

# 그래프를 이용한 과학 탐구 능력 검사 도구의 개발<sup>1)</sup>

문 충 식

(분당중앙고등학교)

## The Development of an Instrument for the Measurement of Science Process Skills Using Line Graphs

Moon, Choong-Sik

(Bundang Jungang High School)

### ABSTRACT

The improvement of process skills has been one of the most important goals in secondary science education. To achieve this goal, it is essential to develop an instrument for evaluating inquiry skills in addition to improving science curricula, inquiry teaching methods, and other educational environment. There are a lot of instruments in testing science process skills in Korea as well as America and Europe. However, it has been difficult to find the instruments that have a characteristic in content and form.

The purpose of this study is to develop a valid and reliable instrument for evaluating science process skills using line graph. This study examined the whole body of scientific process and identified 6 component skills. Three or four items for each component process skills were developed and revised by pilot test and field test. The instrument is considered valid and reliable, for the content validity is 78% and the reliability(KR-20) is 0.82. The discrimination index is 0.57 and difficulty index is 0.47, which also suffice the criteria of good test.

**Key words** : science process skill, graphing skill.

### I. 서 론

과학 교육에서 과학 탐구 능력의 향상은 과학 학습의 중요한 목표가 되어왔다. 1960년대 이후 과학 교육은 결과로 얻어진 이론의 전수보다는 자연을 탐구하는 과정으로 보는 경향이 강하게 대두되어 PSSC, CHEM Study, BSCS, ESCP, IPS 등 많은 혁신적 교육 과정이 개발되었고, 이 영향으로 우리 나라도 제 3차 교육과정부터 과학 탐구 능력의 향상이 과학 교육의 중요한 목표가 되었다. 이와 함께 영국의 APU와 TAPS, 미국의 NAEP의 과학 학력 평가들 등의 과학 탐구 평가들이

만들어졌고(Welch *et al.*, 1981), 과학 탐구 능력 검사 도구는 주로 지필 평가 형태로 외국과 우리 나라에서 개발되고 있고, 논리 사고력과 과학 탐구 능력과의 관계를 알아보는 연구들이 있다(Yap & Yeany, 1988; Baird & Borich, 1987; 김범기, 1989). 다만, APU의 실험 기능 평가와 같이 표준화된 실험 기능 탐구 능력 평가 도구의 개발은 최근에 이루어지고 있다.

과학 탐구 능력은 과학자들이 연구하는 데에 필요한 능력으로, 학생들이 어떤 문제에 부닥쳤을 때 과학적 탐구 방법에 의해 스스로 문제를 해결하는 능력이다(권재술과 김범기, 1994). 즉, 과학의 본질을 탐구라고 본다

\* 1998년 7월 4일 받음.

<sup>1)</sup> 이 논문은 1997년도 한림대학교 부설 한림과학원의 교사 공모 과제 연구비에 의하여 연구되었음.

면, 과학 지식은 탐구 과정의 산물이며, 과학은 어떤 문제의 해답을 찾는 탐구 과정 또는 탐구 방법이라 할 수 있다. 따라서 탐구 학습은 과학적 활동의 산물인 지식보다는 과학적 개념이나 원리를 발견하고 창출해 가는 탐구 과정을 더 중요시한다. Schwab가 1964년에 행한 하버드대의 기념 강연에서 탐구 학습이란 말을 처음으로 사용하였는데, 과학의 학문 체계를 확립하는 데 공헌하는 탐구로서 과학의 체계를 흔들 수 없는 정형 탐구(stable inquiry)와 새로운 개념이나 체계를 창조하고 발견하는 탐구로서 과학의 체계를 근본적으로 바꾸어 놓는 부정형 탐구(fluid inquiry)로 나누었다.

학교 현장에서 탐구 학습이 과학 교육의 대명사처럼 되어 탐구 능력 신장을 강조하고 있으나 과학 탐구 능력 향상이 제대로 이루어지고 있다는 연구보고는 찾아보기 어렵다(Brotherton *et al.*, 1996). 이것은 여러 가지 원인으로 설명될 수 있지만, 본 연구에서는 과학 탐구 능력을 측정할 수 있는 검사 도구의 개발이 미진하여 평가가 제대로 이루어지지 못하는데서 찾아보고자 하는 것이다. 물론 우리 나라에서도 과학 탐구 능력을 측정하는 검사도구가 개발되고 있다(김창식 등, 1992; 나종국, 1993; 이연우, 1989; 이종기, 1988; 이항로, 1991; 권재술과 김병기, 1994). 그러나 거의 모든 과학 탐구 능력 검사 문항들이 문장으로 진술되어 있고 탐구 상황을 분석하거나 자료를 해석하는 문항으로 거의 비슷한 성격을 띠고 있어서 각 검사 도구의 독특한 특징을 발견할 수가 없다.

Robinson(1969)의 과학적 사고 체계의 모형에 의하면, 과학의 사고 체계를 감각적 세계와 논리적 세계로 구분하였으며, 관찰한 사실로부터 가설을 세우고 이를 관찰과 실험을 통하여 검증하는 과정으로 설명하고 있다. 이에 따라 탐구 학습에서 추구하는 탐구과정 요소는 보통, 문제 발상, 가설 설정, 실험 설계, 자료의 수집 및 정리, 자료의 해석 및 분석, 결론 도출, 일반화 등의 과정이다. 이 과정 요소들은 정형화된 것이 아니고 학자에 따라 여러 가지로 세분하고 있다. SAPA에서는 관찰, 공간/시간 관계의 사용, 수의 사용, 측정, 분류, 의사 전달, 예상, 추리, 조작적 정의, 변인 통제, 자료 해석, 가설 설정, 실험 등 13가지의 탐구 과정 요소를 제시하였고, Klofer(1971)는 관찰 및 측정, 문제 발견 및 해결 방안, 자료 해석 및 일반화, 이론적 모델 설정, 검증 및 수정 등 4단계로 구분하였고, 각 항목을 더 구체적인 내용으로 세분하였다. 또, BAPS(Basic Process Skill in Science; Cronins *et al.*, 1985), TIPS(The Test of

Integrated Science Process Skill; Dillashaw and Okey, 1980), TIPS II(Burns *et al.*, 1983), MIPT(Middle Grades Integrated Science Process Skills Test; Padilla & Cronin, 1986), POPS(The Performance of Process Skills; Mattheis *et al.*, 1988) 등과 같은 과학 탐구 검사 문항에서도 비슷한 과학 탐구 과정 요소들을 제시하고 있다.

본 연구에서 개발하고자 하는 선 그래프를 이용한 과학 탐구 능력 검사 문항은 선행 연구에서 선택되고 있는 중요한 탐구 요소들 중에서 가설 설정, 변인 통제, 자료 변환, 예상, 추리, 결론 도출 등의 능력을 선 그래프를 이용하여 측정하는 검사 문항으로 개발하는 것이다. 선 그래프는 그림 요소와 개념 요소를 모두 포함하고 있고, 작은 공간에 많은 정보를 포함시킬 수 있는 중요한 특징을 가지고 있어서 과학의 내용을 전달하거나 실험 결과를 나타내는 중요한 도구로 사용되고 있다. 그러므로 한 실험의 결과를 나타낸 그래프를 보고, 이 실험에서 검증하려고 했던 가설은 무엇이었으며, 실험의 독립 변인과 종속 변인, 통제 변인을 알거나, 그래프를 통하여 예측, 추리하거나 결론을 도출할 수 있는 능력 등을 측정하는 것은 과학의 내용을 보다 더 깊게 이해하게 하는 데에 기여할 수 있다.

본 연구는 선 그래프를 이용한 과학 탐구 능력을 탐구 요소별로 평가 목표를 조작적으로 정의하고, 이 탐구 요소에 따라 문항을 개발하고자 하였다. 개발된 과학 탐구 능력 검사 도구를 예비 또는 현장 검사를 실시하여 난이도, 변별도, 신뢰도, 내용 타당도, 공인 타당도 등을 산출하여 검사 도구의 측정학적 조건을 갖춘 고등학생들의 과학 탐구 능력 평가 도구를 개발하는 데에 그 목적이 있다.

## II. 과학 탐구 능력 검사 도구에 관한 선행 연구

현재까지 개발되어 사용되고 있는 과학 탐구 능력 검사 도구들은 거의 모두 문장으로 진술되어 있고 탐구 상황을 분석하거나 자료를 해석하는 문항으로 거의 비슷한 성격을 띠고 있다. 다만, Padilla와 McKenzie(1986)는 7~12학년용 대상으로 26문항의 그래프를 해석하고 작성하는 능력을 측정하는 TOGS(The Test of Graphing in Science)를 개발하였는데, 부분적으로 탐구 과정 요소를 포함하고 있다.

BAPS(Cronins *et al.*, 1985)는 관찰, 추리, 예상, 측

정, 의사 전달, 분류 등의 6가지 과학 탐구 과정 요소를 측정하기 위해 개발되었으며, 4지 선다형으로 각 요소에 6문항씩 36문항으로 구성되어 있다. 대상 학년은 4~8학년으로 평균은 22.4(표준 편차 5.3), 독해력 FOG 지수는 3.95, 내용 타당도는 95%, 난이도 평균은 0.62, 평균 변별도는 0.34, 신뢰도(KR-20)는 0.78이다.

TIPS(Dillashaw & Okey, 1980)는 SAPA의 통합 탐구 능력인 가설 설정, 조작적 정의, 변인 통제, 연구 계획 수립, 자료 해석 등의 5가지 과학 탐구 과정 요소를 측정하기 위한 도구이며, 4지 선다형의 36문항으로 평균 변별도는 0.40, 평균 난이도도 0.53, 신뢰도(Cronbach's  $\alpha$ )는 0.89이다.

TIPS II (Burns *et al.*, 1983)는 변인 확인, 조작적 정의, 가설 검증, 자료 해석, 실험 설계 등의 5가지 과학 탐구 과정 요소를 측정하기 위한 도구이며, 4지 선다형의 36문항으로 7~12학년을 대상으로 개발되었다. 평균 변별도는 0.35, 평균 난이도는 0.53, 신뢰도(Cronbach's  $\alpha$ )는 0.86이다.

MIPT(Padilla & Cronin, 1986)는 6~9학년을 대상으로 자료 변환, 자료 해석, 변인 통제, 가설 설정, 문제 인식, 실험 방법 등의 6가지 과학 탐구 과정 요소를 측정하기 위해 개발되었으며, 4지 선다형의 40문항으로 평균 난이도는 0.68, 평균 변별도는 0.34, 신뢰도(KR-20)는 0.89이다.

POPS(Mattheis *et al.*, 1988)는 MIPT에 의해 측정된 평가 목표와 문항을 수정·보완하는데 중점을 두었으며, 문제 인식, 변인 확인, 가설 설정, 실험 방법, 자료 변환, 자료 해석 등의 6가지 과학 탐구 과정 요소를 측정하기 위하여 개발되었다. 6~8학년을 대상으로 한 4지 선다형의 21문항으로 평균 변별도는 0.41, 평균 난이도는 0.47, 신뢰도(KR-20)는 0.75이다.

특히, Padilla와 McKenzie(1986)는 7~12학년을 대상으로 한 4지 선다형의 그래프 해석과 작성에 관한 능력을 측정하는 26문항의 TOGS(The Test of Graphing in Science)를 개발하였다. TOGS의 평가 목표는 크게 그래프 작성과 해석 능력으로 나뉜다. 그래프 작성 능력으로는 그래프에 점 찍기, 제시된 문장 자료나 표 자료를 그래프로 변환하는 능력, 가장 적합한 선 그리기, 독립 변인과 종속 변인을 확인하여 적절한 축을 선택할 수 있는 능력, 그래프에 적절한 눈금을 결정하는 능력 등의 5가지가 있고, 그래프 해석 능력으로는 X(Y)값에 대한 Y(X)값 결정하기, 내삽과 외삽을 할 수 있는 능력, 변인들 간의 관계를 기술하기, 두 그래프의

결과를 관련시켜서 일반화시키는 능력 등의 4가지가 있다. 위의 목표 요소들 중에서 자료 변환, 변인 확인, 내삽·외삽, 자료 해석, 일반화 능력 등의 과학 탐구 과정 요소를 포함하고 있다고 할 수 있다. 이 검사 도구의 평균 변별도는 0.43, 평균 난이도는 0.51, 독해력 FOG 지수는 6.2, 신뢰도(KR-20)는 0.83이다.

국내의 과학 탐구 능력 검사 도구로 이종기(1988)는 고등학생을 대상으로 하여 가설 설정, 변인 통제, 실험 설계, 자료의 변형 중 숫자 계산, 실험을 그래프로 나타내기, 추리, 상관 관계 결정, 인과 관계 결정, 예상(내삽과 외삽도 포함), 결론, 일반화 또는 모델 형성, 평가 등의 12가지 과학 탐구 평가 요소를 측정하는 4지 선다형의 36문항을 개발하였다. 이 검사 도구의 평균 변별도는 0.46, 평균 난이도는 0.59, 신뢰도(KR-20)는 0.86이다.

이연우(1989)는 중학교 2학년을 대상으로 자료 분석과 해석 능력에 관한 탐구 기능을 중심으로 과학 탐구 능력 검사 도구를 개발하였다. 20문항의 4지 선다형이고, 평균 정답률은 68.8%, 평균 변별도는 0.39, 신뢰도(KR-20)는 0.69이다. 이항로(1991)는 고등학생을 대상으로 한 5지 선다형의 33문항을 개발하였다. 이 검사 도구는 지구 과학 소재를 중심으로 이종기(1988)의 탐구 요소 중 평가를 뺀 11가지 탐구 요소를 측정하는 검사 문항이다. 이 검사 도구의 평균 변별도는 0.47, 신뢰도(KR-20)는 0.84이다.

은경용(1992)은 초등학교 5, 6학년을 대상으로 하여 관찰, 분류, 문제 인식, 측정, 추리, 가설 설정 및 검증, 변인 통제, 실험 설계 및 수행, 결과 해석, 결론 등 10가지의 탐구 요소를 측정하는 30문항을 개발하였다. 이 검사 도구의 평균 난이도는 0.53, 평균 변별도는 0.42, 신뢰도(KR-20)는 0.78이다.

권재술과 김범기(1994)는 초·중학생을 대상으로 과학 탐구 능력을 측정하는 4지 선다형의 30문항을 개발하였다. 이 검사 도구의 평가 탐구 과정 요소는 관찰, 분류, 측정, 추리, 예상, 자료 변환, 자료 해석, 가설 설정, 변인 통제, 일반화 등의 10가지이며, GALT 검사 도구와의 상관 관계를 이용한 공인 타당도와 요인 분석을 이용하여 구인 타당도를 구하였고, 평균 난이도는 0.59, 신뢰도는 KR-20가 0.74, Spearman-Brown 신뢰도는 0.71이다.

선행 연구들에서 선정된 탐구 과정 요소와 개발 과정 등은 본 연구의 중요한 준거가 되었다. 그러나 선행 연구들의 공통점은 문장으로 진술되어 있는 상황에서 탐

구 능력을 측정하거나 자료를 해석하는 형태를 취하고 있고, 다른 점이 있다면 문장 내용의 상황이나 상황의 복잡성에 있으며 독특한 다른 특징을 찾아보기 어렵다.

### Ⅲ. 연구 방법 및 절차

과학 탐구 능력에 관한 선행 연구의 문헌을 수집·조사하였으며, 과학 교육을 전공하는 대학원생 8명과 의 세미나를 통하여 검사 도구의 개발 준거를 설정하고, 6개의 과학 탐구 과정 요소를 선정하였다. 이를 바탕으로 각 탐구 과정 요소별로 3~5개씩의 문항을 개발하였고, 5차에 걸친 세미나를 통하여 문항을 수정·보완하고 다시, 예비 검사 및 현장 검사를 통하여 문항을 수정·보완하여 선 그래프를 이용한 고등학생들의 과학 탐구 능력 검사 도구를 개발하였다. 그래프를 이용한 과학 탐구 능력 검사 도구 개발 과정은 Fig. 1과 같다.

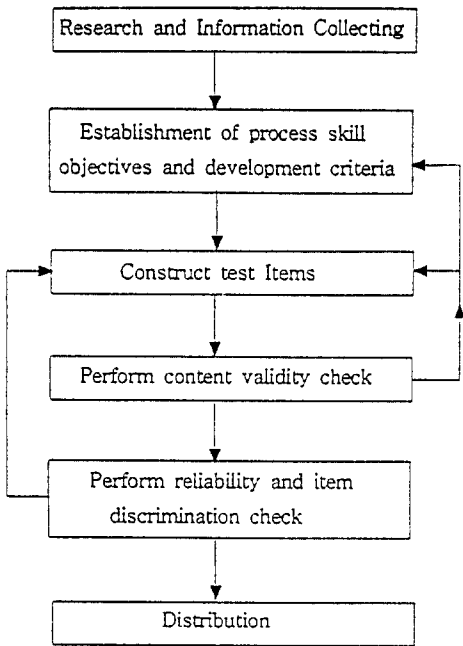


Fig. 1 A flowchart for test development showing feedback loops between steps

#### 1. 연구 대상

선 그래프를 이용한 과학 탐구 능력 검사 도구를 개발하기 위한 표집 대상은 Table 1과 같이 대도시(특별

시 및 광역시), 중·소 도시, 읍·면 소재 지역별로 유층표집하였다.

#### 2. 과학 탐구 요소의 선정 및 검사 도구의 개발 준거

선행 연구에서 제시되고 있는 탐구 과정 요소 중에서 기초 탐구 과정 요소 중에서 관찰은 어떤 현상을 객관적으로 볼 수 있는 능력이고, 측정은 어떤 측정 도구의 정확한 측정, 그리고 분류는 어떤 기준에 의해 제시된 자료를 유목화하는 것이므로 선 그래프로 측정하기 어렵고, 문제 인식은 주로 문장으로 제시되는 상황을 통하여 측정하므로 제외시켰다. 그리고 선 그래프를 이용하는 탐구 능력 검사 문항은 근본적으로 그래프 자료의 해석을 통하여 이루어지므로 자료 해석 역시 제외시켰다. 그 외에 상관 관계, 인과 관계, 일반화 등의 탐구 요소는 선 그래프를 통한 결론 도출의 문항에서 어느 정도 반영할 수 있으므로 제외시켜서, 본 연구에서 선정된 과학 탐구 요소는 가설 설정, 변인 통제, 자료 변환, 예상, 추리, 결론 도출 등 6가지이다. 각 탐구 요소별 평가 목표는 Table 2와 같다.

과학 탐구 능력 검사 도구를 개발하기 위한 준거는 문헌 연구와 세미나의 토의를 거쳐 정하였고, 검사 도구의 개발 준거는 다음과 같다.

- ① 선 그래프를 해석하여 탐구 능력을 측정하는 것이므로 선 그래프를 중심으로 개발하되, 선 그래프와 관련된 지식은 중학교까지 배운 내용을 벗어나지 않게 개발한다.
- ② 검사 도구는 4지 선다형으로 개발하며, 검사 시간은 40분에 해결할 수 있도록 한다.
- ③ 문항의 소재는 중학교까지 배운 내용으로 하되, 암기 지식에 영향을 받지 않도록 탈내용적(content free)으로 개발한다.
- ④ 문항의 내용을 이해하기 위한 독해 수준은 고등학교 1학년용 기준으로 개발한다.
- ⑤ 문제 상황의 이해를 돕기 위하여 문항 속에 실험 상황의 그림을 제시한다.

위와 같은 탐구 요소별 평가 목표와 개발 준거에 따라 탐구 과정 요소별로 3~5개의 문항을 만들었고, 5차의 세미나를 통하여 결론의 도출 문항은 5문항, 추리 문항은 4문항, 그 외의 탐구 과정 요소별로 3문항씩을 개발하여 모두 21문항의 검사 도구 시안을 만들었다.

**Table 1** Subjects of pilot test and field test

Grade level	Pilot test					Field test						
	L-city*		Sm-city		TOT	L-city		Sm-city		Town		TOT
	Male	Female	Male	Female		Male	Female	Male	Female	Male	Female	
Grade 10	54	—	22	75	151	48	56	42	47	—	81	274
Grade 11	—	48	71	23	142	98	—	37	44	16	72	267
Grade 12	58	—	25	71	154	100	—	42	42	24	59	267
Total	112	48	118	169	447	246	56	121	133	40	212	808

\* L-city: Large city, Sm-city: Small and medium city

**Table 2** Six process skill objectives

Generating a hypothesis	Given a line graph of the experimental result, select a testable hypothesis.
Controlling variables	Given a line graph of the experimental result, identify the controlled variables.
Transforming data	Given a experimental data table, identify a line graph with the manipulated and responding variables appropriately assigned.
Prediction	Given a line graph of the experimental result, make a prediction of the probable facts(interpolation and /or extrapolation).
Inference	Given a line graph of the experimental result, make a reasonable judgement about the nonobservable properties of an object or event based on the line graph.
Drawing a conclusion	Given a line graph of the experimental result, draw a conclusion of the experiment, interpreting the graph

개발된 시안을 예비 및 현장 검사를 실시하여 문항을 수정하였다. 예비 검사를 실시하여 산출한 난이도, 변별도, 신뢰도, 타당도를 근거로 5개 문항을 수정하고, 2문항을 삭제하여 최종적으로 19문항을 선정하였다. 현장 검사를 실시하고 3개의 문항을 수정하여 신뢰롭고 타당한 과학 탐구 능력 검사 도구를 개발하였다.

### 3. 연구 결과 분석 및 통계 처리

과학 탐구 능력 검사 도구를 개발하기 위하여 사용된 분석 방법으로는 난이도와 변별도 검증을 통한 문항 분석, 내부 일치 신뢰도, 내용 타당도, 공인 타당도를 통한 타당도 검증 등이 사용되었다.

문항 난이도 계산은 각 문항에 대한 추측 요인을 교정 한 다음 공식(Davis, 1951)을 적용하였다. 난이도 P는

$$P = \left( \frac{R - \frac{W}{K_i - 1}}{N} \right) \times 100$$

여기서 R은 정답수, N은 전체 사례수, K<sub>i</sub>는 답지의 수, W는 오답수이다. 난이도 지수는 대개 20~80%가 적당하며, 평균 난이도는 50% 정도에 머무는 것이 이상적이다(Doran, 1980).

문항 변별도는 상위 집단과 하위 집단의 비율을 각각 27%로 구분하여 분석하는 Johnson(1951)의 변별도 지수를 사용하였다. 이 변별도 지수(D. I. : Item Discrimination Index)는

$$D.I. = \frac{R_u - R_L}{f} \quad (f = 0.27N)$$

여기서 R<sub>u</sub>는 상위 27% 집단의 정답자 수, R<sub>L</sub>는 하위 27% 집단의 정답자 수, f는 각 집단의 학생수, N은 피험자수를 나타낸다. 변별도 지수는 0.2 이상이면 받아드릴 수 있고, 그 값이 클수록 좋다(Aiken, 1979). 내부 일치 신뢰도는 KR-20의 공식을 이용하였으며 맞은 것을 1, 틀린 것을 0으로 채점했을 경우 Cronbach'α와 계

산 방법이 같다. 신뢰도는 0.80이상일 때 집단 평가 검사 도구로써 사용할 수 있는 신뢰로운 검사 문항이라고 할 수 있다(Doran, 1980).

내용 타당도는 과학 교육을 전공하는 대학원생 8명에게 검사 문항과 평가 목표간의 적절성 여부의 평가를 의뢰하여 나타난 CVI(Index of Content Validity) 지수로 구하였고, 최종 개발된 과학 탐구 능력 검사 도구와 TIPSⅡ 검사 도구와의 상관 계수를 이용하여 공인 타당도를 구하였다.

#### IV. 연구 결과 및 논의

과학 탐구 능력 검사 도구를 개발하기 위하여 예비 검사에 447명, 현장 검사에 808명에 대해서 조사하였다. 예비 및 현장 검사를 통하여 타당도, 변별도, 난이도, 신뢰도를 조사·분석하여 문항을 개발하였다.

##### 1. 과학 탐구 능력 검사 도구에 대한 문항 난이도

이 검사 도구의 예비 및 현장 검사의 정답률에 의한 문항의 난이도 분포는 Table 3과 같다. 예비 검사의 평균 난이도는 1학년이 0.53, 2학년은 0.61이고, 3학년은 0.63이며 전체적으로 0.59이며, 현장 검사의 평균 난이

도는 1학년이 0.42, 2학년은 0.49, 3학년은 0.50, 전체적으로는 0.47로 나타났다. 예비 검사와 현장 검사의 난이도 차이는 대상 학생의 차이와 난이도 지수가 현저히 낮은 2문항(9번, 12번)을 제외시킨 데서 비롯되었다. 그리고 문항 난이도가 0.20이하인 5문항(2번, 6번, 15번, 17번, 20번)을 수정·보완하였기 때문이다. Burn 등(1983)이 개발한 TIPSⅡ의 평균 난이도는 0.53, Mattheis 등(1988)이 개발한 POPS의 평균 난이도는 0.47, Padilla와 McKenzie(1988)가 개발한 TOGS의 평균 난이도는 0.51, 이종기(1988)가 개발한 검사 도구의 평균 난이도는 0.59, 권재술과 김범기(1994)가 개발한 검사 도구의 평균 난이도는 0.59로 본 연구의 예비 검사 문항의 평균 난이도와 비슷하였다.

문항의 변별도는 Table 4와 같다. 예비 검사에서 12번 문항의 변별도가 현저히 낮으므로 제외시켰는데, 1학년의 정답률이 현저히 떨어지고 2학년, 3학년의 정답률은 비교적 높은 것으로 나타났으나, 나머지 문항의 변별도는 무난한 것으로 나타났다.

과학 탐구 능력 검사 도구의 예비 및 현장 검사에서의 탐구 과정 요소별 평균 및 표준 편차는 Table 5와 같이 전체적으로 거의 유사하게 나타났으나, 예상과 결론 도출의 평가 문항에서 평균이 비교적 높다. 가설 설정의 평균이 예비 검사보다 현장 검사에서 조금 높은 것은 가

Table 3 Difficulty indices of process skill items

Grade \ Range	Pilot test				Field test			
	10	11	12	Total	10	11	12	Total
0 ~ 0.2	6	3	2	4				
0.21~0.40	6	6	5	6	11	9	8	9
0.41~0.60	4	4	5	4	3	3	3	4
0.61~0.80	5	6	7	6	5	7	8	6
0.81~1.00		2	2	1				
Mean	0.53	0.61	0.63	0.59	0.42	0.49	0.50	0.47

Table 4 Discrimination indices of process skill items

Grade \ Range	Pilot test				Field test			
	10	11	12	Total	10	11	12	Total
0~0.2	1	1	1	1				
0.21~0.40	6	6	3	3	2	2	2	1
0.41~0.60	10	3	6	10	10	9	11	12
0.61~0.80	4	10	7	6	7	8	5	6
0.81~1.00		1	4	1			1	

**Table 5** Mean scores(M) and standard deviation(SD) for process skill test by the grades and subskill

Subtest (items)	Grade	Pilot test(21 items)				Field test(19 items)			
		10	11	12	Total	10	11	12	Total
	M SD	M SD	M SD	M SD	M SD	M SD	M SD	M SD	
Generating a hypothesis(4, 6, 14)		1.12	1.20	1.17	1.17	1.45	1.49	1.50	1.48
		0.88	0.99	1.00	0.96	1.06	1.10	1.10	1.08
Controlling variables(2, 5, 10)		1.59	1.82	1.81	1.74	1.69	1.81	1.75	1.75
		0.95	0.90	0.94	0.93	0.88	0.95	1.00	0.94
Transforming data (3, 7, 15)		1.10	1.46	1.51	1.35	1.36	1.51	1.51	1.46
		0.94	1.09	1.07	1.05	1.07	1.09	1.07	1.08
Prediction (8, 16, 18)		1.94	2.20	2.20	2.11	2.02	2.09	2.06	2.06
		0.86	0.78	0.78	0.82	0.94	0.90	0.93	0.92
Inference (11, 13, 19)		1.71	1.73	1.87	1.77	1.81	1.92	2.04	1.92
		0.84	0.84	0.88	0.86	0.90	0.92	0.97	0.94
Drawing a conclusion(1, 9, 12, 17)		2.53	2.84	2.95	2.77	2.54	2.85	2.95	2.78
		1.10	1.13	1.04	1.10	1.11	1.13	1.04	1.10
Total		10.40	13.53	14.16	12.72	10.79	11.67	11.87	11.44
		5.01	4.54	4.43	4.68	4.31	4.21	4.57	4.39

설 설정의 4번과 6번 문항을 수정·보완한 결과이고, 추리의 평균이 예비 검사보다 현장 검사에서 조금 높은 것은 추리의 13번 문항을 수정·보완한 결과이다.

과학 탐구 요소별로는 가설 설정, 자료 변환의 탐구 과정 요소에 대한 평가 문항은 비교적 어려운 것으로 나타났다. 예상, 추리 결론 도출의 탐구 과정 요소에 대한 평가 문항은 비교적 쉬운 것으로 나타났다.

**2. 과학 탐구 능력 검사 도구의 신뢰도**

검사 도구에 대한 신뢰도는 검사의 동질성의 정도를 나타내는 KR-20 신뢰도를 구하였는데, KR-20 신뢰도는 맞으면 1, 틀리면 0으로 채점할 경우에 Cronbach's α의 신뢰도와 같으므로 이 값으로 구하였다. 신뢰도는 Table 6에 제시된 바와 같이 예비 검사에서는 1학년은 0.76, 2학년은 0.81, 3학년은 0.86, 전체 신뢰도는 0.82 이었고, 현장 검사에서는 1학년이 0.80, 2학년이 0.81, 3

학년은 0.85, 전체 신뢰도는 0.82였다. 문항을 수정·보완한 결과 1학년의 신뢰도가 향상되었다.

이 결과는 Burns 등(1983)의 TIPS II(0.86)나 Padilla와 Cronin(1986)의 MIPT(0.89) 그리고 이종기(1988)의 연구(0.86)보다는 낮으나, Cronin 등(1985)의 BAPS(0.78)나 Mattheis 등(1988)의 POPS(0.75) 그리고 권재술과 김범기(1994)의 연구(0.74)보다는 높다.

**3. 과학 탐구 능력 검사 도구에 대한 타당도 검증**

본 연구에서는 내용 타당도, 정답의 객관도, 문항의 명료성 점검 등과 공인 타당도를 구하였다. 내용 타당도와 정답의 객관도, 문항의 명료성 점검 등은 과학 교육을 전공하고 있는 대학원생 과학 교사 8명이 참여하였고, 공인 타당도(concurrent validity)는 TIPS II와의 탐구 요소별 상관 계수를 구하였다.

**Table 6** Reliabilities(KR-20) for the grades and total

Reliabilities	Grade	Pilot test				Field test			
		1	2	3	Total	1	2	3	Total
KR-20		0.76	0.81	0.86	0.82	0.80	0.85	0.85	0.82

1) 내용 타당도, 정답의 객관도, 문항의 명료성 점검

본 연구의 검사 도구에 대한 내용 타당도는 평가 도구가 측정하고자 하는 평가 목표를 충실히 반영하고 있는지를 측정하는 것으로 적절(1), 부적절(0)의 2단계 평정 방법을 적용하여 다음과 같은 타당도 지수(CVI: Index of Content Validity)를 구하였다.

$$CVI(\%) = \frac{\text{두 평정자가 같은 평가 목표를 측정하는 것으로 판단한 문항수}}{\text{전체 문항수}} \times 100$$

총 168개 응답(8명의 평정자×21문항) 중 78%가 평가 목표와 일치했다. 바람직한 타당도 지수가 80%이상(황정규, 1984)으로 볼 때, 비교적 낮은 타당도 지수 수준이다. 이것은 예상과 추리, 가설 설정과 결론 도출의 구분을 명확하게 이해하지 못한 결과로, 평가 목표를 구체적으로 기술하지 못한데서 비롯되었다. 이런 판단을 줄이기 위하여 평가 목표를 더 상세히 기술하였으며, 가설 설정의 문항을 수정·보완하였다.

과학 탐구 능력 검사 도구의 정답의 객관도 점검은 8명의 과학 교육을 전공하는 대학원생(과학 교사)에게 의뢰하여 작성한 답안을 기초로 분석하였고, 그 결과 검사 문항 개발자와 문항의 정답이 일치하는 비율은 96%였다. 따라서 본 연구 검사 도구의 정답의 객관도는 96%이다.

그리고 문항의 명료성은 세미나를 통하여 토의되었으며, 문항별로 문제가 있거나 더 좋은 표현이 있는 경우에 검사 문항을 수정·보완하였다. 명료성 점검의 준거로는 애매한 표현이나 문장의 문법, 문항 배열 순서, 고등학교 1학년 정도의 독해 수준, 내용 독립성, 문항의 논리성 등이다.

2) 공인 타당도

본 검사 도구의 공인 타당도는 현장 검사 도구와 TIPS II (Burns et, 1983)에서 일치하는 탐구 과정 요소 6가지와의 상관 관계를 구하였다. 가설 설정의 상관 계수는 0.69로 가장 낮았고, 변인 통제는 0.72, 자료 변환은 0.92, 예상은 0.86, 추리는 0.85, 결론의 도출은 0.89, 전체적으로는 0.83이다. 두 검사 도구의 탐구 과정 요소별 상관 계수가 높기 때문에 본 연구의 검사 도구 역시

각 탐구 요소를 측정하는 타당한 검사 도구라고 할 수 있다.

가설 설정과 변인 통제의 상관 계수가 비교적 낮게 나오는 것은 두 검사 도구의 가설 설정 문항 성격이 현저하게 다르기 때문이다. TIPS II의 가설 설정과 변인 통제 문항은 문장으로 진술한 상황에서 검증 가능한 가설을 찾거나 통제해야될 변인을 찾는 데 반해, 본 연구의 가설 설정과 변인 통제 문항은 제시된 가설에서 얻을 수 있는 선 그래프를 찾거나, 선 그래프로 제시된 상황에서 가설을 찾거나 통제된 변인을 찾는 문항으로 구성되어 있다. 나머지 탐구 요소인 자료 변환, 예상, 추리, 결론의 도출 등에서는 TIPS II의 문항에서도 그래프를 이용한 문항이 있다.

V. 개발된 탐구 능력 검사 도구에 대한 기술

1. 그래프를 이용한 과학 탐구 능력 검사(Test of Science Process Skills Using Line Graphs)

본 연구에서 개발한 탐구 능력 측정을 위한 검사 도구의 명칭은 “그래프를 이용한 과학 탐구 능력 검사(Test of Science Process Skills Using Line Graphs)”이며, 영문 약자로는 “TOPUG”라고 했으며, Table 7과 같이 탐구 과정 요소별로 3개 문항씩이지만, 결론의 도출 문항은 4문항으로 모두 19문항으로 구성되어 있다.

Table 7 Number of items of science process skills using line graphs by subskills

Subskills	Number of items
1. Generating a hypothesis	4, 6, 14
2. Controlling variables	2, 5, 10
3. Transforming data	3, 7, 15
4. Prediction	8, 16, 18
5. Inference	11, 13, 19
6. Drowing a conclusion	1, 9, 12, 17

그래프를 이용한 과학 탐구 능력 검사 문항을 각 탐구 과정 요소별로 1문항씩을 [부록]에서 예시하고 있다.



**Table 8** Characteristics of the test of science process skills using line graphs(TOPUG)

Process skills tested	Generating a hypothesis, controlling variables transforming data, prediction inference, drawing a conclusion	
Number of items	19	
Response format	Four alternative multiple choice	
Recommended grade level	10~12	
Time to complete test	40 Minutes	
Discrimination indices	Mean	Range
	0.57	0 ~0.20 ( 0 ) 0.21~0.40 ( 1 Item) 0.41~0.60 (12 Items) 0.61~0.80 (6 Items) 0.81~1.00 (1 Item)
	Mean	Range
	0.47	0 ~0.20 ( 0 ) 0.21~0.40 (9 Items) 0.41~0.60 (4 Items) 0.61~0.80 (6 Items)
	Difficulty indices	

**2. 그래프를 이용한 과학 탐구 능력 검사(TOPUG)의 개관**

본 연구에서 개발한 탐구 능력 검사 도구의 특성은 Table 8과 같다. 본 검사는 선정한 6가지 과학 탐구 과정 요소별로 평가 목표(Table 2 참조)가 기술되었으며, 평가 목표별로 3~4문항씩 총 19문항으로 구성되어 있다. 그리고 고등학교 1, 2, 3학년 학생을 대상으로 했으며, 객관식 4지 선다형 지필 문항으로 개발되었다. 검사 시간은 40분이다. 그리고 내용 타당도 지수는 78%, 정답의 객관도는 96%, 전체 평균 변별도 지수는 0.57, 난이도 지수의 평균은 46.9, 신뢰도(KR-20)는 0.82이다. 이와 같은 결과는 평가 도구가 타당하고, 신뢰롭다는 증거가 된다.

**VI. 요약 및 결론**

본 연구는 올바른 평가가 학습 능력을 향상시킬 수 있다는 가정에서 출발한다. 따라서 본 연구의 목적은 학생들의 탐구 능력을 진단하고, 과학 학습에서 탐구 학습의 효과를 알아보는데 도움을 주기 위하여 탐구 능력을 효

과적으로 측정할 수 있는 평가 도구를 개발하는 것이다. 선행연구에서 개발된 탐구 능력 검사 도구는 거의 문장을 통하여 측정하도록 되어 있으나 본 연구는 선 그래프를 이용한 탐구 능력 검사 도구이므로 탐구 능력 측정의 다양성을 제공하고, 과학의 실험 결과나 이론을 선 그래프로 표현하는 특성으로 볼 때, 유용한 시도라고 할 수 있다. 그러나 관찰, 분류, 측정 등 기초 탐구 능력을 측정하기 위한 검사 문항의 개발에서는 선 그래프 특성상 어려움이 있었다. 기초 탐구 능력은 초등학교를 대상으로 하는 경우가 많으므로 고등학교를 대상으로 한 본 연구의 과학 탐구 능력 검사 도구는 타당하다고 본다.

본 연구에서는 이론과 실제의 차이를 가장 극소화할 수 있어서 교육 자료를 개발할 때 자주 사용하는 R & D 과정을 택하였으며, 그 과정은 자료의 수집 및 정리, 탐구 요소의 선정 및 개발 준거 설정, 과학 탐구 능력 검사 도구 개발, 내용 타당도 및 명료성 점검, 예비 검사, 검사 문항 난이도, 변별도, 신뢰도 검사, 검사 문항 확정(현장 검사), 논문 작성 등의 순서이다.

과학 탐구 요소의 선정은 선행 연구를 토대로 기초 탐구 요소인 관찰, 측정, 분류 등을 제외한 탐구 과정 요소를 6가지, 가설 설정, 변인 통제, 자료 변환, 예상, 추리, 결론의 도출를 선정하고, 각 탐구 과정 요소별로 평가

목표와 검사 도구의 개발 준거를 기술하였다. 이를 근거로 각 과학 탐구 요소당 3~5문항씩 4지 선다형의 21문항을 개발하였다. 그리고 각 문항은 중학교 수준의 내용을 소재로 선택하였으며, 특정 지식이 없이도 해결할 수 있도록 하였다.

개발된 평가 도구는 예비 검사의 결과 분석을 통하여 난이도, 변별도 등의 분석으로 2문항을 제거하였으며 5개의 문항은 수정하였다. 이 결과를 토대로 내용 타당도, 정답의 객관도, 문항의 명료성 등을 과학 교육을 전공하고 있는 대학원생 교사 8명과 5차의 걸친 세미나를 통하여 점검하고 수정·보완하여 최종 19문항을 개발하였다. 그 결과 내용 타당도는 78%, 정답의 객관도는 96%이다.

최종적으로 확정된 과학 탐구 능력 검사도구의 평균 난이도는 0.47, KR-20 신뢰도는 0.82이었다. 공인 타당도를 구하기 위한 TIPSⅡ와의 상관 계수를 구하였는데, 전체적으로 0.83이었다. 이러한 결과는 본 연구를 통해 개발한 과학 탐구 능력 검사 도구가 신뢰롭고 타당한 평가 도구라는 증거가 된다. 이 검사 도구는 탐구 학습의 형성, 진단, 총괄 평가의 자료로 활용될 수 있을 것이며, 탐구 능력을 향상을 위한 참고 자료로도 활용할 수 있을 것으로 기대한다. 과학 탐구 능력 측정은 지필 검사만으로는 그 한계가 있으나 앞으로 다양한 측정 방법 즉, 실기 평가, 실험 보고서, 과제 연구 등의 개발을 통하여 효과적으로 과학 탐구 능력을 측정할 수 있어야 할 것이다.

## 적 요

과학 탐구 능력의 향상은 과학 교육의 중요한 목표가 되어왔다. 이 목표를 달성하기 위하여 과학 교육 과정, 탐구 수업 방법, 교육 환경의 개선이 중요하다. 또 다른 중요한 요인은 과학 탐구 능력을 측정할 수 있는 도구를 개발하는 것이다. 그 동안 과학 탐구 능력 검사 도구는 우리 나라에서 뿐만 아니라 외국에서도 많이 개발되었다. 그러나 대부분은 문장으로 주어진 상황에서 각 탐구 요소를 측정하도록 되어 있어서 각 검사 도구에서의 독특한 특징을 발견할 수가 없다.

본 연구의 목적은 선 그래프를 이용하여 신뢰롭고 타당한 과학 탐구 능력 검사 도구를 개발하는 것이다. 선행 연구 등을 통하여 선 그래프를 이용하여 측정할 수 있는 과학 탐구 요소 6가지(가설 설정, 변인 통제, 자료 변환, 예상, 추리, 결론 도출)를 선정하고, 예비 검사와

현장 검사를 통하여 문항을 수정·보완하고, 최종적으로 4지 선다형의 19문항을 개발하였다.

본 검사 도구의 내용 타당도는 78%, 신뢰도(KR-20)는 0.82, 평균 변별도 지수는 0.57, 그리고 평균 난이도 지수는 0.47이었다. 이 결과는 좋은 검사 문항의 기준에 합당하므로 본 검사 도구는 과학 탐구 능력을 측정할 수 있는 신뢰롭고 타당한 검사 도구라고 할 수 있다.

## 참 고 문 헌

- 권재술, 김범기(1994). 초·중학생들의 과학탐구능력 측정도구의 개발. 한국과학교육학회지, 14(3), 251-264.
- 김창식, 이화국, 권재술, 김영수, 김찬중(1991). 과학 학습 평가. 교육과학사. 113-168.
- 김창식, 권치순, 이화국, 한안진(1992). 과학 탐구실험 능력 평가 모형 및 평가 도구의 개발에 관한 연구. 한국과학교육단체총연합회.
- 나종국(1993). 생활소재를 이용한 과학 탐구 능력 평가 문항 개발. 한국교원대학교 대학원 석사학위 논문.
- 은경용(1992). 국민학생의 과학 탐구능력 측정을 위한 평가도구 개발. 한국교원대학교 대학원 석사학위 논문.
- 이연우(1989). 과학 탐구 능력 측정을 위한 표준화 검사지 개발 -중학교 2학년의 자료 분석과 해석 능력을 중심으로-. 한국교원대학교 대학원 석사학위논문.
- 이종기(1988). 고등학생의 과학 탐구 능력 측정을 위한 평가 도구 개발. 한국교원대학교 대학원 석사학위 논문.
- 이향노(1991). 고등학생의 과학탐구 능력 측정을 위한 평가 도구 개발 -지구과학소재를 중심으로- 한국교원대학교 대학원 석사학위논문.
- 金範基(1989). 物理教育의韓,日比較研究. 廣島大學博士學位論文.
- Baird, W. E., & Borich, G. D.(1987). Validity Considerations for Research on Integrated Science Process Skills and Formal Reasoning Ability, *Science Education*, 71(2), 259-269.
- Brotherton, P. N., & Preece, P. F. W.(1996). Teaching Science Process Skills. *International Journal of Science Education*, 18(1), 65-74.
- Burns, J. C., Okey, J. R., & Wise, K. C.(1983). Integrated Process Skills Test II, Department

- of Science Education University of Georgia Athens, Georgia 30602.
- Dillashaw, G. H., & Okey, J. R. (1980). Test of the Integrated Science Process Skills for Secondary Science Students, *Science Education*, 64, 601-608.
- Doran, R. L. (1980). *Basic Measurement and Evaluation of Science Instruction*, Washington, D.C. : National Science Teachers Association.
- Fyffe, D. W., & Robison, R. W. (1975). The Development of Criterion-validated Test Items for Four Intergrated Science Processes. *Journal of Research in Science Teaching*, 12(4), 415-421.
- Klopfer, L. E. (1971). *Evaluation of Learning in Science in Handbook of Formative and Summative Evaluation of Student Learning*. Bloom, Hastings and Madaus, Eds., New York: McGraw-Hill.
- Mattheis, F. E., Jones, M. L., & Nakayama, G. (1988). A U.S.-Japan Cooperative Study of Science Process Skills in Middle School Students Participating in the Fast Program (Results of Pre-Pilot-Tests in North Carolina).
- Padilla, M. J., & Cronin, L. L. (1986). *Middle Grades Integrated Science Process Skills Test (MIPT)*. Department of Science Education, University of Georgia, Athens, GA.
- Padilla, M. J., & McKenzie, D. L. (1986). The Construction and Validation of the Test of Graphing in Science (TOGS), *Journal of Research in Science Teaching*, 23(7), 571-579.
- Robinson, J. T. (1969). *The Nature of Science and Science Teaching*. California: Wadsworth Publishing Co. 117.
- 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(1), 33-50.
- Wenham, M. (1993). The Nature and Role of Hypotheses in School Science Investigations. *International Journal of Science Education*, 15(3), 231-240.
- Yap, K. C., & Yeany, R. H. (1988). Validation of Hierarchical Relationships among Piagetian Cognitive Modes and integrated Science Process Skills for Different Cognitive Reasoning Levels. *Journal of Research in Science Teaching*, 25(4), 247-281.