

현행 중등학교 과학 교과 지필평가 문항분석

오 현 석(남대문중학교 교사)

이 기 영(한성과학고등학교 교사)

《 요 약 》

본 연구에서는 8개 중·고등학교의 과학 교과 지필평가 문항을 수정 구안된 문항분류틀을 이용하여 현행 학교 지필평가 문항의 출제 경향을 파악하고, 문항분석을 통하여 난이도, 변별도, 신뢰도를 통계적으로 분석하였으며, 학생 점수 분포 유형을 분류하였다.

출제 경향은 전체적으로 지식 영역의 문항 비율이 높으며 탐구과정 영역의 문항 비율이 상대적으로 매우 낮은 것으로 분석되었으며, 대부분의 문항이 교과서적 맥락으로 구성되어 있었다. 문항 표현은 주로 그림이나 문장을 많이 사용하는 것으로 나타났으며, 그래프나 표의 사용 비율은 상대적으로 낮은 것으로 나타났다. 문항 유형과 형식은 학교마다 매우 다르게 나타났다. 난이도는 대체로 낮아 쉬운 문항이 많은 것으로 나타났으며 변별도는 높은 것으로 분석되었다. 단답형 및 서술형 문항의 양호도가 선다형보다 높은 것으로 나타났다. 교사들이 생각하는 난이도와 학생들이 느끼는 난이도 사이에는 상당한 차이가 있었다. 8개 학교의 신뢰도는 모두 양호한 것으로 나타났으며, 학생 점수 분포 유형 분석에서는 정상분포는 없고 부적편포와 양봉분포가 주를 이루었다. 분포 유형에는 난이도, 지역적 특성 등이 영향을 준 것으로 판단된다. 학교 과학 교과 지필평가 문항의 질을 높이기 위해서는 다양한 문항 표현의 사용과 고른 탐구과정 영역이 반영되도록 하여 문항의 난이도를 현행보다 높여야 할 것이며, 선다형 문항의 제작과 배점에 좀 더 세심한 주의를 기울여야 할 것이다.

주제어 : 과학 지필평가, 문항분석, 문항분류틀, 난이도, 변별도, 신뢰도

I. 서론

일반 고등학교 입시 전형이 100% 중학교 과정의 내신 성적으로 결정되고, 대학 입학 전형에서도 고등학교 내신 반영 비율이 높아지고 있는 현실에서 학교 내신 평가의 중요도는 증가되고 있는 추세이다. 일반적으로 학교 현장에서 사용되고 있는 과학 교과의 내신 평가는 중간고사나 기말고사와 같은 지필평가(paper and pencil test)와 실험보고서나 조사 및 발표 등을 포함하는 수행평가(performance assessment)로 나눌 수 있다. 학교 지필평가는 형성평가의

의미로 학생 개인의 성취도를 파악하고 학년말에 사정자료로 사용되는 측면에서는 준거참조 검사(criterion-referenced test)의 성격을 가지지만, 현행 고등학교나 대학교 입시에서 내신 지필평가 점수를 바탕으로 상대적 서열 점수로 변화하여 정보를 제공하는 측면에서는 규준참조 검사(norm-referenced test)라고도 볼 수 있다. 성태제(2002)는 학교에서 한 번 시행으로 끝나는 내신 지필평가는 규준참조검사라는 측면에서 정확한 상대적 서열 정보를 알려주어야 지필평가에 대한 신뢰 수준이 높아진다고 보았다. 그러므로 검사 시행 후 얻은 문항 응답 자료를 분석하여 문항과 검사의 질을 평가하여 타당도와 신뢰도를 추정하는 과정이 필요하다고 하였다. 내신 지필평가가 높은 타당도와 신뢰도를 요구되는데 반하여 실제 학교 현장에서는 출제되는 지필평가 문항에 대한 분석이 거의 이루어지지 못하고 있는 것이 현실이다.

문항분석(item analysis)이라 함은 문항을 작성할 때 개개 문항이 제대로 그 기능을 수행하고 있는지를 확인하는 작업으로, 문항 양호도 분석이라고도 한다(변창진 외, 2001). 기존의 문항분석 연구를 살펴보면 대학수학능력시험 실험평가 문항 중 지구과학 분야의 문항을 분석한 연구(김찬중, 1992), 초등학교 과학 우수 아동 선발을 위한 과학탐구능력 경시대회 문항을 분석한 연구(배영부, 1994), 과학영재교육센터 학생선발문항을 분석한 연구 결과(이상범 외, 1999), 그리고 한국지구과학올림피아드 문항의 질 향상 방안을 위한 문항분석 연구(이기영, 2005)가 있다. 이들 선행 연구들은 거의 대부분이 학생 선발이나 경시대회, 올림피아드와 같은 대규모(large scale)의 표준화된(standardized) 성향을 띄는 검사에서 사용된 문항을 분석하는데 치중되어 있는 반면, 학교 교육과정 중에서 학생들이 가장 많이 접하는 평가인 내신 지필평가 문항분석에 대한 연구는 아직 미흡하다고 할 수 있다. 학교 내신 지필평가에 대한 중요도가 높아지고 있는 현 시점에서 이들 문항의 분석을 통하여 현행 중등학교에서 사용하는 지필평가 문항의 경향과 양호도를 파악함으로써 이 결과를 토대로 문항 제작에 지침이 될 수 있는 기초 자료를 제공하는 연구가 필요하다 하겠다.

따라서 본 연구에서는 8개 중·고등학교의 과학 교과 지필평가 문항을 수정 구안된 문항 분류틀을 이용하여 현행 학교 지필평가 문항의 출제 경향을 파악하고, 문항분석을 통하여 난이도, 변별도, 신뢰도를 통계적으로 분석하여 어느 정도 수준의 문항으로 구성되어 있는가를 알아보고자 한다. 또한 학생 점수 분포 유형을 분석하여 어떤 유형으로 나누어지며, 이 유형은 학교에 따라 어떤 차이를 보이는지 알아보고자 한다. 본 연구는 현 내신 지필평가의 현황을 파악하는 데 도움을 주고, 현장 교사들에게는 양호한 문항을 제작하는데 필요한 지침을 제공할 수 있을 것으로 기대한다.

본 연구에서 알아보고자 하는 문제는 다음과 같다.

첫째, 현행 중등학교 지필평가의 출제 경향은 어떻게 되는가?

둘째, 현행 중등학교 지필평가 검사 및 문항의 양호도는 어느 정도인가?

셋째, 현행 중등학교 지필평가 문항의 신뢰도는 어느 정도인가?

넷째, 현행 중등학교 지필평가에서 학생 점수 분포는 어떤 유형을 나타내는가?

Ⅱ. 연구 방법

1. 연구 대상

본 연구는 <표 1>에서 보는 바와 같이 서울특별시와 인천광역시 그리고 광명시에 소재한 중학교 5개교와 고등학교 3개교의 지필평가 224개 문항으로 총 2,723명의 문항별 점수 자료를 대상으로 하였다. 중학교는 모두 일반 중학교이고, 고등학교는 인문계 고등학교이며, 같은 내신 지필평가를 실시한 동일 학년 내의 2005학년도 1학기 중간고사 자료를 수집하였다. 수집된 지필평가 문항은 선다형(multiple-choice)과 단답 및 서술형(short answer/essay)으로 구분하였다.

〈표 1〉 연구 대상

	중학교					고등학교		
	S	N	G	M	H	J	Y	I
학년	7	8	8	9	7	11	11	11
과목	과학	과학	과학	과학	과학	지구과학	지구과학	지구과학
학생 수	356	319	448	379	453	54	289	425
선다형 문항	20	18	23	25	30	25	24	30
단답 및 서술형 문항	6	7	3	8	-	-	5	-
소재지	서울	서울	서울	서울	광명	인천	서울	서울

2. 출제 경향 분석을 위한 문항분류틀 구안

본 연구에서는 과학 교과 내신 지필평가 문항의 출제 경향을 분석하기 위해 2개의 문항분류틀을 구안하였다. 이 문항분류틀은 우종옥 외(1996)와 홍미영 외(2002)의 연구와 이기영(2005)의 지구과학 올림피아드 문항분석 연구에서 사용한 문항분류틀을 일부 수정한 것으로 내적 문항분류틀과 외적 문항분류틀로 구분된다.

가. 내적 문항분류틀(internal item classification framework)

<표 2>의 내적 문항분류틀은 문항이 포함하는 내적인 요소들을 분석하기 위한 것으로 내용(content), 지식 및 탐구과정(knowledge & Inquiry process), 맥락(context)의 3차원으로 구안되었다. 내용 차원(dimension)은 중학교 문항에서는 물리, 화학, 지구과학, 생물의 내용 영역

(domain)에 따라 교과서 단위 구분 기준에 의해 분류하였다. 고등학교 문항에서는 내용 차원(dimension)은 지구과학의 내용 영역(domain)에 따라 지질(지구물리 포함), 대기, 해양, 천문, 지구환경으로 분류하였다. 각 영역들은 몇 개의 단위(unit)로 세분하였다. 지식 및 탐구과정 차원은 4개의 일반적 탐구과정 영역(문제인식 및 가설 설정, 탐구 설계 및 수행, 자료 분석 및 해석, 결론도출 및 일반화)에 지식 영역을 추가하여 하나의 차원으로 만들었으며, 각 영역들은 몇 개의 요소들로 나누었다. 맥락 차원은 문항의 맥락이 순수하게 교과서적 내용에 의존했는지의 여부와 기술, 사회, 생활환경 등 교과서 외적 내용까지 응용하여 반영하였느냐에 따라 크게 ‘교과서적(textual)’과 ‘탈교과서적(non-textual)’ 2개의 영역으로만 구분하였다.

〈표 2〉 내적 문항분류틀 예시(7차 교육과정 중학교 1학년 대상)

차원	영역	세부 영역
내용	A. 지구의 구조	A1. 대기권 / A2. 지구의 내부
	B. 빛	B1. 빛의 반사와 굴절 / B2. 빛의 분산과 합성
	C. 지각의 물질	C1. 광물 / C2. 암석 / C3. 지표의 변화
	D. 물질의 세 가지 상태	D1. 물질의 상태 변화 / D2. 물질의 상태와 분자 모형
	E. 분자의 운동	E1. 움직이는 분자 / E2. 기체의 압력과 부피 / E3. 기체의 온도와 부피
	F. 생물의 구성	F1. 세포 / F2. 생물의 구성 단계
	G. 상태 변화와 에너지	G1. 상태 변화와 에너지 / G2. 상태 변화와 분자 운동
	H. 소화와 순환	H1. 영양소 / H2. 소화 / H3. 순환
	I. 호흡과 배설	I1. 호흡 / I2. 배설
	J. 힘	J1. 여러 가지 힘 / J2. 힘의 측정과 표시 / J3. 두 힘의 합성
	K. 해수의 성분과 운동	K1. 바닷물의 성분 / K2. 바닷물의 운동
	L. 파동	L1. 파동의 발생 / L2. 소리의 성질 / L3. 파동의 반사와 굴절
지식 및 탐구과정	I. 지식	I 1. 기억 / I 2. 이해 / I 3. 적용
	II. 문제 인식 및 가설 설정	II1. 문제 인식 / II2. 가설 설정
	III. 탐구 설계 및 수행	III1. 변인 설정 / III2. 실험 장치 고안 및 배치 / III3. 관찰, 측정, 분류, 실험 / III4. 실험 절차
	IV. 자료 분석 및 해석	IV1. 정량적 분석 / IV2. 정성적 분석 / IV3. 자료의 변환
	V. 결론 도출 및 일반화	V1. 결론 도출 / V2. 일반화
맥락	a. 교과서적	교과서적인 맥락
	b. 탈교과서적	STS적인 맥락 등

나. 외적 문항분류틀(external item classification frame work)

〈표 3〉의 외적 문항분류틀은 문항을 구성하는 외적 요소들을 분석하기 위한 것으로 문항

표현(item representation), 문항 유형(item type), 문항 형식(item form)의 3개 범주(category)로 구분하였다. 문항 표현 범주는 무엇을 사용하여 발문을 구성하였느냐에 따라 그림, 그래프, 표, 보기로 구분한 것에 이 중 어느 것도 사용하지 않고 오로지 문장으로만 표현된 문장을 추가하여 구분하였으며, 문항 유형 범주는 피험자들이 답을 하는 방식에 따라 선다형, 단답형, 서술형 3개 유형으로 구분하였다. 문항 형식은 하부 문항의 존재 여부에 따라 단독형과 복합형으로 구분하였다.

〈표 3〉 외적 문항분류틀

대범주	소범주	세부 설명
문항 표현	그림	자료가 사진이나 삽화 등의 그림으로 제시된 경우
	그래프	자료가 x와 y축을 가진 그래프로 제시된 경우
	표	자료가 표로 제시된 경우
	보기	자료에 대한 설명이나 해석이 보기로 제시된 경우
	문장	자료가 그림, 그래프, 표나 보기가 없이 문장으로 제시된 경우
문항 유형	선다형	여러 개의 답지 중 1개 또는 그 이상을 고르게 하는 경우
	단답형	단답의 형태로 답만 쓰게 한 경우
	서술형	답에 대한 설명 또는 이유, 풀이과정을 서술하게 한 경우
문항 형식	단독형	하부 문항 없이 문항이 하나로만 구성된 경우
	복합형	문항이 2개 이상의 하부 문항으로 구성된 경우

3. 문항 양호도 분석

문항의 양호도를 분석하기 위해 난이도와 변별도를 산출하였다. 난이도는 문항의 난이도(item difficulty: P)를 사용하였으며, 문항변별도는 변별도 지수(discrimination index: DI)를 사용하였다. 난이도에 의하여 문항을 평가하는 절대적인 기준은 없으나, 본 연구에서는 Cangelosi(1990)의 기준을 적용하였다. 난이도 .25미만은 ‘어려운 문항’으로, .25~.75는 ‘적절한 문항’으로, .75 초과는 ‘쉬운 문항’으로 판단하였다. 문항변별도 또한 평가의 절대적인 기준은 없으나, 본 연구에서는 Ebel(1965)의 기준을 적용하였다. .40 이상은 변별도가 ‘높은 문항’으로, .30~.39는 ‘있는 문항’으로, .20~.29는 ‘낮은 문항’으로, .10~.19는 ‘매우 낮은 문항’으로, .10 미만은 ‘없는 문항’으로 판단하였다. 본 연구에서는 고전검사이론의 가장 기초적인 변별도 지수를 사용하였으나 보다 정확한 문항변별도를 산출하기 위해서는 최신 검사이론에서 사용하는 Point biserial correlation이라든지, 문항반응이론(item response theory)에서의 문항 변별도 모수치(parameter)를 사용할 수 있다(Baker, 1992).

4. 문항의 신뢰도와 점수 분포 유형 분석

신뢰도를 산출하기 위해 본 연구에서는 고전검사이론에서 사용하는 여러 가지 신뢰도 중 문항의 내적일관성신뢰도(internal consistency reliability)인 Cronbach α 를 사용하였다(Cronbach, 1951). 또한 문항 점수 분포 유형을 분석하기 위해서 모든 문항의 배점을 1점으로 환산하여 히스토그램을 그린 다음, 그 분포 형태를 비교하였다. 점수 분포 유형의 판단은 집중경향치를 이용하였다. 집중경향치란, 한 집단을 구성하고 있는 학생들의 특성을 측정하여 이를 점수화했을 때, 이 집단의 점수분포를 하나의 값으로 대표하여 나타내 주는 것이다. 본 연구에서 집중경향치 간의 관계는 평균(M)을 기준으로 하고 최빈치(Mo)가 평균보다 크나 작으나에 따라 분포곡선의 형태가 결정된다. $M(\text{평균}) - Mo(\text{최빈치})$ 에서 그 값이 ‘+’면 정적편포(positive distribution)이고, ‘-’이면 부적편포(negative distribution)가 되며, ‘0’이 나오면 정상분포(normal distribution)가 되며, 봉우리가 두 개인 M자 형태가 되면 양봉분포(bimodal distribution)가 된다.

Ⅲ. 연구 결과 및 분석

1. 문항 출제 경향 분석

가. 내적 문항분류틀 적용 결과

내용 차원에서의 분류는 자료 수집된 8개 학교의 학년과 출제 범위가 모두 달라 일괄적으로 비교하여 분석하기가 어려웠다.

<표 4>는 지식 및 탐구과정 차원에서 문항을 분류한 결과이다. 중·고등학교 모두 전체적으로 지식 영역이 차지하는 비율이 각각 76%(140문항 중 106문항), 74%(84문항 중 59문항)로 가장 높았다. 지식 영역에서 학교별로는 중학교의 경우는 58%~90%로 차이가 많았으나, 고등학교의 경우는 69%~72%로 비교적 차이가 적었다. 또한 지식 영역 중에서 중학교의 경우는 106개의 문항 중 ‘이해’가 47개로 가장 많았고 그 다음은 ‘기억’이 38개를 차지하였으며 ‘적용’은 21개로 가장 적었다. 고등학교의 경우도 중학교의 경우와 거의 비슷하였으나 ‘기억’이 59개 문항 중 29개로 가장 많았다. 탐구과정 중에서는 ‘자료 분석 및 해석’ 영역이 차지하는 비율이 중·고등학교 모두 각각 전체 문항의 14%(19문항)와 17%(20문항)로 가장 높았으며, 중학교의 경우는 학교별로 7%~28%로 상당한 차이를 나타내었다. ‘자료 분석 및 해석’ 영역 중에서는 중·고등학교 모두 ‘정성적 분석’이 가장 많이 출제된 것으로 나타났으며, ‘자료의 변환’에 관련된 문항은 거의 없는 것으로 나타났다. 이것은 대부분의 문항이 실제 수치화된

자료 제시를 통한 문제 해결을 요구하기보다는 그림이나 문장을 통해 문제 해결을 하도록 구성되었음을 의미한다고 볼 수 있다. 한편, ‘문제인식 및 가설설정’ 영역은 5개 중학교에서 단 한 문제도 출제되지 않았고, 고등학교에서도 4%에 불과하였다. 또한 ‘탐구설계 및 수행’이나 ‘결론도출 및 일반화’에 해당되는 문항의 출제 비율도 상대적으로 매우 낮았다. 이것으로 볼 때 출제된 탐구과정 영역의 문항들이 다양한 탐구 기능(inquiry skills)을 요구하도록 제작되지 못하였음을 알 수 있다. 또한 Kempa(1986)의 주장과 같이 탐구과정 요소들 중 탐구설계 및 수행과 같은 실제적 능력(practical ability)보다는 문제인식 및 가설설정, 자료해석 및 분석과 같은 이론적 능력(theoretical ability)을 측정하는 데 치우쳐 있음을 알 수 있다.

전체적으로 볼 때 지식 영역에 대한 문항의 비율이 높으며, 탐구과정 영역의 문항 비율이 상대적으로 매우 낮은 것으로 분석된다. 특히 중학교에서 상대적으로 지식 영역에 많이 치우쳐서 평가하고 있음을 알 수 있다.

〈표 4〉 지식 및 탐구과정 차원에서의 문항 분류 결과

			중학교							고등학교					전체					
			S	N	G	M	H	소개	%	J	Y	I	소개	%	합계	%				
지식 및 탐구 과정	I	I 1	4	8	3	8	15	38	76	9	14	6	29	70		74				
		I 2	9	7	10	12	9	47		7	4	9	20							
		I 3	9	4	2	3	3	21		2	2	6	10							
		소개	22	19	15	23	27	106			18	20	21				59		165	
		%	85	76	58	70	90				72	69	70							
	II	II 1							0		1	1	2	4		1				
		II 2									1	1	1							
		소개									1	2	3					3		
		%									3	7								
	III	III 1							7			1	1	1		5				
		III 2			1	2		3												
		III 3	1	2			1	4												
		III 4			1	2		3												
		소개	1	2	2	4	1	10					1				1		11	
		%	4	8	7	12	3						3							
	IV	IV 1	2	3		2		7	14	2	2		4	24		17				
		IV 2	1	1	7		2	11		4	6	6	16							
		IV 3				1		1												
		소개	3	4	7	3	2	19			6	8	6				20		39	
		%	11	16	28	9	7				24	28	20							
	V	V 1			2	3		5	3	1			1	1		3				
		V 2																		
		소개			2	3		5			1						1		6	
		%			7	9					4									
합계			26	25	26	33	30	140	100	25	29	30	84	100	224	100				

<표 5>는 맥락 차원에서 문항을 분류한 결과를 정리한 것이다. 교과서적 맥락의 문항이 중학교에서 출제 비율이 90%, 고등학교는 98%를 차지한 반면, STS적인 상황 등 탈교과서적 맥락에서 출제된 문항은 중학교에서는 10%, 고등학교에서는 2%에 불과하였다. 대부분의 문항이 교과서적 맥락으로 교과서적인 상황을 토대로 출제된 것을 알 수 있었으며, 중·고등학교 모두 학교별 차이는 크지 않은 것으로 나타났다. 이 결과로 볼 때, 발산적 사고(divergent thinking)나 창의적 사고(creative thinking)를 요구하는 문항의 비율이 현행 학교 지필평가에서 매우 낮은 것으로 판단된다.

〈표 5〉 맥락 차원에서의 문항 분류 결과

		중학교							고등학교					전체	
		S	N	G	M	H	소계	%	J	Y	I	소계	%	합계	%
맥락	교과서	21	25	26	26	28	126	90	25	27	30	82	98	208	93
	탈교과서적	5			7	2	14	10		2		2	2	16	7

나. 외적 문항분류틀 적용 결과

<표 6>은 외적 문항분류틀을 적용한 결과를 나타낸 것이다. 중학교의 경우, 문항 표현 범주에서 문장만을 사용한 비율이 38%로 가장 높았으며, 표를 사용한 비율이 불과 3%로 가장 낮았다. 하지만 학교별로는 큰 차이를 나타내었다. 고등학교의 경우는 문항표현 범주에서 그림을 사용한 비율이 35%로 가장 높았으며, 그래프가 24%, 보기가 21%의 비율을 차지하였다. 중학교와 마찬가지로 표가 3%에 불과할 정도로 가장 표현 비율이 낮았으며, 중학교에서 가장 많이 사용하는 문장은 17% 정도였다. 이것은 대부분의 탐구과정 영역 문항이 그래프나 표를 이용한 정량적 분석보다는 그림이나 문장을 이용한 정성적 분석에 해당된다는 <표 4>에서의 분석과 일치하는 결과이다. 한편, 중학교가 문항 표현이 문장(38%)과 그림(31%)에 치우쳐 있는 것에 비해, 고등학교는 그림, 그래프, 보기, 문장이 비교적 고르게 분포하는 것으로 보아 중학교보다는 고등학교가 좀 더 다양한 형태로 문항을 표현하는 경향이 있음을 알 수 있다.

문항 유형 범주에서는 중·고등학교 모두 학교에 따라 각 문항 유형의 비율이 다양하게 나타났지만, 선다형 문항의 비율이 70%~100%로 가장 높았다. 전부 선다형 문항으로 출제된 학교도 3개교이다. 이렇게 학교마다 문항 유형이 서로 다른 것은 내신평가가 지필평가로만 되어있는 것이 아니라, 실험평가를 포함한 수행평가가 추가로 포함되기 때문인 것으로 판단된다. 또한 학교에 따라 평가 지침이 달라 중간고사와 기말고사에서 문항 유형이 다른 비율이 될 수 있을 것이다.

문항 형식 범주에서는 고등학교 1개교를 제외하고는 모두 단독형이 복합형보다 더 많이 출제되는 것으로 분석되었다. 단독형만으로 모든 문항이 표현된 학교도 2개교이다. 문항 형

식에서는 학교나 교사의 성향에 따라 다양하게 나타나며, 중학교와 고등학교의 차이는 없는 것으로 보인다.

〈표 6〉 외적 문항분류틀 적용 결과

대범주	소범주	중학교(%)						고등학교(%)			
		S	N	G	M	H	계	J	Y	I	계
문항 표현	그림	46	8	12	50	33	31	32	34	37	35
	그래프	15	12	30	3	7	13	20	31	20	24
	표	0	0	8	0	7	3	8	0	3	3
	보기	0	40	4	12	13	15	16	28	20	21
	문장	38	40	46	30	40	38	24	7	20	17
문항 유형	선다형	77	72	88	70	100	81	100	83	100	94
	단답형	8	16	12	18	0	11	0	7	0	2
	서술형	15	12	0	12	0	8	0	10	0	4
문항 형식	단독형	58	84	100	76	87	81	92	48	100	80
	복합형	42	16	0	24	13	19	8	52	0	20

2. 문항 양호도 분석

가. 문항 난이도와 변별도

〈표 7〉은 선다형 문항의 난이도와 변별도를 각각 Cangelosi(1990)와 Ebel(1965)의 기준에 따라 분류하여 정리한 것이다. 전체적으로 난이도가 적절한 문항의 비율이 중·고등학교 모두 가장 높았으나, 학교에 따라서는 많은 차이를 나타내었다. 또한 쉬운 문항의 출제 비율이 상당히 큰 것으로 나타났으며, 어려운 문항이라고 판단되는 문항이 아예 출제되지 않은 곳도 5개 학교가 있었다. 고등학교보다는 중학교 문항이 비교적 더 쉬운 문항으로 구성되어 있는 것으로 나타났으며, 쉬운 문항의 비율이 전체 문항의 50% 이상인 학교가 2개 학교가 있었다. 전반적으로 8개 학교 모두 난이도의 적절한 배분을 고려하지 않고 평가지를 구성하였으며, 대체적으로 쉽게 출제하려는 의도가 있었던 것으로 판단된다.

변별도의 경우는 8개 학교 모두 대체적으로 변별력이 높은 문항으로 구성되어 있는 것으로 분석되었다. 중학교가 고등학교에 비하여 상위권 학생들과 하위권 학생들을 변별할 수 있는 변별도 지수가 다소 높은 것으로 나타났으며, 8개 학교 모두 변별도가 (-)인 문항은 발견되지 않았다. 변별도가 0.10 미만인 문항은 G중학교의 22번 문항과 Y고등학교의 20번 문항으로 각각 문항변별도가 0.00과 0.06으로 나타났다.

〈표 7〉 문항의 난이도와 변별도

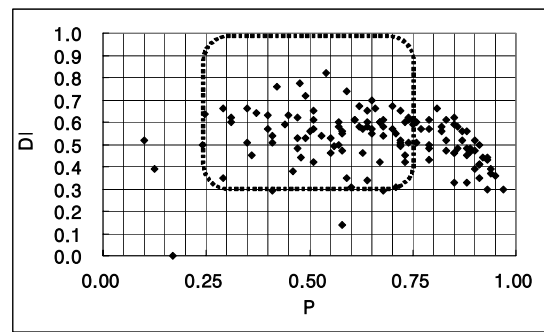
난이도와 변별도		중학교(%)						고등학교(%)			
		S	N	G	M	H	계	J	Y	I	계
난이도(P)	어려운 문항 (<0.25)	0	0	5	0	0	1	4	25	0	9
	적절한 문항 ($0.25 \sim 0.75$)	37	78	85	43	53	59	48	71	67	62
	쉬운 문항 (>0.75)	63	22	10	57	47	40	48	4	33	29
변별도(DI)	≥ 0.30	100	92	92	100	100	97	80	66	90	79
	<0.30	0	8	8	0	0	3	20	34	10	21

나. 문항 양호도

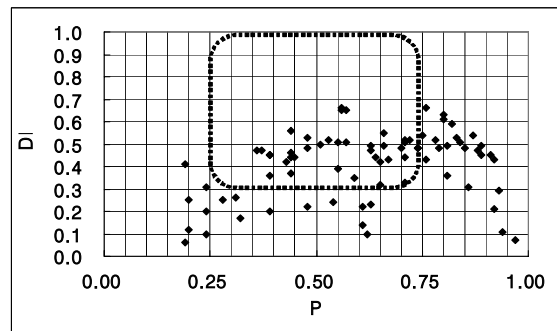
[그림 1]은 난이도와 변별도로 문항의 양호도를 판단한 결과를 나타낸 것이다. 양호도의 판단 기준은 난이도가 $0.25 \sim 0.75$ 이고 변별도가 0.30 이상인 문항을 양호한 것으로 판단하였다. [그림 1]의 (a)는 중학교 문항에 대하여 문항 양호도를 판단한 것으로, 점선의 사각형 안에 찍힌 문항이 양호한 문항이다. 중학교의 경우, 양호한 문항은 전체 140문항 중 82문항으로 59%였다. [그림 1]의 (b)는 고등학교 문항에 대하여 문항 양호도를 판단한 것으로, 점선의 사각형 안에 찍힌 문항이 전체 84문항 중 42문항으로 50%였다.

J고등학교를 제외한 모든 학교에서 양호한 문항이 50% 이상으로 대체적으로 지필평가가 양호한 문항으로 구성되어 있음을 알 수 있었다. 중학교 문항들이 좀 더 고등학교 문항보다 양호한 것으로 분석되며, 양호한 문항 영역으로부터 분산되어 있는 양상도 차이가 있었다. 중학교는 대부분이 변별도는 적절하나 문항이 너무 쉬운 문항들로 구성되어 있는 반면, 고등학교는 변별도가 적절하지 못한 문항이 상대적으로 많이 분포한다.

한편, 단답형 및 서술형과 선다형으로 나누어 살펴보면, 단답형 및 서술형은 전체 29문항 중 양호한 문항은 24문항으로 82%가 양호한 문항이었다. 선다형의 경우는 전체 195문항 중 양호한 문항이 100문항으로 51%가 양호한 문항이었다. 단답형 및 서술형이 선다형에 비하여 양호한 문항의 비율이 높다는 것을 알 수 있다. 이 결과로 판단해 볼 때 교사들이 양호한 문항을 제작하기에 단답형 및 서술형이 선다형보다 더 용이하다고 말할 수 있겠다.



(a) 중학교



(b) 고등학교

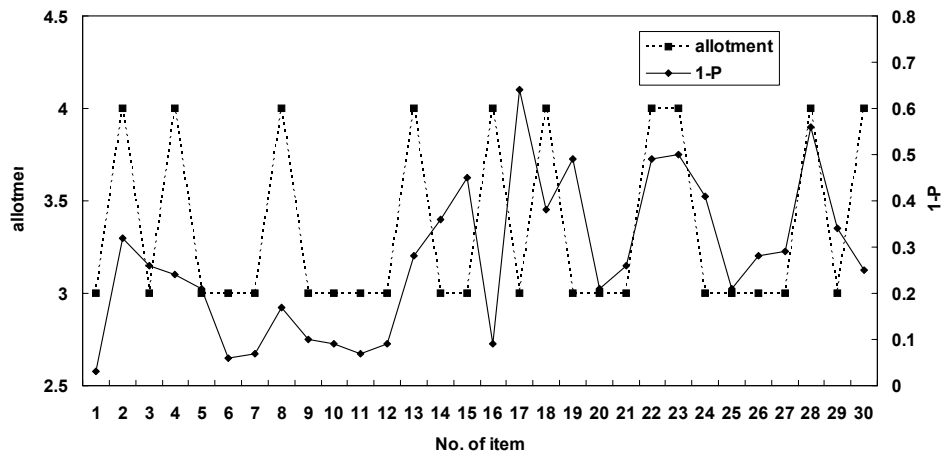
[그림 1] 문항 양호도 판단 결과

다. 문항 난이도와 배점과의 관계 분석

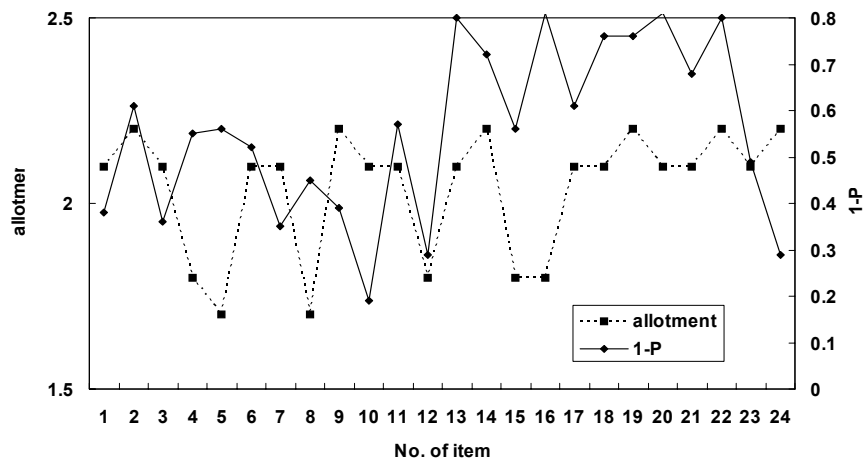
H중학교와 Y, I고등학교의 경우는 문항마다 배점이 달랐다. 문항별 실제 난이도와 교사가 사전에 생각한 난이도가 일치하는지 알아보았다. 배점과 비교하기 쉽게 하기 위하여 1에서 문항 난이도 값을 뺀 값(1-P)을 사용하여 분석하였다¹⁾. [그림 2]와 [그림 3]은 2개 학교의 선다형 문항의 난이도와 배점을 비교한 것이고, [그림 4]는 1개 학교의 단답형 및 서술형 문항의 난이도와 배점을 비교한 것이다. [그림 2]의 H중학교의 경우, 16번 문항은 문항난이도가 0.91로 가장 쉬운 문제인데 문항 배점은 4점으로 교사는 학생들에게 어려울 것이라 예상한 문항이다. 반면, 17번 문항은 문항난이도 값이 0.36으로 30문항 중 가장 어려운 문항인데, 문항 배점은 3점으로 교사는 쉬운 문항으로 예상한 문항이다. 교사가 생각한 것보다 문항이

1) 본 연구에서는 문항의 난이도에 따라 차등 배점을 하였다고 판단하였다. 그 근거로 이원목적분류표를 검토한 결과에서 대부분의 학교가 난이도가 높다고 생각하는 문항에 더 많은 배점을 하였음을 확인하였다. 하지만 언제나 난이도를 기준으로 차등 배점이 되는 것은 아닐 것이다. 경우에 따라서는 출제 단위 내용의 중요도에 따라 차등 배점이 이루어질 수도 있다.

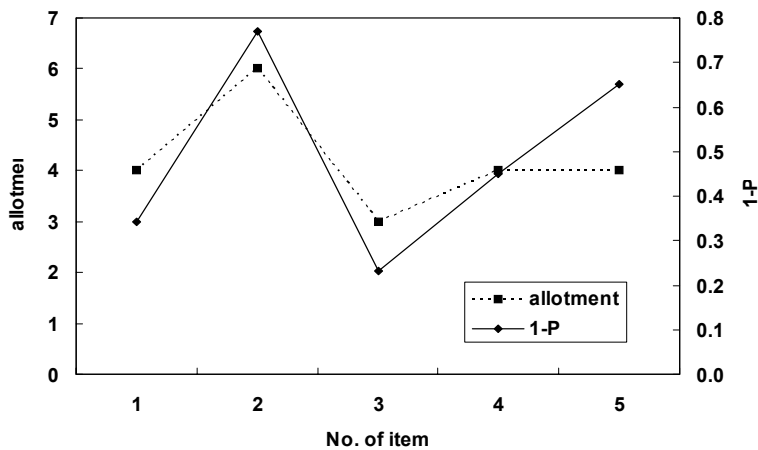
쉬웠던 것은 4문항(13%), 어려웠던 것은 6문항(20%)으로 문항 배점에 문제가 있는 것으로 판단되는 문항은 전체 30문항 중 10문항으로 33% 정도이다. [그림 3]의 Y고등학교의 경우는 대부분의 문항 배점이 분석된 난이도와 일치하지 않는 것으로 나타났다. 반면, [그림 4]의 결과를 보면 [그림 3]과 같은 학교인데 단답형 및 서술형의 경우는 선다형과는 달리 교사가 생각한 문항 난이도와 실제 학생들의 정답률이 대체로 일치하고 있음을 알 수 있다.



[그림 2] H중학교의 선다형 문항 난이도와 배점 비교



[그림 3] Y고등학교의 선다형 문항 난이도와 배점 비교



[그림 4] Y고등학교의 단답형 및 서술형 문항 난이도와 배점 비교

라. 주요 특징적인 문항의 문제점 분석

<표 8>과 <표 9>는 배점, 답지 구성, 그리고 문항 형식에 문제가 있는 문항의 예이다. <표 8>의 16번 문항은 교사가 생각한 것보다 상당히 쉬운 문항으로, 문제점은 답지 구성에 있는 것으로 판단된다. 3번 답지만 ‘보이지 않는다’라는 표현을 사용하고 나머지 답지는 ‘보인다’의 표현을 사용하고 있다. 빛의 굴절에 해당하는 문항이므로 그 현상은 당연히 ‘보인다’의 표현으로 종결되어야 한다. 하지만 3번 답지만 ‘보이지 않는다’는 표현을 사용하여 문제에 대하여 생각하지 않고도 정답을 선택할 수 있는 여지를 남기고 있다. 빛의 굴절이 포함되지 않은 빛의 반사나 합성의 예로 3번 답지를 변경하여 다른 답지와 유사하게 표현하는 것이 적절할 것이다. 17번 문항은 교사가 생각한 것보다 어려운 문항으로, 문제점은 답지 구성에 있는 것으로 판단된다. (ㄷ)항목의 반사 망원경은 대상이 중학교 1학년인 것에 반해, 망원경의 구조와 원리는 중학교 2학년에서 학습하게 되어 있으므로 학생들에게 다소 생소한 항목으로 생각된 것으로 보인다. 마찬가지로, (ㄱ)과 (ㄴ) 항목의 현미경도 아직 학습하지 않은 항목이다. 따라서 정답인 5번이 36%인 것에 반해 (ㄷ), (ㄱ), (ㄴ)으로 묶여 있는 2번 답지에 대한 응답률이 38%로 오히려 높게 나타났다. 특히, 하위권의 48%가 2번 답지에 응답한 것은 눈여겨 볼 사항이다. 또한 발문 형태가 ‘보기에서 모두 고른 것은?’인데, 이처럼 해당 사항을 모두 고르게 지시하는 합답형 문항은 학생들이 내용의 수준을 배제하고도 가장 어려워하는 문항 형태이다. 교사가 쉬운 문항으로 출제할 의도였다면 오히려 16번 문항과 같은 형태를 사용하여야 하고, 역으로 16번 문항은 17번 문항과 같이 보기 중에서 옳은 것을 모두 고르게 하는 합답형이 더 적절할 것이다.

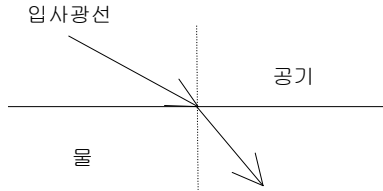
<표 9>는 문항 형식에 문제가 있는 문항의 예이다. 모두 변별력이 있는 문항으로, 3번 문항은 난이도가 0.63으로 양호한 문항이나, 2번과 4번 문항은 각각 문항난이도 0.83과 0.90으로 쉬운 문항이다. 특히 4번 문항은 정답률이 높고 상대적으로 변별력은 떨어지는 문항이다. 이 문항들에서 문제점이 되는 것은 지나치게 반복적인 복합형 문항인 것과 답지의 구성이다. 지구 대기와 내부의 구조를 이용하여 무려 4문항을 연속으로 복합형으로 출제하였다. 이 중에서 2번 문항부터 4번 문항까지 3개의 문항이 지구 대기 구조에 대한 지식 영역 문항이다. 지구 대기와 내부 구조 모두 4개의 층상구조를 가지고 3문항에 걸쳐 4개의 구조 중 답이 한 차례씩 나오기 때문에 실제로 2번 문항은 5지 선다형이 아니라 4지 선다형이 되고, 3번 문항은 3개, 그리고 4번 문항은 2개의 답지만을 가지게 되는 문제점이 발생하게 된다. 따라서 문항의 형식을 바꾸어 복합형이 아닌 단독형 문항으로 수정하여 출제하는 것이 더 적절할 것이다.

그 외 변별도에 문제가 있는 문항으로는 G중학교의 22번 문항과 Y고등학교의 20번 문항으로 각각 0.00과 0.06의 변별도를 보였다. G중학교의 22번 문항의 경우는 속력의 개념을 묻는 문제인데, 주어진 자료가 ‘시간-속도’ 그래프로 해당 학년에 적합하지 않은 자료를 제시함으로 변별력이 전혀 없는 잘못 출제된 문항으로 판단된다. 이상의 분석 결과를 종합해보면, 문제점이 지적되는 문항들은 주로 선다형 문항으로 문항 유형 및 형식, 답지 제작, 자료 제시 등이 적절하지 않은 것으로 분석되었다.

〈표 8〉 배점과 답지 구성에 문제가 있는 문항

문항 번호	P	DI	답지	응답률		
				전체	상위	하위
16	0.91	0.41	1	0.02	0.00	0.08
			2	0.02	0.00	0.06
			3	0.91	1.00	0.76
			4	0.03	0.00	0.06
			5	0.02	0.00	0.05
17	0.36	0.45	1	0.07	0.04	0.08
			2	0.38	0.16	0.48
			3	0.08	0.03	0.15
			4	0.11	0.07	0.15
			5	0.36	0.70	0.13

※ (16~17) 다음 그림은 빛이 공기 중에서 물로 들어 갈 때의 빛의 경로를 나타낸 것이다.



16. 그림과 같은 현상에 의한 예로 옳지 않은 것은?(4점)

- ① 물 속에 가라앉은 동전이 떠 보인다.
 ② 어항 속에 물고기는 실제보다 커 보인다.
 ③ 째깍한 방에 들어가면 아무것도 보이지 않는다.
 ④ 수영장이나 욕조 속의 바닥은 실제보다 얕아 보인다.
 ⑤ 물이 들어 있는 그릇에 젓가락을 담그면 꺾여 보인다.

17. 위와 같은 현상을 이용한 광학기기를 <보기>에서 모두 고른 것은? (3점)

보 기

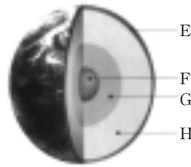
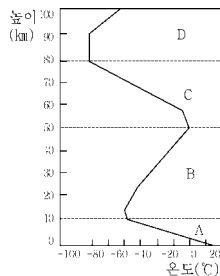
ㄱ. 안경 ㄴ. 돋보기 ㄷ. 사진기 ㄹ. 반사 망원경
 ㅁ. 현미경의 반사경 ㅂ. 현미경의 대물 렌즈

- ① ㄱ, ㄴ, ㄷ ② ㄹ, ㅁ, ㅂ ③ ㄱ, ㄷ, ㄹ, ㅁ
 ④ ㄱ, ㄹ, ㅁ, ㅂ ⑤ ㄱ, ㄴ, ㄷ, ㅂ

〈표 9〉 문항 형식에 문제가 있는 문항

문항 번호	P	DI	답지	응답률		
				전체	상위	하위
2	0.83	0.61	1	0.04	0.00	0.12
			2	0.83	1.00	0.48
			3	0.06	0.00	0.17
			4	0.02	0.00	0.08
			5	0.04	0.00	0.14
3	0.63	0.57	1	0.63	0.95	0.26
			2	0.08	0.00	0.22
			3	0.18	0.02	0.28
			4	0.06	0.02	0.14
			5	0.05	0.01	0.10
4	0.90	0.47	1	0.01	0.00	0.03
			2	0.02	0.00	0.06
			3	0.04	0.00	0.14
			4	0.90	1.00	0.73
			5	0.02	0.00	0.03

※ (2~5) 다음 그림을 보고 문제의 설명에 가장 적당한 것을 보기에서 고르시오.



2. 생명체에 해로운 자외선을 차단해 준다.

- ① A ② B ③ C ④ D ⑤ E

3. 공기를 이동시켜서 태양열을 골고루 퍼뜨려 준다.

- ① A ② B ③ C ④ D ⑤ E

4. 공기가 가장 희박하며, 지역에 따라 오로라가 관측된다.

- ① A ② B ③ C ④ D ⑤ E

3. 학교 지필평가의 신뢰도 분석

<표 10>은 5개 중학교와 3개 고등학교에서 사용된 내신 지필평가 문항의 내적 일관성 신뢰도인 Cronbach α 를 문항 유형에 따라 산출한 것이다. 8개 학교 모두 신뢰도 계수가 양호하게 산출되었다. H중학교와 I고등학교가 신뢰도가 높게 산출된 것은 난이도가 적절하고, 변별도가 높은 양호한 문항이 많았기 때문인 것으로 판단된다. 반면, Y고등학교의 경우는 난이도가 높고, 변별도가 낮은 문항이 다른 학교에 비하여 많았기 때문에 신뢰도가 다른 학교에 비하여 낮은 것으로 분석된다. G중학교의 경우, 단답형 및 서술형의 변별도가 높은 편이나 신뢰도가 선다형에 비해 낮게 산출된 주요 원인은 문항 수가 3문항으로 선다형 23문항에 비하여 상대적으로 문항 수가 적었기 때문인 것으로 판단된다.

〈표 10〉 문항 유형에 따른 신뢰도 계수

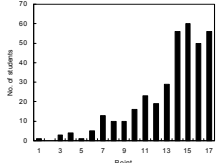
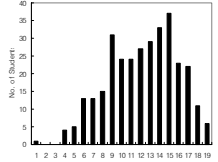
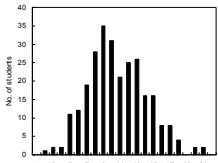
		중학교					고등학교		
		S	N	G	M	H	J	Y	I
신뢰도 계수 (Cronbach α)	선다형 문항	0.814	0.770	0.843	0.845	0.900	0.830	0.605	0.865
	단답 및 서술형 문항	0.716	0.706	0.648	0.797	-	-	0.628	-

4. 문항 점수 분포 유형 분석

<표 11>은 모든 선다형 문항의 배점을 1점으로 하여 히스토그램을 작성한 후 그 유형을 분석한 것이다. 점수 분포 유형은 크게 세 가지(부적편포, 양봉분포, 정적편포)로 나눌 수 있었다. 중학교는 3개 학교가 정상분포를 벗어난 부적편포를 나타내었다. 일반적으로 완전학습의 결과 나타나는 것이 부적편포이나, 쉬운 문항 난이도와 지역적 성향에 따른 학습자의 평균 수준 등이 반영된 것으로 보인다. 참고로 가장 부적편포가 강한 M중학교의 경우는 학군이 발달하고 사교육이 왕성한 양천구 목동 지역에 위치하고 있다. 반면, 2개 중학교와 2개 고등학교에서는 점수 분포 형태가 양봉분포를 나타내어 상위권과 하위권 학생들이 양분되는 경향을 보였다. 문항 난이도와 변별도가 적절하여 정규분포인 가우스분포가 예상되나 실제로는 이질성이 강한 양봉분포를 보이고 있다. 가장 확연한 양봉분포를 보인 N중학교의 경우, 성북구 장위동에 위치하여 가정환경이 상대적으로 열악한 편이고, 사교육으로 인한 보충교육이 부족한 학생의 비중이 높은 편이어서 학습자 수준이 대체로 낮을 것으로 생각되나 실제로는 학습자 수준의 차이가 큰 것으로 나타났다. 이것은 예전에 비해 지역적 성향에 따른 동질성이 많이 희석되고 이질성이 심화되고 있는 지역이 많아지고 있음을 의미한다고 볼 수

있다. 한편, Y고등학교는 연구 대상 8개 학교 중 유일하게 정적분포를 나타내었는데, 이것은 난이도 0.45로 다른 학교에 비해 상대적으로 어려운 문항이 많았기 때문인 것으로 분석된다.

〈표 11〉 문항 점수 분포 유형

유형	학교	난이도(P)	변별도(DI)	분포도
부적분포	S중학교	0.78	0.52	
	M중학교	0.77	0.51	
	H중학교	0.71	0.50	
양봉분포	N중학교	0.62	0.46	
	G중학교	0.56	0.49	
	J고등학교	0.66	0.45	
	I고등학교	0.69	0.45	
정적분포	Y고등학교	0.45	0.31	

IV. 결론 및 제언

본 연구에서는 8개 중·고등학교의 과학 교과 지필평가 문항을 수정 구안된 문항분류틀을 이용하여 현행 학교 지필평가 문항의 출제 경향을 파악하고, 문항분석을 통하여 난이도, 변별도, 신뢰도를 통계적으로 분석하였으며, 학생 점수 분포 유형을 분류하였다.

출제 경향을 분석하기 위해 문항분류틀을 적용한 결과, 전체적으로 지식 영역에 대한 문항의 비율이 높으며 탐구과정 영역의 문항 비율이 상대적으로 매우 낮은 것으로 분석되었다. 탐구과정 중에서는 ‘자료해석 및 분석’에 집중되어 있어 출제된 탐구과정 영역의 문항들이 다양한 탐구 기능을 요구하도록 제작되지 못하였음을 알 수 있었다. 또한 대부분의 문항들이 교과서적 맥락으로 구성되어 있어 발산적 사고나 창의적 사고를 요구하는 문항의 비율이

매우 낮은 것으로 판단된다. 문항 표현은 주로 그림이나 문장을 많이 사용하는 것으로 나타났다. 그래프나 표의 사용 비율은 상대적으로 낮은 것으로 나타났다. 문항 유형과 형식은 학교마다 매우 다르게 나타났다. 현행과 같이 학교나 교사에 따라 문항 유형의 비율이 너무 차이 나지 않도록 하며, 학생들에게 고차원적 사고 기능(high order thinking skills)을 요구하도록 하는 문항이 출제되도록 하기 위해서는 서술형 문항을 일정 비율 이상으로 출제하도록 권장하는 방안을 고려해 보아야 할 것이다.

지필평가 문항의 양호도를 분석한 결과, 난이도는 대체로 낮은 것으로 나타나 난이도의 적절한 배분을 고려하지 않은 것으로 판단되며, 대체로 쉽게 출제하는 경향을 보였다. 8개 학교 모두 전반적으로 변별력 있는 문항을 제작한 것으로 분석되었다. 단답형 및 서술형 문항의 양호도가 선다형보다 높아 교사들이 양호한 문항을 제작하기에 단답형 및 서술형이 더 용이한 것으로 나타났다. 한편, 교사들이 생각하는 난이도와 학생들이 느끼는 난이도 사이에는 상당한 차이가 있음을 확인할 수 있었으며, 특히 단답형 및 서술형 문항보다는 선다형에서 많은 차이가 있었다. 따라서 선다형 문항의 경우, 교사가 임의로 배점하지 않고 동일한 점수로 배점하고, 단답형 및 서술형은 교사가 차등적 배점을 하는 것이 적합하다고 판단된다.

문항 내적 일관성 신뢰도는 8개 학교 모두 양호하였으며, 양호한 문항이 많을수록 신뢰도가 증가함을 확인할 수 있었다. 학생 점수 분포 유형 분석에서는 정상분포는 없고, 부적편포와 양봉분포가 주를 이루었다. 부적편포에는 쉬운 문항 난이도, 지역적 성향, 사교육의 영향 등이 반영된 것으로 보인다. 적절한 난이도를 보인 다수의 학교는 이질성이 강한 양봉분포를 보여 학습자 간 수준 차이가 큰 것으로 나타났다. 또한 주요 특징적인 문항의 문제점을 분석한 결과, 배점과 답지 구성, 그리고 문항 형식에 문제가 있음을 확인하였다.

이와 같은 연구 결과로 볼 때, 학교 과학 교과 지필평가 문항의 질을 향상하기 위해서는 다양한 문항 표현의 사용과 고른 탐구과정 영역이 반영되도록 하여 문항의 난이도를 현행보다 높여야 할 것이다. 또한 적절한 난이도의 문항 출제 비중을 현행보다 높여야 교수·학습 목표의 성취 여부를 확인하여 교수법을 개선하거나 교육과정을 개편하는 등 평가의 교수적 기능을 제대로 수행할 수 있을 것이다. 선다형 문항의 제작에 있어서는 교사들이 문항의 배점이나 표현, 형식에 좀 더 세심한 주의를 기울여야 할 것이다. 더불어 현장의 교사들이 내신 지필평가 문항을 개선하여 양호한 문항을 제작하고, 평가 결과를 올바르게 해석하여 적절한 피드백을 하는 등의 평가 전문성 개발(professional development in assessment)을 위한 교사 연수 프로그램들이 마련되어야 할 것이다.

참 고 문 헌

- 김찬중(1992). 대학수학능력시험 제5차 실험평가 수리탐구 영역에서 지구과학 분야의 문항분석. **한국지구과학학회지**, 13(4).
- 배영부(1994). 국민학교 과학 우수 아동을 위한 과학탐구능력 경시대회의 객관식 문항분석. **한국지구과학학회지**, 15(4).
- 변창진 · 최진승 · 문수백 · 김진규 · 권대훈(2001). **교육 평가**. 학지사.
- 성태제(1996). **문항제작 및 분석의 이론과 실제**. 학지사.
- 성태제(2002). **타당도와 신뢰도**. 학지사.
- 우종옥 · 이항로 · 구창현(1996). 과학 탐구능력 평가 문항 유형 변화에 관한 종단적 연구. **한국과학교육학회지**, 16(3), 314-328.
- 이기영(2005). 한국지구과학올림피아드 문항분석을 통한 문항의 질 향상 방안 연구. **한국지구과학회지**, 26(6), 511-523.
- 이상범 · 이광필 · 최상돈 · 황석근(1999). 과학영재교육센터 학생선발문항분석 및 선발방법에 대한 제언. **한국과학교육학회지**, 19(4), 604-621.
- 이양락(2002). 대학수학능력시험 ‘과학탐구’의 응시자 수와 평균 점수 변화 및 문항에 대한 학생 반응. **한국과학교육학회지**, 22(2), 345-356.
- 홍미영 · 전경문 · 이범홍 · 이양락(2002). 대학수학능력시험 화학II 문항에 대한 학생들의 응답 분석. **한국과학교육학회지**, 22(1), 204-213.
- 홍미영 · 전경문 · 이범홍 · 이양락(2002). 대학수학능력시험 공통과학 중 화학 영역의 문항 및 응시자 응답 분석. **한국과학교육학회지**, 22(2), 378-386.
- Baker, F. B. (1992). *Item response theory: Parameter estimation techniques*. New York: Marcel Dekker.
- Cangelosi, J. S. (1990). *Designing tests for evaluating student achievement*. Longman.
- Cronbach L. J. (1951). Coefficient alpha and the internal structure of test. *Psychometrika*, 16, 297-334.
- Kempa, R. F. (1986). *Assessment in science*. Cambridge University Press.

• 논문접수 : 2006년 4월 15일 / 수정본 접수 : 2006년 5월 15일 / 게재 승인 : 2006년 5월 24일

ABSTRACT

An Exemplary Analysis of Paper and Pencil Test Items of Current Secondary School Science

Hyun-Suk Oh(Teacher, Namdaemun Middle school)

Ki-Young Lee(Teacher, Hansung Science High school)

The purpose of this study is to analyze paper and pencil test items of current secondary school science. We investigated by using modified item classification framework, and yielded item difficulty(P), discrimination index(DI) and reliability(Cronbach α) by applying the classific test theory, and classified students' test score distribution type.

The findings of this study are as follows: (1) In item making trend, the proportion of knowledge domain is much higher than inquiry process domain. Most of items are composed of textual context. For the most part, items are represented by using picture or text, while graph and table is rarely used. Item type and form is various according to school. (2) In case of P and DI, the greater part of items are made easily, and DI is farly high. The proportion of satisfactory item of short answer/essay type is higher than that of multiple choice type. There is big difference between teacher and student in P perception. (3) As a whole, reliability of eight school are satisfactory. In classification of students' test score distribution type, we recognize three distribution types(positive, bimodal, and negative). We inferred that item difficulty and regional characteristics influence to distribution type.

Accordingly, to improve the quality of paper and pencil test items of current secondary school science, teachers need to use various item representation and elements of inquiry process domain impartially, and reduce the proportion of easy item. As well, teachers should be more careful in making multiple choice item and its allotment.

Key Words : science paper and pencil test, item analysis, item classification framework, item difficulty, discrimination index, reliability