

DINA 모형을 활용한 대규모 학업성취도 평가 결과 분석¹⁾

김 성 훈(동국대학교, 교수)

송 미 영(한국교육과정평가원, 연구위원)*

《 요 약 》

우리나라를 비롯한 세계 여러 국가들은 학교 교육의 질을 관리하고 교육 환경 개선에 유용한 정보를 제공하기 위하여 국가수준 혹은 국제수준의 대규모 학업성취도 평가를 주기적으로 실시하고 있다. 주로 전국 규모의 학업성취도 평가 자료는 국가, 지역, 학교 차원의 교육효과 평가 및 교육관련 정책 수립의 기초 자료로 활용된다. 지금까지 대규모 학업성취도 평가 자료에서 학생 개인 또는 집단의 평균적인 학업성취 수준의 정보를 산출해 왔다. 이 연구는 학생들이 습득한 교과 지식의 구조적 정보를 구체적으로 제공하여 추후 교수학습에 유용한 처방적 정보를 제공하는 가능성을 탐색하는 것을 목적으로 한다. 인지진단이론 모형을 적용하여 학생의 지식 및 기능에 대한 숙달 상태를 진단한 결과, 교과 지식 구조에 대한 평가 결과 프로파일을 제공할 수 있었다. 학생의 능력에 대한 상세 진단 결과를 참고하여 학업 능력을 향상시키는 데 필요한 부분을 구체적으로 인식하고 교수학습의 방향을 정하고 정책적 지원을 전개해 나갈 수 있게 될 것이다. 인지상태의 진단을 위해서는 핵심적이고 결정적인 역할을 하는 Q-행렬의 타당성 검토와 확보를 위한 노력이 수반되어야 한다.

주제어 : 인지진단, 대규모 학업성취도 평가, Q-행렬, DINA 모형

1) 이 논문은 한국교육과정평가원의 연구비 지원으로 수행된 '대규모 학업성취도 평가 자료의 활용'(김성훈 등, 2009) 연구보고서의 일부 내용을 발췌하여 재구성한 것이다.

* 교신저자, mysong@kice.re.kr

I. 서론

학교 교육의 성과가 기대한 만큼 이루어지고 있는지, 더 나아가 그 성과가 어느 정도에 이르고 있는지를 점검하는 일은 교육평가의 중요한 역할 중 하나이다. 대규모 학업성취도 평가(large-scale assessment)는 일반적으로 국가의 경쟁력 증진을 위하여 교육의 책무성을 확인하는 차원에서 이루어진다. 우리나라도 학교 교육의 질을 관리하고 교육 환경 개선에 유용한 정보를 제공하기 위하여 국가수준 혹은 국제수준의 대규모 학업성취도 평가를 주기적으로 실시하고 있으며, 그 정도를 강화하고 있다.

대규모 학업성취도 평가 자료를 기반으로 교육효과 평가 및 정책 수립 등 여러 가지 교육적 의사결정을 하게 된다. 교육적 의사결정을 위한 기초 자료로서의 학업성취도 평가 자료는 학생 개인들의 평균적인 학업성취 수준 정보이다. 통상 검사 점수로 표현되는 그 정보는 교육 목표들이 어느 정도 달성되었는지를 요약하는 하나의 측정치이다. 그러므로 검사 점수는 교육 목표들에 내재하는 인지요소들이 어느 정도로 성취되었는가가 집적된 정보라고 할 수 있다. 달리 표현하면, 검사 점수는 학생들이 습득한 지식의 정도를 나타내지만, 그 정도 속에는 학생들이 습득한 교과 지식의 구조적 정보를 구체적으로 보여주지는 않는다. 그러므로 그 점수로만 석차나 서열 또는 자격 인증이나 등급 판정에 직접적 도움을 주지만, 무엇을 어떻게 교수·학습해야 할 것인지와 같은 처방적 판단에는 큰 도움을 주지 않는다. 학생들은 학습 결과를 점수에 의지하게 되고 높은 점수를 받기 위한 노력을 할 뿐, 어떤 개선점이 필요한지에 대해서는 알지 못한다. 학생들에게 전반적인 능력 수준을 나타내는 총합 점수를 제공하는 것보다 세부 정보를 제공한다면 강약점을 알 수 있으며 더 효과적인 학습을 계획할 수 있게 할 것이다. 학생의 학습과 성장을 돕고 교사의 교수 활동과 수업 방법을 개선하는 데 평가 결과가 활용되기 위해서는 검사내 영역별 성취에 대한 구체적인 정보를 제공하여야 한다.

학업성취도 평가 결과의 활용 측면에서 중요성을 인식하고 평가 결과의 제한된 정보를 보완할 필요성에 의해 새로운 접근의 측정이론으로 인지진단이론(cognitive diagnosis theory)이 대두되었다. 1970-80년대에 인지과학에서는 측정·평가 분야에서 학생의 지식과 기능에 대한 이해를 발전시키기 위한 이론적 토대를 만들었고, 1980년대 이후 인지심리학과 컴퓨터 기술의 발달에 힘입어 학생의 내적 인지상태를 진단하기 위한 노력이 활발히 전개되었다(Tatsuoka, 1983, 1985; Frederiksen et al. 1990; Chipman, & Nichols, 1995; Nichols et al, 1995; Embretson, 1990, 1993, 1995; Mislevy, 1993, 1995, 1996, 2006; de la Torre & Douglas, 2004). 그러한 노력은 주로 학생 개인의 진단과 후속 학습이나 지도를 목적으로 하지만, 아직 실제적이고 체계적인 활용에 이르지 못하고 있다. 대규모 학업성취도 평가 자

료를 가지고 인지심리학적 진단의 가능성을 타진한 연구는 있지만(Tatsuoka et al., 2004; Dogan & Tatsuoka, 2008; 김성훈, 2005), 교육상황에 접목하여 실제 활용으로 이어지거나 실제 활용 체계를 개발한 연구는 발견하기 힘들다.

대규모 학업성취도 평가 결과는 타당하고 신뢰로운 양질의 정보를 가진다고 할 수 있다. 그러한 정보의 가치는 점수에만 해당하는 것이 아니라, 그 점수에 내포된 인지상태(knowledge state)에도 해당한다. 인지상태란, 학생이 특정 시점에 형성한 교과 관련 지식의 구조를 의미하는 바로서, 문항 풀이에 요구되는 인지요소(attributes)들 간의 관계를 의미한다. 인지요소란 둘 이상의 문항에 공통적으로 요구되는 능력 요인을 의미하며, 한 검사는 고유의 인지요소 세트를 가지고 있는 것으로 가정한다. 그러므로 한 학생의 인지상태를 진단한다고 함은 그 학생이 (해당 검사에서 측정할 수 있는 모든 인지요소들 중에서) 성취한 인지요소와 성취하지 못한 인지요소들이 무엇인지를 측정함을 의미한다. 만약 개별 학생의 인지상태가 확인되면, 한 학급이나 학교에 속한 학생들의 평균적인 인지상태를 확인할 수도 있다는 논리가 성립한다. 그렇게 된다면, 그 평가 정보는 보다 본질적인 교육 활동에 직접적으로 도움이 될 수 있다. 타당하고 신뢰로운 대규모 학업성취도 평가 자료로부터 학생 개인이나 집단의 지식의 구조에 관한 정보를 찾아낸다면, 교수·학습의 직접적 안내의 질 또한 그만큼 높아진다.

우리나라의 경우 학생들은 초등학교부터 시작하여 고등학교를 졸업할 때까지 수차례 전국 규모의 학업성취도 평가를 경험한다. 전국 규모의 학업성취도 평가는 교육과정에 명시된 목표가 어느 정도로 성취되는지를 평가하는 것이라는 점에서 교육과정 목표를 중심으로 한 개인의 인지 상태에 대한 정보를 담고 있을 뿐만 아니라 후속적 학습과 관련된 판단을 위한 정보의 산출 가능성도 담고 있다. 이 연구는 국가수준 학업성취도 평가 자료에 인지진단이론을 적용함으로써 인지상태를 진단하기 위한 논리적 작업을 거쳐 대규모 학업성취도 평가 자료 분석 및 결과 보고 가능성을 탐색해 보고자 한다.

‘인지상태를 진단하기 위한 논리적 작업’은 문항반응자료로부터 진단해내야 할 인지요소를 추출하고, 그 인지요소들이 어떤 문항들에 요구되는지를 규정하는 것과 관련 있다. 각 인지요소와 문항들 간의 관계는 Q-행렬(Q-matrix)로 나타낼 것이며, 이러한 작업은 가설적이기 때문에 학생들의 인지상태 진단을 위한 이론모형을 개발하는 것으로 이해될 수 있다. ‘인지상태 진단을 위한 자료 분석 및 결과보고’는 어떤 인지심리학적 측정모형을 적용하여 어떤 형태로 보고하고 활용할 것인지를 묻는 것이다. 이 연구에서는 인지심리학적 측정 모형들 중의 하나인 DINA 모형(de la Torre, 2008; 2009)을 적용하여, 학생들이 지닌 인지상태를 진단한다. 그 결과는 개인 수준 및 집단 수준의 인지상태를 보고하는 예와 집단적 인지상태를 비교하는 예를 보여줄 것이다.

II. 인지진단이론

학생들의 학력을 보다 더 향상시키려고 하는 노력과 지적 과정을 이해하려는 요구가 증가하면서 인지진단이론의 현장 적용은 더 중요해지고 있다. 미국의 PSAT/NMSQT™(Preliminary Scholastic Aptitude Test / National Merit Scholarship Qualifying Test)를 예로 들면, 인지진단모형을 통해 'Score Report Plus™'라는 성적표를 제공하여 보다 효과적으로 학생들의 성적을 보고하고 있다. TOEFL(Test of English as a Foreign Language)에서도 인지진단모형을 통해 학생들의 프로파일 성적표를 제공하려고 시도하였다. 또한 국제 학업성취도 평가인 TIMSS(Trends in International Mathematics and Science Study)에서는 인지진단이론의 하나인 규칙장 이론(Rule Space Methodology)을 사용하여 각국 학생들의 상황을 파악하였다(김수진, 2009).

인지진단이론(Cognitive Diagnosis Theory)은 학생의 지식 및 기능에 대한 이해 상태를 상세하게 파악하여 학생의 학습을 돕고자 개발된 측정 이론으로 교과 관련 지식 구조에 대한 평가 결과를 알려주어 교사와 학생에게 적절한 피드백을 제공한다는 장점이 있다. 인지진단모형에 의하면 학생이 습득한 인지요소를 정확하게 진단하고 학생의 학습 발전 및 진전 상황을 정확하게 추정할 수 있다(Embretson, 1990; DiBello, Stout, & Rousses, 1995; Tatsuoka, 1995). 인지요소(attribute)는 특정 문항의 정답을 유추하기 위해 요구되는 학생의 능력이나 인지 과정(cognitive process), 지식(knowledge), 기능(skills) 등을 의미한다. 예를 들면, 수학에 있어서의 덧셈이나 뺄셈과 같은 계산능력, 국어에서 이해하기, 추론하기 등의 인지 과정, 영어에서의 쓰기, 말하기 등을 인지요소라고 할 수 있다. 이러한 인지요소들은 Q-행렬을 사용하여 설명된다.

1. Q-행렬

인지상태를 진단하기 위해서는, 한 검사의 문항과 그 검사가 채는 인지요소의 관계를 미리 구축해야 한다. 인지진단이론에 있어서 문항과 인지요소의 관계는 Q-행렬을 통해 구조화한다. Q-행렬(Embretson, 1984; Tatsuoka, 1983, 1990, 1995; Junker & Sijtsma, 2001)의 원소 q_{jk} 는 문항 j 를 푸는 데 인지요소 k 가 필요하면 1, 그렇지 않으면 0의 값을 갖는다. 문항이 J 개, 인지요소의 수가 K 개이면 Q-행렬은 0 또는 1의 값으로 구성된 $J \times K$ 행렬로 표시된다. 예를 들어 [그림 1]의 Q-행렬에서 문항 1의 풀이에 인지요소 1, 3, 4가 필요하며(Q-행렬의 첫째 행은 (1, 0, 1, 1)로 표현됨), 문항 2는 인지요소 4만, 문항 3은 4개의 인지요소가

모두 요구되며, 문항 4와 문항 5의 풀이에는 인지요소 1, 2가 필요한 것을 나타낸다. 이처럼 Q-행렬은 각 문항에 정답하기 위해서는 어떤 인지요소가 필요한지를 보여준다. 그러므로 Q-행렬의 작성은 각 문항이 진단하고자 하는 인지요소 설계에 해당하며, 인지진단이론의 모형에서 매우 중요한 요소이다.

	인지요소1	인지요소2	인지요소3	인지요소4
문항1	1	0	1	1
문항2	0	0	0	1
문항3	1	1	1	1
문항4	1	1	0	0
문항5	1	1	0	0

[그림 1] 문항과 인지요소 간의 관계를 나타낸 Q-행렬의 예

검사 문항과 인지요소가 [그림 1]의 Q-행렬과 같은 관계를 갖고 있는 경우에 가능한 문항 반응패턴의 예로서 <표 1>를 예상해 볼 수 있다. <표 1>의 문항반응패턴에서 문항의 답을 맞히면 1, 틀리면 0으로 나타낸다. 인지요소의 숙달 여부에 따라 문항반응패턴이 달리 나타날 것이므로, 실제 인지진단 모형을 적용하여 평가 자료를 분석할 때는 문항반응패턴을 가지고 인지요소의 성취여부를 진단해 내게 된다.

<표 1> 인지요소의 숙달 여부에 따라 예상되는 문항반응패턴

학생	숙달한 인지요소				문항반응패턴				
	요소1	요소2	요소3	요소4	문항1	문항2	문항3	문항4	문항5
가	0	0	0	0	0	0	0	0	0
나	1	1	0	0	0	0	0	1	1
다	1	0	1	1	1	1	0	0	0
라	1	1	1	1	1	1	1	1	1

인지상태는 한 검사에서 재고자 하는 모든 인지요소의 숙달 여부로 표현된다. 즉, 한 검사에서 재는 인지요소가 K개라고 할 때, 한 학생의 인지상태는 K개의 1(숙달) 또는 0(미숙달)의 값을 가지는 벡터($\alpha_i = \{\alpha_{ik}\}$, $k = 1, 2, \dots, K$)로 표현된다. 예를 들어 한 검사에서 재고자 하는 인지요소가 4개라고 하고, 한 학생 i(<표 1>의 학생 나)의 인지상태가 $\alpha_i = \{1, 1, 0, 0\}$ 로 진단되었다면, 그 학생은 처음의 두 인지요소 1, 2는 숙달한 것이고, 인지요소 3, 4는 숙

달하지 못한 것으로 이해된다. 그러므로 후속적 교수학습 계획은 인지요소 3, 4를 어떻게 숙달할 것인가에 초점이 모아진다.

2. DINA 모형

DINA(Deterministic Input, Noisy, "And" gate) 모형은 2000년대에 와서 de la Torre 등(de la Torre & Douglas, 2004; de la Torre, 2008, 2009)에 의해서 활발히 연구되고 있는 인지진단이론 모형 중 하나이다. 인지진단이론의 모형은 크게 다음과 같이 세 가지로 구분된다. 첫째는 검사 문항 속성과 측정하는 속성을 통계적으로 평가하여 문항 모수를 추정하고 검사 문항을 평가하는 접근 방식이고, 둘째는 검사 문항에 대한 정보 없이 학생 모수를 추정하여 학생의 능력 상태를 평가하는 접근 방식이며, 셋째는 검사 문항과 학생에 관한 정보를 모두 사용하여 평가하는 접근 방식으로, 처음 두 가지 접근 방식이 문항이나 학생 각각에 대하여 평가할 수는 있지만 학생과 문항이 측정하고 있는 인지요소들의 연관성을 찾을 수는 없다는 문제점을 극복하고자 제안된 모형들이 속한다. DINA 모형은 세 번째 모형에 속한다.

가. 모형의 구조

DINA 모형은 학생의 문항에 대한 정답 여부는 이론적으로 결정할 수 있는 부분과 확률적으로 추정되어야 하는 부분들을 동시에 고려해야 설명할 수 있음을 가정한다. 특정 문항에 요구되는 인지요소를 빠짐없이(conjunctive) 구비한 학생은 정답할 것으로 기대할 수 있지만, 요구되는 인지요소들 중 하나라도 구비하지 못하면 오답할 것으로 기대할 수 있다고 가정한다. 이러한 가정은 학생의 인지상태와 이론적으로 규정한 Q-행렬에 의해서 학생의 잠재반응(latent response)이 결정되는 과정으로 개념화된다. 그러나 학생의 실제 반응은 잡음요인(noise)으로 인하여 기대 또는 잠재반응과 다른 결과로 나타날 수 있다고 가정한다. 즉, 정답할 것으로 기대되는 학생이 오답하면 슬립(slip)한 것으로, 오답할 것으로 기대되는 학생이 정답하면 추측(guessing)한 것으로 본다.

학생(i)의 잠재반응(η_i)은 인지상태(α_i)에 의해서 도달하게 될 (비가시적) 문항반응이다. 인지상태는 인지요소의 유무에 대한 패턴이며, 잠재반응은 인지상태가 만들어내는 문항반응이다. 각 인지요소의 유무는 독립적이라는 가정 아래, 학생(i)의 잠재반응은 식(1)과 같이 인지상태 α_i 와 주어진 Q-행렬에 의해서 결정된다.

$$\eta_i = \eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}}, \quad j = 1, 2, \dots, J. \quad (1)$$

식 (1)은 잡음이 없는 이상적인 상태에서 특정 개인의 특정 문항에 대한 반응은 맞거나 틀리는 형태로 기대할 수 있음을 보여준다. [그림 1]의 1번 문항($q_1 = (1, 0, 1, 1)$)에 대한 잠재반응의 예를 보자. 그 문항에 대해서 인지요소 1, 2, 4를 갖춘 학생($\alpha_i = (1, 1, 0, 1)$)이나 인지요소 2, 3, 4를 갖춘 학생($\alpha_i = (0, 1, 1, 1)$)은 틀릴 것으로 기대하지만(식 1a), 인지요소 1, 3, 4를 갖춘 학생($\alpha_i = (1, 0, 1, 1)$)이나 모든 인지요소를 갖춘 학생($\alpha_i = (1, 1, 1, 1)$)은 맞힐 것으로 기대할 수 있다(식 1b).

$$\eta_{i1} = 1^1 \times 1^0 \times 0^1 \times 1^1 = 0, \quad \eta_{i'1} = 0^1 \times 1^0 \times 1^1 \times 1^1 = 0 \quad (1a)$$

$$\eta_{i''1} = 1^1 \times 0^0 \times 1^1 \times 1^1 = 1, \quad \eta_{i'''1} = 1^1 \times 1^0 \times 1^1 \times 1^1 = 1 \quad (1b)$$

식 (1)이 의미하는 바는, 학생(i)이 어떤 문항(j)에 정답($\eta_{ij} = 1$)하기 위해서는 그 문항을 해결하는 데 필요한 모든 인지요소를 빠짐없이 갖추고 있어야 하며, 그 인지요소 중 하나라도 갖추지 못하면 오답을 할 수밖에 없다는 것이다. 만약 학생이 슬립이나 추측 없이 Q-행렬에 규정된 인지요소를 틀림없이 적용한다면, 실제 반응은 식(1)에 의해서 100% 설명될 것이다. 그러나 인지 과정은 본래 확률적이기 때문에 잠재반응은 이상적으로 기대할 수 있는 반응일 뿐이다(de la Torre, 2008). 실제 학생들은 Q-행렬로 규정한 인지요소를 다 갖추고 있어서 문항의 답을 맞힐 수 있어야 함에도 틀리는 경우나 필요한 인지요소 중 하나 이상의 인지요소를 갖추지 못하여 틀려야 함에도 맞히는 경우가 있다. 전자는 슬립(slip: s_j ; 식 (2))으로, 후자는 추측(guessing: g_j ; 식 (3))으로 개념화된다.

$$s_j = P(X_{ij} = 0 | \eta_{ij} = 1) \quad (2)$$

$$g_j = P(X_{ij} = 1 | \eta_{ij} = 0) \quad (3)$$

한 학생(i)의 인지상태(α_i), 문항 j의 슬립 확률(s_j), 추측 확률(g_j)을 동시에 고려할 때, 문항(j)을 맞힐 확률은 식(4)와 같다.

$$P_j(\alpha_i) = P(X_{ij} = 1 | \alpha_i) = g_j^{1-\eta_{ij}} (1-s_j)^{\eta_{ij}} \quad (4)$$

식 (4)는, 한 학생의 문항 반응은 인지상태와 Q-행렬에 의해서 결정되는 잠재반응(η_{ij})이 일차적으로 결정되고, 다음 단계에서 잠재반응과 추측모수, 슬립모수의 함수적 관계에 의해서 확률적으로 결정됨을 보여준다. 문항에 필요한 인지요소를 빠짐없이 다 갖추어서 잠재반

응이 정답($\eta_{ij} = 1$)으로 결정되면, 실제 정답률은 슬립을 범하지 않으면($s_j = 0, p_j = 1.0$)이지만, 슬립을 범하면 정답률은 그만큼 줄어든다($p_j = 1 - s_j$). 반면 필요한 인지요소 중 하나라도 갖추지 못하여 잠재반응이 오답($\eta_{ij} = 0$)으로 결정되면, 실제 정답률은 추측에 의한 확률(g_j)이다.

이상과 같이 학생의 실제 반응을 잠재반응, 슬립과 추측으로 설명하고자 하는 DINA 모형의 특징은 그 명칭 속에 표현되어 있다. 첫째, 한 학생이 한 문항을 풀 때 기대되는 잠재반응(실제 반응이 아니라 수리적 모형에 의해 기대되는 맞고 틀림의 반응)이 확률적이 아니라 결정적으로 정해지고(Deterministic Input); 둘째, 실제 반응은 각 문항이 갖는 슬립과 추측에 의해서 잠재반응과는 달리 확률적으로 나타나며(Noisy); 셋째, 각 인지요소가 서로 상보적(compensatory)이어서 어느 요소가 결핍된 다른 요소를 보충할 수 있는 것이 아니라 모든 인지요소가 빠짐없이 존재해야만, 즉 결합적(conjunctive)이어야만 정답이 가능하다("And" gate)고 본다.

나. 모수의 추정

DINA 모형에서 한 학생의 실제 반응은 인지상태, Q-행렬, 문항의 슬립 확률과 추측 확률로 결정된다. 이 중에서 Q-행렬은 검사 문항의 내용분석을 통하여 규정되는 것으로서, 자료 분석을 통하여 추정되는 것이 아니라 다른 모수 추정을 위하여 투입(input)되며, 모수 추정의 타당성을 위한 전제조건이 되는 셈이다. 실제 자료 분석을 통하여 추정되어야 하는 것은 각 학생의 인지상태 모수와 각 문항의 슬립 모수와 추측 모수이다.

일반적으로 주변우도(marginal likelihood; 식 (5))가 최대가 되도록 하는 방법을 사용하여 모수를 추정한다.

$$L(X) = \prod_{i=1}^I L(X_i) = \prod_{i=1}^I \sum_{l=1}^L L(X_i | \alpha_l) P(\alpha_l) \quad (5)$$

식 (7)에서 $L(X_i)$ 은 학생 i 의 응답벡터의 주변우도이고, $P(\alpha_l)$ 은 인지상태 α_l 벡터의 사전확률이며, $L = 2^k$ 이다. 인지상태 진단을 전제하는 한 개인의 실제 문항반응은 인지상태가 주어졌을 때 문항에 정답할 조건확률 $L(X_i | \alpha_l)$ 과 인지상태의 결합확률 $P(\alpha_l)$ 에 의해서 결정된다. 주변최대우도에 근거한 모수는 EM 알고리즘을 적용하여 추정한다. 한 학생의 인지상태(α_l)는 비연속적 범주이므로, EM 알고리즘을 적용할 때 $2^k - 1$ 번씩 사후확률 갱신 과정을 거쳐야 한다. 인지요소의 수가 많을 경우에는 주변최대우도 추정법이 컴퓨터 기억 용량의 한계를 넘어서는 제한점이 있다.

컴퓨터 기억 용량의 문제를 극복하기 위하여 de la Torre & Douglas(2004)는 특정 인지상태와 일반능력(θ)의 관계를 식 (6)과 같이 가정하였다.

$$P(\alpha|\theta) = \prod_{k=1}^K P(\alpha_k|\theta) = \prod_{k=1}^K \frac{\exp(\lambda_{0k} + \lambda_1\theta)}{1 + \exp(\lambda_{0k} + \lambda_1\theta)}, \quad (\theta \sim N(0,1), \lambda_1 > 0) \quad (6)$$

식 (6)은 문항반응이론의 1모수 로지스틱 모형 형태로서, 일반능력 θ 가 향상할수록 인지요소를 갖출 확률($P(\alpha = 1|\theta)$)은 1에 근접함을 의미한다. 일반능력(θ)은 인지요소를 포괄하는 상위의 능력이라는 점에서 위의 모형이 추가된 DINA 모형을 상위-DINA(HO-DINA) 모형으로 불린다. 이와 같이 상위 모형을 가정함으로써 $2^k - 1$ 개의 모수 대신 $K+1$ 개, 즉, K 개의 위치(λ_0)와 1개의 기울기(λ_1) 모수를 추정하면 된다. 인지요소의 수가 늘어나더라도 추정에 소요되는 절차와 시간은 기하급수적으로 증가하는 것이 아니라 선형적으로 증가한다. de la Torre & Douglas(2004)는 HO-DINA 모형을 MCMC 알고리즘으로 추정할 때, 모수들은 안정적으로 추정됨을 보여주었다.

III. 연구방법

1. 분석 자료

이 연구를 위하여 2007년 10월에 실시된 국가수준 학업성취도 평가의 중학교 3학년 수학 검사 A형²⁾ 중 30개의 선다형 문항을 선택하였다. 이 검사는 선다형 문항 30개와 구성형 문항(constructed-response item) 6개가 혼합된 검사이다. 혼합형 검사는 문항 유형간 특성의 차이로 인한 유형 효과(format effect)가 발생할 수 있고, 각 문항 유형에서 요구되는 인지행동이 차별화될 수 있기 때문에(노국향, 김신영, 2000; Birnbaum & Feldman, 1998; Thissen, Wainer, & Wang, 1994) 이 연구에서는 문항 유형의 차이를 고려하여 선다형 문항만을 선택하였다. 국가수준 학업성취도 평가의 문항은 여러 차례 수학 교사와 수학 교과전문가들이 검토하여 개발된 것으로 내용 타당도와 신뢰도를 갖추고 있다고 하겠다.

2007년 국가수준 학업성취도 평가는 시·도교육청, 도시화 정도, 학교 규모 등을 고려한 2단

2) 2007년 국가수준 학업성취도 평가는 학업성취도의 변화추이 분석을 위한 검사점수 동등화의 목적으로 교과별 검사도구 2종(A형, B형)을 개발·시행하였고, B형 검사의 문항은 공개하지 않고 있다.

계 층화군집표집설계(A two-stage stratified cluster sample design)에 의해 전국에서 일부 학생을 평가대상으로 선정하여 시행되었다. 이 연구에서는 중학교 3학년생들의 3%(295개 학교, 약 20,000명) 중 A형 검사를 치른 약 15,000명에서 지역을 고려하여 표집된 5,185명의 선다형 문항반응자료를 분석하였다. 분석에 포함된 학생 중 남학생은 2,581명(49.8%), 여학생 2,604명(50.2%)이었고, 성취수준³⁾에 따라 구분하면 우수학력은 1160명(22.4%), 보통학력은 1,868명(36.0%), 기초학력은 1,929명(37.2%), 기초학력 미달은 228명(4.4%)이었다(<표 2> 참조).

〈표 2〉 분석 대상

성취수준 \ 성별	남학생 (2,581명)		여학생 (2,604명)		전체 (5,185명)	
	빈도	비율	빈도	비율	빈도	비율
우수학력	631	24.4%	529	20.3%	1160	22.4%
보통학력	916	35.5%	952	36.6%	1868	36.0%
기초학력	921	35.7%	1008	38.7%	1929	37.2%
기초학력 미달	113	4.4%	115	4.4%	228	4.4%

2. Q-행렬의 개발

검사 도구에 포함된 각 문항과 그 문항을 풀기 위해 요구되는 인지요소를 연결하는 Q-행렬을 작성하기 위하여, 먼저 문항을 해결하는 데 필요하면서 수학적으로 의미있는 인지 행동을 이 연구에서 분석할 인지요소로 추출하였다. 이를 위하여 국가수준 학업성취도 평가 수학과 평가틀, 대학수학능력시험 수리탐구영역의 평가틀에 대한 분석 및 문항 내용에 대한 분석이 이루어졌다.

국가수준 학업성취도 평가는 교육과정에서 규정하고 있는 교육목표를 어느 정도 도달하였는지 분석하는 데 그 목적이 있는 만큼 교육과정의 내용을 반영할 수 있도록 제7차 수학과 교육과정의 내용영역 구분을 사용하고 있다. 중학교 3학년 수학 검사는 중학교 수학 교육과정의 전체 내용영역, 즉 수와 연산, 도형, 측정, 확률과 통계, 문자와 식, 규칙성과 함수 영역에 대하여 측정한다. 국가수준 학업성취도 평가 수학과에서는 행동영역의 평가틀을 계산, 이해, 문제해결, 추론, 의사소통의 다섯 가지로 구분하고 있다. 이에 대한 정의는 대학수학능력시험 수리탐구영역의 평가틀과 대체로 유사하나 다소 차이를 보이기도 하여 이 연구에서는 각

3) 국가수준 학업성취도 평가에서는 개별 학생의 평가 결과를 우수학력, 보통학력, 기초학력, 기초학력 미달의 네 수준으로 구분하여 준거참조평가 방식으로 보고한다. 이 성취수준의 구분은 선다형 문항과 서답형 문항에 대한 응답 결과에 의해 이루어진 것이다.

인지요소에 대한 정의를 다음과 같이 재개념화 하였다. 선다형 문항만이 이 연구의 분석대상 이므로 의사소통능력을 제외한 계산하기, 이해하기, 추론, 문제해결을 적합한 행동영역 인지요 소로 추출하였다. 인지요소의 추출은 국가수준 학업성취도 평가 연구 수학과 또는 대학수학능 력시험 수리탐구영역의 출제 및 분석에 참여한 경험이 있는 수학교육전문가 3명에 의해 이루 어졌다. 이들은 수학교육 박사학위 소지자이면서 수년간 교사로 활동한 경험도 있다.

- 계산: 실수 범위에서 연산의 기본 법칙이나 성질을 적용하여 값을 구하는 능력, 다항식의 덧셈, 뺄셈, 곱셈을 적용하여 식을 간단히 하거나 인수분해하는 능력, 등식의 변형, 간단한 방정 식이나 부등식의 해를 구하는 능력, (수, 대수식 이외에) 알고리즘을 적용하여 해를 구하는 능력
- 이해: 문제에 나타난 수학적 개념, 원리, 법칙 및 그 관련성을 이해하는 능력, 문제에 주어진 기 호, 식, 도식(그림), 그래프, 표의 의미를 해석하는 능력
- 추론: 나열하기, 세어보기, 관찰 등을 통해 귀납적 원리를 발견하는 능력, 주어진 명제의 진위를 판별하는 능력, 증명이나 논리적인 과정을 읽고 이해하는 능력, 수학적 명제를 증명할 수 있는 능력
- 문제해결: 주어진 문제 상황을 수학적으로 표현하는 능력, 문제에 필요한 수학적 개념, 원리, 법칙 을 선택하는 능력, 문제해결을 위한 내용이나 방법, 전략을 선택하고 활용하는 능력, 문제 상황에 맞는 답을 결정하는 능력

검사 문항의 풀이에 요구되는 인지요소를 정한 다음, 문항과 문항이 재고자 하는 인지요 소 사이의 관계를 나타내는 Q-행렬을 작성하였다. 여기서 인지요소는 내용영역, 행동영역 모두에 해당된다. Q-행렬은 인지요소 추출에 참여한 수학교육전문가 1명과 수학교사 4명에 의해 작성되었다. 먼저 행동영역 인지요소에 대한 설명 및 검토 후 각자 문항 하나하나와 인지요소를 연결하였다. 전문가마다 작성한 Q-행렬이 서로 일치하는지 비교하면서 논의를 통한 합의과정을 거쳐 <표 3>과 같은 Q-행렬을 완성하였다. 여기에서 빈 칸의 값은 0이지 만 해독을 쉽게 하기 위하여 표시하지 않았다.

3. 인지상태의 진단

인지요소를 추출하고 이를 각 문항과 연결하여 Q-행렬을 작성한 다음, II장에서 살펴본 DINA 모형에 의하여 학생들의 인지상태를 진단하였다. DINA 모형의 모수는 주변최대우도 방식으로 EM 알고리즘에 의해 추정하였다. 각 문항에서 요구하는 인지요소를 숙달했을 때 와 그렇지 못했을 때의 정답 확률, 학생 개인별 각 인지요소를 숙달할 확률에 따라 판정된 숙달여부를 파악하여 프로파일을 분석하고, 학생 집단에 대한 인지요소별 숙달 비율과 숙달

한 인지요소의 수를 분석하였다.

〈표 3〉 인지진단 분석을 위한 6개 내용영역과 4개 행동영역으로 구성된 Q-행렬

문 항 번 호	내용영역						행동영역			
	수와 연산	문자 와식	규칙 성과 함수	확률 과 통계	도형	측정	계산	이해	추론	문제 해결
1		1					1			
2	1							1		
3						1				1
4		1					1			1
5				1				1		1
6		1								1
7				1				1		
8					1			1		
9				1				1		
10					1			1		
11	1							1		
12			1					1		
13					1			1		
14		1					1			
15	1				1			1	1	
16		1					1			1
17			1					1		1
18	1						1	1		
19					1			1		1
20			1					1		
21				1			1			
22		1				1	1	1		1
23					1			1		
24	1							1		
25		1					1			
26		1					1			1
27				1				1		1
28						1		1		
29					1			1		1
30			1					1		

IV. DINA 모형을 적용한 분석 결과

이 장에서는 DINA 모형을 적용하여 Q-행렬에 정의된 인지요소의 숙달여부에 따른 정답 확률, 즉 문항에 대한 모수를 추정한 결과⁴⁾를 바탕으로 산출한 학생 개인별 인지요소 평가 결과, 그리고 학생 집단에 대한 인지요소 평가 결과를 중심으로 기술한다.

1. 개별 학생의 인지요소에 대한 프로파일

개별 학생이 숙달한 인지요소에 대한 정보를 담고 있는 평가 결과를 제공할 수 있는 프로파일을 작성하였다. 인지진단모형에 의해 평가 자료를 분석하면 학생 개인에게는 점수 또는 등급 이외에 <표 4>와 같은 형태의 프로파일이 제시된다. 이러한 프로파일은 각각의 인지요소에 대해 학생이 숙달했는지 부족한지를 나타내는 것으로, 이를 통해 상세한 피드백이 가능하다. 기존의 평가 결과 보고에서는 각 학생의 종합적인 점수 또는 등급만을 보고한 반면, <표 4>에서는 평가하고자 하는 인지요소별 평가 결과(숙달 여부; 1=숙달, 0=미숙달)를 보고할 수 있음을 보여준다.

<표 4> 인지요소 프로파일의 예

학생	원점수	문항반응패턴	인지요소 상태	
			내용영역	행동영역
갑	15	10100011110111111100000000110	0 0 0 0 1 1	0 0 1 0
을	20	110101110111101111110111000001	1 1 1 0 1 0	0 1 1 0
병	20	111101111110110110000011010111	1 1 0 0 1 1	1 1 0 1
정	25	10110111111110111111110011111	1 1 1 1 1 1	1 1 0 1

또한 <표 4>를 살펴보면 동일한 점수(20점)라도 숙달한 인지요소에 대한 프로파일이 다를 수 있음을 볼 수 있다. 이는 정답한 문항 수가 같더라도 문항반응패턴이 학생마다 다르기 때문에 프로파일 또한 다른 것이다. 이렇게 학생들에게 각각의 프로파일을 제공한다면 점수만 보고하는 것보다 많은 교육적 정보를 담게 되며, 학생들 각자 어떤 능력이 숙달되어

4) 컴퓨터 처리 용량의 한계로 인해 내용영역 Q-행렬에 의한 진단과 행동영역 Q-행렬에 의한 진단을 별도로 분석하였다.

있고 부족한지에 대한 상세한 정보를 얻을 수 있을 것이며, 프로파일의 유형에 따라 학생들이 심화학습 또는 보충학습 해야 할 방향을 설정할 수 있을 것이다.

<표 4>와 같은 프로파일에 따라 교사는 학생들에게 추후 학습 방향에 대한 피드백을 줄 수 있을 것이다. 예를 들어 학생 ‘을’은 내용영역 프로파일이 111010이므로 4째 인지요소인 확률과 통계, 6째 인지요소인 측정 영역에 대한 개념, 원리 등의 지식적 요소의 이해를 높이고, 행동영역 프로파일이 0110, 즉 계산과 문제해결을 숙달하지 못했으므로 계산을 능숙하게 하는 연습과 실생활 문제에서 수학적 식을 끌어내는 능력을 키워야 함을 교사가 제안해줄 수 있을 것이다. 이렇게 학생들에게 각각의 프로파일을 제공한다면 종합적인 점수만 보고하는 것보다 많은 교육적 정보를 담게 될 것이다.

2. 학생 집단의 인지요소 숙달 정도

학생 전체가 어떤 인지요소를 어느 정도 숙달했는지 살펴보고, 학생들의 성취수준별로 숙달한 인지요소를 정리하였다. 이 연구의 분석 대상 중학교 3학년 학생 5,185명의 숙달한 인지요소는 <표 5>와 같다. 중학교 3학년 학생들이 수학적 인지요소를 숙달한 비율은, 내용영역의 경우 47.7%에서 61.7%, 행동영역의 경우 48.2%에서 53.8%로 나타났다. 내용영역은 규칙성과 함수 요소, 행동영역은 문제해결 요소를 가장 덜 숙달한 것으로 분석되었다.

<표 5> 내용영역과 행동영역 인지요소의 숙달 정도

내용영역	빈도	비율(%)	행동영역	빈도	비율(%)
수와 연산	2606	50.3	계산	2787	53.8
문자와 식	2627	50.7	이해	2759	53.2
규칙성과 함수	2474	47.7	추론	2742	52.9
확률과 통계	2743	52.9	문제해결	2500	48.2
도형	3198	61.7			
측정	3073	59.3			

주) 비율은 전체 학생 중 해당 영역을 숙달한 학생의 백분율임.

학생들의 성취수준별로 숙달한 수학적 인지요소가 어떻게 다른지 살펴보았다. 국가수준 학업성취도 평가에서 사용하는 4등급의 성취수준, 우수학력, 보통학력, 기초학력, 기초학력 미달별로 학생들이 인지요소를 숙달한 비율은 <표 6>과 같고, 학생들의 성취수준별로 숙달한 인지요소의 수는 <표 7>과 같다.

〈표 6〉 성취수준별 내용영역과 행동영역 인지요소의 숙달 정도

구분	인지요소	우수학력 (N=1160)		보통학력 (N=1868)		기초학력 (N=1929)		기초학력미달 (N=228)	
		빈도	비율 (%)	빈도	비율 (%)	빈도	비율 (%)	빈도	비율 (%)
내용영역	수와 연산	1160	100.0	1354	72.5	92	4.8	0	0.0
	문자와 식	1160	100.0	1400	74.9	67	3.5	0	0.0
	규칙성과 함수	1158	99.8	1266	67.8	50	2.6	0	0.0
	확률과 통계	1160	100.0	1448	77.5	135	7.0	0	0.0
	도형	1160	100.0	1675	89.7	362	18.8	1	0.4
	측정	1160	100.0	1627	87.1	286	14.8	0	0.0
행동영역	계산	1159	99.9	1501	80.3	127	6.6	0	0.0
	이해	1160	100.0	1502	80.4	97	5.0	0	0.0
	추론	1022	88.1	1091	58.4	604	31.3	25	11.0
	문제해결	1149	99.1	1261	67.5	90	4.7	0	0.0

주) 비율은 각 성취수준의 전체 학생 중 해당 영역을 숙달한 학생의 백분율임.

〈표 7〉 성취수준별 숙달한 내용영역과 행동영역 인지요소의 수

숙달한 인지요소 수		우수학력 (N=1160)		보통학력 (N=1868)		기초학력 (N=1929)		기초학력미달 (N=228)	
		빈도	비율 (%)	빈도	비율 (%)	빈도	비율 (%)	빈도	비율 (%)
내용영역	0	0	0.0	131	7.0	1509	78.2	227	99.6
	1	0	0.0	37	2.0	108	5.6	1	0.4
	2	0	0.0	107	5.7	152	7.9	0	0.0
	3	0	0.0	170	9.1	100	5.2	0	0.0
	4	0	0.0	227	12.2	36	1.9	0	0.0
	5	2	0.2	75	4.0	8	0.4	0	0.0
	6	1158	99.8	1121	60.0	16	0.8	0	0.0
행동영역	0	0	0.0	126	6.7	1191	61.7	203	89.0
	1	0	0.0	172	9.2	594	30.8	25	11.0
	2	6	0.5	341	18.3	112	5.8	0	0.0
	3	138	11.9	415	22.2	28	1.5	0	0.0
	4	1016	87.6	814	43.6	4	0.2	0	0.0

<표 6>에 제시된 바와 같이 성취수준별 수학적 인지요소의 숙달 비율은, 우수학력 학생들의 경우 4개 행동영역 중 추론 영역(88.1%)을 제외한 3개 행동영역과 6개의 내용영역을 거의 숙달하고(99.1~100%) 있는 것으로 나타났다. 보통학력 학생들의 인지요소 숙달 비율은 58.4%에서 89.7%로 숙달된 정도가 높은 편이었고, 기초학력 학생들의 인지요소 숙달 비율은 2.6%에서 31.3%였다. 반면, 기초학력미달 학생들은 추론 영역(11.0%)을 제외한 3개 행동영역과 6개의 내용영역을 거의 숙달하지 못한(0.0~0.4%) 것으로 나타났다. 추론 영역에 대한 결과가 다른 영역에 대한 결과와 다소 차이를 보이는 것은 추론 영역을 요구하는 문항의 수가 1개뿐이어서 상대적으로 오차를 더 보일 것이기 때문이라 예상된다.

<표 7>에 제시된 수학적 인지요소의 수를 성취수준별로 살펴보면, 우수학력 학생들의 경우, 6개의 내용영역을 전부 숙달한 학생은 99.8%, 4개 행동영역을 전부 숙달한 학생은 87.6%인 것으로 나타났다. 보통학력 학생들은, 내용영역은 3개 이상, 행동영역은 2개 이상 숙달한 경우가 많은 편이었고, 기초학력 학생들은 내용영역은 2개 이하, 행동영역은 1개 이하 숙달한 경우가 대부분이었으며, 기초학력미달 학생들은 숙달한 인지요소가 전혀 없는 경우가 내용영역은 99.6%, 행동영역은 89.0%로 나타났다.

국가수준 학업성취도 평가 결과로 학생들에게 4등급의 성취수준이 제공되는데, 우수학력 이외의 성취수준을 보이는 경우는 교육목표에 비추어 더 도달해야 할 부분이 있는 것이며, 성취수준별로 수학적 인지요소를 진단한 결과를 정리한 <표 6>과 <표 7>에서 보듯이 보통학력 이하의 학생들은 수학적 인지요소를 완전히 달성하지 못하고 있다. 성취수준에 따라 수학적 인지요소의 숙달 비율이 다르고, 또한 숙달한 인지요소의 수도 다르므로 주력하여 보충학습이 요구되는 인지요소가 성취수준에 따라 다르다고 하겠다. 학생들이 숙달한 인지요소의 수가 같아도 어떤 수학적 인지요소를 숙달하였고, 그렇지 못하였는지는 다를 수 있다.

학생들이 숙달한 수학적 인지요소의 수에 따라 그 유형을 내용영역과 행동영역에 대해 각각 정리하면 <표 8>과 같다. <표 8>에서 6개 내용영역을 전혀 숙달하지 못한 학생은 36%, 모두 숙달한 학생들은 44.3%이며, 4개 행동영역을 전혀 숙달하지 못한 학생은 29.3%, 모두 숙달한 학생들은 35.4%였다. 학생들의 약 20%(내용영역), 약 35%(행동영역)가 Q-행렬에서 정의한 인지요소 모두 숙달하였거나 전혀 숙달하지 못한 두 가지 유형을 제외한 숙달유형들을 보이고 있다. 숙달한 인지요소의 개수별 학생 비율 이외에도 일부 인지요소를 숙달한 학생들이 구체적으로 어떤 인지요소를 숙달하였는지와 숙달하지 못한 인지요소가 무엇인지에 대한 정보를 알 수 있다. <표 4>의 결과로 학생 개인별 평가 결과를 분석할 수 있다면, <표 5>에서 <표 8>까지 제시된 결과는 중학교 3학년 학생들이 전반적으로 숙달한 수학적 인지요소가 무엇이고, 성취수준별로 어떤 요소를 더 숙달했는지 등 집단에 대한 평가 결과 해석을 가능하게 한다.

〈표 8〉 학생들이 숙달한 인지요소의 수와 그 유형

구분	숙달 요소 수	인지요소 유형	빈도	비율 (%)	숙달 요소 수	인지요소 유형	빈도	비율 (%)
내용 영역	0	0 0 0 0 0 0	1,867	36.0	3	1 0 0 1 0 1	15	0.3
	1	1 0 0 0 0 0	1	0.0		1 0 0 1 1 0	15	0.3
		0 1 0 0 0 0	2	0.0		1 1 0 0 0 1	1	0.0
		0 0 0 1 0 0	1	0.0		1 1 0 0 1 0	3	0.1
		0 0 0 0 1 0	119	2.3		1 1 0 1 0 0	1	0.0
		0 0 0 0 0 1	23	0.4		1 1 1 0 0 0	1	0.0
	2	0 0 0 0 1 1	176	3.4	4	0 0 1 1 1 1	20	0.4
		0 0 0 1 0 1	9	0.2		0 1 0 1 1 1	66	1.3
		0 0 0 1 1 0	17	0.3		0 1 1 0 1 1	39	0.8
		0 0 1 0 0 1	4	0.1		0 1 1 1 0 1	2	0.0
		0 0 1 0 1 0	4	0.1		1 0 0 1 1 1	37	0.7
		0 1 0 0 0 1	2	0.0		1 0 1 0 1 1	8	0.2
		0 1 0 0 1 0	17	0.3		1 0 1 1 0 1	11	0.2
		1 0 0 0 0 1	16	0.3		1 0 1 1 1 0	3	0.1
		1 0 0 0 1 0	13	0.3		1 1 0 0 1 1	39	0.8
		1 0 0 1 0 0	1	0.0		1 1 0 1 0 1	17	0.3
		0 0 0 1 1 1	105	2.0		1 1 0 1 1 0	13	0.3
	3	0 0 1 0 1 1	43	0.8	5	1 1 1 0 1 0	6	0.1
		0 0 1 1 0 1	1	0.0		1 1 1 1 0 0	2	0.0
		0 1 0 0 1 1	16	0.3		0 1 1 1 1 1	15	0.3
		0 1 0 1 0 1	4	0.1		1 0 1 1 1 1	11	0.2
		0 1 0 1 1 0	25	0.5		1 1 0 1 1 1	52	1.0
		0 1 1 0 0 1	1	0.0		1 1 1 0 1 1	2	0.0
		0 1 1 0 1 0	1	0.0		1 1 1 1 0 1	5	0.1
		1 0 0 0 1 1	38	0.7	6	1 1 1 1 1 1	2,295	44.3
행동 영역	0	0 0 0 0	1,520	29.3	2	1 0 1 0	25	0.5
	1	0 0 0 1	59	1.1		1 1 0 0	248	4.8
		0 0 1 0	643	12.4		0 1 1 0	44	0.8
		0 1 0 0	27	0.5	3	0 1 1 1	3	0.1
		1 0 0 0	62	1.2		1 0 1 1	29	0.6
	2	0 0 1 1	48	0.9		1 1 0 1	433	8.4
		0 1 0 1	54	1		1 1 1 0	116	2.2
		1 0 0 1	40	0.8	4	1 1 1 1	1,834	35.4

V. 논의

우리나라에서 평가는 공정하게 객관적으로 실시되는 것에 치중하여, 평가 실시 후의 교육적 처방과 관련해서는 연구나 실행이 미비한 것이 사실이다. 이 연구에서는 대규모 학업성취도 평가 결과를 분석하고 그 결과를 활용할 수 있는 방안을 인지진단이론에 더하여 소개하고 실제 자료에 적용함으로써 그 가능성을 탐색해 보았다. 우리나라 중학교 3학년 학생들이 어떤 수학적 인지요소를 많이 숙달하고 있는지, 성취수준별로 숙달한 인지요소와 그 유형은 어떠한지 파악할 수 있었고, 학생 개인에게 강약점을 알려줄 수 있는 프로파일을 제공할 수 있었다.

이 연구에서 적용한 것처럼 인지진단모형을 바탕으로 평가 결과를 분석하여 그 정보를 교육 현장에서 사용한다면 평가 결과는 보다 내실 있는 학교교육의 모습이 되도록 활용될 수 있을 것이다. 평가 결과를 분석하고 활용하는 틀로서 인지진단모형을 적용하게 되면 전체 점수 또는 수준 이외의 다른 정보, 즉 학생들이 숙달한 인지요소에 대한 정보를 포함하게 된다. 학생의 능력에 대한 상세 진단 결과를 참고하여 학생 입장에서는 학업 능력을 향상시키는 데 필요한 부분을 구체적으로 인식할 수 있게 되고, 교사들은 교수활동에서 역점을 두어야 할 부분을 인식하고 교수의 방향을 정할 수 있을 것이며, 학교나 교육 당국에서는 학생들의 인지요소별 습득 정도에 대한 상세한 이해를 근거로 하여 교육과정과 정책에 대해 평가하고, 교육 지원을 전개해 나갈 수 있게 될 것이다.

인지진단 결과는 학생의 자기평가로 이어지거나 교사의 학생 및 학급 보충지도로 이어질 수 있다. 전자는 별도의 처방을 거치지 않더라도 진단결과의 환류로써 기대할 수 있는 것이고, 교사의 보충지도는 진단평가에 의한 후속적 처방으로 이해될 수 있다. 학생의 자기평가는 학생이 진단된 결과를 확인함으로써 스스로 반성하고 나아가 학습하는 계기를 가진다는 것이다. 자기평가에 의한 학습은 최소한 두 가지 측면에서 가능하다. 그 하나는 진단결과를 학생이 확인하는 순간, 평가결과인 점수를 넘어서서 성취도 평가에 동원된 자신의 인지요소들을 관찰하고 스스로를 이해한다는 것이고, 다른 하나는 스스로의 강점이나 약점을 이해함으로써 후속적 학습정보들을 보다 쉽게 구조화할 것이라는 점이다. 교사의 보충지도는 학생 개인수준이나 학급과 같은 집단 수준에서 이루어질 수 있다. 개인 수준이든 집단 수준이든 보충지도는, 보다 적극적인 처방적 교수활동으로서, 획득되지 못했거나 획득 확률이 낮은 인지요소들을 극복할 수 있도록 보충적인 학습경험을 제공하는 것이다. 진단결과에 근거한 처방으로서의 보충적인 학습경험은, 의사가 환자를 진단하고 처방하는 것이 전체적 하나로 통합되어 있듯이, 평가와 교수활동을 통합할 수 있게 한다. 이러한 진단 및 처방은 평가가 학생을 서열화하는 도구로서가 아니라 교수 및 학습을 위한 핵심적 교육활동을 자리 잡게 한다.

이러한 인지진단 이론을 현장에 적용하기에 앞서 인지요소의 평가가 타당한지에 대한 연구가 선행되어야 한다. 즉 인지요소의 숙달 여부에 따라 학습자를 분류한 결과가 타당한지를 평가해야 한다(Roussos, Templin, & Henson, 2007). 예를 들어, 어떤 학생이 인지요소 중 일부를 숙달하지 못한 것으로 추정된 결과가 정확한지를 점검할 필요가 있다. 또한 검사 이외의 방식으로 인지요소에 대한 진단 결과가 일반화될 수 있는지에 대해서도 평가되어야 한다.

기존 연구들(Hartz, 2002; 김성훈, 2007; 김수진·송미영·김선희, 2008)에서 인지진단이론을 이용하여 학생들의 능력을 정확하게 진단하기 위해서는 Q-행렬이 정교화되어 타당성이 확보되어야 하는 필요성이 논의되고 있다. 학생들의 인지상태에 대한 진단의 정확성은 Q-행렬의 타당성에 달려있다. Q-행렬을 수정해 가면서 반복적으로 분석하고 그 결과들을 비교하여 가장 적합한 Q-행렬을 선택할 수도 있을 것이며, 여러 모형들을 적용하여 분석한 인지진단 결과를 비교하거나 경험적인 데이터 분석(de la Torre, 2008)을 통하여 그 타당성을 검토할 수도 있을 것이다. 또한 Q-행렬에 명시한 문항에 요구되는 인지요소간의 관계가 상보적이나 또는 결합적이나에 따라 여러 인지진단모형 중 더 적절한 모형을 선택할 수 있을 것이다(강태훈, 2009). 예를 들어 ‘문자와 식’과 관련된 문항의 경우 방정식을 잘 모르는 학생이라도 만약 ‘도형’에 뛰어나서 풀 수 있는 가능성이 있다면 두 인지요소간의 관계를 상보적인 것으로 볼 수도 있다. 이럴 경우 DINA 모형보다는 DINO(Deterministic Input, Noisy, “or” gate) 모형이 보다 적합할 수도 있다. 이러한 점에서 모형적합도 분석 등 이론적·통계적으로 여러 인지진단모형 중 주어진 자료에 가장 적합한 모형이 어떠한 것인지를 분석하는 연구가 수행되어야 할 것이다.

인지상태 진단을 위한 평가 자료를 얻기 위해서는 학생들의 능력을 측정할 수 있는 문항을 개발할 때, 각 문항이 측정하는 인지요소를 정하는 것이 필요하다. 인지진단모형은 문항과 여러 인지요소 간의 관계를 가시화함으로써 문항 제작자와 문항 분석자들이 문항을 통해 측정하고자 하는 인지요소의 관점에서 문항의 특성을 논할 수 있게 된다. 주로 문항 풀이에 요구되는 지식과 능력 등은 문항제작자를 통해서 이원분류표의 형태로 정리된다. 이러한 이원분류표에 제시된 문항제작자의 의도가 학생들에게 잘 반영되어 있는가를 보여줌으로써 학생 집단의 특성, 문항과 인지요소의 실제 관계 등에 대해 알 수 있도록 하는 과정이 필요하다. 이러한 방식으로 문항을 개발하고 분석하는 교과 전문가와 평가 전문가와 함께 협력하여 양질의 Q-행렬을 구성하여야 할 것이다. 교육목표에 비추어 각 문항이 어떤 것을 평가하고 있는지에 대한 분석이 문항을 출제할 때부터 고려되고, 문항제작자가 문항과 인지요소를 연결해보는 경험과 자신감이 갖추어질 때 이 연구 결과의 현장 적용성도 높아질 것이라 판단된다.

이 연구는 국가수준 학업성취도 평가의 선다형 문항만을 대상으로 분석하였으나, 구성형

문항을 포함하여서도 학생들의 속성을 파악할 수 있는 모형(Templin, Roussos, & Stout, 2003)을 적용하여 분석할 수 있을 것이다. 대규모 표준화 검사들이 선다형과 구성형 문항으로 이루어진 것을 감안해 볼 때, 구성형 문항과 같은 다분 문항의 분석이 가능한 인지진단 모형을 활용하는 것도 필요하다. 뿐만 아니라 국가수준 학업성취도 평가는 매년 초·중·고 학생들을 대상으로 시행되고 있으므로, 이 연구에 활용된 2007년 자료 이후의 연도별 평가 자료에 인지진단모형을 적용하여 분석하고 그 결과를 비교하는 연구도 수행될 필요가 있다.

참 고 문 헌

- 강태훈(2009). “대규모 학업성취도 평가 자료를 활용한 인지상태 진단 - DINA 모형의 적용 -” 에 대한 토론. 2009학년도 한국교육과정평가원.한국교육평가학회 공동 학술세미나 대규모 학업성취도 평가 자료의 활용 방안 탐색. 연구자료 ORM 2009-31.
- 김성훈(2005). 인지구조모형에 근거한 학생의 지식상태 진단. 교육학연구, 43(1), 81-107.
- 김성훈(2007). 집단수준의 점수해석 기준을 이용한 개인수준의 진단: ACT 수학평가 자료, 학습발달표준, 규칙장이론 활용. 교육평가연구, 20(1), 165-194.
- 김성훈, 김종필, 김준엽, 송미영, 홍세희(2009). 대규모 학업성취도 평가 자료의 활용. 한국교육과정평가원 연구보고 RRE 2009-2.
- 김수진(2009). “대규모 학업성취도 평가 자료를 활용한 인지상태 진단 - DINA 모형의 적용 -” 에 대한 토론. 2009학년도 한국교육과정평가원.한국교육평가학회 공동 학술세미나 대규모 학업성취도 평가 자료의 활용 방안 탐색. 연구자료 ORM 2009-31.
- 김수진.송미영.김선희(2008). Fusion Model에 의한 수학 능력 진단을 위한 Q-행렬의 정교화. 교육평가연구, 21(2), 115-139.
- 노국향, 김신영 (2000). 문항의 형태에 따른 피험자의 인지적.정의적 반응의 차이에 관한 연구. 교육평가연구, 13(1), 181-194.
- Chipman, S.F., & Nichols, P.D.(1995). Introduction. In P. Nichols, S. F. Chipman, & R. L. Brennan(Eds), *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Erlbaum.
- de la Torre, J. (2008). An empirically-based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, 45, 343-362.
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34(1), 115-130.
- de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69, 333-353.
- DiBello, L., Stout, W., & Rousses, L. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dogan. E. & Tatsuoaka, K. (2008). An International Comparison Using a Diagnostic Testing Model: Turkish Student's Profile of Mathematical Skills on TIMSS-R.

- Educational Studies in Mathematics*, 68(3), 263-272.
- Embretson, S. (1984). A general latent trait model for response processes, *Psychometrika*, 49(2), 175-186.
- Embretson, S. E.(1990). Diagnostic Testing by Measuring Learning Processes, Psychometric Considerations for Dynamic Testing. In N., Frederiksen, R., Glaser, A. Lesgold, & M. G. Shafto(Eds.), *Diagnostic Monitoring of Skill and Knowledge Acquisition*. Hillsdale, NJ: Erlbaum.
- Embretson, S. E.(1993). Psychometric Models for Learning and Cognitive Processes. In N. Frederiksen, R.J. Mislevy, & I.I. Bejar(Eds.), *Test Theory for A New Generation of Tests*. Hillsdale, NJ: Erlbaum.
- Embretson, S. E.(1995). A Measurement Model for Linking Individual Learning to Processes and Knowledge: Application to Mathematical Reasoning. *Journal of Educational Measurement*, 32(3), 227-294.
- Frederiksen, N., Glaser, R., Lesgold, A., & Shafto, M. G.(Eds.).(1990). *Diagnostic Monitoring of Skill and Knowledge Acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hartz, S. (2002). A Bayesian framework for the Unified Model for assessing cognitive abilities: blending theory with practice. Doctoral thesis, The University of Illinois at Urbana-Champaign.
- Junker, B.W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 258-272.
- Mislevy, R. J. (1993). Foundations of a New Test Theory. In N. Frederiksen, R.J. Mislevy, & I.I. Bejar(Eds.), *Test Theory for A New Generation of Tests*. Hillsdale, NJ: Erlbaum.
- Mislevy, R. J. (1995). Probability-Based Inference in Cognitive Diagnosis. In P. Nichols, S. F. Chipman, & R. L. Brennan(Eds), *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Erlbaum.
- Mislevy, R. J. (1996). Test Theory Reconceived. *Journal of Educational Measurement*, 33(4), 379-416.
- Mislevy, R. J. (2006). Cognitive psychology and educational evaluation. in Robert L. Brennan (Eds.), *Educational measurement*, 4th Eds, (pp.257-306), CT: American Council on Education and Praeger Publishers.

- Nichols, P., Chipman, S. F., & Brennan, R. L.(Eds.)(1995). *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Erlbaum.
- Roussos, L. A., Templin, J. L.,& Henson, R. A.(2007). Skills Diagnosis Using IRT-Based Latent Class Models. *Journal of Educational Measurement*, 44(4), 293-311.
- 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, & 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, & R. L. Brennan (Eds.), *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tatsuoka, K. K., Corter, J. E., & Tatsuoka, C. (2004). Pattern of diagnosed Mathematical content and process skills in TIMSS-R across a sample of 20 countries. *American Educational Research Journal*, 41(4), 901-926.

· 논문접수 : 2011년 1월 1일/ 수정본 접수 : 2011년 3월 9일/ 게재승인 : 2011년 3월 11일

ABSTRACT

Diagnosis of knowledge states using large scale assessments – An application of DINA model –

Sung-Hoon Kim

(Dongguk University)

Mi-Young Song

(Korea Institute for Curriculum and Evaluation)

This study intended to show that large-scale assessments also were viable to provide individual-level and group-level diagnostic information, which could guide teaching and learning. The DINA model was applied to the Mathematics test of the 2007 National Assessment of Educational Achievement in Korea, which followed two steps, Q-matrix development and model parameter estimation. For Q-matrix development, 3 Mathematics education experts, using the assessment framework, identified 10 attributes, 6 in content domain and 4 in behavior domain, and then mapped them to each of the 30 multiple-choice items. For parameter estimation, two computer-runs, one with Q-matrix containing content attributes and the other with Q-matrix containing behavior attributes, were executed. Item parameter estimates showed that some items had high slip and guessing probabilities, which implied a necessity of Q-matrix refinement. Examinee parameter estimates showed that there were knowledge state differences among the individuals of identical scores, and that the individual knowledge states could be summarized into group-level diagnoses.

Key Words : cognitive diagnosis, large-scale assessment, Q-matrix, DINA model