

## 물리 탐구 실험의 평가를 위한 도구의 개발과 분석

김미경 · 오희균 · 박종원  
(전남대학교)

(1995년 9월 17일 받음)

### I. 연구의 배경과 목적

1960년대 추진된 새로운 과학개혁운동에 의해 만들어진 학문중심 교육과정에서는 지식의 전수나 답습이 아닌 지식의 창조, 지식의 구조화, 탐구 중심 혹은 발견 중심 교육방법을 강조하였다(Bruner, 1966; Schwab, 1966).

이러한 과학 교육의 시대적 요청에 따라, 우리나라의 학교교육계에서도 1973년(중학교)과 1981년(고등학교)에 교육과정의 전면적인 개편을 통해 새로운 교육사조를 도입하였으며, 제5차 교육과정에서는 과학탐구력의 배양이 과학 교육의 주요 목표로 설정되었다(교육부, 1988). 더욱이 1994년도부터 시행되고 있는 대입 수학능력시험에도 과학적 탐구력이 포함되어 있다.

평가 부분에 있어서는 현행 교육과정(5차교육과정)에서도 교과 목표와 더불어 평가상의 유의점을 설정하고, 평가관점 및 방법을 제시하여 인지적 영역인 지식과 이해뿐 아니라 탐구수행능력, 정의적 영역 및 실험과 실습기능 등을 고루 평가하도록 하고 있다. 그러나 교육 현장에서는 실험에 대한 평가가 잘 이루어지고 있다고 보기 힘들다. 예를 들면, 학교에서 실제로 실시하는 실험 평가는 실험보고서에 의하거나 교사의 주관적인 판단에 의존하는 경향이 강한데 이러한 방법은 객관성이 결여되어 있어서 그 결과에 대한 신뢰도나 타당도를 인정받기 힘들다. 따라서 이러한 평가 방법은 과학적 탐구능력의 향상에 별로 도움을 주지 않는다고 할 수 있다.

그러므로 교육현장에서는 객관적이고 타당성이 있으면서도 교사가 쉽게 적용할 수 있는 탐구실험능력의 평가방법이 요구되고 있다. 물론 여러 연구들을 통해 과학적 탐구능력을 평가하기 위한 문항들이 개발되고 적용된 바 있다(TIPS II, 1985; 이종기, 1988; 이연우와 우종욱, 1991; 정완호 외, 1993). 그러나 그러한 평가도구들 중 대부분이 특정 내용과

무관하게 개발되었기 때문에 물리 분야 중 힘과 운동 분야와 같이 특정 실험에 관련된 탐구능력을 평가하는 데에는 실제적인 어려움이 있다.

본 연구의 목적은 현장 과학교육에서 효과적이고 간편하게 사용할 수 있는, 그리고 보다 더 특정 실험에 관련된 평가 도구를 개발하는데 있다. 이를 위해 과학 교과서(고등학교 물리)에서 실험 내용을 선정하여 실험을 실시한 후에 탐구 요소별로 학생의 탐구능력을 평가할 수 있는 지필형 문항을 개발하고자 한다.

자세한 연구 목표는 다음과 같다.

- 1) 탐구 실험 평가를 위한 평가틀을 개발한다.
- 2) 고등학교 실험 내용을 선정하여 실험 안내서 및 보고서를 개발하여 실험을 실시한다.
- 3) 실험 평가틀에 기초하여 평가 문항을 개발하여 적용한다. 개발한 평가도구를 실험을 실시한 집단에 적용하여 실험을 잘한 학생과 잘하지 못한 학생을 구분할 수 있는 문항을 추출한다.
- 4) 개발한 평가도구를 실험을 실시하지 않은 집단에도 적용하여, 실험을 실시한 집단과 실시하지 않은 집단을 구분할 수 있는 문항을 추출한다.

물론 종합적으로 과학적 탐구능력을 평가하기 위해서는 선다형 지필 문항만으로는 한계가 있다. 본 연구에서 개발된 문항은 선다형 지필 문항이므로 교사가 쉽게 적용할 수 있으나 전반적이고 종합적인 능력을 평가한다고 보기는 힘들다. 따라서, 앞으로 여러 가지 형태의 평가문항 개발이 계속되어야 할 것이다.

### II. 선행 연구

1960년대 이후 탐구학습이 주류를 이루면서 여러 학자들은 탐구학습을 평가하기 위한 평가틀을 개발하고 이 평가틀

에 맞추어 평가도구도 개발되었다. 이러한 평가도구들 중 몇 가지를 살펴보면 다음과 같다.

Burns 등은(Burns, Okey & Wise, 1985) 과학적 탐구 사고력 측정을 위해 TIPS를 약간 수정하여 TIPS II를 개발하였는데, TIPS II는 조작적으로 정의하기, 변인통제하기, 검증 가능한 가설 선택하기, 그래프로 나타내기 및 변인들 간의 관계 세우기, 적절한 가설 검증 방법 선택하기와 같은 5개 목표에 따라, 총 36개 문항으로 이루어져 있다. 본 검사는 총 459명의 학생들을 대상으로 실시하였는데 평균 19.14, 난이도는 0.53, 변별도는 0.35, 신뢰도는 0.86으로 나왔다.

이종기(1988)는 우리 나라 고등학생을 대상으로 TSIS(Test of Science Inquiry Skills)를 개발하였다. 이 도구는 12개 탐구요소(가설설정, 변인통제, 실험설계, 자료의 변형 및 숫자 계산, 실험값을 그래프로 나타내기, 추리, 상관관계 결정, 인과관계 결정, 예상(내삽과 외삽 포함), 결론, 일반화 또는 모델형성, 평가)를 선정하고 각 요소당 3문항씩 모두 36문항으로 구성되어 있다. 평가 시간은 50분간, 적용 대상은 고등학교 1,2,3 학년이며, 변별도는 0.46, 문항 난이도는 58.9%, 신뢰도(K-R 20)는 0.86이었다.

이연우와 우종욱(1991)은 중학교 2학년 8학생을 대상으로 하여 과학 탐구능력 측정을 위한 표준화 검사지를 개발하였다. 개발된 검사지는 탐구 능력 중에서 데이터 해석 및 분석 능력에만 한정하였으며, 구체적인 요소로 추리, 관계 설정, 원인설명, 예상을 선정하여, 각 요소별로 5개 항목씩 총 20문항으로 되어 있다. 신뢰도는(KR-20) 0.69, 변별도 평균은 0.39로 나타났다.

이 무(1992)는 1990년 12월 19일 중앙교육 평가원에서 실시한 과학교육 적성시험 1차 실험 평가를 가지고 객관식 문제가 갖는 추측의 요인을 최소화하면서 측정하고자 하는 과학적 탐구 사고력을 적어도 주관식 평가에 준할만큼 타당하고 신뢰롭게 평가할 수 있는 객관식 문항으로 두번 선택 평가 문항과 세번 선택 평가 문항을 개발하여 분석 연구하였다.

정완호 등(정완호 외, 1993)은 국민학교 고학년 학생들에게 적용 가능한 과학탐구과정요소를 관찰, 분류, 문제 인식, 측정, 추리, 가설 설정, 변인 조절, 실험, 결과 해석, 결론의 10가지로 선정하고 그에 따른 탐구과정 평가 목표를 진술한 다음 총 30개의 평가 문항을 개발하였다. 개발결과, 변별도 평균은 0.42, 신뢰도(KR-20)는 0.87로 나타났다.

우종욱 등(우종욱 외, 1991)은 과학적 탐구능력 측정을 위한 행동요소를 관찰, 측정, 가설설정, 실험 설계 및 변인 통제, 추리, 예상(내삽, 외삽 포함), 자료처리 및 해석, 조작적 정의, 모델형성 또는 일반화 및 결론 내리기로 추출하고 각 요소에 대한 평가 목표를 상세히 제시하였다. 그리고 1992

년에는 그에 대한 보완적 후속 연구로서 과학적 탐구사고력 신장을 위한 과학 탐구과정 모형을 문제인식 및 가설설정, 탐구의 설계, 탐구의 수행, 자료의 해석, 결론도출 및 평가로 제시하고 이 모델로부터 16개의 행동요소를 추출하여, 각 요소들을 조작적으로 정의하였다(우종욱 외, 1992).

박승재(1991)는 과학적 탐구 사고력 평가들을 과학적 탐구 사고력, 탐구 상황, 과학 개념을 포함하는 3차원적 평가틀로 제시하고, 각 차원에 대한 범주와 소범주 및 주요 요소를 제시하였다. 즉 과학적 탐구 사고력은 문제 파악 및 정보 수집, 관찰실험 및 자료수집, 정보자료의 이용분석, 정보자료의 종합 및 해결, 검증비판과 평가 및 판단과 같이 5개 범주로 나누고, 다시 15개의 소범주로 나누어 제시하였다.

김은진과 김영수(1992)는 박승재(1991)의 과학적 탐구 사고력 평가틀에 기초하여 대학수학능력 시험을 위한 1,2,3차 실험 평가 문항을 분석하였다. 분석결과 신뢰도는 0.45-0.56, 1차와 3차의 경우에는 변별도 평균이 각각 0.32, 0.23이었으며, 2차의 경우에는 변별도 평균이 -0.10이었다. 그리고 시험 평가 결과가 학교 과학성적과 상관이 있는 것으로 나타났다.

김대식, 이기중, 박종원(1993)은 대학 수학능력시험 실험 평가 문항을 이용하여 학생들이 문제를 해결하는 과정을 분석하여 학생들의 사고 과정 속에서 탐구기능과 개념이 어떻게 나타나고 적용되는지 밝혔다.

김상철과 권재술(1994)은 대학수학능력 시험을 위한 실험평가 문항과 '94년도 1차 수학능력 시험 문제의 수리 탐구 영역(II) 중에서 과학탐구 문제 A형을 분석하여 각 문제들이 박승재(1991)의 3차원적 평가를 중 어디에 속하는지를 정리하였다.

### III. 연구 방법

#### 1. 대상

연구 대상은 전남에 소재하는 A 고등학교 3학년 학생으로 선정하였다. 평가 결과를 실험을 실시하지 않은 경우와 비교하기 위해 추가로 B고 48명, C고 33명, D고 30명의 학생들을 선정하였다. 연구 대상을 실험실시 집단과 실험미 실시 집단으로 나누어 제시하면 <표 1>과 같다.

#### 2. 실험내용 선정과 실험 안내서 개발

실험 내용은 고등학교 2학년 물리 II "힘과 운동" 단원 중에서 기본적인 중요하다고 판단되며 서로 내용이 연관

되는 실험 3개를 선정하였다: (실험1)시간 기록계에 의한 운동의 기록, (실험2)힘과 가속도의 관계, (실험3)자유 낙하 운동. 각 실험에 대해 실험 안내서를 개발하여 실험을 실시하였다.

<표 1> 연구 대상 집단

실험	실험 실시 집단		실험 미실시 집단	
	대상	학생수	대상	학생수
실험1	A고 3학년	158	B고 3학년	48
실험2	A고 3학년	157	C고 2학년	33
실험3	A고 3학년	157	D고 2학년	30

### 3. 탐구 실험 평가를 개발

탐구실험 평가틀은 허명의 과학탐구 평가표(허 명,1984) Tamir 등(1978)등이 개발한 실험구조 및 과제 분석표, Klopfer(1971)의 과학탐구과정 목표 분류틀, TIPS II, WPS, SAPA II, NSTA, NAEP가 제시한 탐구과정요소(과

학학습 평가 참고, 1991), BSCS에 제시된 과학탐구과정의 교수 목표(Mayer et al., 1978), 영국의 APU평가틀, 과학행동 분류표(최종락 외,1975) 등을 참고하여 고등학생에게 학교 수업시간에 실험을 실시하고 직접 평가할 수 있는 5개 영역을 선정하고 16개의 세부탐구과정 요소를 선정하였다.

탐구과정 영역과 세부요소는 <표 2>와 같다.

선정한 16가지의 세부탐구과정요소에 대하여 평가 문항 개발에 대한 준거로 이용할 수 있도록 각 탐구과정요소에 대한 평가 목표를 진술하였다. 각 요소에 대한 평가 목표는 다음과 같다.

#### A. 문제 파악 및 예측, 가설의 설정

##### A1. 목표 인식/기술, 탐구 문제의 인식/설정

- 주어진 상황을 통해 실험목표를 인식할 수 있다.
- 주어진 실험장면을 보고 어떤 실험인지 알아낼 수 있다.
- 자연 현상이나 사물의 관찰에서 규칙성을 찾거나 변화의 원인을 규명하고 해결하는데 필요한 탐구 문제를 찾아내거나 설정할 수 있다.

##### A2. 예측

- 주어진 조건이나 정보들로부터 실험결과를 예측할 수 있다.

<표 2> 탐구 과정 영역과 세부 탐구 과정 요소

탐구 과정 영역		세부 탐구 과정 요소	
A	문제 파악 및 예측, 가설의 설정	A1	목표 인식/기술, 탐구문제 인식/설정
		A2	예측
		A3	가설설정, 여러 가설 중 검증 가능한 가설 선택
B	탐구의 설계	B1	변인의 조작적 정의, 관찰/측정 방법 및 방법의 고안
		B2	독립변인과 종속 변인의 설정/구분, 변인 통제
		B3	이상화 조건이나 가정의 설정
		B4	실험 단계의 설정, 실험시 주의사항
C	정보의 수집	C1	실험 기구의 선정
		C2	관찰/측정 결과의 기록/기술
		C3	자료의 정리 및 분류
D	분석 및 처리	D1	기호나 도표의 사용 및 해석, 내삽/외삽
		D2	그래프 사용 및 해석, 그래프의 변형, 내삽/외삽
		D3	개념/원리/법칙/정보/자료 등에 의한 정성적 설명
		D4	정량적 처리
E	일반화 및 적용	E1	일반화된 결론 도출, 개념의 형성 및 발견
		E2	결과의 적용, 새로운 상황에서의 예측

- 주어진 탐구 문제에 대해 가능한 답을 예측할 수 있다.
- A3. 가설설정, 여러 가설 중 검증 가능한 가설설정
  - 가설과 가설이 아닌 진술을 구분할 수 있다.
  - 문제 상황에서 검증가능한 가설을 설정할 수 있다.
- B. 탐구의 설계
  - B1. 변인의 조작적 정의, 관찰/측정 방법 및 방법의 고안
    - 측정이나 관찰을 통해 조작적 정의를 필요로 하는 변인이나 용어를 찾아내어 조작적으로 정의할 수 있다.
    - 필요할 경우 개념이나 절차 등도 조작적으로 정의할 수 있다.
    - 관찰 또는 측정에서 오차를 줄일 수 있는 방법을 고안할 수 있다.
    - 실험의 용이성을 위해 적절한 관찰 및 측정방법을 도입할 수 있다.
  - B2. 독립변인과 종속변인의 설정/구분, 변인 통제
    - 실험에 영향을 미치는 변인들을 찾아낼 수 있다.
    - 가설을 검증하기 위해 통제해야 할 변인을 찾을 수 있다.
    - 변인들간의 관계를 알기 위한 변인 통제를 할 수 있다.
  - B3. 이상화 조건이나 가정의 설정
    - 실험의 편리성을 위해 적절한 조건이나 가정을 사용할 수 있다.
  - B4. 실험 단계의 설정 및 실험시 주의 사항
    - 실험 목표에 맞는 실험단계를 설계할 수 있다.
    - 실험 수행상의 유의점을 인식할 수 있다.
- C. 정보의 수집
  - C1. 실험 기구의 선정
    - 실험을 수행하는데 필요한 실험기구와 재료를 선정할 수 있다.
  - C2. 관찰/측정, 결과의 기록/기술
    - 오감을 사용하여 물질의 정성/량적 성질을 파악할 수 있다.
    - 적절한 언어로 관찰을 기술할 수 있다.
    - 측정도구에 나타난 측정치를 바르게 읽을 수 있다.
    - 적절한 측정 단위를 쓸 수 있다.
  - C3. 자료의 정리 및 분류
    - 적절한 자료와 적절하지 못한 자료를 분리할 수 있다.
    - 관찰이나 측정 등을 통해 수집된 자료를 특정 성질이나 기준에 따라 구분/분류할 수 있다.
    - 문제를 해결하는데 필요한 자료를 선택할 수 있다.
- D. 분석 및 처리
  - D1. 기호와 도표 사용 및 도표 해석, 내삽이나 외삽
    - 실험자료를 보다 명료하고 조직화된 도표로 나타낼 수 있다.
    - 도표로 나타난 자료에서 특징이나 규칙성을 찾아낼 수 있다.
    - 주어진 도표로부터 내삽이나 외삽을 통해 다른 결과를 추리할 수 있다.
  - D2. 그래프 사용 및 해석, 그래프의 변형, 내삽이나 외삽
    - 주어진 자료나 도표로부터 그래프를 그릴 수 있다.
    - 그래프로 나타난 자료에서 특징이나 규칙성을 찾아낼 수 있다.
    - 그래프로 나타난 자료에서 추리하거나 예상할 수 있다.
    - 그래프를 다른 형태로 변형할 수 있다.
    - 주어진 그래프를 통해 내삽이나 외삽을 통해 다른 결과를 추리할 수 있다.
  - D3. 개념/원리/법칙/정보/자료 등에 의한 정성적 설명
    - 주어진 정보나 자료로부터 규칙성을 설명할 수 있다.
    - 주어진 자료에서 귀납이나 연역을 통해 결론을 내릴 수 있다.
    - 도표나 그래프 등의 자료를 이용하여 주어진 현상을 설명할 수 있다.
    - 얻어진 결과를 원리나 법칙을 이용하여 설명할 수 있다.
  - D4. 정량적 처리
    - 도표나 그래프를 보고 정량적인 계산을 할 수 있다.
    - 주어진 자료로부터 정량적인 계산을 할 수 있다.
    - 법칙이나 공식을 이용하여 정량적인 계산을 할 수 있다.
- E. 일반화 및 적용
  - E1. 일반화된 결론 도출, 개념의 형성 및 발전
    - 실험 결과가 보여주는 규칙성을 토대로 보다 일반적인 결론을 내릴 수 있다.
    - 얻어진 결과들을 일반화하고 이론적 모델을 설명할 수 있다.
    - 두 가지 이상의 관찰, 실험 결과에서 공통된 특성을 찾아내거나 보다 일반적인 결론을 내릴 수 있다.
  - E2. 결과의 적용, 새로운 상황에서의 예측
    - 주어진 결과를 통해 가능한 사실이나 현상을 예측할 수 있다.
    - 내려진 결론을 다른 문제 상황에 적용할 수 있다.

4. 탐구 실험 평가 문항 개발

탐구 실험 평가 문항은 학생들이 수행한 실험 전반에 관해 탐구실험능력을 평가할 수 있도록 탐구과정요소별로 해당하는 문제를 출제하였고 각 문항은 현장에서 쉽게 적용할 수 있고, 앞으로 문제은행 구축에 사용될 수 있도록 선다형 객관식 문항으로 출제하였다.

각 실험당 탐구과정요소별 문항 수는 <표 3>과 같다.

<표 3> 각 실험에서의 탐구과정요소별 평가 문항수

탐구과정 영역	세부 영역	실험1 문항수	실험2 문항수	실험3 문항수
A	A1	2	2	2
	A2	.	.	.
	A3	.	.	.
B	B1	2	2	1
	B2	1	4	.
	B3	1	2	1
	B4	3	2	2
C	C1	3	2	3
	C2	6	.	2
	C3	1	3	1
D	D1	10	4	4
	D2	8	6	5
	D3	3	.	1
	D4	3	.	2
E	E1	.	2	5
	E2	1	3	4
계		44	30	29

\* 합계가 세부영역별 문항수를 더한 것보다 적은 이유는 1개 문항이 2개 이상의 영역에 속할 수 있기 때문이다.

탐구과정요소 각각에 대한 문제의 예는 다음과 같다.

B. 탐구의 설계

B2. 독립변인과 종속변인의 설정/구분, 변인통제  
 <실험 2-8> '힘과 가속도의 관계' 실험에서 독립변인은 (변화시켜 주어야 할 양은)?

- ① 물체에 작용한 힘의 크기    ② 가속도
- ③ 수레의 질량                    ④ 수레를 당기는 시간

B3. 이상화 조건이나 가정의 설정

<실험 1-4> '운동의 기록' 실험에서 반드시 가정해야 할 조건은?

- ① 시간기록계가 일정한 주기로 타점을 찍는다고 가정한다.
- ② 종이 테이프를 당길 때 힘이 일정하다고 가정한다.
- ③ 종이 테이프의 두께가 일정하다고 가정한다.
- ④ 종이 테이프의 타점간격이 일정하다고 가정한다.

D. 분석 및 처리

D3. 개념/원리/법칙/정보/자료/ 등에 의한 정성적 설명  
 <실험 3-22> '자유낙하 운동' 실험에서 자유낙하 운동이 왜 등가속도운동이라고 할 수 있는가?

- ① 시간에 따라 구간별 평균 가속도가 거의 일정하므로
- ② 시간에 따라 이동거리가 증가하므로
- ③ 시간에 따라 구간별 평균 속도가 일정하므로
- ④ 시간에 따라 구간별 평균 가속도가 일정하게 증가하므로

5. 탐구 실험의 실시 및 평가의 실시

실험은 충분한 탐구가 이루어지도록 각 실험마다 연속 2 시간씩으로 하여 총 6시간이 소요되었고, 평가는 3개 실험에 대해 각각 실시하였고 각 평가에 걸린 시간은 50분씩이었다.

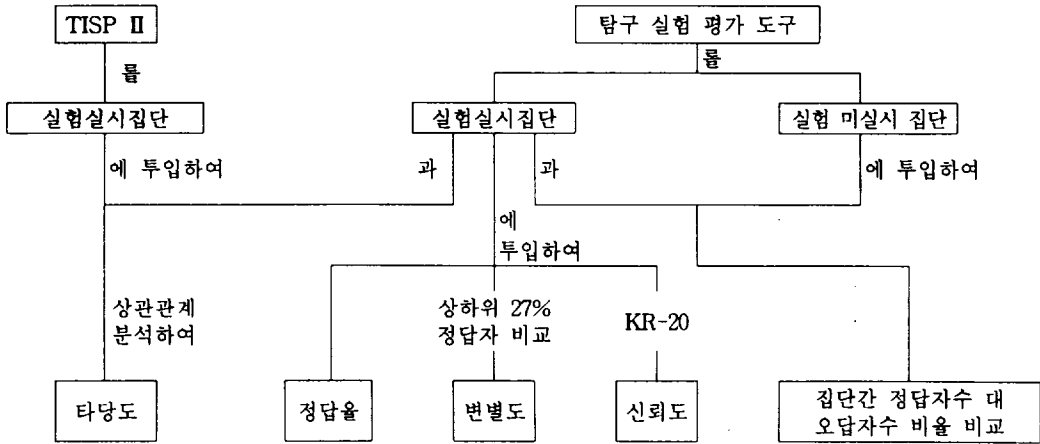
6. TIPS II 평가 실시

본 연구에서는 실험을 실시한 집단에 개발된 탐구 평가 문항뿐 아니라 TIPS II도 투입하여 두 평가 도구간 상관관계를 조사함으로써 본 연구에서 개발한 탐구 평가도구의 공인타당도(concurrent validity)를 조사하였다.

7. 분석 방법

전체적인 분석 내용을 그림으로 나타내면 <그림 1>과 같다.

개발한 실험 탐구 도구를 실험 실시 집단과 실험 미실시 집단에 각각 투입하여, 실험 실시 집단에 투입한 결과로부터 문항의 정답율과 변별도, 그리고 신뢰도를 구했고, 실험 미실시 집단에 투입한 결과와 비교( $\chi^2$ -검증)함으로써, 문항별로 실험실시 집단의 정답비율이 실험 미실시 집단의 정답비율보다 유의미하게 높은지를 분석하였다. 그리고, 실험 실시 집단에 추가로 TIPS II를 투입하여 탐구실험 평가의 결과와의 상관정도를 통해 탐구실험 평가도구의 타당도를 분석하였다. 이러한 분석을 통해 알 수 있는 결과들은 다음과 같다 : 첫째, 타당도를 통해 탐구 실험 평가 도구가 얼마나 탐구 실험 능력을 올바르게 측정할 수 있는 문항인지 알 수 있고, 둘째, 신뢰도를 통해 탐구 실험 평가 도구의 적용결과



<그림 1> 분석도

가 얼마나 일관적인지를 알 수 있고, 세제, 정답율을 통해 문항별/탐구과정 영역별 난이도를 알 수 있고, 네제, 변별도를 통해 실험을 실시한 후에 실험을 잘 수행한 학생과 그렇지 못한 학생을 구별할 수 있는 문항/영역을 알 수 있고, 다섯째, 실험실시 집단과 미 실시 집단간 정답자수 대 오답자수 비율을 비교하여( $\chi^2$ -검증) 실험을 실시한 학생과 그렇지 않은 학생을 구별할 수 있는 문항/영역이 무엇인지 알 수 있다.

#### IV. 결과

##### 1. TIPS II 에 의한 타당도

본 연구에서 개발한 탐구실험 평가도구와 TIPS II와의 상관계수의 조사 결과는 <표 4>와 같다.

<표 4> 탐구실험 평가 도구와 TIPS II와의 상관 계수

비교도구	상관계수
실험 1 - TIPS II	0.546
실험 2 - TIPS II	0.613
실험 3 - TIPS II	0.622

본 연구에서 상관 계수는 0.546 ~ 0.622로 나타나 탐구실험 평가도구가 어느 정도는 타당하다고 볼 수 있다.

##### 2. 실험 실시 집단의 정답율/ 변별도/ 신뢰도

실험실시 집단의 정답율, 변별도, 신뢰도를 영역별로 평

균하여 제시하면 <표 5>와 같다

정답율은 실험1, 2, 3 모두 평균 0.75이상으로 나타나 실험1의 E영역을 제외하고는 실험을 수행한 학생이면 대부분 평가문항을 잘 맞춘 것으로 나타났다. 변별도가 0.2이상인 경우를 실험을 잘하는 학생과 못하는 학생을 구별할 수 있는 문항이라고 보면 변별도는 실험 1,2,3 모두 평균 0.27이상으로 나타나, 전체적으로 정답율이 비교적 높으면서도 실험을 잘하는 학생과 그렇지 못한 학생을 구별하는데 사용할 수 있는 것으로 나타났다. 물론 영역별 및 문항별로는 변별도가 낮은 영역이나 문항이 있다. 예를 들면 C영역은 실험 1,2,3(단, 문제 3의 경우 변별도가 특별히 높게 나온 1개 문항을 제외하면)모두 변별도가 낮아, 이 영역에 대한 문항은 실험을 한 학생이면 모두 잘 맞출 수 있는 것으로 나타났다. 각 실험평가 문항에서 변별도가 0.2이상인 문항수를 보면 실험 1의 경우에는 전체 문항 44개 중 22개로 나타났고, 실험 2의 경우에는 전체 문항 30개 중 25개, 실험 3의 경우에는 전체 문항 29개 중 22개로 나타났다. 신뢰도는 모두 0.77이상으로 높은 편이었다.

##### 3. 실험실시 집단과 실험미 실시 집단간 비교

실험실시 집단과 실험 미 실시 집단간의 정답자수와 오답자수를 비교하면( $\chi^2$ -검증) 어떤 문항이나 어떤 탐구 영역이 실험을 한 학생과 실험을 하지 않은 학생을 구분하는 데 사용될 수 있는 지 알 수 있다.

먼저, 실험 실시한 집단과 실시하지 않은 집단을 구분할 수 있는 문항의 예를 들면 다음과 같다.

<표 5> 각 실험의 영역별 정답률, 변별도, 신뢰도

영역	실험 1			실험 2			실험 3		
	정답률	변별도	신뢰도	정답률	변별도	신뢰도	정답률	변별도	신뢰도
A	0.84	0.08	0.77	0.8	0.46*	0.77	0.91	0.24*	0.86
B	0.78	0.29*		0.75	0.38*		0.77	0.25*	
C	0.87	0.12		0.9	0.13		0.87	0.3*(0.19) <sup>1)</sup>	
D	0.7	0.32*		0.81	0.35*		0.82	0.44*	
E	0.2	0.4*		0.82	0.32*		0.7	0.58*	
평균	0.75	0.27*		0.81	0.31*		0.793	0.36*(0.34) <sup>1)</sup>	

\* 표시한 것은 변별도가 0.2이상인 것

1) 괄호 안의 값은 특별히 변별도가 높게 나온 21번 문항(변별도:0.79)을 제외한 경우의 변별도이다.

첫째, 실험 3의 1번 문항과 같이 실험실시 집단의 정답자수와 오답자수가 <표 6>과 같은 경우이다. 이 경우에는 실험을 한 경우 잘 맞추는 학생이 많고 실험을 하지 않는 경우에는 맞추지 못하는 학생이 많다는 것을 의미한다.

<표 6> 실험 3의 1번 문항의 집단간 정답자수와 오답자 수

	실험실시집단	미 실시집단
정답자수	142	12
오답자수	15	18

p < .001

둘째, 실험 3의 6번 문항과 같이 실험실시 집단의 정답자수와 오답자가 <표 7>과 같은 경우이다. 이 경우에는 실험을 실시한 경우와 실시하지 않은 경우 모두 정답자수가 많으나 실험을 실시하면 훨씬 더 많은 학생이 정답을 맞출 수 있다는 것을 의미한다.

<표 7> 실험 3의 6번 문항의 집단간 정답자수와 오답자수

	실험실시집단	미 실시 집단
정답자수	151	21
오답자수	6	9

p < 0.001

셋째, 실험 3의 26번 문항과 같이 실험실시 집단의 정답자수와 오답자수가 <표 8>과 같은 경우이다. 이 경우에는 오답자수가 정답자수보다 많지만 실험실시 집단의 학생이 미 실시 집단의 학생보다 좀 더 덜 틀린다는 것을 의미한다.

<표 8> 실험 3의 26번 문항의 집단간 정답자수와 오답자수

	실험실시집단	미 실시집단
정답자수	80	1
오답자수	97	29

p < 0.001

실험을 실시한 학생과 실시하지 않은 학생을 구분할 수 없는 경우는 실험 2의 21번 문항과 같이(<표 9>), 실험실시 집단의 정답자수(122)에 대한 오답자수(35)의 비율과 미 실시 집단의 정답자수(22)에 대한 오답자수(11)의 비율이 거의 같게 나타난 경우라고 하겠다.

<표 9> 실험 2의 21번 문항의 집단간 정답자수와 오답자수

	실험실시집단	미 실시집단
정답자수	122	22
오답자수	35	11

p = 0.0783

마지막으로 실험실시 집단의 정답자수 비율보다 미 실시 집단의 정답자수 비율이 더 많거나, 반대로 실험실시 집단의 오답자수 비율이 미 실시 집단의 오답자수보다 더 많은 경우에는 좋은 문제가 아니라고 하겠다. 본 연구에서 그러한 경향을 보인 문항으로는 실험 1의 28, 29, 30, 34번 문항과 실험 3의 5번 문항이 있었다. 그러나 그러한 경향이 통계적으로 유의한 것이 아니었으므로 문항을 삭제하지는 않았다.

각 실험에 대해  $\chi^2$ -검증결과를 나타내면 <표 10>과 같다.

<표 10> 실험실시 집단과 미 실시 집단의 정답자수 대 오답자수 비교결과 ( $\chi^2$ -검증)

영역	실험		
	실험1	실험2	실험3
A	10.47*	7.72*	44.10*
B	13.97*	11.46*	2.24 (7.57*) <sup>2)</sup>
C	9.02*	30.27*	19.49*
D	4.27 (7.73*) <sup>1)</sup>	10.47*	29.96*
E	2.53	7.49*	23.48*

- 1) 괄호 안의 숫자는 실험을 실시한 경우보다 실시하지 않은 경우에 오답자수에 대한 정답자수 비율이 높은 문항 28, 29, 30, 34번을 제외한 결과임
  - 2) 괄호 안의 숫자는 실험을 실시한 경우보다 실시하지 않은 경우에 오답자수에 대한 정답자수 비율이 높은 문항 5번을 제외한 결과임
- \* 표시한 것은  $p < .01$  인 경우임.

전반적으로 각 실험 평가 문항은 실험을 한 학생과 하지 않은 학생을 구분하는 데에는 적절한 것으로 나타났다. 문항별로 살펴보면  $\chi^2$ -검증 결과  $p < .01$ 인 문항의 개수가 실험 1의 경우에는 전체 44개 문항 중에서 24개 문항이었고, 실험 2의 경우에는 30개 문항 중에서 22개 문항, 실험 3의 경우에는 29문항 중에서 24개 문항이었다. 즉 이 문항들은 실험을 한 학생과 하지 않은 학생을 구분하는데 사용할 수 있는 것으로 나타났다.

### V. 요약 및 결론

본 연구에서는 특정 물리 실험에 대해 학생들의 탐구실험 능력을 지필로 손쉽게 측정할 수 있는 평가도구를 개발해서 평가도구의 특성을 분석하였다.

탐구실험 평가도구의 타당도를 점검하기 위해 실험을 실시한 집단에 TIPS II를 투입하고 탐구실험 평가도구와 TIPS II와의 상관관계를 조사한 결과 실험 1의 경우는 0.546, 실험 2의 경우는 0.613, 실험 3의 경우는 0.622로 만족할 만한 정도의 타당도가 있는 것으로 나타났다. 따라서 본 실험 평가도구가 학생들의 탐구실험 평가능력을 평가하는데 타당한 문제들이라고 할 수 있다.

전체적으로 실험 1의 정답률은 0.75, 변별도는 0.29, 신뢰도는 0.7648이었다. 실험 2의 정답률은 0.81, 변별도는 0.4031,

신뢰도는 0.7675이었고, 실험 3의 정답률은 0.79, 변별도는 0.40, 신뢰도는 0.862이었다. 문항별로 변별도가 0.2보다 큰 문항을 선택하면 실험을 실시한 학생들 중에서도 실험을 잘 한 학생과 그렇지 않은 학생을 구분하는데 사용할 수 있다 하겠는데, 실험 1의 경우 총 44개 문항 중 27개 문항, 실험 2의 경우에는 총 30개 문항 중 25개 문항, 실험 3의 경우에는 총 29개 문항 중 22개가 변별도가 0.2이상으로 나타났다.

영역별로는 C영역이 정답률이 높고 변별도가 낮아, 실험을 한 학생이면 문제를 모두 잘 푸는 것으로 나타났다.

본 도구를 실험 미 실시 집단에 투여하여 실험 실시 집단과의 정답자수 대 오답자수 비율을 비교하기 위해  $\chi^2$ -검증을 실시하였다. 전반적으로 각 실험에 대한 평가 도구가 실험을 실시한 집단과 실시하지 않은 집단을 구분하는데 사용할 수 있는 것으로 나타났다. 문항별로 살펴보면, 실험 1의 경우에는 총 44개 문항 중 24개 문항, 실험 2의 경우에는 30개 문항 중 22개 문항, 실험 3의 경우에는 29개 문항 중 24개 문항을 그러한 목적으로 사용할 수 있는 것으로 나타났다.

본 연구에서는 현장에서 실험 수업 후에 학생의 탐구 실험 능력을 평가하기 위해 문항을 개발하였을 때, 어떤 문항이 실험을 잘 한 학생과 그렇지 않은 학생을 구분하는데 사용할 수 있는지, 또 어떤 문항이 실험을 한 학생과 그렇지 않은 학생을 구분하는데 사용할 수 있는지를 알기 위한 분석방법과 결과를 제시하였다. 본 연구에서 사용한 방법으로 과학탐구 실험능력을 측정할 수 있는 문항들을 문제은행화한다면 목적에 따라 적절한 문항들을 손쉽게 추출하여 학교 현장에서 쉽게 적용할 수 있겠다.

본 연구에서는 연구의 한계상 고등학교 물리과목 중 일부 실험만을 가지고 문항을 개발했으나 이러한 방법으로 다른 실험 다른 과목에서도 추후 평가문항을 개발하여 사용할 수 있을 것이다. 고등학생의 과학탐구실험 능력을 측정하는데 본 연구에서 개발한 지필 검사 문항만으로는 충분하다고 할 수 없다. 따라서 보다 교사들이 쉽게 사용할 수 있는 방법을 고안하여 실험 보고서, 실기시험, 과제 등을 통해 학생의 탐구실험 능력을 평가할 수 있도록 다양한 평가 도구, 점검표(checklist)등이 보완되어야 할 것이다.

### 참 고 문 헌

- 교육부(1988). 고등학교 과학과 교육과정해설. 국정교과서 주식회사
- 김대식, 이기종, 박종원(1993). 고등학생의 탐구 사고과정 조사. 물리교육, 11(2), 140-157.
- 김상철, 권재술(1994). 대학수학능력시험에서 과학탐구 영



- 역의 분석. 한국과학교육학회지, 14(2), 214-224.
- 김은진, 김영수(1992). 대학수학능력시험 실험 평가 문제의 분석. 한국과학교육학회지, 12(1), 75-92.
- 김창식, 이화국, 권재술, 김영수, 김찬중(1991). 과학학습 평가. 교육과학사.
- 박승재(1991). 과학적 탐구 사고력 평가. 서울대학교 사범대학 물리 교육과 물리학습 연구실.
- 우종옥, 이항로, 이경훈(1991). 대학수학능력 시험의 수리·탐구영역 중 지구과학 교과에 관련된 탐구능력 측정 을 위한 행동요소의 추출과 평가목표의 상세화 연구 I. 한국과학교육학회지, 11(1), 83-96.
- 우종옥, 이경훈, 이항로(1992). 대학수학능력시험의 자연과 학 탐구능력 평가를 위한 행동요소의 추출과 평가 목 표의 상세화 연구 II. 한국과학교육학회지, 12(2), 81-96.
- 정완호, 허명, 은경용(1993). 국민학생의 과학 탐구능력 측 정을 위한 평가도구 개발. 한국과학교육학회지, 13(1), 80-91.
- 이 무(1992). 과학적 탐구사고력 평가 문항형태에 관한 연 구. 박사학위논문. 서울대학교.
- 이연우, 우종옥(1991). 과학탐구능력 측정을 위한 표준화 검 사지 개발 -중학교 2학년의 자료 분석과 해석능력을 중심으로-. 한국과학교육학회지, 11(1), 59-72.
- 이종기(1988). 고등학생의 과학 탐구능력 측정을 위한 평가 도구 개발. 석사학위 논문. 한국교원대학교.
- 최종락, 오대섭, 박봉두, 손진호, 김학수(1975). 과학교육의 성취도 평가. 경북사대교육연구지, 17, 131-156.
- 허 명(1984). 과학 탐구 평가표 개발. 한국과학교육학회지, 14(1), 57-63.
- Bruner, J.S., Goodnow, J.J. and Austin, G.A.(1966). *A study of thinking*. New York, John and sons, Inc.
- Burns, J.C., Okey, J.R., and Wise, K.C. (1985). Development of an intergrated process skills test. *Journal of Research in science Teaching*, 12(2), 169-177.
- Klopfer, L. (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*(pp.559-642). McGraw-Hill : new York.
- Mayer,W.V.(1978). *Biology teachers handbook*, 3rd ed NY: John Wiley and Sons.
- Schwab,J.J.(1966). The teaching of science as inquiry. In J. Schwbab & P. Brandwein (Ed.), *The teaching of science*(pp.3-103). Cambridge: Harvard University Press.
- Tamir, P., Nussinovitz, R., and Friedlor, Y.(1982). The design and use of a practical tests assessment inventory. *Journal of Biological Education*, 16, 42-50.

(ABSTRACT)

## The Development and Analysis of a Test for Assessment of Physics Inquiry Experiment.

Mi Kyung Kim · Hee Gyun Oh · Jongwon Park  
(Chonnam National University)

The purpose of this study is to develop and analyze a test for the assessment of physics inquiry experiment.

To do this, three experiments about 'analysis of motion', 'the relation of force and acceleration', and 'free fall motion' in high school physics textbooks were chosen, and 5 scientific inquiry domains and 16 science process skills have been specified.

For each experiments, testsheet of 29~44 questions for assessing students' ability about physics inquiry experiment were developed on the basis of the scientific inquiry processes developed earlier.

After instruction about 3 experiments mentioned above, a test was administered to the students who took experiment.

After the administration of a test, the ratio of correct answers, discrimination index, and reliability were analyzed. Using the ratio of correct answers, we can determine item difficulty. Through the D.I(discrimination index), we can find which items can discriminate the students who took experiment well from those who took experiment badly, and we can also find the stability of a test result by the reliability analysis.

The test developed in this study were also administered to the students who did not take experiments, and the results were compared with the those of the students who took experiments. With the comparison by chi-square method, we could find which items can discriminate the students who took experiments from those who did not take experiments.