

# 수행평가와 선다형 지필 평가에 의한 변인확인 능력 평가의 차이

이 항 로  
대전과학고등학교

## A Difference of Identifying Variable Skills Assessment between Performance and Multiple Choice Items

Lee Hang-Ro  
(Taejon Science High School)

### ABSTRACT

Since 1960's, aims of science education have changed from attainment of scientific concepts, principles and laws to improvement of science process(or inquiry) skills. According to the science education philosophy like this, our nation has adopted improvement and evaluation of science process(or inquiry) skills in science education.

The purpose of this study was to examine the performance of high school students on 4 types of multiple choice items used to assess students ability to identify independent and dependent variables. Stimulus materials were either a question focusing on the relationship between two variables, a hypothesis, a description of an experiment, or a description of results of an experiment. Student performance on these item types was compared to this performance on a standard Piagetian interview task of variable identification.

The results of the study included : (1) the "hypothesis" type was the most difficult, while the "question" type appeared to be the easiest ; (2) the "procedure" item type had a higher correlation with the total interview than any other item type. Among conclusions reached in this study was that although all four item types operated similarly, they did not correlate very highly with the performance assessment by interview.

**Key words** : identifying variable skills, multiple choice item, item type, performance assessment, structured interview.

### 1 서론

제 3차 교육 과정 때부터 탐구 실험을 통한 과학 탐구능력 신장을 중심 축으로 하는 탐구 과학 교육

철학이 혁신적으로 도입되었으며, 그 후 개정된 과학 과 교육과정에서도 일관되게 과학 탐구능력 신장을 강조하여 왔다. 특히, 제 6차 교육과정에서는 탐구능력을 강조하는 공통과학 교과를 신설하였고 교과 내

\*1999년 1월 8일 받음

용 체계를 내용 영역과 탐구 영역으로 양분하여 탐구 영역에서는 실제 실험(concrete or hands-on)을 통한 과학 탐구능력 신장을 강조하고 있다.

이와 같은 철학을 기초로 하여 그 동안 과학 탐구능력 신장을 중심으로한 과학 학습 결과의 평가 방법과 도구 개발에 관한 연구가 꾸준히 국내외적으로 수행되어 왔다. 그 대표적인 예로 Dillashaw & Okey(1980)는 중·고등학생들을 대상으로 과학 탐구능력을 측정할 수 있는 평가도구인 TIPS(Test of Integrated Process Skills)를 개발하였으며, Cronin과 Padilla(1986), Mckenize & Padilla(1986), Smith & Welliever(1990)도 이와 유사한 과학 탐구능력 평가도구를 개발하였으며, 특히 최근에는 다양한 평가 방법을 제시한 Jones(1994), Sabra와 George(1994), David 등(1995), Janet 등(1996)의 연구가 있다. 우리나라에서도 과학 탐구능력 평가도구의 개발에 대한 필요성에 의해 이종기(1988), 이연우(1991), 이항로(1992), 김범기와 권재술(1994), 남정희(1996), 정 철(1997) 등의 연구가 있는데, 이 연구들은 비형식적 검사도구와는 달리 현장 검증을 통한 평가도구의 구비 조건을 갖추어 완성한 평가도구를 개발하여 보고한 바 있다. 또한 다수의 학생을 대상으로 평가해야 하는 현실적인 문제를 해결하기 위한 이 무(1991)의 평가 방법에 대한 연구가 있었다.

이들 연구에 포함된 탐구능력 측정 요소는 연구자에 따라 다소 다르나 모두 과학 탐구능력 평가도구라고 지칭하고 있다. 이들이 개발 제시한 평가도구를 구성하는 탐구능력 요소들 중, 변인 확인 능력 측정에 있어서는 대부분 하나의 실험에 대한 기술(description)을 제시한 문항에서 학생들로 하여금 독립변인과 종속 변인을 찾아내도록 하고 있다.

이와 같은 문항은 해결하고자 하는 과학 탐구 주제를 계획하고 실행하는 데 있어서 학생 스스로 참여하는 학생 중심의 탐구 학습 결과를 평가한 다기보다는 교사 중심의 평가관에서 나온 방법이며, 비교적 문항을 개발하고 관리하는 데는 편리할지 모르나 학생들이 알고 있는 것을 정확하게 평가할 수 없다고 지적되고 있다(Jones, 1994).

선다형 문항으로 과학 탐구능력을 평가할 때는 실험 상황에 대한 문항 진술의 모호함이나 추상성 때문에 변인 확인에 실패할 경우는 탐구능력 평가 방법 및 결과의 해석에 대한 문제가 제기될 수도 있다.

이러한 문제점이 있음에도 불구하고 현실적으로 대부분의 교사들은 선택형 지필 평가 유형을 가장 많이 사용하고 있는 실정이다. 이와 같은 선다형 지필 평가에서는 평가자가 설정한 정답을 선택하면 1점, 오답지를 선택하면 0점을 부여하는 이분적(dichotomous) 점수 체계(scoring system)를 가지며, 이와 같은 문항들을 이분문항(binary item)이라 부른다(성태제, 1991).

그러나 이와 같이 이분적 점수 체계를 갖는 과학 문제를 해결하는 사고 과정을 녹음한 내용을 근거로 작성한 프로토콜 분석에 의한 문제 해결 유형에 대한 연구 결과에 의하면, 선다형으로 평가할 경우 동일한 문제 해결 과정을 거쳤음에도 불구하고 개인에 따라서는 성공자가 실패자로 또는 실패자가 성공자로 평가되는 경우가 많은 것으로 나타났다(이항로, 1998). 이와 같은 연구 결과는 과학 학습 결과를 평가하는 방법에 따라 해석이 달라질 수 있음을 암시하는 동시에 그 동안의 평가 방법에 대한 대안적 평가(alternative assessment) 방법의 연구가 필요한 것으로 해석할 수 있다.

이러한 문제점을 보다 개선하기 위한 대안적 평가 방법으로 수행평가 방법의 적용이 강조되고 있는 실정이다. 수행평가(performance assessment)란 피험자 스스로 자신의 지식이나 기능을 나타낼 수 있도록 산출물을 만들거나 답을 작성(구성)하도록 요구하는 측정 평가 방식이다. 수행 검사에서는 학생이 배우고자 하는 지식이나 기능을 측정하거나 평가함에 있어서 선택형 검사와 같이 정답을 선택할 수 있는 능력이 곧 '지식을 안다'거나 '기능을 습득했다'고 가정하는 것을 부정하고, 학생이 답안을 작성하거나 행동으로 나타내는 것을 통해 지식이나 기능을 보다 직접적으로 측정 평가하고자 한다. 이러한 특성 때문에 수행검사라는 용어 대신에 진정한(authentic), 진짜(true), 적절한(appropriate), 직접적인(direct), 대안적인(alternative), 또는 지적인(intelligent) 검사

(assessment)라는 용어로도 두루 사용되고 있다(김민호, 1996; 백순근, 1995; Mattson, 1995; Frederiksen, 1990; Jongma, 1989; Wiggins, 1989). 이러한 수행평가의 종류로는 서술형(주관식) 검사, 논술형 검사, 구술 시험, 실기 시험, 실험법, 면접법, 관찰법, 포트폴리오(portfolio)법 등이 있다.

과학 탐구능력을 평가할 때도 교사에 따라서는 지필 검사, 실험 보고서 평가, 관찰 평가, 면접법, 실험 과제 평가 등 다양한 평가 방법을 사용할 수 있다(허명, 1996)는 견해는 수행평가의 중요성에 대한 함의로 볼 수 있다.

그러나 기존의 다양한 이분법적 선다형 평가 방법과 수행평가 방법을 이용했는데도 불구하고 이 두 방법에 의한 평가 결과가 상이하게 나타난다면 평가 결과의 신뢰성과 방법에 대한 의문을 제기하지 않을 수 없다. 실제로 관찰을 통한 수행평가와 선다형 지필 평가를 실시하고 있음에도 불구하고, 선다형 문항의 경우 문항의 표현 방법에 따라 많은 차이가 있음이 보고된 바 있다(Shaw et al., 1988).

결국 이와 같은 연구 결과들은 동일한 과학 탐구능력을 평가하더라도 평가 전략에 따라 평가의 결과에 차이가 날 수 있어 연구 결과의 해석이 달라질 수 있기 때문에 보다 객관적이고 정당한 결과 해석을 위해서는 다양한 평가 방법에 대한 연구가 필요하다.

이상과 같은 연구의 필요성에 의해 본 연구에서는 과학 실험 주제를 선정한 다음 주제에 관련된 독립변인과 종속변인 확인 능력을 선다형 지필 평가 방법과 학생들이 직접 실험을 수행하는 과정을 관찰하는 과정에서 변인확인 능력의 인지 여부를 면접을 통해 판정하는 수행평가 방법을 사용하여 이 둘간의 상호 관련 정도를 밝혀보고자 하였다.

본 연구에서 이용한 선다형 지필 평가 문항 유형의 종류는 실험 주제에 대한 질문을 제시하고 변인을 찾도록 하는 질문형, 실험 주제에 대한 가설을 제시하고 변인을 찾도록 하는 가설형, 실험 주제에 대한 실험 과정을 제시하고 변인을 찾도록 하는 과정형, 실험 주제에 대한 실험 결과를 제시하고 변인을 찾도록 하는 결과형 문항으로 문항의 표현 양식에서만 다르고 측정하고자 평가 목표는 모두 동일하다. 또한 선

다형 문항에 포함된 두 가지의 실험 주제를 직접 실험하는 과정을 관찰하면서 변인 확인 능력의 유무를 알아보는데 신뢰도를 높이기 위하여 미리 작성된 원고에 근거한 구조화된 면접법을 이용하였다.

선다형 지필 평가 방법과 수행평가 방법간의 상호 관련성을 분석하기 위한 본 연구에서는 다음과 같은 연구 문제를 설정하였다.

첫째, 선다형 문항 유형에 따라 변인확인 능력 평가는 차이가 있는가?

둘째, 문항 유형에 따른 선다형 평가와 수행 평가간의 상관 관계는 어떠한가?

셋째, 독립변인 확인능력과 종속변인 확인 능력 평가간에는 차이가 있는가?

## II. 연구의 방법 및 절차

본 연구는 과학 탐구능력을 구성하는 하위 요소들 중에서 변인 확인 능력을 네 가지 유형의 선다형 지필 평가와 수행평가의 한 방법인 면접에 의한 평가 결과간의 공통점과 차이점 및 상관 관계를 알아보고자 하였다. 고등학교 지구과학 교과서의 실험 주제 중에서 독립 변인과 종속변인의 관계가 뚜렷한 2종의 실험 주제를 선정하였다. 각 주제당 변인 확인 능력을 측정하는 5지 선다형 문항 16개를 개발한 다음, 실험 주제를 해결할 수 있는 실험 기구를 제작하였다. 제작한 실험 기구는 학생들이 실험을 직접하는 과정을 관찰하면서 면접을 통해 변인 확인 능력의 인지 여부를 알아보는 데 이용되었다. 다음으로 선다형 지필 평가와 면접에 의한 평가 결과들 간의 상호 관계를 기술하였다. 본 연구의 전체적인 진행 과정을 도식으로 나타내면 다음 fig 1과 같다.

### 1. 변인 확인 능력 측정을 위한 실험 주제 선정

현행 과학과 교육과정을 근거로 개발된 고등학교 지구과학 교과서에 기술된 실험 주제들 중에서 본 연구에서 선정한 실험 주제는 다음 표 1과 같다.

위와 같은 실험 주제를 선정한 이유는 지구과학 과목에서 다루는 대상은 거시적이고 많은 수의 변인들

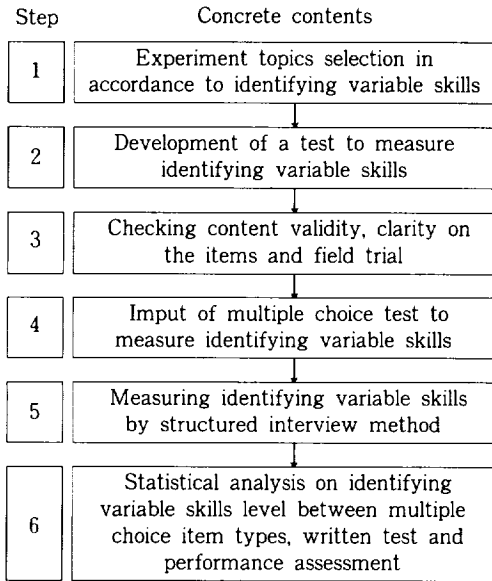


fig. 1. Research procedure used in this study

이 작용하는 경우가 많기 때문에 실험실에서 종속 변인에 영향을 주는 독립 변인들을 통제하기 어렵지만, 이 두 실험은 종속 변인과 독립 변인의 관계가 비교적 명확하고 실험실에서 재현하기 쉽기 때문이다.

## 2 변인 확인 능력 평가 도구 개발

위에서 선정된 실험 주제와 관련된 변인 확인 능력을 측정하기 위해 선다형 16 문항으로 구성된 변인 확인 능력 평가 도구를 개발하였다. 본 연구에서 개발한 변인 확인 능력 평가 도구의 문항 구성은 다음 표 2와 같다.

8문항(하나의 문항 유형 당 2 문항)은 독립 변인을 확인하는 능력을, 나머지 8문항(하나의 문항 유형 당 2 문항)은 종속 변인을 확인하는 능력을 측정하도록 구성되어 있다. 그러므로 검사 도구는 독립 변인에 대한 8문항의 하위 검사와 종속 변인에 대한 검

Table 1. Experimental topics selected on this study

Experimental topics	Content related with earth science	Independent variable	Dependent variable
Parallax measurement	Distance to a star	Distance	Parallax, the number of panel
Pendulum period measurement	Gravity of a planet	Length of pendulum, weight of bulb	Period of pendulum

Table 2. Item number distribution of identifying variable skills

Variable and item type	Identifying variable								Total of item
	Independent variable				Independent variable				
	Item type				Item type				
	Q	H	P	R	Q	H	P	R	
Parallax measurement	1	1	1	1	1	1	1	1	8
Pendulum period measurement	1	1	1	1	1	1	1	1	8
Total of items	2	2	2	2	2	2	2	2	16
	8				8				

Q : Question type, H : Hypothesis type, P : Procedure type, R : Result type

**Table 3.** Distinction of groups in test ordering

Group	Abbreviate word	Comparison
A	WID	W : multiple choice items I : identifying independent variable by interview D : identifying dependent variable by interview
B	WDI	
C	IDW	
D	DIW	

사 8문항의 하위 검사(subtest)로 이등분할 수 있다.

### 3. 변인 확인 능력 측정 도구의 타당도, 문항의 명료성 및 신뢰도 점검

#### 1) 변인확인 능력 평가 도구의 타당도, 문항의 명료성 점검

본 연구자가 1차로 개발한 평가 도구를 구성하는 문항에 대한 적절성과 명료성을 점검 받기 위해 작성한 체크리스트를 천문학자, 과학교육학자, 현장 지구 과학 교사에게 의뢰한 후 이들의 지시 사항을 근거로 수정·보완하여 완성하였다. 또한 완성한 평가 도구를 구성하는 문항들이 변인확인 능력이라는 탐구능력 평가 목표에 부합되는지를 점검하기 위해 작성한 내용 타당도 점검표를 의뢰하여 점검한 결과 내용 타당도 지수(CVI)는 98%로 나타났다. 여기서 체크리스트는 문항의 적절성과 명료성을 점검 받기 위한 항목들을 기술한 내용이고, 내용 타당도 점검표는 문항들이 평가하고자 하는 목표가 변인 확인 능력과 합치되는 정도를 기술한 내용으로 구성되어 있다.

#### 2) 평가 도구의 신뢰도 산출

내용 타당도와 문항의 명료성이 검증된 변인확인 평가 도구를 독립 변인과 종속 변인에 대한 조작적 정의를 학습한 일반계 고등학교 79명에게(1학년 40명, 2학년 39명으로 남자 39명, 여자 40명) 투입하여 평가도구의 신뢰도를 산출한 결과 0.81이었다(KR-20).

### 4. 변인확인 능력 평가 피험자 집단 구분

위에 제시한 과정을 거쳐 완성된 변인 확인 능력

평가 도구를 신뢰도 산출 대상자와는 다른 98명의 학생(과학고등학교 학생 88명, 과학실험평가 대회에 참가한 10명)에게 투입하여 얻은 결과를 통계적 분석에 이용하였다.

문항 유형별 선다형 지필 검사와 수행평가의 한 방법인 면접을 통한 변인 확인 능력 성취도 검사를 받는 순서에 따른 영향을 최소화시키기 위해 피험자들을 다음 표 3과 같이 4집단으로 구분하고 그 중의 한 집단에 무선적으로(randomly) 할당하였다.

집단 A의 검사 순서는 지필 평가 → 면접에 의한 독립 변인 확인능력 평가 → 면접에 의한 종속 변인 확인능력 평가이고, 집단 B의 검사 순서는 지필 평가 → 면접에 의한 종속 변인 확인능력 평가 → 면접에 의한 독립 변인 확인능력 평가이고, 집단 C의 검사 순서는 면접에 의한 독립 변인 확인능력 평가 → 면접에 의한 종속 변인 확인능력 평가 → 지필 평가이고, 집단 D의 검사 순서는 면접에 의한 종속 변인 확인능력 평가 → 면접에 의한 독립 변인 확인능력 평가 → 지필 평가의 순이다.

동일한 학생들이 변인 확인 능력 선다형 평가도구에 관련된 2개의 실험 주제를 실제로 실험을 수행하면서 해결해 가는 전체 과정의 관찰과 면접을 통한 수행평가(performance assessment)에서 얻은 변인 확인 능력 성취도를 선다형 평가 방법에 의한 성취도와의 상호 관계 분석에 이용하였다.

### 5. 수행평가에 의한 변인 확인 능력 평가의 실시 및 점수 부여

학생들에게 진자의 길이가 다른 3개의 실과 질량이 다른 3개의 추로 구성된 단진자 실험 세트와 시차 측

정 장치를 주었다. 단진자 실험에서 측정된 진자의 길이와 추의 질량값 변화에 따른 주기의 변화, 물체까지의 거리 변화에 따른 시차와 배경 관지 개수의 변화 값을 측정하여 도표로 작성한 후 독립변인과 종속변인을 기록하도록 하였다.

각각의 실험 주제를 해결해 가는 과정을 본 연구자가 직접 관찰하였으며, 측정 자료의 신뢰도를 높이기 위하여 미리 원고(manuscript)를 작성한 다음 이를 근거로 면접을 실시하는 구조화된 면접법(structured interview method)을 이용하여 변인들의 인지 여부를 확인하고 기록하였다.

이때 실험 결과를 기록한 각 변인들의 측정값과 변인 이름, 면접을 통해 확인된 변인들이 모두 일치하는 경우만을 변인확인 능력을 인지하고 있는 것으로 판정하였다.

각 실험 주제에 관련된 변인 확인 능력에 대한 성취도 점수 부여는 어떠한 변인도 확인하지 못하였으면 0점, 하나의 변인을 확인하였으면 1점, 두 개의 변인을 확인하였으면 2점, 2개 이상의 변인을 확인하였으면 3점을 부여하였다.

### 6. 집단간 변인 확인 능력 성취 수준에 대한 통계적 검증

4 집단에서의 지필 검사에 의한 변인 확인 능력 성취도 수준과 수행평가 방법인 면접에 의한 성취도

수준간의 차이에 대한 통계적 검증을 실시한 후, 4 집단을 다시 한 집단으로 합하여 분석하였다. 전체 검사와 하위 검사에 대한 기술통계치와 Spearman 상관 계수를 구하였다. 4가지 문항 유형에 대한 학생들의 성취 수준 차이, 종속 변인과 독립 변인에 대한 하위 검사간의 차이 그리고 선다형 지필 평가와 수행 평가에 의한 변인확인 능력에 차이가 있는지를 확인하기 위하여 분산 분석을 실시하였다.

## III. 연구 결과 및 논의

선다형 지필 평가 문항의 유형에 따른 변인 확인 능력의 성취도 차이를 검증한 결과, 선다형 문항을 통한 변인 확인 능력 평가와 관련 실험 주제를 수행하는 과정을 관찰하는 동안 면접을 통해 얻은 변인 확인 능력 평가간의 상관 관계를 분석한 결과, 문항 유형에 따른 독립 변인과 종속 변인 확인 능력 성취도에 대한 차이를 검증한 결과를 기술하고자 하였다.

### 1. 선다형 지필 평가 문항 유형에 따른 변인 확인 능력 성취도 차이

본 연구 대상 학생들은 다음 table 4에 제시된 바와 같이 “가설형” 문항에서 4점 만점중 3.16점으로 최저 값을, “질문형” 문항에서 3.60 점으로 최고 값을 나타냈다.

Table 4. Mean scores and standard deviations

Measure	N	Maximum score	$\bar{X}$	SD	Range
Total Variable Identification Test	98	16.00	13.5204	3.6361	0.00-16.00
Independent Subtest	98	4.00	7.0150	2.0926	0.00- 4.00
Dependent Subtest	98	4.00	6.4694	1.9541	0.00- 4.00
Question Type	98	4.00	3.6020	.9165	0.00- 4.00
Hypothesis Type	98	4.00	3.1633	1.0716	0.00- 4.00
Procedure Type	98	4.00	3.4286	.9844	0.00- 4.00
Results Type	98	4.00	3.3265	1.0232	0.00- 4.00
Total Interview Score	98	6.00	5.2347	.9823	0.00- 6.00
Independent Interview Score	98	3.00	2.6531	.5198	0.00- 3.00
Dependent Interview Score	98	3.00	2.5816	.7169	0.00- 3.00

**Table 5.** Analysis of variance in the score on four item types by multiple choice item

Source of variation	Sum of squares	D.F.	Mean squares	F Ratio	F Prob.
Between groups	5,232	3	1,744	1,760	.154
Within groups	384,561	388	.991		
Total	389,793	391			

**Table 6.** Analysis of variance in the score on the identifying variables between four groups by multiple choice item

Source of variation	Sum of squares	D.F.	Mean squares	F Ratio	F Prob.
Between groups	23,724	3	7,908	.597	.618
Within groups	1244,450	94	13,239		
Total	1268,173	97			

**Table 7.** Analysis of variance in the score on the identifying variables between four groups by performance assessment

Source of variation	Sum of squares	D.F.	Mean squares	F Ratio	F Prob.
Between groups	21,450	3	7,150	1,493	0.132
Within groups	73,611	94	.783		
Total	95,061	97			

질문형 평가, 가설형 평가, 과정형 평가, 결과형 평가 점수간의 차이가 있는지를 검증하기 위해 분산분석을 실시한 결과는 다음 table 5와 같다.

F값은 1.760이고, 유의도가 .154이므로 '평가 문항 유형간의 점수 차이가 없다'는 귀무 가설은 유의 수준 .05에서 채택된다. 즉, 평가 문항 유형간의 차이는 없는 것으로 판단된다. 따라서 선다형 지필 평가 방법으로 변인확인 능력을 측정하는 경우는 문항 유형에 따라 많은 차이가 나지는 않으나 그 결과가 완전하게 일치한다고 볼 수는 없다.

## 2 선다형 평가와 수행평가의 집단별 차이

1) 선다형 평가에 의한 집단간의 변인확인 능력 차이 분석

선다형 평가 방법에 의한 변인확인 능력이 집단별로 차이가 있는지를 검증하기 위해 분산분석을 실시한 결과는 다음 table 6과 같다.

F값은 .597이고, 유의도가 .618이므로 '변인확인능력이 집단간에는 차이가 없다'는 귀무 가설은 유의 수준 .05에서 채택된다. 따라서 선다형평가 의한 변인확인 능력은 집단간에 차이가 없는 것으로 나타났다.

2) 수행평가에 의한 집단간의 변인확인 능력 차이 분석  
수행평가 방법에 의한 변인확인 능력간의 점수가 집단별로 차이가 있는지를 검증하기 위해 분산분석을 실시한 결과는 다음 table 7과 같다.

F값은 1.493이고, 유의도가 .132이므로 '변인확인 능력이 집단간에는 차이가 없다'는 귀무 가설은 유의 수준 .05에서 채택된다. 따라서 수행평가에 의한 독

립변인과 종속변인 능력은 집단간에 차이가 없는 것으로 나타났다.

### 3. 선다형 평가와 수행평가 간의 상관 관계

선다형 지필 평가 문항 유형에 따른 성취도와 수행평가의 면접에 의한 성취도간의 상관 계수는 다음 table 8에 제시한 바와 같이 0.01~0.13인 것으로 나타났다. 선다형 문항 유형과 변인확인 능력 총점간에는 전체적으로 매우 높은 상관을 보이는 반면, 선다형에 의한 독립변인 확인 능력 평가와 종속변인 확인 능력 평가간의 상관은 .61정도로 낮다고 볼 수 있다. 선다형과 수행평가의 면접에 의한 변인확인 능력 평가간의 상관은 .078로 매우 낮은 상관을 보이고 있다. 이는 평가 방법에 따라 학생들이 다르게 분류될 수 있음을 의미하는 것으로 해석할 수 있다. 또한 수행

평가에 의한 독립변인 확인 능력과 종속변인 확인 능력간에는 .243으로 매우 낮은 상관 관계를 보이고 있어 독립변인 인지 여부가 반드시 종속변인 인지 여부를 결정한다고 볼 수 없다.

### 4. 문항 유형에 따른 독립 변인과 종속 변인에 대한 성취도 차이

선다형 지필 검사에서의 독립 변인과 종속 변인에 대한 평균 점수는 8점 중에서 7.01점과 6.47점으로 매우 비슷하지만, 이 두 하위 검사간의 상관 계수(r)는 0.61로 낮은 편이다(table 8).

수행평가인 면접을 통한 독립 변인 확인 능력 성취도와 종속 변인 확인 능력 성취도의 평균은 각각 3점 중 2.65와 2.58이었으며, 이들 성취도간의 상관관계는 유의 수준 0.05에서 0.24로 매우 낮은 것으로 나

**Table 9.** Analysis of variance in the score on two variables by performance evaluation

	Sum of squares	D.F.	Mean squares	F Ratio	F Prob.
Score between groups	1,540	2	.770	5,514	.002
Within groups	48,307	95	.508		
Total	49,847	97			

**Table 8.** Matrix of correlation coefficients for all measures

	T	I	D	Q	H	P	R	TI	II	DI
T	1.00									
I	.906**	1.00								
D	.891**	.614**	1.00							
Q	.892**	.865**	.733**	1.00						
H	.935**	.864**	.815**	.745**	1.00					
P	.883**	.709**	.884**	.749**	.741**	1.00				
R	.931**	.870**	.800**	.774**	.914**	.705**	1.00			
TI	.078	.079	.060	.013	.077	.130	.055	1.00		
II	.004	.016	-.011	.010	.002	.047	.049	.706**	1.00	
DI	.104	.097	.090	.011	.104	.144	.111	.858**	.243*	1.00

\*\* significant at .01 level \* significant at .05 level

T : total variable identification test

I : independent subtest

D : dependent subtest

Q : question type

H : hypothesis type

P : procedure type

R : results type

TI : total interview

II : independent interview

DI : dependent interview



타났다. 또한 이들에 대한 분산분석 결과는 다음 table 9와 같다.

F값은 5.514이고, 유의도가 .002이므로 '수행평가에 의한 변인 확인 점수간에 차이가 없다'는 귀무 가설은 유의 수준 .05에서 기각된다. 따라서 수행평가에 의한 독립변인과 종속변인 점수간에는 차이가 있는 것으로 나타났다. 이는 두 개의 독립변인과 종속변인이 내재된 문제를 해결하는 과정에서 어느 한쪽의 인지 여부와 다른 한 쪽의 인지 여부와는 별개의 문제일 수도 있음을 함의하는 것으로 볼 수 있다.

### N. 결론 및 제언

본 연구를 통해 밝혀진 연구 문제별 결과는 다음과 같다.

첫째, 선다형 문항 유형에 따라 변인확인 능력 평가는 차이가 있는가?

선다형 지필 평가 문항 유형에 따라 학생들의 변인 확인 능력 성취 수준이 비슷할지라도, 이들은 수행평가 방법인 면접을 통한 변인 확인 능력 성취도와 상관 관계가 매우 낮은 것으로 나타났다.

둘째, 문항 유형에 따른 선다형 평가와 수행 평가간의 상관 관계는 어떠한가?

선다형 평가 문항 유형에 따른 변인 확인 능력 성취도와 수행평가 방법인 면접에 의한 변인 확인 능력 성취도간의 상관 관계는 매우 낮은 것으로 나타났다. 만일 교사가 면접을 통해 학생들의 탐구능력을 평가한다면 학생들은 지필 평가에서는 잘 할 수 있었어도 문제를 해결하는 과정을 정확하게 기술하는 데 어려움을 가질 수도 있고 그렇지 않을 수도 있다. 따라서 면접과 선다형 문항들을 함께 하는 평가를 필요로 한다

셋째, 독립변인 확인능력과 종속변인 확인 능력 평가간의 차이가 있는가?

독립 변인과 종속 변인 성취도 수준간에 상관 관계가 낮은 것으로 보아 하나의 변인을 인지하고 있는 학생이 다른 변인을 인지하고 있다는 보증을 할 수 없는 것으로 판단된다.

본 연구에 나타난 결과에 의하면 과학 교육학자들이나 현장의 과학 교사가 학생들의 과학 탐구능력을

정확하게 평가하고 있지 못하고 있을지도 모른다는 판단을 내릴 수 있다. 특히, 독립 변인과 종속 변인 확인 능력과 관련된 평가의 경우는 더욱 그러하다고 볼 수 있다. 독립 변인과 종속 변인들간의 상관 관계가 낮고 문항 유형들 간에 유의미한 차이가 없기 때문에 변인 확인 능력을 평가할 때 한 가지 방법보다 그 이상의 방법을 가능한 한 사용하여야 함을 의미하며, 과학 탐구능력 평가에 있어서 표준화 검사와 같은 평가 방법을 이용하여야 함을 암시하고 있다.

이상과 같은 본 연구의 결과로부터 다음과 같은 결론을 내릴 수 있다.

선다형 지필 평가 문항 유형에 따른 변인 확인 능력 평가 결과가 유사할지라도 수행평가인 면접에 의한 결과와는 상관 관계가 높지 않으며 독립변인의 성취 유무가 종속 변인의 성취 유무를 설명하지 못하는 것으로 나타났다. 따라서 과학 탐구능력을 평가하고자 할 때는 다양한 선다형을 이용하는 것은 물론이고 문제를 해결해 가는 과정에 중점을 두는 수행평가 방법도 병행하는 것이 바람직한 것으로 해석할 수 있다. 이상과 같은 연구 결과로 볼 때 학생들이 과학 탐구 문제를 실제로 수행하는 과정을 중심으로 종합적으로 평가하는 수행평가 결과와 기존의 선다형 지필 평가 방법에 의한 결과간에 차이가 나는 원인을 밝히기 위한 문제 해결시의 사고 과정을 중심으로 한 분석적이고 심층적인 연구가 필요하다.

### 적 요

본 연구에서는 과학 탐구능력을 구성하는 하위 요소들 중에서 변인확인 능력을 측정하는 평가 문항의 유형에 따른 차이와 최근에 한층 더 강조되고 있는 수행평가 방법중의 하나인 면접법을 통한 변인확인능력 평가 결과간의 상관 관계를 알아보려고 하였다.

연구 결과에 제시한 바와 같이 선다형 지필 평가 문항 유형별 변인확인 능력은 대체로 높은 상관관계를 나타냈으나, 선다형에 의한 변인확인능력 평가 결과와 수행평가 방법에 의한 평가 결과는 매우 낮은 상관 관계를 나타냈다.

특히, 독립변인 확인 능력과 종속변인 확인 능력은

거의 상관 관계가 없는 것으로 나타나 두 능력 중 어느 한쪽의 인지 여부가 다른 한쪽의 인지 여부를 결정해 주지는 못하는 것으로 볼 수 있다. 이는 과학교육 활동에서 이 두 능력을 동시에 강조하는 과학과 교수 학습 전략의 적용이 필요함을 함의한다.

또한 선다형에 의한 평가 결과와 수행평가에 의한 평가 결과가 상이한 결과를 보이는 것으로 보아 선다형만을 이용한 과학 탐구능력 평가와 그에 관련된 연구보다는 수행평가 방법과 같은 과정 평가를 이용한 정성적 방법을 병행하는 것이 평가의 결과와 연구 결과의 해석에 대한 타당성을 한층 높여줄 것으로 해석된다.

### 참 고 문 헌

김민호(1996). 국민학교 수학 수행평가 결과에 관한 분석. 한국교원대학교 대학원 석사학위 논문, 6-14.

김범기, 권재술(1994). 초·중학생들의 과학 탐구능력 측정 도구의 개발. 한국과학교육학회지, 14(3), 215-264.

남정희(1996). 화학 실험에서 과학 탐구 과정 기능의 평가도구 개발. 한국교원대학교 대학원 박사 학위 논문.

백순근(1995). 부분점수를 이용한 다단계형 자료의 분석 연구. 교육평가연구, 8(1), 59-75.

성태제(1991). 문항반응이론 입문. 서울 : 양서원, 14-15.

이 무(1991). 과학적 탐구 사고력 평가 문항 형태에 관한 연구. 서울대학교 대학원 박사 학위 논문.

이연우(1991). 중학생의 과학 탐구능력 측정을 위한 표준화 검사지 개발-자료해석 능력을 중심으로-. 한국교원대학교 대학원 석사 학위 논문

이종기(1988). 고등학생의 과학 탐구능력 측정을 위한 평가도구의 개발. 한국교원대학교 대학원 석사 학위 논문

이항로(1991). 고등학생의 과학 탐구능력 측정을 위한 평가도구의 개발-지구과학 소재를 중심으로-. 한국교원대학교 대학원 석사 학위 논문.

이항로(1998). 과학 탐구능력과 개념 이해도가 대학

수학능력시험 지구과학 문제 해결에 미치는 영향. 한국교원대학교 대학원 박사 학위 논문.

정 철(1997). 구조화된 문항을 이용한 지구과학 탐구 능력 평가도구 개발. 한국교원대학교 대학원 석사 학위 논문.

허 명(1996). 과학교육 평가의 문제점 및 개선 방향. 제 1회 과학교육 연합학술대회 보고서, 1-13

Cronin, L.L., & Padilla, M.J.(1986). *The development of a middle grades integrated science process skills test*. Paper presented at the meeting of the National Association for Research in Science Teaching, San Francisco, CA.

David L. Radford, Linda L. Ramsey and William C. Deese.(1995). *Demonstration Assessment, Science and Children*, October, 52-55.

Dillashaw, E.G., & Okey, J.(1980). Test of integrated process skills for secondary students. *Science Education*, 64, 601-608.

Frederiksen, N.(1990). Introduction in N. Frederiksen, R. Glaser, A. Lesgold & M. G. Shafto(Eds). *Diagnostic monitoring of skill and knowledge acquisition*. Hillsdale, NJ : Erlbaum, i x - x v ii.

Janet K. Kleinheider(1996). Assessment Matters. Adapt this model for an alternative assessment tool. *Science and Children*, January, 23-41.

Jones, M.G.(1994). *Assessment Potpourri. Science and Children*, October, 14-17.

Jongsma, K.S.(1989). Portfolio assessment. *The Reading Teacher*, 3(73), 264-265.

Mattson, L.(1995). *Portfolios in the major*. Sonoma State University, California.

Mckenize, D.L., & Padilla, M.J.(1986). The construction and validation of the test of Graphing in Science(TOGS). *Journal of Research in Science Teaching*, 23, 571-579.

Sabra Price and George E. Hein.(1994). Scoring Active Assessment, *Science and Children*, October, 26-29.

Shaw, Edward L., Jr and Others(1988). *Assessing the Skills of Variable Identification with Multiple Choice Items*, ED 292 663.

Smith, K.A., & Welliever, P.W.(1990). The

development of a science process assessment for fourth-grade students. *Journal of Research in Science Teaching*, 27(8), 727-738.

Wiggins, G.(1989). A true test: toward more authentic and equitable assessment. *Phi Delta Kappan*, May, 703-713.

**[부록]**

**변인 확인 능력 측정을 위한 면접 원고와 변인 확인 능력 평가도구**

※ 다음 그림은 시차에 의한 별까지의 거리 측정을 위한 실험의 모식도이다. 물음에 답하시오(1~ 8)



**1. 시차 측정 면접 원고**

교사 : “여기에 시차를 측정하는 실험 기구가 있습니다.”

학생 : 여러분은 시차에 대하여 충분히 생각해 주시기 바랍니다.

교사 : “나는 이 측정 장치를 이용하여 물체까지의 거리를 측정하려고 합니다. 내가 물체를 앞뒤로 이동시킬 때 어떤 일이 일어나는지 자세히 보기 바랍니다.”

학생 : 여러분은 물체가 앞뒤로 이동하는 것을 자세히 보시기 바랍니다.

교사 : “20cm, 40cm, 60cm인 곳에 물체가 있습니다.”

학생 : 여러분은 시차에 대해 충분히 생각해 주시기 바랍니다.

교사 : “자, 우리는 40cm인 곳에 놓여 있는 물체를 양 눈을 번갈아 감으면서 보았습니다. 만일 60cm인 곳에 놓여 있는 물체를 양 눈을 번갈아 가면서 본다면 20cm인 곳에 놓여 있는 물체를 볼 때와 무엇이 다르겠습니까? 종속 변인은 무엇입니까?”

학생 : 충분히 생각하고 대답해 주시기 바랍니다.

교사 : 학생이 하나의 변인만 지적하면 “그 이외에 또 다른 변인은 없다고 생각합니까?”라고 질문한다.

교사 : “지금 10cm인 곳에 놓여 있는 물체를 양 눈을 번갈아 감으면서 보면 어떻게 달라지겠습니까? 독립 변인은 무엇입니까?”

학생 : 충분히 생각하고 대답해 주시기 바랍니다.

교사 : 학생이 하나의 변인만 지적하면 “그 이외에 또 다른 변인은 없다고 생각합니까?”라고 질문한다.

**2. 시차 측정 관련 변인확인 능력 평가도구**

1. 철수는 “물체까지의 거리는 시차에 어떤 영향을 주는가?” 라는 질문에 답하고 싶었다. 만일 그가 측정을 하면서 이 질문에 대답한다면, 종속 변인은 무엇일까?

- ① 시차
- ② 물체까지의 거리
- ③ 물체의 크기
- ④ 시차 내에 보이는 배경의 나무 판지 수
- ⑤ 물체의 색

2. 철수는 “물체까지의 거리는 시차에 어떤 영향을 주는가?” 라는 질문에 답하고 싶었다. 만일 그가 측정을 하면서 이 질문에 대답한다면, 독립 변인은 무엇일까?

- ① 물체까지의 거리
- ② 시차
- ③ 물체의 크기
- ④ 시차 내에 보이는 배경의 나무 판지 수
- ⑤ 물체의 색

3. 철수는 “물체까지의 거리가 멀수록 시차는 감소할 것이다.”라는 가설을 검증하기 위해 실험한 결과 “거리가 멀었을 때 시차는 감소했다.” 이 실험에서 종속 변인은 무엇일까?

- ① 물체까지의 거리
- ② 시차
- ③ 물체의 크기

④ 시차 내에 보이는 배경의 나무 판지 수

⑤ 물체의 색

4. 철수는 "물체까지의 거리가 멀수록 시차는 감소할 것이다."라는 가설을 검증하기 위해 실험을 했다.

이 실험에서 독립 변인은 무엇일까 ?

① 시차

② 물체까지의 거리

③ 물체의 크기

④ 시차 내에 보이는 배경의 나무 판지 수

⑤ 물체의 색

5. 철수는 위의 그림과 같은 실험 방법을 고안했다. 그는 물체까지의 거리를 달리하면서 시차의 변화를 자세히 관찰하였다. 그는 물체까지의 거리를 10cm, 20cm, 30cm로 변화시키면서 이 실험을 계속했다. 이 실험에서 종속 변인은 무엇일까 ?

① 물체의 색

② 물체까지의 거리

③ 시차

④ 시차 내에 보이는 배경의 나무 판지 수

⑤ 물체의 크기

6. 철수는 위의 그림과 같은 실험 방법을 고안했다. 그는 물체까지의 거리를 달리하면서 시차의 변화를 자세히 관찰하였다. 그는 물체까지의 거리를 20cm,

40cm, 60cm로 변화시키면서 이 실험을 계속했다.

이 실험에서 독립 변인은 무엇일까 ?

① 물체의 색

② 시차

③ 물체까지의 거리

④ 시차 내에 보이는 배경의 나무 판지 수

⑤ 물체의 크기

7. 철수는 시차 실험에서 물체까지의 거리를 증가시킬 때 시차의 크기는 감소했다. 이 실험에서 종속 변인은 무엇일까 ?

① 시차 내에 보이는 배경의 나무 판지 수

② 물체까지의 거리

③ 물체의 크기

④ 시차

⑤ 물체의 색

8. 철수는 시차 실험에서 물체까지의 거리를 증가시킬 때 시차의 크기는 감소했다. 이 실험에서 독립 변인은 무엇일까 ?

① 시차 내에 보이는 배경의 나무 판지 수

② 시차

③ 물체의 크기

④ 물체까지의 거리

⑤ 물체의 색