

# 인지 진단 이론을 활용한 중학교 과학 시험 결과의 분석

김지영\* · 김수진<sup>1</sup>

이화여자대학교\* · <sup>1</sup>한국교육과정평가원

## Analysis of Test Result at Secondary Science Using Cognitive Diagnosis theory

KIM, Ji Young\* · Kim, Soo Jin<sup>1</sup>

Ewha Womans University · <sup>1</sup>Korea Institute for Curriculum and Evaluation

**Abstract:** The purpose of this study is to search effective assessments methods by using the Fusion model of Cognitive diagnosis theory. Attributes are skills or cognitive processes that are required to perform correctly on a particular item. After test items were developed, item's attributes were decided and Q-matrix about item's attributes was made. After testing, the result was analyzed according to gender and achievement level. The results of the analysis showed that students mastered 'Interpreting data' best, and 'synthesizing' worst among the five attributes.

Female students showed higher ability than male students in 'recalling.' Students of high achievement level mastered more scientific attributes than students of low achievement level. Conventional assessments only provided a single summary score but Cognitive diagnosis modeling provided useful information by estimating individual knowledge states by assessing whether an examinee has mastered specific attributes measured by the science test. The skill profiles can offer a skill level of strong, weak, or mixed for each student for each skill. Therefore, the skill profiles will provide useful diagnostic information in addition to single overall scores.

**Key words:** scientific attributes, cognitive diagnosis theory, Fusion Model, Q-matrix

### I. 서 론

인지과학은 측정·평가 분야에서 기본적인 지식과 인지 기능(cognitive skills)의 이해와 더불어 측정 내용에 대한 인지적 정보를 줄 수 있는 이론적 토대를 만들어 왔으며, 피험자에 대한 학습 정보 보안을 위하여 인지 진단 이론(Cognitive Diagnosis Theory)을 발전시켰다(Chipman *et al.*, 1995). 인지 진단 이론은 검사에 의해 측정되는 특정 속성(attribute)을 학습자가 어느 정도 숙달했는지 알려줄 수 있는 이론이다(DiBello *et al.*, 1995; Tatsuoka, 1983, 1995; Hartz, 2003). 여기서 특정 속성은 학생들이 문항을 정확히 푸는데 필요한 능력(ability), 지식(knowledge), 기능(skills)이나 인지적 과정(cognitive process) 등과 같이 다양하다(Tatsuoka, 1983). 인지 진단 이론은 개별 학생의 지식 및 기능에 대한 숙달 상태를 진단하여, 검사에 의해 측정되는 유용한 정보를 학생이나 교사, 학부모 등에게 제공한다는 장점이 있다(김수진 외, 2008). 또한 학생이 습득한 속성을 정확하게

진단할 뿐 아니라 학생의 학습 발전 및 진전 상황을 정확하게 추정해 준다(DiBello *et al.*, 1995; Embretson, 1990; Tatsuoka, 1983, 1995).

과학 학습에서 이루어지는 평가는 학생의 인지적 영역과 정의적 영역에 대한 유용한 정보를 제공하여 학생 개개인의 과학 학습과 전인적인 성장을 돕고 교사의 교수 활동과 수업 방법을 개선하는 데 활용되어야 한다(교육인적자원부, 2007). 그러나 전통적으로 학교에서 이루어진 과학 평가는 학력의 측정에 목적을 두고 실시되었으며 그 결과를 점수로 제공해 주는 것이 대부분이었다. 숫자로 요약된 점수만을 가지고 알 수 있는 정보에는 한계가 있으므로 학생들이 과학을 잘 하기 위해 어떤 개선점이 필요한지에 대해서 구체적으로 알기 어렵다. 인지 진단 이론의 목적은 학생 개개인이 어떤 속성을 습득 또는 숙달하였는지 알려 주어 교사와 학생에게 적절한 피드백을 제공하는 것이므로(DiBello *et al.*, 1995; Embretson, 1990; Tatsuoka, 1983, 1995), 과학 교육에서 이 이론에 기초하여 평가 결과를 분석할 경우 학생들이 과학 문제

\*교신저자: 김지영(nicefine@treechal.com)

\*\*2009.03.06(접수) 2009.04.24(1심통과) 2009.08.29(2심통과) 2009.10.10(3심통과) 2009.10.13(최종통과)

를 풀 때 필요한 능력에 대한 정보를 교사와 학생들에게 제공해 줄 수 있을 것이다.

인지 진단 이론을 이용하여 평가 결과를 해석한 연구로는 수학·과학 성취도 추이 변화 국제비교 연구인 TIMSS(Trends in International Mathematics and Science Study)를 이용해 각국 학생들이 문제를 풀 때 필요한 여러 가지 속성을 분석한 연구가 있으며(Tatsuoka *et al.*, 2004), TIMSS-R 결과를 가지고 터키와 미국 학생들이 수학 문제를 풀 때 필요한 속성의 숙달 정도를 비교한 연구가 있다(Dogan & Tatsuoka, 2008). 우리나라에서도 인지 진단 이론을 바탕으로 한 몇 가지 연구가 진행되었다(김명화, 1997; 김성호, 송미영 1997; 김성훈, 박영선, 2004; 김성훈, 2005; 김성훈, 2007, 김수진, 송미영, 김선희, 2008). 특히, 김수진 등(2008)은 중학교 3학년 학생들의 수학 문제를 풀 때 필요한 능력을 진단하기 위해 Hartz(2002)의 RUM(Reparameterized Unified Model)이라고도 불리는 Fusion Model을 적용하여 학생들의 속성 숙달 정도를 분석하였다. 그러나 과학 교육에서 인지 진단 모형을 활용하여 학생들의 과학 능력을 진단하고 평가하는 연구는 거의 없는 실정이다. 학생들이 과학 문제를 풀 때 필요한 속성의 숙달 여부를 평가하여 학생들에게 이와 관련된 정보를 주기 위한 하나의 방법으로 인지 진단 이론이 활용될 수 있을 것이다.

이에 본 연구에서는 인지 진단 이론에 근거하여 과학 지필 평가 결과를 분석하고 그 결과를 활용할 수 있는 방안을 모색해 보고자 한다. 즉, 인지 진단 이론의 한 모형인 Fusion Model을 이용하여 과학 지필 평가 결과를 분석하는 방법을 제안하고 학생들이 어떤 과학적 속성을 숙달하고 있는지 파악해 보고자 한다. 또한 이러한 분석한 결과가 제공해줄 수 있는 정보와 분석의 의의를 찾아보고자 한다. 특히 본 연구에서는 현장에서 이러한 평가 결과의 분석이 이용될 수 있도록 학교에서 실시되는 중간고사와 기말고사 시험을 활용하여 분석을 실시해 보고자 한다.

## II. 연구 방법 및 절차

### 1. 검사 도구 및 연구 대상

본 연구의 검사 도구는 모두 과학 교육 전공자로 5년 이상의 경력을 가진 중학교 과학 교사 2인과 과학

교육 전문가인 연구자가 함께 개발하였다. 개발된 문항은 중학교 2학년 2학기 중간고사와 기말고사의 객관식 문항으로 중간고사에 16개, 기말고사에 15개 포함되었다. 전체 31개 문항 모두 5지 선다형으로 개발하였다. 평가 대상이 중학교 2학년 이므로 중간고사와 기말고사에 포함된 문항의 내용은 ‘식물의 구조와 기능’ 단원, ‘자극과 반응’, ‘지구의 역사와 지각 변동’, ‘전기’ 단원에서 골고루 출제하였으며 마지막 단원인 ‘혼합물의 분리’ 단원은 앞부분의 내용만을 포함하였다.

3명의 출제자가 출제하고자 하는 문항수의 2배가 되는 문항을 전통적인 문항 개발 방법에 의하여 개발한 후 평가 목표를 잘 반영하고 있다고 판단되는 문항을 선정하였고 일부 문항의 경우 평가 목표에 맞추어 내용을 수정하였다. 이 연구에서의 내적 신뢰도(Cronbach's  $\alpha$ )는 .88이었다.

검사 대상은 서울 소재 중학교 2학년 학생 466명으로 남학생이 239명 여학생이 227명 이었다. 검사는 2008년 10월 초와 12월 초에 각각 실시되었다.

### 2. Fusion Model을 활용한 분석 절차

본 연구는 인지 진단 이론을 활용하여 과학 시험 결과를 해석하기 위해서 Fusion Model을 활용하였다. 이 모형을 활용하여 인지 진단을 하기 위해서 먼저 학교의 과학 수업시간에 학습한 내용을 잘 평가할 수 있는 평가 목표를 세우고 그 목표에 맞추어 문항을 개발하였다. 시험을 실시한 후 각 문항에 알맞은 속성을 부여하였다. 속성이 정해지면 다음으로 Q-행렬이 작성되는데, Q-행렬은 문항과 문항이 측정하는 속성 사이의 관계를 나타내는  $K \times n$ 행렬이다(Tatsuoka, 1983, 1990, 1995). Q-행렬의 원소는 속성을 습득하는 것이 문항을 푸는 데 요구되면 1, 그렇지 않으면 0의 값을 가지며, 여기서  $K$ 는 측정하고자 하는 문항의 수를,  $n$ 은 속성의 수를 가리킨다. Table 1은 Q-행렬의 예를 나타낸다.

Table 1  
The example of Q-matrix

	attribute 1	attribute 2
Item 1	1	0
Item 2	0	1
Item 3	1	1

Table 2

Item patterns expected by mastered attributes

Master attributes		Item response patterns		
attribute 1	attribute 2	Item 1	Item 2	Item 3
0	0	0	0	0
1	0	1	0	0
0	1	0	1	0
1	1	1	1	1

문항 1을 풀기 위해서는 속성1이 필요하고, 문항 2를 풀기 위해서는 속성2가 필요하다. 문항 3의 경우에는 두 속성이 모두 필요하다. 학생들이 문항에 성공적으로 답하기 위해서는 그와 관련된 속성을 모두 숙달하고 있어야 한다. Table 1의 경우에서 가능한 문항 반응 패턴으로 Table 2를 예상할 수 있다. 문항 반응 패턴은 문항을 맞으면 1, 틀리면 0으로 나타낸다.

Q-행렬을 바탕으로 Table 2와 같은 문항 반응 패턴에서 문항 모수, 각각의 속성에 대한 숙달 정도를 나타내는 학생 모수를 추정한다. 문항 모수는 두 가지로 측정되는데, 하나는 문항의 풀이에 요구되는 모든 속성을 숙달한 학생이 그 문항을 풀 때 그 속성을 정확히 적용할 확률( $\pi_i^*$ )이다. 다른 하나는 속성을 숙달했으면서 속성이 추출되는 문항에 정답을 맞힐 확률에 대한 어떤 속성을 숙달하지 못하면서 그 속성이 추출된 문항에 정답을 맞힐 확률의 비율( $I_{ik}^*$ )로, 속성과 관련된 문항의 변별도에 해당하는 것이다. 학생 모수는 학생이 각 속성을 숙달하였는지를 나타내는 과학적 속성 추정치이며, 이 추정치에 따라 속성의 숙달 여부를 결정하게 된다(Hartz, 2003; 김수진 등, 2008). 어떤 속성을 측정하는 문항의 수가 너무 적을 경우 속성의 숙달 여부를 판단하기 어렵다. 따라서 여러 가지 많은 속성의 숙달 여부를 알아보기 위해서는 그 만큼 많은 문항수가 필요하다(de la Torre, 2008). 학생들의 여러 가지 많은 과학적 속성을 측정하기 위해서는 문항수가 많아야 하는데 학교 시험 시간이 정해져 있고 그 시간 안에 풀 수 있는 문항의 수는 어느 정도 한계가 있으므로 학교 시험으로 많은 인지적 속성을 측정하는 데에는 한계가 있다.

### 1) 문항의 속성 추출

문항이 측정하고 있는 속성을 판단하기 위해서 본 연구에서는 Table 3의 TIMSS 2007 인지영역 평가

틀(Mullis *et al.*, 2005)을 참조하였다. TIMSS의 인지 영역은 알기, 적용하기, 추론하기의 3개 영역으로 구성되어 있는데 각각의 영역 내에는 Table 3과 같이 각각의 하위요소가 있다. 이처럼 TIMSS의 평가틀은 인지 영역 내에 세부적인 하위 요소가 있어, 너무 포괄적이지 않으면서 특수한 문항에 한정되지 않도록 속성 추출을 가능하게 하는 데 도움이 된다. 이 평가틀은 문항의 설계시 행동 영역을 나타내는 것이며, 평가 결과의 분석에서 각 문항이 어떤 속성을 갖고 있는지의 근거로도 활용될 수 있다.

문항의 속성 추출은 연구자를 포함한 두 명의 과학 교육 전문가와 문항을 출제했던 과학 교사 2명이 함께 참여하여 진행하였다. 연구에 사용된 문항의 내용

Table 3

Evaluation framework of cognitive domain at TIMSS 2007

cognitive domain	Elements
Knowing	Recall/Recognize Define Describe
	Illustrate with Examples Use Tools and Procedures
Applying	Compare/Contrast/Classify Represent/Model Relate
	Extract/Apply Information Find Solutions Explain
Reasoning	Analyze/Interpret/Solve Problems
	Integrate/Synthesize Hypothesize/Predict Design/Plan Draw Conclusions Generalize Evaluate Justify

을 분석한 후 TIMSS 2007의 인지 영역 평가들에서 제시된 각 하위요소의 정의를 바탕으로 각 문항의 풀이에서 요구되는 능력으로 적합한 것을 추출하였다. 모두 TIMSS의 인지 영역의 하위 요소에 제시된 정의를 정확하게 숙지한 후 문항 푸는데 필요하다고 판단되는 속성을 부여하도록 한 후 의견 일치가 되지 않는 문항에 대해서는 토론을 거쳐 속성을 결정하였다. 이러한 과정을 거쳐 본 연구에서는 검사지에 속한 문항에서 평가하고 있다고 판단할 수 있는 속성인 회상/인식하기, 정보해석하기, 해답 찾기, 설명하기, 통합/종합하기, 결론도출하기의 6개를 선정하였다.

Table 3의 인지 영역 하위 요소 중에서 본 연구의 검사 도구에 포함된 문항에서 측정하고 있는 능력으로 추출된 6가지 속성을 TIMSS의 평가들에서는 다음과 같이 설명된다.

- 회상/인식하기: 과학적 사실, 관계, 과정과 개념에 대하여 정확하게 진술한다. 특정 생물체나 물질, 과정의 성질이나 특징을 안다.
- 정보해석하기: 과학 개념이나 원리에 비추어 문장, 글 또는 그래프 형태의 정보를 추출하여 적용한다.
- 해답찾기: 개념을 직접적으로 적용하는 문제에서 정량적이거나 정성적인 해답을 찾기 위해 관계나 식 등을 이용한다.
- 설명하기: 과학 개념, 원리, 법칙이나 이론을 이용하여 관찰 사실이나 자연 현상에 대하여 설명하거나 이유를 제시한다.
- 통합/종합하기: 서로 다른 몇 개의 요인이나 개념들을 고려하여 문제를 해결한다. 다른 과학 영역에 해당하는 개념들을 서로 연결하거나 여러 분야에 걸친 통합된 개념이나 주제를 이해한다.
- 결론도출하기: 증거나 과학 개념 이해를 바탕으로 타당하게 추론한다. 제기된 문제나 가설에 부합하는 적절한 결론을 도출한다. 원인과 결과에 대한 이해를 보여준다.

본 연구에서는 회상/인식하기를 회상하기로, 해답 찾기는 과학적 문제를 풀기 위해 수학적 관계나 식을 이용하는 것으로 한정하여 계산하기로 이름을 바꾸어서 명명하였다. 정보 해석하기는 표나 그래프나 문장으로 제시된 정보를 해석하는 것으로 한정하여 자료 해석하기로, 통합/종합하기는 종합하기 속성으로 명

명하였다. 다른 중요한 과학적 속성이 문항을 통해 평가될 수 있지만, 여기서는 검사 문항에서 측정하고 있는 6개의 속성만을 살펴보고 있다.

예를 들어 살펴보면 Fig 1에 제시된 문항의 경우 그래프에 제시된 정보를 해석해야 하므로 정보해석하기와 그래프에서 주어진 정보로부터 정량적인 해답을 찾기 위해 옴의 법칙을 이용해 문제의 답을 계산해야 하므로 계산하기의 속성이 부여되었다.

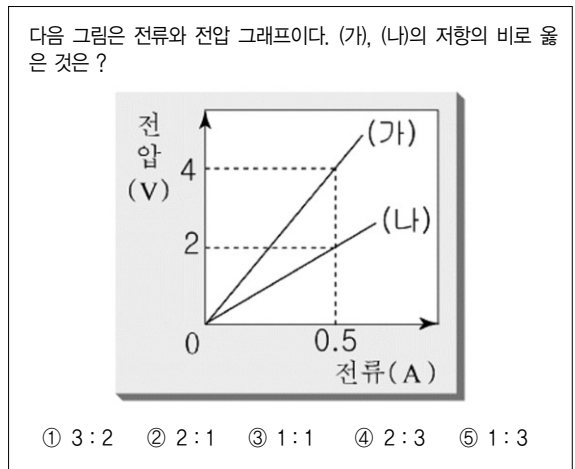


Fig. 2 Example of item

## 2) Q-행렬 작성

Fusion Model을 활용하기 위해서는 각각의 문항과 속성을 연결하는 Q-행렬이 필요하다. 연구진이 각각의 문항에 부여한 속성을 토대로 Q-행렬을 작성한 후 행렬의 적절성을 평가하기 위해 Fusion Model의 문항 모수의 수렴 여부와 문항모수 추정치를 확인하였다. 분석결과 6개 속성에 대한 문항 모수의 수렴이 적절치 않음이 나타났다. 다시 협의과정을 거쳐 3인이 문항의 속성을 재검토한 후 ‘결론 도출하기’ 속성을 제외한 5개 속성의 Q-행렬을 확정하였고, 문항과 속성의 연결을 수정하였다. 5개의 속성으로 Q-행렬을 작성한 후 문항 모수의 수렴 여부와 문항모수 추정치를 확인한 결과 6 가지 속성일 때보다 문항 모수가 더 잘 수렴하는 것으로 나타났다. 이에 본 연구에서는 5가지 속성의 과학 능력을 포함하여 분석을 실시하였다.

31개의 문항과 회상하기, 정보 해석하기, 계산하기, 설명하기, 종합하기 등 5개의 속성들을 연결하기 위해 작성된 Q-행렬은 Table 4와 같다.

**Table 4**  
Q-matrix of item

	Recalling	Interpreting data	Computing	Explaining	Synthesizing	Total
1			1	1		2
2			1		1	2
3			1		1	2
4				1	1	2
5		1	1			2
6	1			1		2
7	1					1
8				1		1
9	1					1
10	1					1
11	1	1			1	3
12				1	1	2
13		1				1
14	1	1				2
15	1				1	2
16	1				1	2
17	1					1
18	1					1
19				1		1
20				1		1
21	1	1				2
22	1			1		2
23	1					1
24					1	1
25				1		1
26	1					1
27			1			1
28		1				1
29			1		1	2
30		1				1
31				1		1
합계	14	7	6	10	9	46

### 3) 분석 방법

Q-행렬과 문항 반응 패턴을 바탕으로 Arpeggio 컴퓨터 프로그램(Hartz *et al.*, 2002)의 MCMC (Markov Chain Monte Carlo) 추정법을 사용하여 Fusion Model의 문항 모수 및 학생 모수를 추정하였다. 분석 결과를 토대로 검사에 참여한 학생 전체가 어떤 속성을 잘 숙달했는지 살펴보고, 학생들의 과학

성취 점수에 기초한 성취 수준별로 숙달한 속성과 속성의 수를 조사하였다.

MCMC 추정의 경우 모수 추정치의 자기상관계수 (autocorrelaiton)를 기초로 수렴 여부를 결정하는데, 본 연구에서 5가지 속성으로 Q-행렬을 작성하여 추정한 문항 모수와 학생 모수의 자기상관계수가 대부분 0.2보다 적게 나타나 수렴하였다.  $\pi_i^*$  문항 모수

추정치는 0.92~0.98의 높은 값을 가진 것으로 나타나 Q-행렬에 명시된 과학적 속성이 강하게 요구되는 문항들이라 할 수 있었다.  $r_{jk}^*$  문항 모수 추정치는 대체로 0.28~0.81로 각 문항과 관련된 각각의 속성을 잘 변별한다고 할 수 있었다. 문항 모수 추정을 통해 본 연구에 사용된 전체 문항은 5가지 속성을 요구하면서 각 속성을 잘 변별한다고 할 수 있다(김수진 외, 2008; Hartz, 2002).

## IV. 연구 결과

### 1. 전체 학생들의 속성 숙달 정도

#### 가. 전체

본 연구에 참여한 전체 466명 학생들의 검사 점수는 평균은 21.55이고 표준편차는 7.12이었다. 연구에 참여한 학생들이 숙달한 속성을 나타내면 Table 5와 같다.

연구에 참여한 중학교 2학년 학생들이 각 과학적 속성을 숙달할 기대 확률은 46.0%에서 65.2%임을 볼 수 있다. 학생들의 60.5%가 '회상하기' 를, 65.2%가 '정보 해석하기' 를, 46.6%가 '계산하기' 를, 59.0%가 '설명하기' 를 46.0%가 '종합하기' 를 숙달하고 있는 것으로 나타났다. 전체적으로 '정보해석하기', '회상하기', '설명하기', '계산하기', '종합하기' 의 순으로 학생들이 숙달한 속성이 드러났다. 이는 5

개의 과학적 속성 중에서 '정보 해석하기' 가 가장 쉬우며 '종합하기' 가 가장 어려운 것임을 보여준다.

과학 문제를 푸는 데 있어서 학생들에게 필요한 능력 중에서 전체적으로 학생들이 어떠한 속성을 숙달하는데 어려움을 느끼는지 판단할 수 있으며 교사는 이러한 점들을 참고하여 교수 학습의 지도 계획을 세울 수 있을 것으로 판단된다.

#### 나. 성별

본 연구에 참여한 학생들의 남학생과 여학생의 점수 차이를 분석한 결과는 Table 6과 같다.

남학생의 평균은 21.34로 여학생의 평균인 21.78보다 조금 낮았으나 그 차이는 통계적으로 유의하지 않았다. 남학생과 여학생이 숙달한 속성에 차이가 있는지 알아보기 위해 분석을 실시한 결과는 Table 7과 같다.

전체 속성의 숙달 정도를 성별로 비교한 결과 '자료 해석하기', '계산하기', '설명하기', '통합하기' 의 숙달 비율에서 성별로 유의한 차이가 나타나지 않았으나 '회상하기' 의 숙달 비율에서 유의한 차이가 나타났다. '회상하기' 에서 여학생의 숙달 비율은 65.2%로 남학생의 숙달 비율인 56.1%에 비해 유의하게 높았다. 여학생이 남학생에 비해 '회상하기' 능력을 더 많이 가진 것을 알 수 있다. 남학생과 여학생의 전체 점수에서는 차이가 나타나지 않았으나 인지 진단 이론으로 분석한 결과 여학생과 남학생의 인지 속성의 숙

**Table 5**

*The distribution of students that mastered attributes*

Attribute	Numbers of students (%)	
	Mastering	Non-mastering
Recalling	282(60.5)	184(39.5)
Interpreting data	304(65.2)	162(34.8)
Computing	217(46.6)	249(53.4)
Explaining	275(59.0)	191(41.0)
Synthesizing	218(46.0)	248(54.0)

**Table 6**

*The analysis of score difference between male and female*

Gender	Numbers of students	Mean	S.D	t
Female	227	21.78	6.79	.661
Male	239	21.34	7.43	
Total	466	21.55	7.12	

Table 7

The analysis about distribution difference of student that mastered attributes by gender

					Numbers of students (%)
Attribute	Gender	Mastering	Non-mastering	Total	$\chi^2$
Recalling	Female	146(65.2)	79(34.8)	227(100)	4.06*
	Male	134(56.1)	105(43.9)	239(100)	
Interpreting data	Female	150(66.1)	77(33.9)	227(100)	.139
	Male	154(64.4)	85(35.6)	239(100)	
Computing	Female	106(46.7)	121(53.3)	227(100)	.003
	Male	111(46.4)	128(53.6)	239(100)	
Explaining	Female	134(59.0)	93(41.0)	227(100)	.000
	Male	141(59.0)	98(41.0)	239(100)	
Synthesizing	Female	105(46.3)	122(53.7)	227(100)	.049
	Male	113(47.3)	126(52.7)	239(100)	

Table 8

The distribution of mater attributes by achievement level

Attribute	Achievement level	Mastering	Non-mastering	Total	$\chi^2$
Recalling	High	150(100.0)	0(0.0)	150	327.23***
	Medium	126(78.3)	35(21.7)	161	
	Low	6(3.9)	149(96.1)	155	
Interpreting data	High	150(100.0)	0(0.0)	150	309.21***
	Medium	137(85.1)	23(14.9)	161	
	Low	17(11.0)	138(89.1)	155	
Computing	High	146(97.3)	4(2.7)	150	290.84***
	Medium	71(44.1)	90(55.9)	161	
	Low	0(0.0)	155(100)	155	
Explaining	High	150(100.0)	0(0.0)	150	344.08***
	Medium	124(77.0)	37(23.0)	161	
	Low	1(0.6)	154(99.4)	155	
Synthesizing	High	148(98.7)	2(1.3)	150	286.00***
	Medium	66(41.0)	95(59.0)	161	
	Low	4(2.6)	151(97.4)	155	

달 정도에 차이가 있는 것을 알 수 있다.

#### 다. 성취수준

학생들의 성취수준별로 숙달한 과학적 속성이 어떻게 다른지 살펴보기 위해 학생들의 성취수준을 상, 중, 하로 구분하였다. 상위 33%, 하위 33%를 기준으로 학생들의 성취 수준을 구분하였는데, 동점자가 있어 정확히 33%로 구분되지 않았다. 상 수준은 27점 이상을 받은 학생 150명으로 32.2%, 중 수준은 19~26점을

받은 학생 161명으로 34.5%, 하 수준은 18점 이하를 받은 학생 155명으로 33.3%에 해당하였다.

성취수준별로 학생들이 숙달한 속성을 조사한 결과는 Table 8과 같다.

상 수준 학생들은 5개의 속성 중에서 '회상하기', '정보 해석하기', '설명하기'의 속성을 모두 숙달하고 있는 것으로 나타났다. 하지만 하 수준의 학생들은 '회상하기'를 3.9%, '정보 해석하기'를 11.0%, '계산하기'를 0%, '설명하기'를 0.6%, '종합하기'를 2.6%

만 숙달하고 있어 '정보 해석하기'를 가장 많이 숙달하고 있었고 '계산하기'를 숙달한 학생은 하나도 없었다. 중 수준의 학생들은 '정보 해석하기'를 85.1%로 가장 많이 숙달하였고, '종합하기'를 41.0%로 가장 적게 숙달하였다. 모든 속성은 성취수준별로 숙달한 빈도 사이에 유의한 차이가 나타났다.

학생들의 성취수준별로 숙달한 속성의 개수를 살펴보면 Table 9와 같다.

하 수준의 학생들은 숙달한 속성이 전혀 없는 경우가 많았고, 1개의 속성만 숙달하고 있는 경우가 15.5%, 2개의 속성을 숙달한 경우가 1.3%이었다. 중 수준의 학생들은 속성을 3개 숙달하고 있는 경우가 29.2%로 가장 많았다. 다음으로 4개의 속성을 숙달하고 있는 경우가 25.5%, 2개의 속성을 숙달하고 있는 경우가 19.3%였다. 5개의 속성을 모두 숙달하고 있는 경우는 18.0%에 해당하였다. 속성을 하나도 숙달하지 못한 경우는 0.6%였다. 성취수준 상, 중, 하에 따라 숙달한 속성의 개수에는 차이가 있는 편이지만, 숙달한 속성의 개수가 학생들의 성취수준을 말한다고 할 수는 없다.

위의 평가 결과는 학생들이 전반적으로 숙달한 과학적 속성이 무엇이고, 성취 수준별로 어떤 속성을 많

이 숙달했는지 등 집단에 대한 평가의 결과 해석을 가능하게 한다.

## 2. 학생 개인별 과학적 속성에 대한 프로파일

과학적 속성을 5개로 구성한 뒤 Q-행렬에 대한 Fusion Model을 이용하여 과학적 능력에 대한 각 속성별 모수를 추정한 결과를 예시하면 Table 10과 같다. 각 학생들의 과학 성취 점수는 20점으로 모두 같지만 과학적 속성 추정치가 각기 다른 것을 볼 수 있는데, 이는 각 학생들의 문항반응 패턴이 다르기 때문이다.

Table 10에 제시된 학생들의 과학적 속성 추정치를 기초로 학생 개인별 과학적 속성의 숙달 상태에 대한 프로파일을 작성한 결과는 Table 11과 같다. 각 속성 추정치가 .50 이상이면 주어진 속성을 숙달한 것으로 판정하고 숫자 1로, .50 미만이면 주어진 속성을 숙달하지 않은 것으로 판정하고 숫자 0으로 나타내었다.

일반적으로 점수만 제공되는 평가에서는 점수가 20점인 학생들의 경우 모두 같은 등수를 차지하게 된다. 그러나 Table 11의 프로파일 정보를 보면 점수가 20점인 학생들이 숙달한 인지적 속성의 수는 2~3개 정

**Table 9**  
Numbers of master attributes by achievement level

The number of mastered attribute	High	Medium	Low	Total(%)
0	0(0.0)	1(0.6)	129(83.2)	130(27.9)
1	0(0.0)	12(7.5)	24(15.5)	36(7.7)
2	0(0.0)	31(19.3)	2(1.3)	33(7.1)
3	0(0.0)	47(29.2)	0(0.0)	47(10.1)
4	6(4.0)	41(25.5)	0(0.0)	47(10.1)
5	144(96.0)	29(18.0)	0(0.0)	173(37.1)

**Table 10**  
Estimates about scientific attributes of student that take score 20

Name of students	The pattern of item response	Recalling	Interpreting data	Computing	Explaining	Synthesizing
권대*	01011111010110111111001101000	.86	.69	.06	.77	.01
홍두*	011111111111110100110000111000	.63	.63	.54	.01	.17
황연*	000111011111001110111110101010	.85	.25	.01	.83	.16
김수*	100111110001111101011100101011	.86	.87	.12	.82	.02
김신*	0001111111110011001001111111100	.89	.01	.49	.72	.86
윤홍*	0010111100110110111111001011110	.02	.58	.52	.49	.15



Table 11

Profile about scientific attributes of student that take score 20

Name of students	The pattern of item response	Recalling	Interpreting data	Computing	Explaining	Synthesizing
권대*	0101111110101101111111001101000	1	1	0	1	0
홍두*	0111111111111110100110000111000	1	1	1	0	0
황연*	000111011111001110111101101010	1	0	0	1	0
김수*	1001111110001111101011100101011	1	1	0	1	0
김신*	0001111111110011001001111111100	1	0	0	1	1
윤홍*	0010111100110110111111001011110	0	1	1	0	0

Table 12

Profile about scientific attributes of student that take score 20

Name of students	Total score	Recalling	Interpreting data	Computing	Explaining	Synthesizing
공정*	29	1	1	1	1	1
김다*	22	1	1	1	1	0
김혜*	13	0	0	0	0	0
류현*	19	1	1	0	0	0
이은*	21	0	1	0	1	0
맹은*	19	1	0	0	0	0
김성*	25	1	1	0	1	1
김인*	24	1	1	1	1	1
김홍*	16	1	0	0	0	0
박찬*	25	1	1	0	1	0

도이며 각 학생별로 숙달한 속성이 다른 것을 알 수 있다. 학생들의 총점이 동일하지만 과학적 속성의 숙달 여부가 다른 것을 볼 수 있다.

프로파일은 각각의 기능에 대해 학습자가 숙달을 했는지 안 했는지를 나타내는 것으로, 총점에서 알 수 없는 학생 개개인의 인지적 속성에 대한 구체적인 정보를 얻을 수 있다.

Table 12는 과학 성취 점수가 다른 학생들에 대한 프로파일을 나타낸 것이다.

기존의 평가 결과가 각 학생의 점수만을 보고한 반면, Table 12에서 제시하고 있는 프로파일에는 문항을 풀 때 사용하는 과학적 속성의 숙달 여부가 제시되어 있어 학생 개개인의 강점과 약점을 구체적으로 파악할 수 있다.

학생의 점수와 속성의 숙달 여부를 살펴보면 총점이 29로 높은 공정\* 학생의 경우 모든 속성을 숙달하고 있는 것으로 나타났고 총점이 13점으로 낮은 김혜\*

학생의 경우 모든 속성을 숙달하지 못한 것으로 나타났다. 대체로 학생들의 총점이 높을수록 숙달한 속성의 수가 많음을 알 수 있다. 그러나 24점을 받은 김인\* 학생은 5가지 속성을 모두 숙달한 것으로 나타났으나 25점을 받은 김성\* 학생이나 박찬\* 학생의 경우 4가지 속성만 숙달한 것으로 나타났다. 따라서 과학 총점이 높다고 해서 학생들이 숙달한 과학적 속성의 수가 반드시 많다고 할 수는 없다. 이처럼 과학의 총점만으로 학생의 과학 능력을 판별하기 어려우나 인지 진단 이론을 활용하여 분석할 경우 학생들이 숙달한 능력에 대한 정보를 얻을 수 있다. 예를 들어 Table 12에서 점수 25점을 받은 박찬\* 학생의 경우 계산하기와 종합하기의 속성을 숙달하지 못하였다. 교사는 이 학생에게 정량적인 해답을 찾기 위해 관계나 식을 이용하는 과학 문제들을 많이 풀어보는 경험 갖도록 또한 여러 개념들이 관련된 문제의 경우 개념들을 서로 연결하여 문제를 풀 수 있는 능력을 키워야 함을

제안할 수 있다.

본 연구의 결과에서 제시한 프로파일과 같은 정보가 학생들에게 제시되면 학생들 스스로 자신이 과학 능력 중에서 부족한 부분이 무엇인지 알 수 있을 것이다. 또한 교사는 학생 개인에 대한 프로파일을 통해 학생의 과학적 능력에 대한 피드백이 가능할 것으로 판단되며 학생들의 부족한 능력을 보완하기 위한 지도 계획을 세울 수 있을 것으로 판단된다.

#### IV. 결론 및 제언

본 연구에서는 학교 교육에서 과학 시험 결과가 의미 있게 활용되고 교사와 학생 및 학부모에게 유용한 교육적 정보를 제공할 수 있도록 인지 진단 이론을 이용하여 결과를 분석하였다.

과학적 속성은 문항을 통해 측정될 수 있으므로 문항을 제작하고 평가를 실시한 후 결과를 분석하여 학생들이 어떤 속성을 많이 숙달하고 있는지, 성취수준별로 숙달한 속성과 속성의 수는 어떠한지 알 수 있었다. 5개의 속성 중에서 학생들이 가장 많이 숙달한 속성부터 차례대로 기록하면 정보 해석하기, 회상하기, 설명하기, 계산하기, 종합하기의 순인 것으로 나타났다. 이러한 결과에서 정보 해석하기가 가장 숙달하기 쉽고 종합하기가 가장 어려운 인지 속성임을 알 수 있다. 서로 다른 몇 개의 요인이나 개념을 고려하여 문제를 푸는 능력인 종합하기의 능력은 여러 개의 요인을 고려하여 문제를 풀어야 하므로 학생들이 숙달하기 어렵하다고 할 수 있겠다. 숙달하기 쉬운 속성과 어려운 속성들이 있으므로 학생들의 과학 문제 풀기를 지도할 때 정보 해석하기, 회상하기, 설명하기, 계산하기, 종합하기의 순서로 지도하는 것이 학생들의 이해에 효과적일 것으로 판단된다.

연구 결과 학생들의 성취 점수는 성별에 따라 유의한 차이가 나타나지 않았으나 인지 진단이론으로 분석한 자료에서 남학생과 여학생의 숙달의 차이에 대한 정보를 제공할 수 있었다. 즉, 여학생이 남학생에 비해 회상하기 능력의 숙달이 유의하게 더 높은 것으로 나타났다. 회상하기는 문제를 풀 때 문제와 관련된 과학적 사실이나 관계 및 개념 등을 정확히 떠올리는 기능인데 여학생이 이 기능의 숙달 정도가 남학생에 비해 더 높음을 알 수 있다. 학교의 정규교사를 분석한 결과에서는 여학생이 회상하기의 능력의 숙달이

남학생에 비해 유의하게 더 높은 것으로 나타났으나 학교의 정규교사가 아닌 진단 평가, 형성 평가 등의 시험이나 국가수준 학업성취도 평가와 같은 대규모 표준화 검사에서도 이러한 차이가 나타나는지 알아볼 필요가 있겠다.

학생들의 과학 성취 수준별로 상, 중, 하로 구분하여 인지 능력의 숙달 여부를 비교한 결과 성취 수준이 높을수록 숙달한 인지 능력이 많은 것으로 나타났다. 그러나 개별 학생들의 점수와 인지 숙달을 비교한 결과(Table 12)에서 알 수 있듯이 성취 점수가 높다고 해서 반드시 인지 능력의 숙달 빈도가 높다고 할 수는 없다. 점수로 알 수 있는 학생들의 성취 수준 외에도 인지 숙달에 대한 정보는 인지 속성의 숙달 여부와 관련지어 학생들을 구분할 수 있는 정보를 제공해 줄 수 있다.

특히 본 연구의 결과에서 학생 개개인별로 과학적 속성의 숙달 여부를 판단할 수 있는 프로파일을 얻을 수 있었다. 학생들의 경우 인지적 속성에 대한 정보를 제공받을 수 있다면 자신에게 부족한 인지적 속성에 대한 정보를 얻을 수 있을 것이다. 교사의 경우도 프로파일 결과를 통해 학생들에 대하여 정보를 더 얻을 수 있고 그 정보를 바탕으로 수업과 평가를 연계함으로써 기존의 수업을 개선된 방향으로 발전시킬 수 있을 것이다. 평가를 실시하였을 때 그 결과가 전체 점수 이외의 다른 정보, 즉 학생들이 숙달한 속성에 대한 정보를 포함하는 것은 보다 의미 있는 평가 결과로 교육 현장에 피드백 될 수 있을 것이다.

본 연구의 결과를 바탕으로 몇 가지 제언을 해보고자 한다.

과학 수업에서 수준별 수업을 계획할 때 교사는 프로파일이 동일한 유형에 따라 학생들을 분류하여 지도할 수 있을 것으로 판단된다. 본 연구 결과 학생 개인에게 무엇이 강점이고 약점인지에 대해 알려줄 수 있는 프로파일을 얻었다. 이 프로파일은 학생 스스로 자신의 능력을 개선시키는 데 필요한 부분을 구체적으로 인식할 수 있게 해줄 뿐 아니라 교사가 학생들이 부족한 능력이 무엇인지 파악할 수 있다. 따라서 학생들의 프로파일별로 수준을 분리하여 그룹 또는 개별 지도를 시도할 수 있을 것이다. 기존에 이루어지고 있던 과학 내용 지식의 숙달 여부에 따른 보충교육 뿐 아니라 인지적 속성의 숙달 여부에 따른 보충이 이루어질 수 있을 것으로 판단된다.

인지 진단 이론이 현장에서 활용되기 위해서는 교사들에게 과학과 관련된 인지적 속성에 대한 정확한 정의 및 인지적 능력을 신장시키기 위한 방법에 대한 정보를 제공해 주는 것이 필요하겠다. 더 나아가서 다양한 인지적 능력을 측정할 수 있는 과학 문항들을 개발할 필요가 있겠다. 이러한 문항은 인지적 능력을 측정하기 위해 사용될 수 있을 뿐 아니라 그러한 인지적 능력을 기르기 위한 문항으로도 활용될 수 있을 것으로 판단된다.

학습자들의 인지 능력을 정확히 파악하기 위해서는 인지 속성의 바탕이 되는 Q-행렬을 명세화 하는 것이 중요하다. 학습자는 Q-행렬의 인지 속성을 분석하여 얻은 정보를 통해, 학습이 필요한 영역에 비중을 두는 법을 배우게 된다. 따라서 인지 속성과 문항 간의 관계를 설명하는 Q-행렬의 구성은 학습자들을 정확히 진단하는 데 있어서 매우 중요하다. Q-행렬을 명세화 하기 위해서 본 연구에 TIMSS의 평가틀을 기초로 사용한 것과 같이 기존 연구를 활용해서 문항에서 측정하고 있는 인지 속성을 추출하여 Q-행렬을 구성하여야 할 것이다. 또한 전문가들의 검토 작업을 거쳐 Q-행렬의 정교성을 확보하고 문항 모수의 수렴 여부와 문항모수 추정치 등을 확인하여 Q-행렬의 정확도 및 타당화를 확인할 필요가 있다. 그 외에도 Fusion Model이 아닌 De la Torre(2008) 등의 연구에서 사용한 다른 모형을 활용하여서도 Q-행렬의 정확도 및 타당화를 확인할 수 있겠다.

학생들이 평가결과를 통해 자신을 스스로 진단할 수 있도록 평가 결과가 구체적으로 제시되기 위한 연구가 지속적으로 이루어져야 하겠다. 본 연구에서는 객관식 문항을 통해 학생들의 인지적 속성을 5가지로 구분하여 인지적 속성의 숙달 여부에 대한 평가 결과를 제시할 수 있었다. 그러나 학생들이 과학 문제를 풀기 위해서는 더 다양한 인지적 속성이 필요하므로 다양한 인지적 속성들을 평가할 수 있는 문항을 통해 학생들에게 평가 결과를 제시할 수 있는 방안이 대해 계속적으로 연구가 진행될 필요가 있겠다. 또한 객관식 문항 뿐 아니라 서술형 문항을 통해서도 학생들의 속성을 파악하는 연구가 진행될 필요가 있겠다. 특히 과학에서 중요하게 여기고 있는 것의 하나로 과학 탐구 능력이 있는데 과학 탐구능력의 구성요소로는 문제인식 관찰, 측정, 분류, 예상, 변인통제, 결론도출, 일반화와 같은 탐구 능력을 측정할 수 있는 문항들을

통해 탐구 능력의 숙달여부와 관련된 결과를 제시할 수 있는 방안이 대한 연구도 함께 이루어져야 하겠다

## 국문 요약

본 연구는 인지 진단 이론을 활용하여 과학 시험 결과를 분석하는 방안을 모색하고자 하는 목적 하에 실시되었다. 중학교 2학년을 대상으로 한 시험 문항을 개발하여 평가를 실시하였고, Fusion Model을 이용하여 시험 결과를 분석하였다. 결과를 분석하기 위해 문항별로 학생들이 문항을 풀기 위해 숙달해야 할 속성을 판별하여 Q 행렬을 작성한 후 분석을 실시하여 속성에 대한 문항 모수의 수렴을 확인하고 학생 전체와 성별, 성취수준별로 숙달한 속성과 그 속성의 개수를 분석하였다. 전통적인 평가가 학생들의 성취도 총점만을 제공하는 반면 인지 진단 이론은 평가에서 측정된 과학적 속성의 숙달 여부에 대한 정보를 제공해주었다. 여학생은 남학생에 비해 회상하기의 능력이 유의하게 높은 것으로 나타났다. 분석결과 회상하기, 정보 해석하기, 계산하기 설명하기, 종합하기의 5개 속성 중에서 학생들이 가장 많이 숙달한 속성은 정보 해석하기였으며 가장 적게 숙달한 속성은 종합하기였다. 분석 결과 얻어진 프로파일은 교사들에게 학생 개인의 과학적 능력에 대한 정보를 구체적으로 제공해 줄 수 있으므로 이 결과는 학생들의 과학 학습에 대한 진단과 처방, 추후 학습 지도에 유용한 정보로 활용될 수 있을 것이다.

## 참고 문헌

- 교육인적자원부 (2007). 과학과 교육과정. 교육인적자원부 고시 제 2007 - 79호 [별책 9].
- 김성호, 송미영(1997). 범주형 자료의 인과모형에 의한 능력상태 추정과 교육적 활동. *교육평가연구*, 10(1), 183-212.
- 김성훈(2005). 인지구조모형에 근거한 학생의 지식상태 진단. *교육학연구*, 43(1), 81-107.
- 김성훈(2007). 집단수준의 점수해석 기준을 이용한 개인수준의 진단: ACT 수험평가 자료, 학습발달표준, 규칙장이론 활용. *교육평가연구*, 20(1), 165-194.
- 김성훈, 박영선(2004). 오개념 진단 및 교정을 위한 규칙장 모델의 알고리즘 개발. *교육평가연구*,

17(2), 183-195.

김수진, 송미영, 김선희 (2008). Fusion Model에 의한 수학 능력 진단을 위한 Q-행렬의 정교화. *교육평가연구*, 21(2), 115-139.

김명화(1997). 학습자의 인지 진단 방법에 관한 연구-분수의 인지과정모델에 Bayesian 추론망 모델 적용-. 고려대학교 박사학위 논문.

Chipman, S.F, Nichols, P., Brennan, R.L. (1995). Introduction, In P. D. Nichols, S. F. Chipman, and R. L. Brennan (Eds.), *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

DiBello, L., Stout, W., & Rousses, L. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, and R. L. Brennan (Eds.), *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

De la Torre, J.(2008). An Empirically Based Method of Q-matrix Validation for the DINA Model: Development and Application. *Journal of Educational Measurement*, 45(4), 343-362.

Dogan, E. & Tatsuoka, K. (2008). An international comparison using a diagnostic testing model: Turkish students' profile of mathematical skills on TIMSS-R. *Educational Studies in Mathematics*, 68(3), 263-272.

Embretson, S. (1990). Diagnostic testing by measuring learning processes: Psychometric considerations for dynamic testing. In N. Frederiksen, R. L. Glasser, A. M. Lesgold, and M. G. Shafto (Eds.), *Diagnostic monitoring of skills and knowledge acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hartz, S.(2002). A Bayesian framework for the Unified Model for assessing cognitive abilities: bledning theory wit practice. Doctoral thesis, The University of Illinois at Urbana-Champaign.

Hartz, S., Roussos, L., & Stout, W. (2002). *Skills Diagnosis: Theory and Practice*. User Manual for Arpeggio software. ETS.

McMillan, J. H. (2004). *Classroom assessment principle and practice for effective instruction*. Boston: Allyn & Bacon.

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A., & Erberber, E. (2005). *TIMSS 2007 Assessment Framework*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston: Boston College.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20(4), 345-354.

Tatsuoka, K. K. (1990). Toward integration of item response theory and cognitive error diagnoses. In N. Frederiksen, R. L. Glasser, A. M. Lesgold, and M. G. Shafto (Eds.), *Diagnostic monitoring of skills and knowledge acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tatsuoka, K. K. (1995). Architecture of knowledge structure and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman, and R. L. Brennan (Eds.), *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tatsuoka, K. K., Corter, J. E., & Tatsuoka, C. (2004). Patterns of Diagnosed Mathematical Content and Process Skills in TIMSS-R Across a Sample of 20 Countries. *Educational Research Journal*, 41(4), 901-926.