

과학적 탐구 사고력의 3단계 선다형 평가 연구

이 무

(서울대학교)

Development of the Three-tier Test Items for the Thinking Skills of the Scientific Inquiry

Lee, Moo

(Seoul National University)

ABSTRACT

In order to assess students' higher mental abilities, such as scientific inquiry thinking skills, the essay type items would be more adequate than the multiple choice items. However, due to the present condition in which a huge number of students take the examination at the same time, it is inevitable to use the multiple choice type. For this reason, it is necessary to develop a new type of multiple choice items which can reduce the disadvantages of the traditional multiple choice type and can achieve a similar level of validity as subjective type assessment. The three-tier multiple choice test items which can be used for a large sample of students and especially for scientific inquiry thinking abilities, are proposed and examined.

The three-tier multiple choice test items asked firstly conclusion or the results of calculation or experimental apparatus, secondly the processes of calculation or of developing conclusion, thirdly asking relevant scientific concepts.

For the item analysis, 1 point was given to the correct answer, while 0 point was given to the wrong one. The data were processed through the computer program developed in Turbo C 2.0 language with an IBM compatible personal computer.

The average score in the sub-items asking for scientific concepts was lower than that in the sub-items asking for results or processes. The score of guessing by chance in the three-tier multiple choice items was only 0.13%, so that the probability of making correct answers by just guessing would be extremely low.

The three-tier multiple choice items, even if they are objective items, are thought to assess thinking skills of the scientific inquiry meaningfully excluding the possibility of guessing by chance.

Key words : scientific inquiry, thinking skill, three-tier multiple choice, score of guessing.

I. 서론

과학적 탐구 사고력 요소의 부분적인 그리고 종합

적인 적용은 탐구 사고력 요소에 대한 이해를 바탕으로 지적인 적용 측면을 뜻하는데 지필 검사로 어느 정도 평가할 수 있다. 그러나 실험적 적용은 실제로 지적 활동

*1998년 10월 15일 받음

이 관찰이나 실험을 하는 과정에 적용되는 것을 뜻하므로 실험 활동의 직접 관찰과 면접 또는 실험보고서로 얼마간 평가할 수 있으며, 새로운 추구 방법 창안도 경연 대회 등에서 일부 시도할 수 있겠으나 일반적으로 다인수 지필 고사로는 곤란하며 현재의 대학 입시 평가 상황으로는 더욱 곤란하다. 대학수학능력시험에서는 과학 기초 개념 이해와 과학적 탐구 사고력 요소의 부분적 및 종합적인 적용 수준의 평가가 가능하다고 하였다. 과학적 탐구 사고력은 역학, 전자기학, 원자물리학, 유전학, 생태학 등의 자연 과학을 발전시켜 근거있게 체계적으로 자연 현상을 이해하고 설명하며 예언하고 통제할 수 있을 뿐만 아니라 광범위하게 구체적으로 응용할 수 있다는 점에서 실질적이며 실제적이다. 한국의 사회 문화적 풍토 속에 분석적 지성이 미흡하고 실질적이고 실제적인 면이 미약한 것을 고려할 때 실증성과 어울리는 과학적 탐구 사고력은 더욱 중요한 의의가 있다.

관찰이나 면담, 주관식 문항이 학생의 성취뿐 아니라 사고 과정 등을 보다 심도 있게 알 수 있는 방법임은 분명하다(Osborn & Gilbert, 1980 ; Freyberg & Osborn, 1985). 따라서 이러한 문항은 과학적 탐구 사고력과 같은 고등 정신 능력을 측정하기 위한 효과적인 방법으로 생각된다. 이에 비해 객관식 문항 형태는 단순한 상지력만을 측정할 위험성이 있고, 학생이 응답할 때 무모한 정답 선택 요인, 즉 추측 요인이 크게 작용할 가능성이 있다(황정규, 1989). 그러나 현재의 평가 여건이 많은 수의 학생을 대상으로 평가해야 하므로 채점의 편의성과 객관성을 고려하여 경우에 따라서는 객관식 평가가 이루어질 수 밖에 없는 실정이다. 그러므로 객관식 문항의 단점인 추측 요인을 최소화하면서 효과적으로 탐구 사고력을 평가할 수 있는 방법의 고안이 필요하다. 추측의 요인을 줄이는 한 방안으로 과거 대학 입학 학력 고사의 4지 선다형과는 달리 대학수학능력시험에서는 답지 수를 하나 늘려 5지 선다형으로 하고 있다(교육부, 1991). 답지 수를 하나 더 늘림으로써 추측의 요인을 줄일 수 있다는 것은 실험적으로 밝혀졌다(이무와 박승재, 1995).

과학적 탐구 사고력을 평가하기 위해서는 단순한 기능적 측면만을 다루어서는 안된다. 또한 객관식 평가 문항으로 과학적 결과만을 물었을 때, 그 결과에 이르기까지의 탐구 과정을 평가할 수도 없다. 따라서, 과학적 탐구 사고력을 효과적으로 평가하기 위해서는 동일한 문제 상황에 대하여 과학적 결과뿐 아니라, 결과에 이르기까지의 탐구 과정을 평가할 수 있어야 하는데, 이것은 2

단계 선다형 평가로 연구된 바 있다(이무와 박승재, 1996). 그러나 일반적으로 탐구 사고력 문제는 그 어떤 과학적 개념과 밀접하게 관계된 경우가 많으므로 한 문제 상태에서 서로 관련있는 과정, 결과, 개념을 비록 선다형이지만 3번을 물어 모두 정답을 하는 경우에 점수를 부여하는 방안을 생각할 수 있다.

이를 위해서 본 연구에서는 과학적 탐구 사고력의 3요소를 단계적으로 물어 보아 과학적 탐구 사고력을 종합적으로 평가할 수 있는 3단계 선다형 문항 형태를 제안하고, 이에 대한 문항을 개발하고 투입하여 그 도입 가능성과 효과성을 추구하였다.

II. 연구 방법

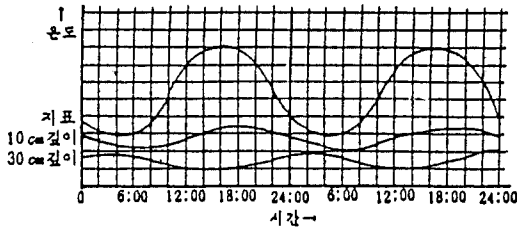
1. 도구 개발

3단계 선다형 문항이란 한 문항당 세 개의 소문항을 연결하여 한 개의 소문항에서는 과학적 결과를 묻고, 또한 개의 소문항에서는 결과에 이르기까지의 탐구 과정을 물으며, 또 하나의 소문항에서는 문제 상태와 관계되는 과학 개념의 이해를 묻는 것이다. 연구 문항은 과학 교육 교수 3명, 교육개발원 연구원 1명, 과학교육 박사 과정 중인 과학교사 6명에 의한 3차에 걸친 검토를 통하여 개발하였다. 개발된 문항 형태는 다음과 같다.

1) 3단계 선다형 문항(Three-tier multiple choice item of 5 options asking a result, process and concept, MrMpMc)

첫번째 소문항에서는 결론, 계산치, 실험 기구 선택 등의 결과를 묻고, 두번째 소문항에서는 어떤 계산치나 결론이 나온 과정을 묻은 다음, 세번째 소문항에서는 관계되는 과학적 개념을 묻은 것이다. 그러나 순서는 임의적이다. 또한, 추측만으로 답을 택한 확률을 구하기 위하여 첫번째, 두번째, 세번째 소문항 모두의 답안지에 ① 아주 확실, ② 확실, ③ 보통, ④ 확신없음, ⑤ 확신 전혀 없음의 5단계 확신도 표시를 위한 리커트 척도(Kerlinger, 1986)를 답안지에 포함시켰다. 그 한 예를 제시하면 다음과 같다.

◎ 다음은 한 학생이 토양의 깊이에 따른 온도의 변화를 알아보기 위하여 지표, 지표에서 10 cm 및 30 cm 깊이 온도계를 설치한 후 48시간 동안의 온도를 측정하여 작성한 그래프이다.



- (1) 이 그래프로부터 얻을 수 있는 결론 중 가장 합당한 것은?
- ① 온도 변화는 토양의 깊이가 깊을수록 커진다.
 - ② 온도 변화는 토양의 깊이가 깊을수록 작아진다.
 - ③ 온도 변화의 주기는 시간이 지남에 따라 커진다.
 - ④ 온도 변화는 토양의 깊이와는 무관하다.
 - ⑤ 온도는 시간에 관계없이 일정하다.

- (2) 위와 같이 응답한 근거는 무엇인가?
- ① 지표에서의 온도 변화가 가장 작기 때문.
 - ② 지표에서의 온도 변화의 주기가 제일 크기 때문이다.
 - ③ 30 cm 깊이에서 온도 변화가 제일 작기 때문.
 - ④ 30 cm 깊이에서 온도 변화의 주기가 제일 크기 때문이다.
 - ⑤ 온도 변화의 주기가 깊이에 반비례하기 때문.

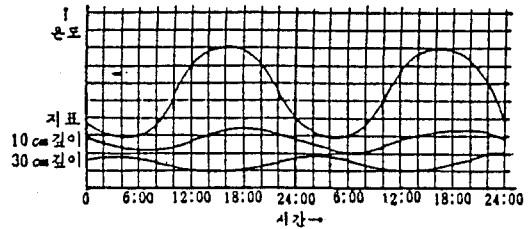
- (3) (1)의 응답과 같은 현상이 생기는 이유는?
- ① 시간이 지날수록 태양복사에너지의 영향을 적게 받기 때문.
 - ② 토양이 깊을수록 태양복사에너지의 영향을 적게 받기 때문.
 - ③ 시간이 지날수록 기후의 영향을 많이 받기 때문이다.
 - ④ 토양이 깊을수록 기후의 영향을 많이 받기 때문이다.
 - ⑤ 토양이 깊을수록 지열의 영향을 많이 받기 때문이다.

2) 3단계 주관식 문항(Three-tier subjective item asking a result, process and concept, SrSpSc)
 첫번째 소문항에서는 결론, 계산치, 실험 기구 선택

등의 결과를, 두번째 소문항에서는 어떤 계산치나 결론이 나온 과정을, 세번째 소문항에서는 관계되는 과학적 개념을 주관식으로 물은 것이다.

3단계 선다형 문항과 이에 대응하는 3단계 주관식 문항은 각각 8문항을 개발하였다. 다음은 이 문항의 한 예이다.

◎ 다음은 한 학생이 토양의 깊이에 따른 온도의 변화를 알아보기 위하여 지표, 지표에서 10 cm 및 30 cm 깊이에 온도계를 설치한 후 48시간 동안의 온도를 측정하여 작성한 그래프이다.



- (1) 토양의 깊이에 따른 온도 변화에 대하여 이 그래프로부터 얻을 수 있는 결론은?
- (2) 위와 같이 응답한 근거는 무엇인가?
- (3) (1)의 응답과 같은 현상이 생기는 이유는?

2. 연구 대상

문항 형태에 따른 정답률, 확산도에 따른 정답률 등을 비교 분석하기 위하여는 동일 집단내 다른 형태의 문항을 반복 투입하여 결과를 분석하는 것이 이상적인 것으로 생각될 수도 있지만 반복 검사에 대한 학생들의 부정적 태도에 의한 영향을 배제해야 한다.

일반적으로 일반계 고등학교 2학년은 학생들의 선택에 따라 인문 사회 계열과 자연 계열로 나누고, 각 계열의 학생들은 1학년 성적의 전교 석차순으로 학급별 성적이 고르게 분포되도록 분반된다. 이 점을 고려하면 2학년 동일 계열의 각 학급은 통계적으로 대등한 집단이라고 볼 수 있다.

본 연구에서는 본 연구의 문항이 과학에 관한 것이므로 과학의 이수 단위가 많고 우열 학급으로 나누지 않아서 대등하다고 판단되는 일반계 고등학교 2학년 자연계 3개 학급을 선정하여 3단계 선다형 문항을 2개 학급에, 3단계 주관식 문항을 1개 학급에 동일 여건 하에서 투입하였다.

3. 분석 방법 및 내용

문항 분석을 위하여 IBM 호환 PC에 선다형 문항은 정답과 학생 답안지의 응답을 입력하고, 주관식 문항은 정답을 1, 오답 또는 무응답은 0으로 채점하여 입력한 다음, Turbo C 2.0으로 프로그램을 작성하여 전산 처리하였다.

분석한 내용은 다음과 같다.

- (1) 3단계 선다형 문항과 3단계 주관식 문항의 각 채점 방법에 따른 각 문항 형태에 대한 평균과 표준편차, 100점 환산 평균과 100점 환산 표준편차, 변별도 및 신뢰도를 구하여 비교하였다.
- (2) 3단계 선다형 문항에서 채점 방법에 따른 추측 점수를 구하였다.
- (3) 3단계 선다형 문항에서 상위 50%, 하위 50%로 양분하여 상위 집단과 하위 집단의 추측 점수를 비교하였다.

III. 결과 분석

연구 문항 형태인 3단계 선다형 문항과 3단계 주관식 문항은 각각 총 8문항으로 다음과 같은 방법으로 채점하여 분석하였다.

한 문항당 결론, 계산치, 실험기구의 선택 등의 결과를 묻는 (1)번 소문항, 어떤 계산치나 결론이 나온 과정을 묻는 (2)번 소문항, 관계되는 과학적 개념을 묻는 (3)번 소문항을 각각 분리 채점하여 24문항에 해당하는 처리 결과를 “분리”로 나타내고, 각 문항에서 (1)번 소문항만 모은 8문항에 대한 처리 결과는 “(1)결과”,

(2)번 소문항만 모은 8문항의 처리 결과는 “(2)과정”, (3)번 소문항만 모은 8문항의 처리 결과는 “(3)개념”으로 나타내었다. 또한 결과와 과정을 묻는 (1)번 소문항과 (2)번 소문항을 연결 채점한, 즉 (1)번 소문항과 (2)번 소문항 모두 맞을 경우에만 정답으로 처리한 결과는 “(1)(2)”, 결과와 개념을 묻는 (1)번 소문항과 (3)번 소문항을 연결 채점한 처리 결과는 “(1)(3)”, 과정과 개념을 묻는 (2)번 소문항과 (3)번 소문항을 연결 채점한 처리 결과는 “(2)(3)”으로 나타내었고, (1)번 소문항과 (2)번 소문항 및 (3)번 소문항을 연결 채점하여 모두 맞아야 정답으로 처리한 결과는 “(1)(2)(3)”으로 나타내었다.

1. 3단계 선다형 문항과 3단계 주관식 문항의 비교

3단계 선다형 문항과 3단계 주관식 문항의 각 채점 방법에 따른 각 문항 형태에 대한 평균과 표준편차, 그리고 100점 환산 평균과 100점 환산 표준편차, 변별도 및 신뢰도는 Table 1, Table 2와 같다. 평균과 표준편차는 분리 채점의 경우는 24점 만점에 대한 것이고 나머지 채점 방법은 8점 만점에 대한 것이다. 3단계 선다형 문항은 2개 학급에 투입하여 사례수가 99명이고 3단계 주관식 문항은 49명이었으며, 투입 직전의 학교 기말 고사 학급 평균은 3단계 선다형 문항의 2개 학급이 각각 66.8점과 65.8점이고 3단계 주관식 문항의 학급은 67.2점으로 통계적으로는 대등한 집단이라고 할 수 있다.

Table 3은 채점 방법에 따른 3단계 선다형 문항과 3단계 주관식 문항의 100점 환산 평균점을 비교한 것이다.

3단계 선다형과 주관식에서 결론, 계산치, 실험 기구

Table 1 Scores according to the scoring methods in MrMpMc

Scoring methods	Separate scoring	(1)	(2)	(3)	(1)(2)	(1)(3)	(2)(3)	(1)(2)(3)
Number of items	24	8	8	8	8	8	8	8
Mean	10.40	3.58	3.95	2.88	2.78	1.73	1.82	1.43
S.D.	3.64	1.62	1.59	1.31	1.64	1.16	1.32	1.18
Percentile score	43.35	44.70	49.37	35.99	34.72	21.59	22.73	17.93
S.D. of percentile score	15.16	20.26	19.89	16.41	20.53	14.52	16.51	14.78
D.I.	0.38	0.51	0.50	0.40	0.51	0.35	0.40	0.37
KR20	0.66	0.45	0.37	0.14	0.51	0.34	0.37	0.38

(1) = results, (2) = process, (3) = concepts

Table 2 Scores according to the scoring methods in SrSpSc

Scoring methods	Separate scoring	(1)	(2)	(3)	(1)(2)	(1)(3)	(2)(3)	(1)(2)(3)
Number of items	24	8	8	8	8	8	8	8
Mean	7.41	2.65	2.82	1.94	2.00	1.41	1.49	1.16
S.D.	4.20	1.45	1.83	1.43	1.44	1.31	1.44	1.27
Percentile score	30.87	33.16	35.20	24.24	25.00	17.60	18.62	14.54
S.D. of percentile score	17.49	18.13	22.82	17.93	18.03	16.35	18.06	15.84
D.I.	0.44	0.46	0.58	0.45	0.45	0.41	0.45	0.35
KR20	0.80	0.39	0.61	0.48	0.47	0.49	0.56	0.53

(1) = results, (2) = process, (3) = concepts

Table 3 Comparison between MrMpMc and SrSpSc

Scoring method	Separate scoring	(1)	(2)	(3)	(1)(2)	(1)(3)	(2)(3)	(1)(2)(3)
MrMpMc	43.35	44.70	49.37	35.99	34.72	21.59	22.73	17.93
SrSpSc	30.87	33.16	35.20	24.24	25.00	17.600	18.62	14.54
Difference	12.48	11.54	14.17	11.75	9.72	3.99	4.11	3.39
t-test	4.37	3.35	3.85	3.95	2.80	1.50	1.37	1.27

($t=1.65$, $\alpha=0.05$, $t=1.29$, $\alpha=0.10$)

(1) = results, (2) = process, (3) = concepts

의 선택 등의 결과를 묻는 (1)결과 평균점과 어떤 계산이나 결론이 나온 과정을 묻는 (2)과정 평균점보다 관계되는 과학적 개념을 묻는 (3)개념 평균점이 많이 떨어진다는 것, 또한 (1)(2)평균점보다 과학적 개념을 묻는 (3)번 소문항이 포함된 (1)(3)평균점과 (2)(3)평균점이 떨어진다는 것은 표나 그래프 등을 보고 결과를 구하고 과정을 찾는 것보다는 관계되는 과학적 개념을 설명하고 생각해 내는 것이 더 어렵기 때문이라고 판단된다.

한편 각 채점 방법에 대한 3단계 선다형과 주관식의 100점 환산 평균점 차를 비교하면 각 소문항별로 분리 채점하여 한 번 선택 문항과 같은 (1)결과 평균점 차는 11.5점, (2)과정 평균점 차는 14.2점, (3)개념 평균점 차는 11.8점으로 차이가 많으나, 두 개의 소문항을 연결 채점하여 두 번 선택 문항과 같은 (1)(2)평균점 차는 9.7점, (1)(3) 평균점 차는 4.0점, (2)(3)평균점 차는 4.1점으로 과학적 개념을 묻는 (3)번 소문항이 포함된 (1)(3)평균점 차와 (2)(3)평균점 차가 현저히 감소했으며, 특히 세 개의 소문항을 연결 채점하여 3단계 선다형 문항과 같은 (1)(2)(3) 평균점 차는 3.4점으로 객관식

평균점이 주관식 평균점에 근접해 감을 볼 수 있다 (Fig. 1).

2. 3단계 선다형 문항에서 추측에 의해 얻은 점수 비교

3단계 선다형 문항에서 추측에 의해 맞을 확률을 구하기 위하여 확신도의 자신 없음과 전혀 없음에 표했으며 맞은 사례수를 각 채점 방법에 따라 조사하여 1명당 맞은 개수와 그 값을 100점으로 환산한 점수를 계산하였다. 그 결과는 Table 4와 같다.

1명당 맞은 점수가 각각을 분리 채점한 (1)결과는 4.2점, (2)과정은 4.9점, (3)개념은 7.2점으로 큰 값을 보이나, 두 개의 소문항을 연결 채점한 (1)(2)는 1.4점, (1)(3)은 0.9점, (2)(3)은 0.4점으로 그 값이 현저히 줄었으며, 특히 세 개의 소문항을 연결 채점한 (1)(2)(3)은 0.1점으로 추측으로 맞을 확률이 거의 없다는 것을 보여준다. 또한 Fig. 2는 분리 채점했을 때 평균점이 낮은 (3)개념에서 추측으로 맞을 확률이 높다는 것을 보여준다.

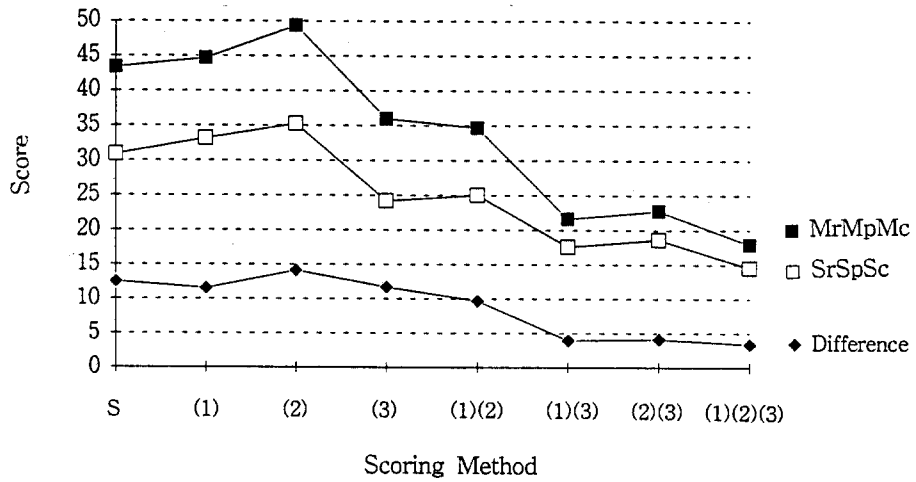


Fig. 1 Comparison between MrMpMc and SrSpSc

Table 4 Scores by Guessing in MrMpMc

Scoring methods	Separate scoring	(1)	(2)	(3)	(1)(2)	(1)(3)	(2)(3)	(1)(2)(3)
Correct answers without certainty for the class	129	33	39	57	11	7	3	1
Correct answers without certainty per student	1.30	0.33	0.39	0.58	0.11	0.07	0.03	0.01
Percentile score of correct answers without certainty	5.43	4.17	4.92	7.20	1.39	0.88	0.38	0.13

(1) = results, (2) = process, (3) = concepts

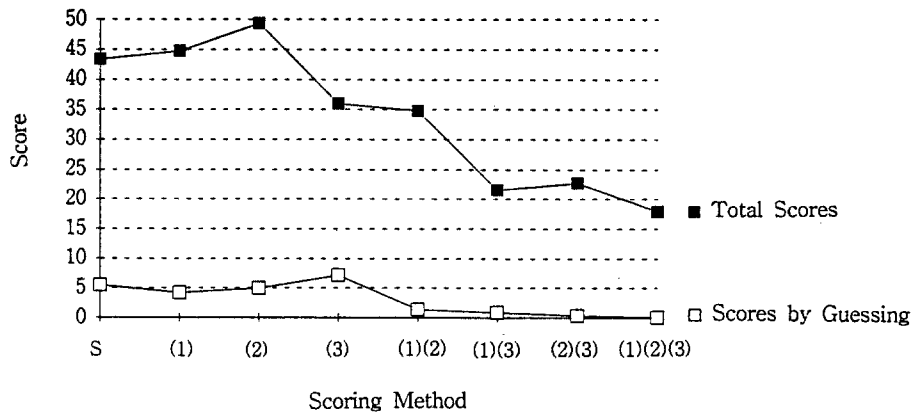


Fig. 2 cores by Guessing in MrMpMc

Table 5 Comparison between students of upper 50% and students of lower 50% in MrMpMc

	Scoring method	Separate scoring	(1)	(2)	(3)	(1)(2)	(1)(3)	(2)(3)	(1)(2)(3)
Mean	Upper 50%	55.08	57.50	63.25	44.50	50.00	29.50	32.75	26.75
	Lower 50%	31.38	31.63	35.20	27.30	19.13	13.52	12.50	8.93
Score of Guessing	Upper 50%	6.17	5.00	4.00	9.50	2.00	1.00	0.78	0.25
	Lower 50%	4.68	3.32	5.87	4.85	0.77	0.77	0.00	0.00
Share of Guessing(%)	Upper 50%	11.20	8.70	6.32	21.35	4.00	3.39	2.29	0.93
	Lower 50%	14.91	10.50	16.68	17.77	4.03	5.70	0.00	0.00

(1) = results, (2) = process, (3) = concepts

3. 상하위 집단의 비교

3단계 선다형 문항에서 상위 집단과 하위 집단의 추측으로 얻은 점수를 비교하기 위하여 3단계 선다형 문항과 3단계 주관식 문항을 각각 상위 50%와 하위 50%로 양분하여 각 집단의 평균점, 추측으로 얻은 점수, 추측 점유율(이무와 박승재, 1995)을 구한 결과는 Table 5와 같다.

Table 5에서 각각을 분리 채점한 (1)결과, (2)과정, (3)개념의 경우에 상하위 집단 모두 개념의 추측 점유율이 결과나 과정의 추측 점유율보다 높은 것은 관련된 과학적 개념을 생각하는 것이 어렵다는 것을 나타내며 평균 점수가 낮을수록 추측으로 맞을 확률이 크다는 것을 보여준다.

또한 두 개의 소문항을 연결 채점한 (1)(2), (1)(3), (2)(3)의 경우에 상하위 집단 모두 추측 점유율이 크게 감소하였고 상하위 집단의 추측 점유율 차가 적음을 알 수 있다. 특히 세 개의 소문항을 연결 채점한 경우의 추측 점유율은 거의 없음을 알 수 있다.

IV. 결 론

본 연구에서 결과를 묻는 소문항만의 평균점과 과정을 묻는 소문항만의 평균점보다 과학적 개념을 묻는 소문항만의 평균점이 많이 떨어졌다. 또한 결과와 과정을 묻는 소문항을 연결 채점한 평균점보다, 과학적 개념을 묻는 소문항이 포함된 경우, 즉 결과와 개념을 묻는 소문항을 연결 채점한 평균점과 과정과 개념을 묻는 소문항을 연결 채점한 평균점이 떨어졌다. 3단계 선다형 문항과 3단계 주관식 문항의 비교에서 각 소문항별로 분

리 채점한 경우보다 두 개의 소문항을 연결 채점한 경우의 평균 점수 차가 많이 감소했으며, 특히 세 개의 소문항을 연결 채점한 경우와의 평균 점수 차는 100점 만점 당 3.39점으로 3단계 선다형 문항의 평균 점수와 3단계 주관식 문항의 평균 점수가 거의 같았다. 또한 확신도를 포함한 3단계 선다형 문항에서 추측에 의해 얻은 점수는 100점 만점 당 0.13점으로 추측으로 맞을 확률이 극히 낮음을 볼 수 있었다.

본 연구의 3단계 선다형 문항은 과학적 결과와 그 결과가 나온 과정 또 관계된 과학적 개념을 함께 물어 봄으로써 탐구 사고력을 의미있게 평가할 수 있는 문항 형태로 객관식 문항 형태이지만 주관식 응답에 준하여 객관식 평가의 단점인 추측의 요인을 거의 제거할 수 있고 전산 처리가 가능하여 다인수의 학생을 대상으로 평가할 수 있는 문항 형태이다.

적 요

본 연구의 3단계 선다형 평가 문항은 과학적 결과와 그 결과가 나온 과정 또 그러한 과정에 관계된 과학적 개념을 함께 물어 봄으로써 과학적 탐구 사고력을 의미있게 평가할 수 있고 객관식 문항형태이지만 객관식 평가의 단점인 추측의 요인을 주관식 응답에 준하여 거의 제거할 수 있다고 판단된다. 오랜 기간동안 1단계 4지 선다형 또는 5지 선다형 시험에 익숙해 있던 학생들은 3단계 선다형 문항에 매우 적응하기 어렵겠고 교사들 또한 문항을 제작하는데 많은 시간과 노력이 필요하겠지만 평가가 학생들의 학습에 미치는 영향을 고려한다면 3단계 선다형 문항의 현장 도입은 질실히 필요하다. 이의 도입은 학생들의 탐구 사고력의 신장에 큰 기여를 할 것으로 판단된다.

참고 문헌

- 교육부(1991). 대학 입학 시험 개선안. 교육부.
- 이무, 박승재(1995). 과학적 탐구 사고력 선다형 평가 방법에서 4지와 5지 문항의 비교, 과학교육연구논총 제20권 제1호, 서울대학교 과학교육연구소.
- 이무, 박승재(1996). 과학적 탐구 사고력 2단계 선다형 평가 연구, 과학교육연구논총 제 21 권 제1호, 서울대학교 과학교육연구소.
- 황정규 (1989). 학교 학습과 교육 평가. 교육 과학사, 서울.
- Freyberg, P., & Osborne, R. (1985). *Learning in Science*. Auckland Heinemann, London.
- Kerlinger, F. F. (1986). *Foundations of Behavioral Research*, pp.455-456, Holt, Rinehart and Winston.
- Osborne, R. J., & Gilbert, J. K. (1980). A Method for Investigating Concept Understanding in Science. *European Journal of Science Education*, Vol. 2., pp. 311-321.