과정. 서울, 대한교과서 주식회사.

김명환·권오범·김병업·김재경·임정환, 1981, 초등과학 교육에서 탐구능력의 평정 척도 작성과 그 적용에 관한 연구. 대구 교육대학 논문집.

김종진, 1974, 듀이와 부르너의 교육 방법 분석. 서울대학교 교육대학원 석사학위 논문.

김주훈·이양락, 1984, 국민학교 자연과 평가의 원리와 실제. 한국교육개발원 연구보고, TR 84-7.

문교부, 1973, 1982, 1987, 국민학교 교육과정. 서울, 교학도서주식회사.

문교부, 1988, 고등학교 교육과정. 서울, 대한교과서 주식회사.

박승재, 1989, 과학 학습평가 - 국민학교 -. 연구보고' 88, 문교부.

박승재, 1991, 과학적 탐구 사고력 평가. 서울대학교 사범대학 물리교육과 물리학습 연구실.

송영호·채남식, 1984, 과학적 탐구학습을 위한 탐구과정 요소의 평가에 관한 연구. 전주 교육대학 과학교육연구 논문집, 10 : 1~37.

우종욱·정완호·권제술·최병순·정진우·허명, 1991, 국민

- 학교 자연 교과서 개발체제 분석 및 평가 연구.
한국교원대학교 과학교육연구소.
- 이 무, 1992, 과학적 탐구 사고력 평가 문항 형태에 관한 연구. 서울대학교 박사 학위 논문.
- 이연우, 1989, 과학 탐구 능력 측정을 위한 표준화 검사지 개발 -중학교 2학년의 자료 분석과 해석능력을 중심으로-. 한국교원대학교 석사학위 논문.
- 이종기, 1988, 고등학생의 과학 탐구능력 측정을 위한 평가 도구 개발. 한국교원대학교 석사학위 논문.
- 이항노, 1991, 고등학생의 과학 탐구 능력 측정을 위한 평가 도구 개발 -지구과학 소재를 중심으로-. 한국교원대학교 석사 학위 논문.
- 장남기 외, 1990, 탐구과학교육론. 서울, 교육과학사.
- 황정규, 1989, 학교학습과 교육평가. 서울, 교육과학사.
- Bruner, J.S., Goodnow, J.J., & Austin, G.A., 1956, A Study of thinking. New York.
- Davis, F.B., 1951, Item selection technique, in E.F. Lindquist(Ed.) Educational Measurement. Washington, D.C. : American Council on Education.
- Dewey, J., 1933, How We Think. Boston:D.C. Heath and Co.
- Dillashaw, G.H., & Okey, J.R., 1980, A Test of the Integrated Science Process Skills for Scondary Science Students. Science Education, 64.
- Doran, R.L., 1978, Measuring the 'Processes of Science' Objectives. Science Education, 62 : 19~30.
- Hoyt, C.J., 1941, Test reliability estimated by analysis of variance. psychometrika, 6 : 1953~1960.
- Hur, M., 1984, Evaluation of Inquiry Activity in Science Curricula. Dictoral Dissertation, Teachers College, Columbia University.
- Klopfer, L. E., 1971, "Evaluation of Learning in Science," In B.S. Bloom, J.T., Hasting and G.F. Madaus(Eds.), "Handbook on Formative and Summative Evaluation of Student Learning", McGraw-Hill : New York.
- Moliter, L.L., & George, K.D., 1976, Development of a Test of Science Process Skills. Journal of Research in Science Teaching, 13: 405~412.
- Nelson, M.A., & Abraham, E.C., 1973, Inquiry Skill Measures, Journal of Research in Science Teaching, 10 : 291~297.
- Nitko, A. J., 1983, Educational Test and Measurement an Introduction. New York, Harcourt Brace Jovanovich, Inc.
- Padilla, M.J., Cronin, L.L., & Twiest, M., 1985, The Development and Validation of a Test of Basic Process Skills. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching.
- Schwab, J.J., 1964, The Teaching of science. Cambrige, MA : Harvard University Press.
- Smith, K.A., Welliver, 1990, The Development of a Science Process Assessment for Fourth-Grade Students. Journal of Research in Science Teaching, 27(8) : 717~738.
- Welch, W.W., Klopfer, L.E., Aikenhead, G.S., & Robinson, J.T., 1981, The Role of Inquiry in Science Education: Analysis and Recommendations. Science Education, 65: 33~50.

(ABSTRACT)

Development of a Test of Science Inquiry Skills Elementary School Students

Wan-Ho Chung Myung Hur

(Korea National University of Education) (Ewha Womans University)

Kyong-Yong Eun

(Korea National University of Education)

The purpose of this study is to develop an reliable instrument for evaluating science inquiry skills through an R&D procedure.

A total of ten science inquiry skills were selected for the development of the instrument, ie, observation, classifying, recognizing problems, measuring, inferring, formulating hypothesis, controlling variables, experimenting, interpreting and drawing a conclusion.

And three items were developed for each science inquiry skill, totaling up to thirty items.

The content areas of developed items are divided into three categories, material and energy, life and environment, and the earth and the space. There are 10 items for each category.

The content validity and the objectivity of developed items were checked, verified in the two field trials and revised according to the analysis of items by seven experienced specialists with the degree of doctors in science education and five teachers who were studying for their master's degree in science education.

According to the results of the second field trial, the content validity of the instrument was 91.1%, reliability(K-R 20) 0.78, difficulty index 49.13%, discrimination index 0.42, mean achievement 61.85%, standard deviation 5.11, and standard error 2.40.

Considering the above results, the developed instrument in this study is regarded as a appropriate instrument for evaluating science inquiry skills of elementary school students.