



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

교육학석사학위논문

수학적 역량의 인지와 내용 영역에  
대한 인지진단 평가모형 적합성 연구

2018년 2월

서울대학교 대학원

수학교육과

유 민 경

# 수학적 역량의 인지와 내용 영역에 대한 인지진단 평가모형 적합성 연구

지도교수 유 연 주

이 논문을 교육학석사학위논문으로 제출함

2017년 12월

서울대학교 대학원

수학교육과

유 민 경

유민경의 석사학위논문을 인준함

2017년 12월

위 원 장 \_\_\_\_\_ (인)

부 위 원 장 \_\_\_\_\_ (인)

위 원 \_\_\_\_\_ (인)

# 수학적 역량의 인지와 내용 영역에 대한 인지진단 평가모형 적합성 연구

수학교과에서 전통적인 교육관은 개념, 원리, 법칙과 같은 내용의 습득을 강조해왔다. 그러나 내용중심 학습의 한계가 지적되면서 인지영역을 포함하는 학습자의 역량 함양이 교육과정의 목표가 되었고 계속해서 그 중요도가 강조되고 있다. 인지진단평가는 학생의 문항반응을 토대로 내용적, 인지적 요소의 숙달여부에 관한 프로파일을 추론함으로써 역량 평가에 중요한 정보를 제공할 수 있다. 그러나 인지요소의 관계성 설정에 따른 인지진단모형의 다양성에도 불구하고, 추정하는 인지요소의 특성을 고려하여 인지진단모형의 적합성을 판정하는 연구는 거의 전무하다. 따라서 이 연구는 세 가지 대표적인 인지진단모형 (DINA, DINO 그리고 ACDM)을 적용하여 모의실험과 실제 데이터 TIMSS 2007 각각에서 인지요소의 특성에 따른 평가모형 적합성을 비교분석하였다.

결과에 따르면, 검사에서 측정하는 인지요소 간 관계성과 동일한 구조를 가정하는 인지진단모형의 추정이 가장 적합한 것으로 나타났고, 검사와 추정 모형이 설정하는 관계성이 상이할수록 추정이 적합하지 못한 것으로 나타났다. 절대 적합도 지수를 통한 적합성 검증에서는 검사지의 문항 수준에 관계없이 올바른 모형을 적용하는 것이 추정의 적합성을 좌우하는 결정적 요소로 나타났고, 상대 적합도 지수에서는 어떤 모형을 선택하는가 보다 높은 문항 수준의 검사를 사용하여야 추정 적합성을 높일 수 있는 것으로 나타났다. 따라서 인지진단평가를 통해 유용한 정보를 얻기 위해서는 적절한 인지진단모형의 선택과 함께 검사 목적에 맞는 적합한 문항을 사용해야 함을 알 수 있다.

또한 각 검사지를 구성하는 문항마다 측정하는 인지요소가 다르고,

그에 따라 인지요소의 관계성 또한 달라질 수 있음을 확인하였다. TIMSS 2007 4번 검사지에 대한 미국 학생들의 응답데이터에 각 인지진단모형을 적용하여 추정한 결과, RMSEA 값이 가장 낮게 나타나는 모형은 문항별로 상이하였다. 25개의 문항 중 과반 이상을 차지하는 15개 문항의 RMSEA 값은 ACDM 모형에서 가장 낮았다. 이러한 결과는 검사지에 관한 여러 추정치가 ACDM 모형에서 가장 적합하게 나타난 것과 일관성이 있다. 또한 각 문항에서 추정 적합성이 높게 나타난 인지진단모형이 가정하는 인지요소 관계성과 문제의 풀이과정이 일치하는 것으로 나타났다.

**주요어:** 역량, 인지 영역, 내용 영역, 인지진단모형, 적합도

**학 번:** 2016-21567

# 목 차

I. 서론 .....	1
1. 연구의 목적 및 필요성 .....	1
2. 연구문제와 연구의 중요성 .....	4
II. 이론적 배경 .....	5
1. 수학적 역량의 인지와 내용 영역 .....	5
1. 1. 인지 영역 .....	5
1. 2. 내용 영역 .....	6
2. 인지진단모형(Cognitive Diagnosis Model, CDM) .....	7
2. 1. DINA .....	12
2. 2. DINO .....	14
2. 3. ACDM .....	16
3. 모형 적합도 지수 .....	19
3. 1. 절대 적합도 지수 .....	19
3. 2. 상대 적합도 지수 .....	21
4. 인지프로파일 추정 방법 및 분류 정확도 .....	23
4. 1. 인지프로파일 추정 방법 .....	23
4. 2. 인지프로파일 분류 정확도 .....	26
III. 연구방법 .....	29

1. 연구대상 .....	29
2. 연구절차 .....	31
2. 1. 각각의 인지진단모형에 대한 모의실험 설계 .....	31
2. 2. 여러 인지진단모형을 혼합한 모의실험 설계 .....	39
2. 3. TIMSS 2007 데이터 .....	49
2. 4. TIMSS 2007 데이터에 기반한 모의실험 설계 .....	54
<b>IV. 연구 결과 .....</b>	<b>56</b>
1. 각각의 인지진단모형에 대한 모의실험 결과분석 .....	56
1. 1. DINA 모형으로 생성한 응답데이터의 적합도 지수 ..	56
1. 2. DINO 모형으로 생성한 응답데이터의 적합도 지수 ..	61
1. 3. ACDM 모형으로 생성한 응답데이터의 적합도 지수 ..	65
2. 여러 인지진단모형을 혼합한 모의실험 결과분석 .....	69
3. TIMSS 2007 데이터 결과 .....	74
4. TIMSS 2007 데이터에 기반한 모의실험 결과분석 .....	80
<b>V. 논의 및 제한점 .....</b>	<b>83</b>
1. 논의 및 결론 .....	83
2. 제한점 및 제언 .....	84
<b>참고문헌 .....</b>	<b>86</b>
<b>Abstract .....</b>	<b>91</b>

## 표 목 차

〈표 II-1〉 2015 교육과정 중학교 수학 내용체계 일부 .....	5
〈표 II-2〉 김희경 외(2013)의 문항과 인지요소 간 관계 예시에 대한 Q행렬 .....	9
〈표 II-3〉 인지진단모형의 분류 .....	12
〈표 III-1〉 인지요소 수가 3개일 때 모의실험 설계 구조 .....	32
〈표 III-2〉 인지요소 3개를 측정하는 검사지의 Q행렬 .....	34
〈표 III-3〉 인지요소 5개를 측정하는 검사지의 Q행렬 .....	36
〈표 III-4〉 인지요소 8개를 측정하는 검사지의 Q행렬 .....	38
〈표 III-5〉 인지요소 수가 3개, 5개, 8개일 때 모의실험 설계구조 .....	40
〈표 III-6〉 인지요소 3개를 측정하는 혼합 검사지의 Q행렬 .....	42
〈표 III-7〉 인지요소 5개를 측정하는 혼합 검사지의 Q행렬-1 .....	44
〈표 III-8〉 인지요소 5개를 측정하는 혼합 검사지의 Q행렬-2 .....	45
〈표 III-9〉 인지요소 8개를 측정하는 혼합 검사지의 Q행렬 .....	48
〈표 III-10〉 영역에 따른 문항 분포 .....	50
〈표 III-11〉 Lee, Park & Tylan(2011)에서 사용한 인지요소 틀 ..	52
〈표 III-12〉 Lee, Park & Tylan(2011)에서 사용한 Q행렬 .....	53
〈표 III-13〉 TIMSS 2007 데이터의 RMSEA 값에 따른 모의실험 설계구조 .....	55
〈표 IV-1〉 인지요소 3개일 때 DINA 모형으로 생성한 응답데이터의 적합도 지수 .....	57
〈표 IV-2〉 인지요소 5개일 때 DINA 모형으로 생성한 응답데이터	

의 적합도 지수 .....	58
<표 IV-3> 인지요소 8개일 때 DINA 모형으로 생성한 응답데이터 의 적합도 지수 .....	58
<표 IV-4> 인지요소 3개일 때 DINA로 생성한 응답데이터 인지요 소 분류 정확도 .....	59
<표 IV-5> 인지요소 3개일 때 DINA로 생성한 응답데이터 프로파 일 분류 정확도 .....	59
<표 IV-6> 인지요소 5개일 때 DINA로 생성한 응답데이터 인지요 소 분류 정확도 .....	59
<표 IV-7> 인지요소 5개일 때 DINA로 생성한 응답데이터 프로파 일 분류 정확도 .....	60
<표 IV-8> 인지요소 8개일 때 DINA로 생성한 응답데이터 인지요 소 분류 정확도 .....	60
<표 IV--9> 인지요소 8개일 때 DINA로 생성한 응답데이터 프로파 일 분류 정확도 .....	60
<표 IV-10> 인지요소 3개일 때 DINO 모형으로 생성한 응답데이터 의 적합도 지수 .....	61
<표 IV-11> 인지요소 5개일 때 DINO 모형으로 생성한 응답데이터 의 적합도 지수 .....	62
<표 IV-12> 인지요소 8개일 때 DINO 모형으로 생성한 응답데이터 의 적합도 지수 .....	62
<표 IV-13> 인지요소 3개일 때 DINO로 생성한 응답데이터 인지요 소 분류 정확도 .....	63

<표 IV-14> 인지요소 3개일 때 DINO로 생성한 응답데이터 프로파일 분류 정확도 .....	63
<표 IV-15> 인지요소 5개일 때 DINO로 생성한 응답데이터 인지요소 분류 정확도 .....	63
<표 IV-16> 인지요소 5개일 때 DINO로 생성한 응답데이터 프로파일 분류 정확도 .....	64
<표 IV-17> 인지요소 8개일 때 DINO로 생성한 응답데이터 인지요소 분류 정확도 .....	64
<표 IV-18> 인지요소 3개일 때 DINO로 생성한 응답데이터 프로파일 분류 정확도 .....	64
<표 IV-19> 인지요소 3개일 때 ACDM모형으로 생성한 응답데이터의 적합도 지수 .....	65
<표 IV-20> 인지요소 5개일 때 ACDM모형으로 생성한 응답데이터의 적합도 지수 .....	66
<표 IV-21> 인지요소 8개일 때 ACDM모형으로 생성한 응답데이터의 적합도 지수 .....	66
<표 IV-22> 인지요소 3개일 때 ACDM으로 생성한 응답데이터 인지요소 분류 정확도 .....	67
<표 IV-23> 인지요소 3개일 때 ACDM으로 생성한 응답데이터 프로파일 분류 정확도 .....	67
<표 IV-24> 인지요소 5개일 때 ACDM으로 생성한 응답데이터 인지요소 분류 정확도 .....	67
<표 IV-25> 인지요소 5개일 때 ACDM으로 생성한 응답데이터 프	

로파일 분류 정확도 .....	68
<표 IV-26> 인지요소 8개일 때 ACDM으로 생성한 응답데이터 인 지요소 분류 정확도 .....	68
<표 IV-27> 인지요소 8개일 때 ACDM으로 생성한 응답데이터 프 로파일 분류 정확도 .....	68
<표 IV-28> 인지요소 3개일 때 여러 모형으로 생성한 응답데이터 의 RMSEA 값 .....	70
<표 IV-29> 인지요소 5개일 때 여러 모형으로 생성한 응답데이터 의 RMSEA 값-1 .....	71
<표 IV-30> 인지요소 5개일 때 여러 모형으로 생성한 응답데이터 의 RMSEA 값-2 .....	72
<표 IV-31> 인지요소 8개일 때 여러 모형으로 생성한 응답데이터 의 RMSEA 값 .....	73
<표 IV-32> TIMSS 2007 4학년 미국 응답데이터의 적합도 지수 ..	74
<표 IV-33> TIMSS 2007 4학년 미국 응답데이터의 문항별 RMSEA 값 .....	80
<표 IV-34> TIMSS 2007 데이터에 기반한 모의검사 적합도 지수 ·	81
<표 IV-35> TIMSS 2007 데이터에 기반한 모의검사 RMSEA 값 ....	81

## 그림 목 차

[그림 II -1] 김희경 외(2013)의 문항과 인지요소 간 관계 예시 ....9	
[그림 IV -1] TIMSS 2007 4번 검사지 1번 문항 .....76	
[그림 IV -2] TIMSS 2007 4번 검사지 25번 문항 .....77	
[그림 IV -3] TIMSS 2007 4번 검사지 12번 문항 .....78	



# I. 서론

## 1. 연구의 목적 및 필요성

평가는 의도된 교육과정이 교육 현장을 통해 학생의 학습으로 성취되었는지 확인하기 위한 도구적 목적을 지닌다. 따라서 평가의 내용과 방법은 최근 수학교육의 목표, 즉 교육과정이 담고 있는 교육관을 반영할 수 있어야 한다. 시대에 따라 요구되는 인재상이 자연스럽게 변화하는 것과 같이, 교육관 역시 변화되어 왔다. 이에 우리나라의 국가차원 교육과정 또한 여러 차례에 걸쳐 교육개정을 발표했고, 수학과 교육과정도 꾸준히 변화하며 교육과정 내실화 및 교육의 질 개선에 힘써왔다. 영역간의 연계성이나 위계성 등과 같은 수학 과목 고유의 특성은 시대적 요구에 따라 변하지 않는 개념이며 개정을 통해 변화시켜야 할 내용에서 제외된다. 교육개정의 최근 동향에서 눈여겨볼만한 변화는 개념, 원리, 법칙과 같은 내용 중심의 전통적인 교육관 대신, 이해하기, 표현하기, 추론하기 등과 같은 인지 및 기능 숙달을 포함한 역량(competence) 함양의 자리매김이라 할 수 있다.

역량은 실제 생활에서의 수행과 밀접한 관련이 있는 복잡한 능력 개념으로 학습을 통해 습득 가능한 것으로 여겨져(Koepfen et al., 2008; Weinert, 2001) 최근 연구 및 실천 현장에서 중요한 개념으로 주목받고 있다(e.g., Klieme, Hartig, & Rauch, 2008). 학생의 역량 함양이 교육이 달성해야 할 목표로 여겨짐에 따라 적절한 역량평가를 가능케 하는 평가 도구 또한 요구되고 있다. 적절한 역량평가는 교육 프로그램, 기관, 제도의 선진화, 교육 실천의 최적화, 그리고 학생 개개인의 학습 과정을 지원하는 데 핵심적인 역할을 할 수 있다(Koepfen et al., 2008). 특히 학생 개개인의 학습 지도에서 학생들의 강점과 약점을 파악하며 개인의 필요에 맞는 피드백을 제공한다는 목적을 달성하기 위해서는 충분히 세분화된 정보를 제공할 필요가 있다(e.g., Black & Wiliam, 1998; Wiliam,

2006).

기존의 선행연구에 따라 수학교과에서 학생의 수행능력은 내용 영역(content domains)과 인지 영역(cognitive domains)으로의 분류가 가능하다(Harks et al., 2014). 이는 수학교과 지식의 차원성에 따른 평가틀을 구성하기 위한 여러 시도의 결과물로 받아들여지고 있다(Bloom, 1956; Csapó, 2010; Borg, 1986). 평가의 목적을 달성하기 위해서는 내용 영역과 인지 영역으로 세분화된 각 요소에 대한 진단을 가능하게 하는 선진적인 평가 도구를 사용해야 하며, 이러한 관점에서 학생의 능력을 여러 요소의 다차원적 결합으로 설정하는 인지진단모형(Cognitive Diagnosis Model, CDM)의 적용이 가능하다. 인지진단모형은 평가에서 나타난 학생의 응답반응을 토대로 문항을 해결하는데 필요한 인지요소의 숙달 또는 미숙달 여부를 프로파일 형태로 제공하는 통계 추정 모형이다(Tatsuoka, 1995). 여기서 인지요소는 학생들이 문항의 정답을 맞히는 데 필요한 능력(ability), 기능(skills), 지식(knowledge)이나 인지과정(cognitive process) 등을 포함한다(Tatsuoka, 1983).

인지진단모형은 학생들의 능력을 세세하게 평가한다는 점에서 학생과 교사 모두에게 학생의 학업상태에 관한 자세한 피드백을 줄 수 있다는 강점을 지닐 뿐 아니라 정보처리 이론이 뒷받침하는 평가, 교육과정, 교육학을 통합하는 관점을 제시할 수 있는 잠재력을 가진 것으로 여겨진다(de la Torre et al., 2016). 인지진단모형을 통한 평가결과가 학생 개인의 인지요소 숙달 및 미숙달 여부를 나타낸다는 점은 학생이 스스로의 학업상태를 돌아보고 관리할 수 있게 한다. 또한 교사가 학급 전반의 진단결과를 취합하면 학급이 어떤 인지요소의 숙달에 어려움을 겪는지 쉽게 파악할 수 있어 수업 진행에 대한 전략적 도움을 얻을 수 있다(de la Torre et al., 2016). 인지진단모형은 높은 수준의 정보력 제공과 여러 이론의 통합이라는 관점에서 여러 관심과 기대를 받아왔으며 Tatsuoka(1983) 이후 30여 년간 활발히 개발되어 그 종류 또한 매우 다양하다. 그러나 다수의 선행 연구와 꾸준한 개발 및 사용에도 불구하고 인지진단모형은 여러 관점에서 여전히 발달의 초기 단계로 여겨지고 있

다(Harks et al., 2014).

인지진단모형으로 학생의 능력을 추정하기 위해서는 응답데이터가 필요하고, 응답데이터의 기반이 되는 검사지의 각 문항은 여러 인지요소의 숙달을 요구한다. 이 때 숙달해야하는 인지요소가 크게는 내용영역에 해당하는지 인지영역에 해당하는지, 작게는 개념 습득에 해당하는지 문제 해결전략에 해당하는지 등에 따라 문항이 요구하는 인지요소 중 일부만 숙달해도 높은 확률로 문항을 해결할 수도 있고, 모든 인지요소를 숙달해야만 문항을 맞힐 확률이 높아질 수도 있다. 현재까지 개발된 각각의 인지진단모형은 이러한 차이에 입각해 인지요소의 관계성을 상이하게 설정하고 있으며 올바른 분석을 위해서는 추정하고자 하는 인지요소의 특징에 따라 적절한 평가모형을 선정할 수 있어야 한다. 그러나 다양한 평가모형의 개발에도 불구하고 연구에서 사용되는 평가모형들은 대표적인 몇 가지 종류의 반복에 불과하며 여러 인지진단모형 중 어떤 모형을 선택하는 것이 적합한가에 관한 선행연구는 매우 부족한 실정이다.

따라서 이 연구는 검사 및 문항에서 측정하는 인지요소의 특성에 따라 어떤 인지진단모형의 적용이 적합한가를 판별할 수 있는 연구적 기반을 마련하고자 한다. 연구 목적에 부합하는 결과를 얻기 위해 인지요소의 관계 설정이 뚜렷하게 상이한 모형을 선정하고, 모의실험을 통해 인지요소의 관계 설정이 다양한 모형 간 추정치를 비교한다. 또한 모의실험을 통해 얻은 결과를 토대로 실제 데이터인 수학과학 성취도 추이변화 국제 비교 연구(Trends in International Mathematics and Science Study: 이하 TIMSS) 2007년도 응답데이터에 대한 추정치의 의미를 해석한다. TIMSS 검사지는 의도된 교육과정이 교사를 통해 실행된 교육과정을 통해 학생에게 성취된다는 평가모형을 따라 학생의 성취단계 진단에서 사용되었으며, 검사를 통해 측정하는 요소를 내용, 인지 영역으로 구별한 평가틀을 기반으로 하고 있어 이 연구에서 인지진단모형을 적용하기에 적합하다고 판단하였다.

## 2. 연구문제와 연구의 중요성

평가는 교육과정의 방향성 및 수업 상황의 개선을 위해 학습목표의 달성 여부를 확인할 수 있어야 한다. 수학교과에서 학생이 숙달해야 할 역량은 크게 내용 영역과 인지 영역으로 분류 가능하며 이에 따라 다양한 능력을 동시에 고려할 수 있는 평가도구의 적용이 필요하다. 인지진단모형은 내용과 인지라는 다른 성격의 영역 내에 있는 여러 요소들의 숙달 여부를 동시에 파악할 수 있다는 강점이 있어 역량평가를 위한 도구로 사용할 수 있다. 그러나 추정하는 인지요소의 관계성을 상이하게 설정하는 다양한 인지진단모형 중 어떤 모형을 선택하는 것이 적합한가에 대한 판단은 영역 전문가에게 맡겨져 있다. 인지진단모형을 적용한 많은 연구가 있어왔음에도 불구하고 다수의 인지진단모형을 동일한 데이터에 적용하여 비교분석한 연구는 분야를 막론하고 매우 부족하고(Lee & Sawaki, 2009) 따라서 검사지에 따라 어떤 인지진단모형을 선택해야 하는지 판단하는 데 어려움을 겪는다. 이 연구는 하나의 응답데이터를 여러 인지진단모형으로 추정해봄으로써 응답데이터가 추정하는 인지요소의 특징에 따라 어떤 인지진단모형을 사용하는 것이 가장 적합한지에 대해 살펴보고자 한다. 특히 수학교과가 다양한 내용 및 인지 영역으로 이루어져 있음에도 어떤 인지진단모형으로 추정하는 것이 적절한가에 대한 선행연구의 논의가 개략적인 수준에 그쳐왔다. 그러므로 이 연구에서는 검사 도구별 적합성과 더불어 검사를 구성하는 개별 문항 수준의 적합성까지 논의하고자 한다. 이를 위한 구체적인 연구 질문은 다음과 같다.

1. DINA, DINO, 그리고 ACDM 각각의 인지진단 평가모형은 어떤 특성의 수학적 역량을 평가하기에 적합한가?

2. 문항에서 측정하는 수학적 역량의 특성에 적합하지 않은 평가모형의 선택은 어떤 영향을 미치는가?

## II. 이론적 배경

### 1. 수학적 역량의 인지와 내용 영역

수학은 개념, 원리, 법칙의 이해와 기능 습득을 통한 여러 현상의 수학적 해석 및 합리적인 문제해결 능력을 기르는 교과로 여겨진다(한국교육과정평가원, 2014). 이는 수학문제를 해결하기 위해 개념, 원리, 법칙과 같은 내용지식과 이해하기, 추론하기 등과 같은 기능을 활용할 수 있어야 함을 의미한다. 수학교과 내용지식 및 기능은 <표 II-1>과 같이 하나의 영역 내에서 내용요소와 기능이라는 명칭 하에 연결되어 학습되는 것이 일반적이다. 그러나 두 요소는 습득 및 활용 과정에서 상이한 성격을 나타내므로 이를 이해함으로써 학생들의 문제해결과정을 보다 구체적으로 파악할 수 있다.

<표 II-1> 2015 교육과정 중학교 수학 내용체계 일부

영역	핵심개념	내용 요소	기능
문자와 식	다항식	문자의 사용과 식의 계산 식의 계산 다항식의 곱셈과 인수분해	표현하기 계산하기 문제 해결하기
	방정식과 부등식	일차방정식 일차부등식과 연립일차방정식 이차방정식	이해하기 활용하기 검토하기

#### 1.1 인지 영역

인지 영역의 대표적인 분류는 Niss(2003)을 따른다. 그는 인지 영역을 (a) 수학적 사고, (b) 문제해결능력, (c) 수학적 모델링, (d) 기호로 나타내

기, (e) 의사소통, (f) 도구의 활용 그리고 (g) 수학적 추론으로 구별했다. Niss의 분류는 수학적 소양을 위한 평가틀로써 PISA 2003부터 2009까지 채택되었다. PISA의 최근 수정된 평가틀은 Niss의 영역을 약간 수정하여 이를 인지과정(cognitive process)의 폭넓은 기저를 이루는 기본 역량으로써 정의한다. 그 영역은 다음과 같다. (a) 주어진 상황을 수학적으로 공식화하기, (b) 수학적 개념, 사실, 절차, 그리고 추론 사용하기, 그리고 (c) 결과 해석, 적용, 그리고 평가하기. 마지막으로 TIMSS의 연구는 인지영역을 (a) 이해하기, (b) 적용하기, 그리고 (c) 추론하기로 분류하였다. 우리나라 2015 개정의 수학과 수학과목의 내용체계에서는 기능이라는 명칭의 범주 하에 (a) 계산하기, (b) 이해하기, (c) 문제 해결하기, (d) 설명하기, (e) 판별하기, (f) 증명하기, (g) 구별하기 등으로 분류하고 있다.

인지 영역은 수학적 내용지식의 습득여부를 확인하는 영역이 아니라 문제해결을 위해 습득한 내용지식을 사용할 수 있는지를 진단하는 영역이다. 예를 들어, 피타고라스의 공식이 무엇인지 답하기 위해서는 내용지식의 습득으로 충분하지만 주어진 상황에 맞게 공식을 적용할 수 있는지는 해석, 적용, 계산 등의 능력을 필요로 한다. 인지 영역은 이러한 사고과정의 숙달정도를 진단하기 위한 영역이며 안다, 모른다는 이분법적인 분리가 다소 모호한 측면이 있다.

## 1.2 내용 영역

(a) 수와 연산, (b) 대수, (c) 기하, (d) 측정, 그리고 (e) 자료해석과 확률은 교육과정에 사용되는 내용 영역의 대표적인 예시들이다(National Council of Teachers of Mathematics, 2000). 또한 이러한 내용 영역은 시험의 개발과 TIMSS 같은 대규모 평가의 시험데이터를 분석하는 데 사용된다(e.g., Mullis et al., 2009). 다른 분류로는 PISA의 평가틀로 주어진 (a) 치환과 관계, (b) 공간과 도형, (c) 양, 그리고 (d) 불확정성이 있다.

우리나라 2015 개정의 수학과 수학과목의 내용체계에서는 영역, 핵심

개념, 내용 요소라는 명칭의 범주 하에 내용 영역을 나타내고 있으며, 영역에서 내용 요소의 순서로 좀 더 세세하게 분류하고 있다. 예를 들어, 문자와 식이라는 영역 하에 다항식, 방정식과 부등식이라는 두 가지 핵심 개념이 존재하고, 다항식 범주 하에 다항식의 연산, 나머지정리, 인수분해, 그리고 방정식과 부등식 범주 하에 복소수와 이차방정식, 이차방정식과 이차함수, 여러 가지 방정식과 부등식으로 분류하고 있다. 영역을 기준으로 살펴보면 (a) 문자와 식, (b) 기하, (c) 수와 연산, (d) 함수, (e) 확률과 통계의 분류를 나타낸다.

수학교과에서 내용 영역은 개념, 원리, 정리와 같은 내용지식을 의미한다. 예를 들어, 다항식의 계산 원리 습득은 내용지식에 포함되는데 다항식과 다항식 사이의 계산 방법을 아는 지식이 선행되고 나서야 다항식 간 계산을 할 수 있기 때문이다. 이로 인해 계산 원리는 내용 영역, 계산하기는 인지 영역으로 분리되어 받아들여진다. 내용 영역은 내용 지식에 관한 영역이므로, 인지 영역에 비해 안다, 모른다고 이분화 하기에 적합하다.

## 2. 인지진단모형(Cognitive Diagnosis Model, CDM)

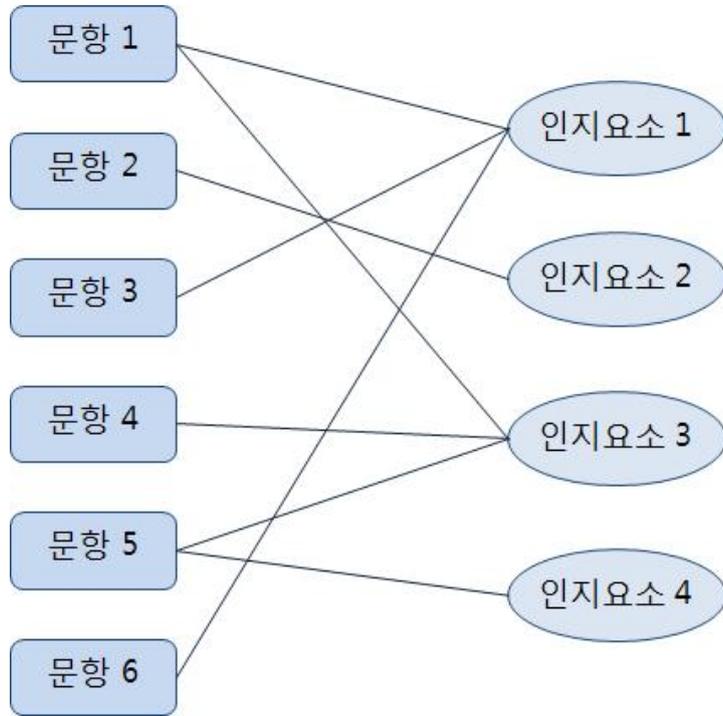
인지진단모형에서 추정하는 인지요소는 인지 영역과 내용 영역의 여러 요소들, 학생의 능력, 지식, 기술, 인지과정 등의 역량을 포함한다. 인지진단모형은 학생들이 함양해야 할 역량을 구성하는 여러 인지요소들을 측정하여 줄 뿐 아니라 동시에 측정하기 때문에 교육개정의 동향과 그 변화에 대해 충분히 선진적인 평가도구로 여겨진다. 앞서 살펴 본 인지 및 내용 영역은 각각의 숙달, 즉 내용적 지식의 숙달여부와 기능적, 인지적 지식의 숙달여부는 구분하여 진단할 필요가 있으면서 또한 함께 측정할 수밖에 없는 구조를 지닌다. 단순히 내용지식을 아는지 모르는지와 문제 해결, 추론과 같은 인지를 할 수 있는지 여부를 동일한 수준의 사고과정으로 고려하는 것은 직관적으로도 맞지 않으며 학생이 어떤 사고

과정을 더 학습해야하는지에 대한 정보도 주지 못한다. 또 한편으로는 내용 영역 없이 사고 기능을 물을 수 없는데다가 각 내용 영역별로 요구되는 인지 영역의 성격 또한 상이하다. 따라서 인지진단모형이 하나의 검사지, 그리고 검사지를 구성하는 각 문항을 통해 여러 인지요소의 숙달여부 각각에 대한 정보를 제공하면서도 한 검사 내에서 동시에 진단하여 주는 것은 인지진단모형이 국내에서도 주목받고 또 적용될만한 충분한 강점이라 할 수 있다.

인지진단모형은 검사지에 대한 학생의 문항반응을 토대로 인지요소 숙달여부에 관한 프로파일을 추론한다. 그렇기 때문에 각각의 인지요소들은 문항을 통해 추정되는 것이며, 각 문항과 인지요소 간의 관계가 명확해야 한다. 이러한 관계는 Q행렬(Tatsuoka, 1983)로 나타난다. Q행렬은 문항을 행, 인지요소를 열로 두어 문항이  $n$ 개, 인지요소가  $K$ 개라 할 때  $n \times K$ 행렬로 표현된다. Q행렬의 원소  $q_{ij}$ 는 문항  $i$ 를 해결하는데 인지요소  $j$ 가 필요할 경우 1, 그렇지 않을 경우 0의 값을 가진다. 특히 Q행렬의 행벡터, 즉 문항  $i$ 에서 측정하는 인지요소의 열에 해당하는 원소는 1, 그렇지 않은 열에 해당하는 원소는 0의 값을 가지는 벡터를  $q$ 벡터라 한다. Q행렬 및 원소의 정의는 다음과 같다.

$$Q = (q_{ij}), \quad q_{ij} = \begin{cases} 1 & (\text{문항 } i \text{를 해결하는데 인지요소 } j \text{가 필요한 경우}) \\ 0 & (\text{문항 } i \text{를 해결하는데 인지요소 } j \text{가 필요하지 않은 경우}) \end{cases}$$

[그림 II-1]의 예를 살펴보자. 우선 [그림 II-1]는 6개의 문항으로 구성된 검사지를 통해 총 4개의 인지요소 숙달여부를 추정하는 경우이다. 이는 각 문항이 측정하는 인지요소가 무엇인지를 보여주는 것으로 문항을 해결하는데 필요한 인지요소와 문항은 선으로 연결되어 있고 필요하지 않은 경우는 연결되어 있지 않다. 즉, 문항 1을 해결하기 위해서는 인지요소1과 3의 숙달이 필요하고, 문항 2를 해결하기 위해서는 인지요소 2의 숙달이 필요함을 의미한다. 이에 따라 Q행렬은 <표 II-2>로 나타내어진다.



[그림 II-1] 김희경 외(2013)의 문항과 인지요소 간 관계 예시

<표 II-2> 김희경 외(2013)의 문항과 인지요소 간 관계 예시에 대한 Q행렬

	인지 1	인지 2	인지 3	인지 4
문항 1	1	0	1	0
문항 2	0	1	0	0
문항 3	1	0	0	1
문항 4	0	0	1	0
문항 5	0	0	1	1
문항 6	1	0	0	0

Q행렬의 분류가 잘못된 경우, 즉 문항을 해결하는데 필요한 인지요소를 필요하지 않다고 설정하거나 혹은 필요하지 않은 인지요소를 필요하다고 설정하는 경우, 모수 추정과 프로파일 추정에 오류가 생기기 때문에 정확한 Q행렬을 사용하는 것이 중요하다. 모의실험을 통해 Q행렬 분류 오류에 관해 조사한 Rupp & Templin(2008)의 연구에 의하면, 문항 해결에 필요한 인지요소를 필요하지 않다고 설정한 경우에는 필요한 인지요소를 모두 숙달한 학생들의 정답률이 과추정되어 문항수준이 실제보다 쉬운 수준으로 나타났고, 반대의 경우에는 문항 해결에 필요한 인지요소를 모두 숙달하지 못한 학생들의 정답률이 과추정되어 문항수준이 실제보다 어려운 수준으로 나타나는 경향을 보였다. 또한 Kunina-Habenicht, Rupp & Wilhelm(2012)의 연구에 따르면 Q행렬의 잘못된 분류로 인한 타당성 결여는 학생들의 인지요소 프로파일 분류 정확도를 특히 감소시키는 것으로 나타났다.

실제 데이터에서는 참(true) Q행렬이 무엇인지 정확히 알 수 없지만 모의실험은 생성 Q행렬(generating Q-matrix) 및 생성 모형(generating model)이 있기 때문에 이와 같은 연구가 가능하고, 이러한 결과를 실제 데이터를 이해하는데 사용할 수 있다. 즉, 인지진단모형을 적용한 평가 결과가 유용하려면 모의실험에서는 올바른 Q행렬을, 실제 데이터에서는 타당한 Q행렬을 적용해야 한다. 이 연구에서도 마찬가지로 모의실험은 생성 Q행렬과 생성 모형을 가지고 있기 때문에 정확한 Q행렬과 평가모형이 무엇인지 알고 추정치를 비교할 수 있지만, TIMSS 2007 검사지의 정확한 Q행렬이 무엇인지는 알 수 없기 때문에 여러 전문가를 통해 질적 분석을 시행한 기존의 선행연구를 참고하여 타당성이 보증된 것으로 여겨지는 Q행렬을 적용하여 분석한다. 이 연구에서 사용한 Q행렬 및 데이터 분석에 대한 자세한 설명은 연구방법에 있다.

앞서 언급한 바와 같이, Tatsuoka의 Rule Space Model(RSM; Tatsuoka, 1983) 이후 다양한 종류의 인지진단모형이 개발되어왔다. 몇 가지 인지진단모형의 예시로는 Reparameterized Unified Model(RUM), Compensatory Reparameterized Unified Model(C-RUM), reduced

noncompensatory reparameterized unified model(rRUM), Deterministic Input, Noisy “and” gate(DINA), Deterministic Input, Noisy “or” gate(DINO), Noisy Inputs, Deterministic, “and” gate(NIDA), Noisy Inputs, Deterministic, “or” gate(NIDA), Generalized DINA(G-DINA), General Diagnosric Model(GDM), Log-linear Cognitive Diagnosis Model(LCDM), Hierarchical Diagnositic Classification Model(HDCM), Attribute Hierarchy Method(AHM), 등이 있다.

이처럼 다양한 인지진단모형은 측정하는 인지요소 사이의 관계를 다양하게 설정하고 있다. 따라서 인지진단모형을 적용하여 평가하는 인지진단평가(Cognitive Diagnosis Assessment, CDA)를 시행할 때는 타당한 Q행렬의 사용뿐만 아니라 측정하는 인지요소의 특성과 그로인한 관계성을 고려하여 적합한 인지진단모형을 선택할 수 있어야 한다.

인지요소 간의 관계성은 크게 비보상적 모형과 보상적 모형 두 가지로 분류가능하다. 비보상적 모형과 보상적 모형의 분류는 문항 해결에 필요한 인지요소 중 일부의 숙달이 다른 인지요소의 미숙달을 보상할 수 있는지 여부를 기준으로 한다. 비보상적 모형은 문항을 맞추기 위해서는 문항 해결에 필요한 인지요소의 숙달이 선행되어야 하고, 숙달한 인지요소가 다른 인지요소의 미숙달을 보상해주지 못한다고 가정한다(Henson, Templin, & Willse, 2009). 보상적 모형은 어떤 인지요소의 미숙달이 다른 인지요소의 숙달에 의해 보충될 수 있다고 가정하여 필요한 인지요소의 일부, 극단적으로는 하나만 숙달하여도 성공적인 문항 해결능력을 보일 수 있다고 가정한다(DiBello, Roussos & Stout, 2007). 이 연구에서는 모든 인지요소를 숙달해야만 높은 정답률을 가질 수 있다고 가정하는 극단적 비보상적 모형 DINA와 하나의 인지요소만 숙달해도 높은 정답률을 보장받는다고 가정하는 극단적 보상적 모형 DINO 그리고 숙달한 인지요소 수에 따라 조금씩 더 높은 정답률을 갖는다고 가정하는 보상적 모형 ACDM을 모의실험을 통해 먼저 비교한다. ACDM 모형의 경우 문항 해결에 필요한 인지요소 중 일부의 숙달과 전체의 숙달에 대한 정답률 차이를 설정할 수 있지만, 일부의 숙달로도 일정수준의 정답률에 도달할 수

있다는 점에서 보상적 모형으로 분류가능하다. 이 연구에서 사용하는 세 가지 모형의 비보상적, 보상적 모형 분류표는 <표 II-3>과 같다.

이 연구는 모의실험을 통해 생성모형과 동일한 모형으로 추정된 경우, 그리고 다른 모형을 통해 추정된 경우의 각 검사지 및 문항 추정치가 어떻게 다른지를 DINA, DINO, 그리고 ACDM의 세 가지 인지진단모형으로 비교하는데, 특히 각 인지진단모형이 가정하는 인지요소의 관계성, 즉 비보상성과 보상성을 기준으로 비교 분석한다. 또한 TIMSS 2007 데이터를 DINA, DINO, 그리고 ACDM 모형으로 추정하고, 모의실험에서 생성모형과 다른 모형 간 추정치 차이의 시사점을 적용하여 세 가지 인지진단 모형에서 얻은 추정치를 바탕으로 TIMSS 2007 검사지 및 문항의 인지요소 사이의 관계성을 해석한다.

<표 II-3> 인지진단모형의 분류

비보상적 모형	보상적 모형
DINA	DINO
	ACDM

## 2.1 DINA(Deterministic-Input, Noisy-And-Gate)

DINA 모형은 비보상적 모형으로 필요한 인지요소를 모두 숙달해야만 문항을 맞출 수 있다고 가정한다. 이러한 가정은 하나의 인지요소라도 숙달하지 못한 경우에 모든 인지요소를 미숙달한 것과 같은 확률을 부여하게 한다. 즉, 피험자의 정답률은 문항해결에 필요한 인지요소를 모두 숙달한 집단의 경우  $1 - s_i$ , 그리고 하나라도 숙달하지 못한 집단의 경우  $g_i$ 로 이분화 된다. 이는 DINA 모형의 대표적인 특징이면서 동시에 한계점으로 여겨지기도 한다(de la Torre et al., 2016). DINA 모형의 문항반응함수는 다음과 같다(Junker & Sijtsma, 2001).

$$P(X_{ij} = 1 | \alpha_j) = (1 - s_i)^{n_{ij}} g_i^{(1 - n_{ij})}$$

여기서  $X_{ij}$ 는 피험자  $j$ 의 문항  $i$ 에 대한 정답여부를 표현하는 것으로, 정답을 맞히면 1, 틀리면 0으로 나타낸다.  $\alpha_j$ 는 피험자  $j$ 의 인지요소 벡터로 각각의 인지요소를 숙달했을 경우를 1, 숙달하지 않았을 경우를 0으로 표현한다. 또한  $s_i, g_i$  각각은 slip과 guess 모수로 문항  $i$ 에서 필요한 인지요소를 모두 숙달했으나 틀릴 확률을  $s_i$ 로, 문항  $i$ 에서 필요한 인지요소를 숙달하지 못했음에도 맞힐 확률을  $g_i$ 로 나타낸다.  $n_{ij}$ 은 피험자  $j$ 가 문항  $i$ 를 푸는데 필요한 모든 인지요소를 숙달했는지 여부를 판별하는 것으로  $n_{ij} = \prod_{k=1}^K \alpha_{jk}^{q_{ik}}$ 의 식을 따른다.  $\alpha_{jk}$ 는 피험자  $j$ 가 인지요소  $k$ 를 숙달하면 1, 미숙달하면 0으로,  $q_{ik}$ 는 문항  $i$ 를 해결하는데 인지요소  $k$ 가 필요하면 1, 필요하지 않으면 0으로 나타낸다. 즉, 문항  $i$ 를 푸는데 필요한 인지요소  $k$ 를 피험자  $j$ 가 숙달했을 때는  $\alpha_{jk}^{q_{ik}}$ 가 1의 값을 갖고 필요하지만 숙달하지 못한 경우 0의 값을 가진다. 인지요소  $k$ 가 필요하지 않은 경우는  $\alpha_{jk}^{q_{ik}}$ 에서 지수가 0이므로 1의 값을 얻어 필요한 인지요소를 모두 숙달했는가를 판별하는 데 영향을 미치지 않는다.

수학교과에서 DINA를 분석틀로 사용한 예시로는 혼합계산과 같이 어느 하나의 인지요소만 숙달하지 못해도 올바른 값을 계산하기 어려운 경우가 있다. 이규진(2015)은 초등수학의 분수와 소수 진단평가 데이터를 분석하는데 DINA를 적용하였다. 이규진(2015)은 분수, 소수의 사칙연산 계산능력, 혼합계산 순서 이해, 그리고 문제해결능력으로 인지요소를 설정하였다. 이러한 인지요소들은 하나만 숙달하지 못해도 값을 제대로 계산하기 어렵기 때문에 문항을 해결하기 위해서는 모든 인지요소의 숙달이 요구되고 따라서 비보상적 모형인 DINA로 추정하기에 적합하다.

수학교과에 여러 내용 및 인지 영역이 존재함에도 불구하고 기존의 선행연구들은 수학교과의 여러 시험에 적용한 결과에 입각해 수학교과의 인지진단평가가 DINA 모형과 같은 비보상적 평가모형으로 추정하기 적합하다고 판단해왔다(Roussos, Templin & Henson, 2007). 그러나 수학교과 내에서 시행되는 시험은 다양한 인지 및 내용 영역을 포함하기 때문에 특정 인지진단모형의 추정이 항상 적합하다고 결론짓기에는 어려운 점이 있다. 이 연구의 질문 또한 이러한 의문점에서 출발하며 대표적인 비보상적 모형 DINA와 보상적 모형 DINO, 그리고 ACDM 모형을 통한 심층적 분석을 통해 수학교과 내의 여러 인지요소와 인지진단모형의 적합성이 갖는 연관성을 살펴본다.

## 2.2 DINO(Deterministic-Input, Noisy-Or-Gate)

Templin과 Henson은 대학생들의 ‘도박성향’에 관한 결과분석에 적합한 모형으로 DINO를 소개하였다. 도박을 하게 되는 여러 가지 이유들을 인지요소라 할 때, 대학생들은 여러 인지요소 중 하나만 충족되어도 도박을 하기에 충분한 이유를 가진 것처럼 행동하는 경향을 드러냈다. 이처럼 일부의 만족, 극단적으로 하나의 만족이 높은 확률을 보장하는 경우는 기존 DINA 모형의 ‘and’로는 설명될 수 없었고, 학생들의 도박성향에 따른 결정을 올바르게 분석은 DINA의 ‘and’가 아닌 ‘or’가정이 필요하다는 주장 하에 개발된 모형이 DINO 모형이다.

DINO 모형은 DINA 모형과 마찬가지로 문항별 slip, guess 모수를 가진다. DINA 모형과의 유일한 차이점은 문항에 관여된 한 가지 인지요소만 알고 있어도 문항을 해결할 수 있다는 점이다. DINO 모형은 하나의 숙달된 인지요소가 다른 인지요소의 미숙달을 보상하기 때문에 보상적 모형으로 분류된다. DINO 모형의 문항반응함수는 다음과 같다(Templin & Henson, 2006).

$$P(X_{ij} = 1 | \alpha_j) = (1 - s_i)^{w_{ij}} g_i^{(1-w_{ij})}$$

주어진 문항반응함수에서  $w_{ij}$ 를 제외한 다른 기호는 전부 DINA 모형의 기호와 동일하며 slip 모수  $s_i$ , guess 모수  $g_i$ 의 의미만 달라진다. DINO 모형의 경우 문항 해결에 필요한 인지요소는 하나로 충분하기 때문에, 필요한 인지요소 중 하나를 숙달했음에도 실수하여 틀릴 확률이  $s_i$ , 필요한 인지요소를 단 하나도 숙달하지 못했으나 추측으로 맞힐 확률이  $g_i$ 가 된다. 여기서  $w_{ij}$ 는 피험자  $j$ 가 문항  $i$ 를 푸는데 관여된 인지요소 중 하나라도 숙달했는지에 대한 여부를 나타내며

$w_{ij} = 1 - \prod_{k=1}^K (1 - \alpha_{jk})^{q_{ik}}$ 로 표현된다. 피험자  $j$ 가 문항  $i$ 를 해결하는데 필

요한 인지요소  $k$ 중 하나라도 숙달( $\alpha_{jk} = 1$ )하면  $\prod_{k=1}^K (1 - \alpha_{jk})^{q_{ik}} = 0$ 이 성립하고 따라서  $w_{ij} = 1$ 을 얻는다. 즉,  $w_{ij}$ 는 피험자  $j$ 가 문항  $i$ 를 해결하는데 필요한 인지요소 중 하나라도 숙달하면 1, 하나도 숙달하지 못하면 0으로 나타난다.

DINO 모형은 피험자의 정답률을 문항 해결에 필요한 인지요소 중 어느 하나라도 숙달한 경우  $1 - s_i$ , 전혀 숙달하지 못한 경우  $g_i$ 로 이분화한다. DINO 모형은 DINA 모형과 정반대의 의미로 극단적인 함수이며 DINO 모형의 이러한 가정은 DINA 모형과 마찬가지로 특징이자 한계점으로 평가되기도 한다(de la Torre et al., 2016).

DINO 모형의 경우 하나의 인지요소만 숙달하면 맞힐 수 있는 문제에 적용하기 적절하다. 따라서 한 문제를 해결하는 데 여러 가지 전략을 사용할 수 있는 문제에서 여러 전략을 인지요소로 설정한 경우를 예로 들 수 있다. 우리나라 교육과정에서 함수는 중학교에서 다뤄지며, 함수의 종속성을 이해시키기 위해 함수의 정의를 배우기 전에 초등수학에서 정비례, 반비례를 학습한다. 어느 한 변량에 대한 다른 변량의 값을 표에

채워보라는 문제가 주어졌을 때, 피험자는 한 변량의 증가량이 동일 할 때 다른 변량의 증가량이 동일함을 이용해 값을 채워 넣을 수도 있고, 한 변량에 적당한 값을 곱하여 다른 변량의 값이 나온다는 규칙을 이용할 수도 있다. 이 때 피험자는 변화량을 이용하든지 곱의 관계를 이용하든지 간에 어느 하나의 전략만 안다하더라도 문제를 해결하기 위한 충분한 지식을 가진다고 할 수 있다.

### 2.3 ACDM(Additive Cognitive Diagnosis Model)

앞서 살펴본 DINA 모형은 문항을 해결하기 위해 필요한 모든 인지요소가 숙달되어야 한다고 가정한다. 이 가정은 DINA 모형의 특징인 동시에 인지요소의 관계를 극단적으로 설정한다는 측면에서 한계를 갖기도 한다. 이에 de la Torre(2011)는 필요한 인지요소 각각의 숙달이 정답률에 영향을 주는 보다 일반적인 G-DINA(generalized DINA) 모형을 제안하였다. G-DINA 모형은 문항에서 측정하는 인지요소 각각의 숙달에 관한 모수 뿐 아니라 문항 해결에 필요한 인지요소 중 일부를 숙달 했을 때의 특정 조합에 대한 모수도 설정가능하다. link 함수를 identity로 설정한 경우의 문항반응함수는 다음과 같다(de la Torre, 2011).

$$P(X_{ij} = 1|\alpha_j) = \delta_{i0} + \sum_{k=1}^{K_i^*} \delta_{ik} \alpha_{jk} + \sum_{k'=k+1}^{K_i^*} \sum_{k=1}^{K_i^*-1} \delta_{ikk'} \alpha_{jk} \alpha_{jk'} + \dots + \delta_{i12\dots K_i^*} \prod_{k=1}^{K_i^*} \alpha_{jk}$$

여기서  $K_i^*$ 는 문항  $i$ 를 해결하는데 필요한 인지요소들의 집합을 의미하고,  $\alpha_{jk}$ 는 피험자  $j$ 가 인지요소  $k$ 를 숙달했는지 여부를 나타낸다. 숙달한 경우를 1, 숙달하지 못한 경우를 0으로 나타낸다. 덧붙여  $\delta_{i0}$ 는 문항  $i$ 의 intercept 값,  $\delta_{ik} \alpha_{jk}$ 항은 피험자  $j$ 가 인지요소  $k$ 를 숙달했을 때 정답률에 미치는 영향,  $\delta_{ikk'} \alpha_{jk} \alpha_{jk'}$ 항은 피험자  $j$ 가 인지요소  $k$ 와  $k'$  모두

를 숙달했을 때 정답률에 미치는 영향,  $\delta_{i12\dots K_i^*} \prod_{k=1}^{K_i^*} \alpha_{jk}$  항은 피험자  $j$ 가 문항  $i$ 를 해결하는데 필요한 모든 인지요소를 숙달했을 때 정답률에 미치는 영향을 나타낸 계수이다.

예를 들어, 문항  $i$ 의 해결을 위해 필요한 인지요소가  $k_1, k_2, k_3$ 라 하자. 피험자  $j_1$ 이 인지요소  $k_1$ 만 숙달한 경우,  $\alpha_{j_1 k_1} = 1$ ,  $\alpha_{j_1 k_2} = 0$ ,  $\alpha_{j_1 k_3} = 0$  이므로 정답률은 다음과 같다.

$$P(X_{ij_1} = 1 | \alpha_{j_1}) = \delta_{i0} + \delta_{ik_1}$$

즉, 피험자  $j_1$ 의 정답률은 문항  $i$ 의 intercept값에  $k_1$ 을 숙달하여 얻는 모수의 합으로 나타난다. 피험자  $j_2$ 가 인지요소  $k_1, k_2$ 를 숙달했다면  $\alpha_{j_2 k_1} = 1$ ,  $\alpha_{j_2 k_2} = 1$ ,  $\alpha_{j_2 k_3} = 0$  이므로 정답률은 다음과 같다.

$$P(X_{ij_2} = 1 | \alpha_{j_2}) = \delta_{i0} + \delta_{ik_1} + \delta_{ik_2} + \delta_{ik_1 k_2}$$

피험자가 필요한 인지요소를 두 개 숙달한 경우에는, 각각의 인지요소 숙달에 해당하는 모수  $\delta_{ik_1}$ ,  $\delta_{ik_2}$  뿐 아니라 두 인지요소의 숙달에 대한  $\delta_{ik_1 k_2}$  값을 추가로 얻는다. 만약 피험자  $j_3$ 가 필요한 인지요소 모두를 숙달했다면  $\alpha_{j_3 k_1} = 1$ ,  $\alpha_{j_3 k_2} = 1$ ,  $\alpha_{j_3 k_3} = 1$ 의 값을 갖고, 정답률은 다음과 같다.

$$P(X_{ij_3} = 1 | \alpha_{j_3}) = \delta_{i0} + \delta_{ik_1} + \delta_{ik_2} + \delta_{ik_3} + \delta_{ik_1 k_2} + \delta_{ik_1 k_3} + \delta_{ik_2 k_3} + \delta_{ik_1 k_2 k_3}$$

피험자  $j_3$ 가  $j_2$ 보다 숙달한 인지요소의 수는 하나 더 많지만, 정답률

에 나타나는 모수의 수는 4개가 더 많다. 이처럼 한 문항에서 측정하는 인지요소가 3개만 되어도 문항에서 추정해야하는 모수의 값은  $2^3 = 8$ 개가 된다.

de la Torre(2011)는 이와 같이 필요한 인지요소를 여러 개 숙달했을 때 더해지는  $\delta_{ikk'}, \dots, \delta_{i12\dots K_i^*}$ 와 같은 상호 계수(interaction effects)를 0으로 설정하고 각각의 인지요소 숙달로 인한 각각의 주 효과( $\delta_{ik}$ , main effect)만을 고려하는 ACDM(additive CDM)을 정의하였다. ACDM 모형은 G-DINA 모형 보다는 간단하고 DINA 모형 보다 일반적이라는 특징이 있다. 모형의 복잡도(model complexity)가 높을수록 충분한 환경에서는 정확한 추정치를 구할 수 있지만, 한편으로는 정확하고 유의미한 추정을 위한 조건 또한 까다로워지는 경향이 있고 추정 시간이 오래 걸리는 단점이 있다. 반대로 모형이 지나치게 단순하면 평가에 대한 정확한 피드백이 어려울 수 있다는 점에서 ACDM 모형의 유용성을 기대할 수 있다. ACDM의 문항반응함수는 다음과 같다.

$$P(X_{ij} = 1 | \alpha_j) = \delta_{i0} + \sum_{k=1}^{K_i^*} \delta_{ik} \alpha_{jk}$$

ACDM은 필요한 인지요소를 하나도 숙달하지 못한 경우의 정답률을  $\delta_{i0}$ 로 두고, 숙달한 인지요소  $k$ 가 있을 때마다 해당 모수  $\delta_{ik}$ 만큼씩 더 높은 정답률을 갖게 하는 구조이다. 특히  $\delta_{ik}$ 의 값을 인지요소 번호  $k$ 에 따라 다르게 설정할 수 있어 문항 해결에 필요한 인지요소 사이에 상이한 가중치를 설정할 수 있다는 특징이 있다. ACDM의 경우 문제해결에 필요한 인지요소 중 일부만 숙달했다 하더라도 일정수준의 정답률을 갖기 때문에 다른 인지요소의 미숙달을 보완해준다는 측면에서 보상적 모형으로 분류된다. 이 연구에서는 ACDM이 보상적 모형으로 분류된다는 사실보다는 극단적 비보상적 모형인 DINA 모형과 극단적 보상적 모형인

DINO 모형보다 비교적 일반적인 분류가 가능하다는 측면에서 DINA, DINO 모형과 함께 비교분석을 진행할 대상으로 선정하였다. 특히, ACDM 모형은 선형적인(linear) 평가모형으로 학생들이 인지요소 각각을 숙달했을 때 그 인지요소들 사이의 종합적인 상호작용으로 문항을 해결할 것이라는 가정을 포함하고 있다. 이는 극단적인 DINA 모형과 DINO 모형이 고려하지 못하는 일반적인 인지과정에 대한 설명을 가능하게 할 것으로 기대된다.

### 3. 모형 적합도 지수

인지진단모형은 개인의 수학적 역량에 대한 숙달 프로파일을 추정함으로써 유용한 정보를 제공한다. 그러나 추정 결과를 의미 있게 받아들이기 위해서는 추정의 타당도를 검증해야 한다. 이러한 모형의 타당도는 모형 적합도 지수(model fit index)를 통해 판정될 수 있다(Hu et al., 2016). 모형 적합도 지수에는 각 평가모형이 기준 하한선을 통과하는지 판단하기 위해 사용하는 절대 적합도 지수(absolute fit index)와 여러 모델 중 비교적 더 나은 모형을 찾기 위해 적용하는 상대 적합도 지수(relative fit index)가 존재한다. 이 연구는 절대 적합도 지수인 MADcor(the mean absolute deviation correlation), SRMSR(standardized root mean square residual; Maydeu-Olivaras, 2013), RMSEA(Root mean square error of approximation; Kunina-Habenicht, Rupp & Wilhelm, 2009)와 상대 적합도 지수에 해당하는 deviance, AIC(Akaike's information criterion; Akaike, 1974), BIC(Bayesian information criterion; Schwarz, 1978)를 사용해 모형 적합도(model fit)를 살펴본다.

#### 3.1 절대 적합도 지수(absolute fit index)

상대 적합도 지수를 통해 모형 사이의 우위를 비교하기에 앞서 각 모

형의 추정치가 타당한지, 즉 비교 대상이 될 만한 유효성을 가지는지를 확인하는 것은 매우 중요하다(Chen, Torre & Zhang, 2013). 어떤 모형이 더 적합한가를 판정하는 상황에서 타당성이 없는 모형을 동일한 비교 선상에 두는 것은 적절하지 않다. 따라서 이 연구는 절대 적합도 지수와 상대적합도 지수를 함께 사용하여 각 모형이 비교 대상이 될 수 있는지를 여부를 판별한 뒤 모형 간 비교하여 분석하는 과정을 따른다. 이 연구에서 사용하는 MADcor, SRMSR 그리고 RMSEA에 대해 살펴보도록 하겠다. 먼저 MADcor는 다음과 같이 정의된다.

$$\text{MADcor} = \frac{2}{J(J-1)} \sum_{j' < j} |r_{j'j} - \hat{r}_{j'j}|$$

여기서  $r_{j'j}$ 는 문항  $j'$ 와  $j$ 에 대해 측정된 상관계수이고  $\hat{r}_{j'j}$ 은 상관계수의 예측값을 의미한다. MADcor 값은 측정치와 예측값인 각각의 상관계수  $r_{j'j}$ ,  $\hat{r}_{j'j}$  사이의 차이의 평균을 나타낸다. 또 다른 절대 적합도 지수 SRMSR의 정의는 다음과 같다.

$$\text{SRMSR} = \sqrt{\frac{2}{J(J-1)} \sum_{j' < j} (r_{j'j} - \hat{r}_{j'j})^2}$$

사용하는 기호들은 MADcor에서의 의미와 동일하고 구조 또한 유사하다. 단, MADcor 값은 상관계수의 측정값과 예측값 차이에 대한 평균값을 계산하는 대신 SRMSR은 두 상관계수 값 차이에 대한 제곱근의 평균을 나타낸다. 다음으로 문항 별 RMSEA의 정의는 다음과 같다.

$$\text{RMSEA}_j = \sum_{k=1}^K \sum_{d=1}^D \pi(a_{kd}) \left[ P(X=1|a_{kd,j}) - \frac{N(X=1|a_{kd,j})}{N(a_{kd,j})} \right]^2$$

먼저  $a_{kd}$ 는 인지요소  $k$ 에 대한 숙달 수준  $d$ 를 나타내는 것으로, 여기서  $d$ 는 숙달과 미숙달의 두 가지 수준만 가진다.  $\pi(a_{kd})$ 는 인지요소  $k$ 에서 숙달 수준  $d$ 를 갖는,  $a_{kd}$ 에 해당하는 학생들의 비율을 나타낸다.  $P(X=1|a_{kd,j})$ 는 문항  $j$ 에서  $a_{kd}$ 에 해당하는 학생에게 기대되는 정답률(predicted probability to solve)을 의미한다.  $N(X=1|a_{kd,j})$ 은 문항  $j$ 에서  $a_{kd}$ 에 해당하는 학생이 실제로 문항을 해결한 수를 나타내고,  $N(a_{kd,j})$ 는 문항  $j$ 에서  $a_{kd}$ 에 해당하는 학생 중 답을 표시한 학생의 수를 의미한다. 즉, 문항  $j$ 에 대한  $RMSEA_j$ 값은 각 인지요소의 숙달여부로 구분한 각각의 학생 그룹 내에서 기대 정답률과 측정된 정답률 사이의 차이를 제곱 꼴로 구한 것이다.  $RMSEA$ 는 문항별  $RMSEA_j$ 의 평균을 문항 전체 집합  $J$ 에서 구한 값이다. 이 연구는  $RMSEA_j$ 가 문항별 값을 갖는 것에 주목하여, 검사지 전체에 대한  $RMSEA$ 값 뿐 아니라, 문항별  $RMSEA_j$  값 또한 살펴본다.

### 3.2 상대 적합도 지수(relative fit index)

절대 적합도 지수는 잘 못 적용된 모형을 탈락시킬 수 있는 좋은 지표이다. 그러나 응답반응에 대한 절대적인 지수이기 때문에 모형 간 비교는 고려하지 않으며(Rupp, Templin & Henson, 2010) 절대 적합도 지수를 통해 검증된 모형은 하나의 모형이 아니라 여럿일 경우가 많다(Chen, Torre & Zhang, 2013). 따라서 절대적 기준을 통과한 여러 모형 간 적합도 비교는 상대 적합도 지수를 통해 이루어질 필요가 있다. 즉, 절대 적합도 지수는 추정에 적절하지 못한 모형을 가려내는 도구인 반면 상대 적합도 지수는 모형 간 적합도를 비교함으로써 각 응답 데이터에 좀 더 유용한 모형이 무엇인지 판별하는 데 사용한다. 이 연구에서 사용하는 deviance의 정의는 다음과 같다.

$$\text{deviance} = -2 \times \log(\text{marginal maximum likelihood})$$

복잡한 모형일수록 주어진 데이터에 대한 적합 수준이 높아지기 쉽고 이로 인해 deviance 값은 작아지는 경향이 있다(김명연, 강태훈, 2016). deviance는 AIC와 BIC 지수에서도 사용하는데, deviance 값에 더해 각기 다른 벌칙(penalty)을 고려하여 좀 더 유용한 정보를 제공한다. 먼저 AIC의 정의는 다음과 같다.

$$\text{AIC} = \text{deviance} + 2 \times [\text{the number of parameters in the model}]$$

AIC는 동일한 deviance를 갖는 모형이라면 좀 더 적은 모수를 사용하는, 즉 더 간단한 모형이 낮은 AIC 값을 갖게 함으로써 더 나은 적합도를 갖는 것으로 판별하는 지수이다. 모형이 복잡할수록 추정치 자체는 좋게 나올 수 있기 때문에 모형의 효율에 대한 정보를 함께 제공하려는 측면에서 모수에 관한 벌칙을 정의하였다. 이 연구에서는 인지진단모형에 적용하므로 모수는 문항반응함수에서 추정하는 각각의 문항모수(item parameter)와 인지요소패턴 추정 모수(skill class parameter)를 포함하여 각 인지진단평가가 추정하는 모수 전체를 의미한다. 그러나 AIC는 모수의 수만 고려하고 추정에 사용하는 표본의 수는 고려하지 않는다. 만약 조금 더 복잡한 모형과 간단한 모형이 같은 AIC를 가진다는 정보밖에 없다면 연구자는 간단한 모형이 더 유용한 모형이라고 결론지을 수 있다. 그러나 복잡한 모형이 훨씬 더 적은 표본을 사용했다면 간단한 모형이 더 유용하다는 결론은 더 이상 유효하다고 보기 어려워진다. 이러한 논의에서 한층 더 자유롭게 정의된 지수가 BIC이고 그 정의는 다음과 같다.

$$\text{BIC} = \text{deviance} + 2 \times [\text{the number of parameters}] \times \log[\text{sample size}]$$

BIC는 deviance에 모수의 수와 표본의 수를 함께 사용하는 지수이다.

모수의 수가 비슷한, 즉 복잡도가 비슷한 모형에서 같은 deviance를 가진다면 좀 더 적은 표본으로 그 deviance를 얻은 모형이 더 유용하다고 볼 수 있고, BIC는 그러한 선택이 가능하게끔 돕는다. 그러나 BIC를 사용하는 경우 큰 수의 표본을 사용한 모형은 복잡성에 대한 벌칙 요소를 크게 부여받기 때문에 간단한 모형이 선호받기 쉬운 경향이 있다. 정의에서 오는 차이로 인해 AIC와 BIC의 선택은 일치하지 않을 수 있으며 따라서 이 연구는 두 지수를 모두 고려할 뿐 아니라 deviance 또한 참고하여 비교한다.

#### 4. 인지프로파일 추정 방법 및 분류 정확도

인지진단모형의 추정치가 일차원적인 여타 다른 평가모형 보다 교육과정, 교사의 수업 실행, 학생의 학업 성취 등에 더욱 유용한 정보를 제공할 수 있는 이유는 기본적으로 인지진단모형이 다차원적인 인지프로파일을 추정하기 때문이라는 사실에 있다. 따라서 인지프로파일을 어떻게 추정할지에 대한 방법을 선택하고 추정한 인지프로파일의 분류 정확도(classification accuracy)를 어떻게 확인할 것인지에 대한 논의가 필요하다.

##### 4.1 인지프로파일 추정 방법

Heubner & Wang(2011)에 따르면 학생들의 숙달 패턴을 전반적으로 정확하게 추정하기 위해서는 MLE(Maximum Likelihood Estimation)와 MAP(Maximum A Posteriori) 방법의 사용이 적절하다. 이에 이 연구는 학생들 개개인이 자신의 실제 프로파일과 가장 가까운 인지프로파일을 제공받도록 한다는 설정 하에 MLE, MAP 방법으로 인지요소 숙달여부를 추정하였다. MLE와 MAP는 R 프로그램의 CDM 패키지(Robitzsch, Kiefer,

George, & Uenlue, 2014)가 제공하는 기본 추정치이기도 하다.

MLE, MAP 추정법은 기본적으로 지역 독립성(local independence)을 가정하고 각 피험자의 인지프로파일을 추정하기 위해 인지요소 패턴에 따른 피험자의 응답반응을 적용해 다음과 같은 우도함수(likelihood function) 식을 구한다(반재천 & 김선, 2016).

$$L(X=x|\alpha_l) = \prod_{j=1}^J [P_j(\alpha_l)]^{x_{ij}} [1 - P_j(\alpha_l)]^{1-x_{ij}}$$

우도함수에서  $\alpha_l$ 은 각 인지요소패턴  $(0, \dots, 0)$ 부터  $(1, \dots, 1)$ 중 특정 인지요소패턴을,  $P_j(\alpha_l)$ 은 인지요소 패턴이  $\alpha_l$ 일 때 문항  $j$ 에 대한 정답률을,  $x_{ij}$ 는 피험자  $i$ 가 문항  $j$ 를 맞게 풀었는지 여부를 나타낸다.  $x_{ij}$ 는 피험자  $i$ 가 문항  $j$ 를 해결한 경우 1, 그렇지 않은 경우 0으로 나타난다. 즉, 피험자  $i$ 가 가지는 특정 인지프로파일  $\alpha_l$ 이 있을 때, DINA, DINO, 그리고 ACDM 모형 중 추정에 사용하는 인지진단모형의 문항반응함수에 따라 정답률  $P_j(\alpha_l)$ 의 값이 결정되고, 피험자  $i$ 의 문항반응  $x_{ij}$  또한 결정된다. 그러나 실제로는 피험자  $i$ 의 인지요소프로파일을 알 수 없기 때문에 추정을 사용한다.

우선 MLE 추정방법은 가능한 인지요소패턴  $(0, \dots, 0)$ 부터  $(1, \dots, 1)$ 중 위의 우도함수에 대입했을 때 우도함수 값이 최대가 되게끔 하는 인지요소패턴  $\alpha_l$ 을 찾아 추정치로 결정한다. MLE의 식은 다음과 같다.

$$\hat{\alpha}_{MLE} = \arg \max_l \{L(X=x|\alpha_l)\}$$

다음으로 MAP 추정방법을 이해하기에 앞서 베이저안 방법(Bayesian Method)을 적용하여 다음과 같은 식이 성립함을 안다.

$$P(\alpha_l | X = x) = \frac{P(\alpha_l, X = x)}{P(X = x)}$$

다시 한 번 베이저안 방법을 사용하면 다음의 식 또한 성립한다.

$$P(\alpha_l, X = x) = L(X = x | \alpha_l)P(\alpha_l)$$

인지요소 패턴에 대해 모두 더해지면 다음의 식을 얻는다.

$$P(X = x) = \sum_{m=1}^L L(X = x | \alpha_m)P(\alpha_m)$$

마지막으로 위에서 얻은 식들을 종합하면, 다음의 식을 얻는다.

$$P(\alpha_l | X = x) = \frac{L(X = x | \alpha_l)P(\alpha_l)}{\sum_{m=1}^L L(X = x | \alpha_m)P(\alpha_m)}$$

이렇게 얻은 식은 인지요소패턴에 대한 사전분포(prior distribution)가 있을 때, 인지요소패턴에 대한 사후분포(posterior distribution)를 구할 수 있음을 나타낸다. 이 때 인지요소패턴은  $\alpha_l$ , 인지요소패턴에 대한 사전 분포는  $P(\alpha_l)$ , 인지요소패턴에 대한 사후분포는  $P(\alpha_l | X = x)$ 으로 표현되었다( $l = 1, \dots, m$ ).

MAP 추정방법은 베이저안 방법을 적용하여 구한 최종 식을 통해 사후 분포  $P(\alpha_l | X = x)$ 가 최대가 되게 하는 인지요소패턴을 피험자의 추정치로 사용한다. 이를 표현한 식은 다음과 같다.

$$\hat{\alpha}_{MAP} = \arg \max_l \{L(\alpha_l | X = x)\}$$

## 4.2 인지프로파일 분류 정확도

인지프로파일의 추정방법을 결정한 다음에는 추정방법을 통해 얻은 인지프로파일 추정치의 분류 정확도를 조사할 수 있다. 분류 정확도를 조사한다는 것은 실제 프로파일과 추정된 프로파일을 비교한다는 의미이다. 그러나 실제 데이터에서는 학생의 프로파일에 관한 참(true) 값을 알 수 없고 따라서 직접적인 조사가 불가능하다. 이에 모의실험에 한해 인지요소 프로파일 추정의 정확도를 비교한다. 실제 데이터에서 인지프로파일의 분류 정확도를 조사할 수는 없지만, 모의실험에서의 각 평가모형이 나타내는 적합도 지수와 인지프로파일 분류 정확도 사이의 관련성 등을 확인할 수 있고, 이에 따른 시사점을 적용하면 실제 데이터에서의 분류 정확도 또한 일정수준 이상으로 나타날 것으로 추측 가능하다.

인지진단모형에서 분류 정확도는 크게 두 가지 값을 계산하는데, 하나는 전체 피험자 중 인지 프로파일이 정확히 일치하는 피험자의 비율(the attribute profile classification rate: 이하 the attribute profile rate), 즉 피험자의 모든 인지요소에 대한 숙달여부를 정확하게 추정한 비율이고, 다른 하나는 각각의 인지요소를 기준으로 전체 피험자의 숙달여부가 맞게 추정된 비율(the marginal attribute classification rate: 이하 the marginal attribute rate)로, 예를 들어 인지요소1에 대한 the marginal attribute rate이라하면, 인지요소1에 대한 각 피험자의 숙달 진위여부와 동일하게 추정해낸 비율을 의미한다. The attribute profile rate은 검사를 통해 측정하는 인지요소 전체에 대한 숙달여부를 올바르게 추정해야 하므로 일반적으로 높은 값을 갖기 어렵고 한 두 개만 잘 못 측정할지라도 비율을 계산하는 데 제외된다는 특징이 있다. 특히 the marginal attribute rate보다 이론적으로 작거나 같을 수밖에 없으며 실제로는 훨씬 더 작게 측정되기 쉽다. 즉 각 사람에 대한 정확도 비율인 the

attribute profile rate은 충분히 높은 수준의 추정을 했음에도 불구하고 터무니없이 작은 값이 나올 수 있다는 문제점이 있다. 따라서 이 연구는 해밍 거리(hamming distance)를 도입한다. the marginal attribute rate가 인지요소에 대한 marginal rate을 계산하듯이, 각 사람에 대한 marginal rate을 측정한다. 전체 패턴을 올바르게 추정해야만 하는 the attribute profile rate과 두 marginal rate을 비교함으로써 marginal하지 않은 비율과 marginal한 비율의 차이, 그리고 두 marginal rate 사이의 일관성 등을 확인한다.

해밍 거리는 곱집합 위에서 정의되는 거리 함수로, 같은 길이의 두 문자열에서 대응되는 위치의 값이 일치하지 않는 개수를 의미한다. 예를 들어, 길이가 같은 두 문자열 ‘happy’, ‘apply’ 에서 해밍 거리를 구하려면 대응되는 위치 값이 일치하지 않는 수를 세어야 하고, 첫 번째, 두 번째, 네 번째 자리가 일치하지 않기 때문에 해밍 거리 3을 가진다. 길이가 같은 문자열의 다른 예를 살펴보자. 두 문자열 (1, 1, 0, 1, 0), (0, 1, 0, 1, 1)에서는 첫 번째, 다섯 번째 자리의 수가 일치하지 않고 따라서 해밍 거리 2를 가진다. 동일한 길이  $n$ 의 두 벡터  $a = (a_1, a_2, \dots, a_n)$ ,  $b = (b_1, b_2, \dots, b_n)$ 에 대한 해밍 거리는 다음과 같이 정의할 수 있다.

$$d(a,b) = |a_1 - b_1| + |a_2 - b_2| + \dots + |a_n - b_n|$$

해밍거리를 이용해 피험자 개개인의 인지요소 프로파일 참값과 추정치를 비교할 경우, 해밍거리는  $n$ 개의 인지요소 중 일치하지 않는 인지요소의 수, 즉 잘못된 추정의 수를 알려준다. 즉,  $d(a,b)$ 는 인지요소  $n$ 개 중 틀린 추정의 수를 나타내고,  $n - d(a,b)$ 는 인지요소  $n$ 개 중 올바른 추정의 수를 나타낸다. 따라서  $\frac{n - d(a,b)}{n} = 1 - \frac{d(a,b)}{n}$ 는 인지요소  $n$ 개 중 올바르게 추정한 비율을 의미하게 된다. 이로써  $1 - \frac{d(a,b)}{n}$ 를 통

해 각 피험자에 대해 올바르게 추정한 인지요소의 비율을 얻을 수 있고, 이를 the marginal attribute profile classification rate(: 이하 the marginal profile rate)이라 하겠다.

### Ⅲ. 연구방법

#### 1. 연구대상

이 연구는 모의실험과 실제 데이터 분석을 모두 수행한다. 모의실험을 통해 조사한 생성모형과 추정모형 간 적합성에 대한 시사점은 실제데이터를 분석한 추정치가 의미하는 바를 해석하는 데 유용하게 사용될 수 있다. 먼저 모의실험에서는 가상의 피험자 1000명에게 각 인지진단모형이 설정하는 인지요소 관계를 요구하는 문항으로 구성된 검사지가 주어지는 상황을 가정하여 응답데이터를 산출한다. 검사지는 높은 문항수준과 낮은 문항수준의 두 종류가 있다고 가정한다. 인지진단모형의 종류가 세 가지이고 문항수준이 두 종류이기 때문에 총 6가지 종류의 검사에 대한 실험이며 각 검사마다 1000명의 피험자가 응시하는 셈이다. 모의실험의 설계 방법 및 추정 방법에 대한 더 자세한 설명은 연구절차에서 밝힌다.

이 연구에서 사용한 실제 데이터는 TIMSS 2007(Mullis et al., 2008) 데이터이다. TIMSS 2007 연구의 주요 목적은 연구 참여국 학생들의 수학 및 과학 성취도를 국제적인 수준에서 파악하고 이전 연구와의 변화추이를 비교하며 성취도에 영향을 미치는 관련 변인들을 다양하게 고려하여 각국의 교육 질 개선에 도움이 되는 정보를 제공한다(김희경 외, 2008). TIMSS 2007의 평가도구는 수학, 과학 두 과목의 성취도 검사도구와 교육과정 및 학교, 교사, 학생의 배경변인 설문지로 구성되며 이 연구에서 사용하는 데이터는 수학 성취도 검사도구에 한정한다. TIMSS는 국제 성취도 변화 추이를 측정하는 데 초점을 두고 있어 이전 주기의 검사지 문항 중 일부를 다시 사용한다. 이에 따라 이전 주기에서 다시 사용하는 공통 문항 군 한 개, 그리고 새로 개발된 문항 군 한 개를 합쳐 검사지 하나를 구성한다. 또한 이전 주기와의 변화 추이를 고려할 뿐 아니라 시행되는 검사지 사이의 연계성을 보장하기 위해 검사지 간 하나의 문항군

이 겹치도록 분배한다. TIMSS 2007 역시 마찬가지로의 방식을 따른다.

TIMSS 2007 수학 성취도 평가는 4학년 학생들과 8학년(우리나라는 중학교 2학년) 학생들을 대상으로 하는데, 이 연구에서는 4학년 수학시험을 본 학생의 성취를 대상으로 한다. TIMSS 2007은 14개의 문항군을 통해 14개의 검사지를 구성하였는데, 이 연구에서 사용한 검사지는 4, 5번 문항군으로 구성된 4번 검사지이다. 이러한 학년 및 검사지 선정방식은 기본적으로 TIMSS 2007 선행연구 Lee, Park & Tylan(2011)의 제안을 따른다. 4번 검사지는 15개의 선다형 문항을 포함하는 25개의 문항으로 구성되어 있는데, Lee, Park & Tylan(2011)에 따르면 4번 검사지의 문항들이 1점 아니면 0점, 즉 맞혔는지 틀렸는지의 이분화(dichotomous) 되는 채점방식을 따르는 문제의 비중이 가장 높다는 사실을 알 수 있다. 선행 연구나 이 연구에서 사용하는 인지진단모형은 다분화(polytomous)된 점수를 입력하는 구조가 아니라 이분화(dichotomous)된 점수를 입력하여 추정하는 구조이기 때문에 이분화 하는 채점방식을 적용한 문항의 비율에 따른 선정은 적절한 기준이라 할 수 있다. 4번 검사지에서 2점, 1점, 0점으로 다분화된 채점방식을 사용한 문항은 단 두 개만 존재한다(문항 번호 M041275와 M031247). 이 경우는 Lee, Park & Tylan(2011)의 방식에 따라 다분화 채점 방식에서 2점을 맞은 학생에게는 1점을, 다분화 채점 방식에서 1점, 0점을 맞은 학생에게는 0점을 부여하였다. 이러한 방식은 부분점수 1점을 맞은 학생은 전체를 맞히지 않았기 때문에 정답자 보다는 오답자로 분류하는 것이 적절하다는 측면에서 기인한다.

우리나라 4학년은 TIMSS 2007 수학 성취도 평가에 참여하지 않았기 때문에 이 연구에서는 선행연구와 일관되게 미국학생의 데이터를 분석한다. 분석한 미국학생의 데이터는 US 표본뿐만 아니라 벤치마크 주로 선정된 Minnesota와 Massachusetts 학생들의 표본도 포함한다. 부분적으로 결측값을 낸 학생과 달리 검사지 전체에서 답이 인식되지 않은 학생의 경우는 표본에서 제외하였다. 이에 US 표본에서 564명, Minnesota 표본에서 132명, Massachusetts 표본에서 127명을 포함하여 총 823 명의 미국

학생의 응답데이터를 대상으로 인지진단모형을 적용하였다.

## 2. 연구절차

### 2.1 각각의 인지진단모형에 대한 모의실험 설계

연구의 목적이 모형 간 추정 적합성의 비교에 있는 만큼, 실제 데이터를 분석하고 분석한 결과에 대해 진단하기에 앞서 동일한 응답 데이터에 대한 모형 사이의 추정 차이를 좀 더 면밀히 살펴보려는 의도를 가지고 모의실험을 설계하였다. 인지요소 사이의 관계를 극단적인 비보상적 관계로 설정하는 DINA 모형, 극단적인 보상적 관계로 설정하는 DINO 모형, 그리고 적은 숙달과, 그보다 많은 일부의 숙달, 전부의 숙달을 구분하는, 인지요소 사이의 가중치 차등을 설정하는 ACDM 모형 각각을 생성모형으로 설정하고, 높은 문항수준과 낮은 문항수준의 두 가지 문항수준의 검사지를 구성한다. 즉, 하나의 Q행렬에 대해 총 6가지 검사도구 및 응답데이터를 구성한다. 동일한 Q 행렬을 토대로 서로 다른 문항 수준, 인지진단모형을 적용해 생성한 응답데이터는 다시 세 모형 모두에 의해 추정되며, 이에 따라 적합도 지수를 포함하는 몇 가지 결과 값 (evaluation)과 피험자의 인지요소프로파일 정확도를 계산한다. 이러한 설계는 모형의 이론적 차이가 추정치를 통해 상이하게 나타나는지 확인하고 문항수준이 낮은 경우에도 동일한 경향이 유지되는지를 살펴보기 위함이다. 결과 분석은 이론적 배경에서 설명한 여러 지수를 통해 이루어진다.

인지진단모형은 여러 인지요소의 숙달여부를 동시에 추정하려는 목적으로 설계된 모형이긴 하지만 적은 문항 수로 너무 많은 인지요소를 추정하면 오류가 생겨 신뢰할 만한 결과를 얻지 못하고 의미 없는 추정에 그칠 수 있다. 그러나 검사 내에서 확인할 수 있는 인지요소 수가 적으면 평가 적용이 제한적일 수밖에 없다. 이에 이 모의실험은 인지요소의

수를 가능한 다양하게 설정하였다. 동일한 피험자 수와 비슷한 수준의 문항 수를 사용하는 검사 내에서 측정하는 인지요소의 수만 많아지면 상대적으로 부정확한 추정치를 얻을 수 있음을 고려하여 숙달여부를 판별할 인지요소의 수에 따라 적절한 q벡터 및 문항 수를 설정하되 측정하는 인지요소 수의 증가가 평가모형의 적합성 판정에까지 영향을 미치는지 확인한다. 모의실험에서 설정한 인지요소의 수는 3개, 5개, 그리고 8개이고 측정하는 인지요소의 수가 많더라도 한 문항에서 측정하는 인지요소의 수는 비슷한 수준을 유지하도록 설계하였다. 각 검사에서 피험자는 1000명이고 모의실험을 통해 좀 더 정확한 정보를 얻고자 1000명에 대한 검사를 1000번 반복하여 시행하여 결과 값의 평균을 사용한다. <표 III-1>은 검사에서 측정하는 인지요소 수가 3개인 경우의 모의실험 설계 구조를 간략히 나타낸 것이다. 인지요소의 수가 5개, 8개인 경우의 구조도 이와 동일하다.

<표 III-1> 인지요소 수가 3개일 때 모의실험 설계 구조

인지요소 수	생성모형 (1000번 반복)	문항수준 (slip, guess 모수로 결정)	추정 (1000번 반복)
3	DINA	high	DINA
			DINO
			ACDM
		low	DINA
			DINO
			ACDM
	DINO	high	DINA
			DINO
			ACDM
		low	DINA
			DINO
			ACDM
ACDM	high	DINA	
		DINO	
		ACDM	
	low	DINA	
		DINO	
		ACDM	

### 1) 인지요소가 3개인 경우

세 개의 인지요소를 추정하는 검사를 시행할 때, 피험자는 각각의 인지요소를 숙달했는지 안했는지에 따라 다음과 같은  $2^3 = 8$ 가지 패턴을 가질 수 있다.

(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0), (1, 1, 1)

문항은 (0, 0, 0)을 제외한 나머지 7가지 패턴의  $q$ 벡터를 3번씩 측정하도록 설계하여 각각의 인지요소에 대한 정보를 얻을 수 있도록 하였다. 만약, 문항 수가 많다 하더라도 각 인지요소를 측정하는  $q$ 벡터의 종류가 하나뿐이라면, 예를 들어 세 가지 인지요소 중 첫 번째와 두 번째 인지요소를 측정하는  $q$ 벡터가 (1, 1, 0)만 사용된 경우, 첫 번째 인지요소와 두 번째 인지요소는 항상 같이 측정되기 때문에 각각의 숙달여부에 대해서는 판정하기 어려워진다. 즉 각 인지요소에 대해 적정한 문항 수를 가지는 것도 중요하지만, 측정하는  $q$ 벡터의 인지요소 조합을 교차적으로 설정하는 것 또한 매우 중요하다. 세 가지 인지요소를 추정하는 경우의 Q행렬은 <표 III-2>와 같다.

모의실험에서 피험자의 인지요소 패턴 참값은 가능한 8개의 패턴이 동일한  $\frac{1}{8} = 0.125$ 의 확률로 선정되도록 설정하였으며 0.125의 확률에 따른 1000번의 시행으로 피험자 1000명의 인지요소 패턴을 결정하였다. 피험자 1000명의 응답 데이터를 만들어내기 위해서는 생성 모형의 종류에 관계없이, 피험자의 인지요소 패턴과 Q행렬 그리고 slip, guess 모수를 필요하다. 3개의 인지요소가 갖는 인지요소 패턴의 종류와 설정한 Q행렬은 생성모형에 관계없이 위에서 설명한 방식을 동일하게 적용하였다. 검사도구의 수준, 즉 문항수준(item quality)을 판단하는 기준으로 여겨지는 slip, guess 모수는 크게 두 분류로 나누어서 적용하였다. slip, guess 모수

가 큰 값을 가진다고 가정하면 알면서도 틀리거나 모르면서도 맞출 확률을 높아지기 때문에 문항수준이 낮아지고, 모수가 작은 값을 가지면 알면서도 틀리거나 모르면서도 맞출 확률이 낮아지기 때문에 문항수준이 높아진다고 할 수 있다. 높은 문항수준은 0.0에서 0.2사이의 값을, 낮은 문항수준은 0.1에서 0.3사이의 값을 사용하였다.

〈표 III-2〉 인지요소 3개를 측정하는 검사지의 Q행렬

문항	인지요소1	인지요소2	인지요소3	합계
1	1	0	0	1
2	1	0	0	1
3	1	0	0	1
4	0	1	0	1
5	0	1	0	1
6	0	1	0	1
7	0	0	1	1
8	0	0	1	1
9	0	0	1	1
10	1	1	0	2
11	1	1	0	2
12	1	1	0	2
13	1	0	1	2
14	1	0	1	2
15	1	0	1	2
16	0	1	1	2
17	0	1	1	2
18	0	1	1	2
19	1	1	1	3
20	1	1	1	3
21	1	1	1	3

## 2) 인지요소가 5개인 경우

5개의 인지요소를 측정하는 경우, 각 인지요소의 숙달여부에 따라 다

음과 같은  $2^5 = 32$ 개의 인지요소 패턴이 존재한다.

$$(0, 0, 0, 0, 0), (0, 0, 0, 0, 1), (0, 0, 0, 1, 0), \dots, (1, 1, 1, 1, 1)$$

이 때  $(0, 0, 0, 0, 0)$ 을 측정하는 문항은 아무의미가 없으므로 3개의 인지요소를 측정할 때와 같은 방법으로 각  $q$ 벡터를 측정하는 문항을 3개씩 배정한다고 가정하면  $31 \times 3 = 93$ 개의 문항으로 이루어진 검사도구를 구성할 수 있다. 학교에서 일반적인 정기고사가 대략 25문제 내외이고 연구에서 사용하는 TIMSS 2007의 4학년 수학 4번 검사지도 25문제로 구성되었음과 비교해보면 문항 수가 약 4배나 많고, 따라서 검사 시간도 3시간 또는 그 이상 필요하다. 모의실험은 실제 데이터에서 조정하기 어려운 실험 조건을 더 유연하게 조정할 수 있다는 장점이 있지만 현실성이 과도하게 떨어지는 도구의 사용은 활용도가 거의 없다는 점에 유의할 필요가 있다. 따라서 3개의 인지요소를 사용할 때와 같은 방식으로 Q행렬을 설정하기보다 다양한  $q$ 벡터 조합을 이용하되 좀 더 적은 수의 문항을 사용하도록 조정하였다. 인지요소 패턴 중 하나의 인지요소만 측정하는 패턴은  $5C1 = 5$ 개, 두 인지요소를 측정하는 패턴은  $5C2 = 10$ 개, 세 인지요소를 측정하는 패턴 또한  $5C3 = 10$ 개임을 이용하여 각 패턴을  $q$ 벡터로 설정하고 하나의 인지요소만 측정할 때의 패턴에 대해서만 두 문항씩 측정하는 검사지를 고려하면  $5 \times 2 + 10 + 10 = 30$ 개의 문항으로 이루어진 검사지를 구성할 수 있다. 이를 설명하는 Q행렬은 <표 III-3>로 나타내었다.

<표 III-3> 인지요소 5개를 측정하는 검사지의 Q행렬

문항	인지요소1	인지요소2	인지요소3	인지요소4	인지요소5	합계
1	1	0	0	0	0	1
2	1	0	0	0	0	1
3	0	1	0	0	0	1
4	0	1	0	0	0	1
5	0	0	1	0	0	1
6	0	0	1	0	0	1
7	0	0	0	1	0	1
8	0	0	0	1	0	1
9	0	0	0	0	1	1
10	0	0	0	0	1	1
11	1	1	0	0	0	2
12	1	0	1	0	0	2
13	1	0	0	1	0	2
14	1	0	0	0	1	2
15	0	1	1	0	0	2
16	0	1	0	1	0	2
17	0	1	0	0	1	2
18	0	0	1	1	0	2
19	0	0	1	0	1	2
20	0	0	0	1	1	2
21	1	1	1	0	0	3
22	1	1	0	1	0	3
23	1	1	0	0	1	3
24	1	0	1	1	0	3
25	1	0	1	0	1	3
26	1	0	0	1	1	3
27	0	1	1	1	0	3
28	0	1	1	0	1	3
29	0	1	0	1	1	3
30	0	0	1	1	1	3

### 3) 인지요소가 8개인 경우

8개의 인지요소를 추정하는 경우, 가능한 인지요소 패턴은 각 인지요소의 숙달여부에 따라 다음과 같은 형태로  $2^8 = 256$ 개가 존재한다.

$$(0, 0, 0, 0, 0, 0, 0, 0), (0, 0, 0, 0, 0, 0, 0, 1), \dots, (1, 1, 1, 1, 1, 1, 1, 1)$$

인지요소가 8개인 경우 역시  $(0, 0, 0, 0, 0, 0, 0, 0)$ 을 측정하는 문항은 아무의미가 없다. 앞서 논의한 바와 같이 가능한 모든  $q$ 벡터를 고려하는 것은 비현실적임을 감안하여 각각의 인지요소를 측정할 수 있는, 하나의 인지요소만 측정하는  $q$ 벡터와 두 개의 인지요소를 측정하는  $q$ 벡터를 사용하겠다. 이 때 문항의 수는  $8C_1 + 8C_2 = 8 + 28 = 36$ 개로 다소 많지만 일반 중, 고등학교의 정기고사를 기준으로 한 시간 반 정도 소요된다고 볼 수 있으므로 시행 가능한 수준의 검사 길이라 할 수 있다. 이에 따른 Q행렬은 <표 III-4>로 나타내었다.

<표 III-4> 인지요소 8개를 측정하는 검사지의 Q행렬

문항	인지1	인지2	인지3	인지4	인지5	인지6	인지7	인지8	합계
1	0	0	0	0	0	0	0	1	1
2	0	0	0	0	0	0	1	0	1
3	0	0	0	0	0	1	0	0	1
4	0	0	0	0	1	0	0	0	1
5	0	0	0	1	0	0	0	0	1
6	0	0	1	0	0	0	0	0	1
7	0	1	0	0	0	0	0	0	1
8	1	0	0	0	0	0	0	0	1
9	0	0	0	0	0	0	1	1	2
10	0	0	0	0	0	1	0	1	2
11	0	0	0	0	0	1	1	0	2
12	0	0	0	0	1	0	0	1	2
13	0	0	0	0	1	0	1	0	2
14	0	0	0	0	1	1	0	0	2
15	0	0	0	1	0	0	0	1	2
16	0	0	0	1	0	0	1	0	2
17	0	0	0	1	0	1	0	0	2
18	0	0	0	1	1	0	0	0	2
19	0	0	1	0	0	0	0	1	2
20	0	0	1	0	0	0	1	0	2
21	0	0	1	0	0	1	0	0	2
22	0	0	1	0	1	0	0	0	2
23	0	0	1	1	0	0	0	0	2
24	0	1	0	0	0	0	0	1	2
25	0	1	0	0	0	0	1	0	2
26	0	1	0	0	0	1	0	0	2
27	0	1	0	0	1	0	0	0	2
28	0	1	0	1	0	0	0	0	2
29	0	1	1	0	0	0	0	0	2
30	1	0	0	0	0	0	0	1	2
31	1	0	0	0	0	0	1	0	2
32	1	0	0	0	0	1	0	0	2
33	1	0	0	0	1	0	0	0	2
34	1	0	0	1	0	0	0	0	2
35	1	0	1	0	0	0	0	0	2
36	1	1	0	0	0	0	0	0	2

## 2.2 여러 인지진단모형을 혼합한 모의실험 설계

앞 절에서 설명한 첫 번째 모의실험에서는 하나의 검사에 해당하는 문항반응을 생성하기 위해 각각의 인지진단모형을 사용하였다. 이 절에서는 한 검사의 문항반응을 생성함에 있어 세 인지진단모형을 혼합하여 사용한다. 이러한 설계는 실제 시험에서 각 문항이 요구하는 인지요소가 무엇인지에 따라 여러 인지요소의 결합 방식, 그들의 관계성이 항상 같지 않을 수 있다는 의문점을 반영한 것이다. 검사에서 측정하는 인지요소들 전체를 하나의 집합이라 하면, 각 문항에서 측정하는 인지요소들은 그 집합의 부분집합이라 할 수 있다. 만약 어떤 부분집합에 계산능력에 관한 인지요소만 있다면, 모든 인지요소를 아는 것이 정확한 결과 계산에 중요하게 작용할 것이다. 한 편, 특정 문항에서 요구하는 인지요소 부분집합이 여러 문제해결전략을 나타내는 것이라면, 그 중 일부만 알아도 문항을 해결할 수 있다. 학교현장에서 매 수업마다 진행되는 형성평가와 같이 상대적으로 적은 범위를 묻는 시험은 검사지 내 인지요소 관계의 통일성을 가질 수 있지만, 학교에서 중요하게 다뤄지는 단원평가나 정기고사와 같이 더 넓은 영역을 아우르는 많은 인지요소를 요구하는 시험일수록 문항 별로 요구하는 인지요소의 결합방식이 상이할 수 있다.

이론적 배경에서 설명한 바와 같이, Q행렬은 문항이 측정하는 인지요소에 따라 문항과 인지요소 사이의 관계를 나타낸 것으로, 문항을 따라 q벡터 단위로 나눌 수 있다. 이를 이용하면 Q행렬 전체를 하나의 생성모형에 적용하는 대신, Q행렬을 구성하는 세 개의 q벡터 집합으로 나누고 각 집합에 대해 별개의 인지진단모형이 응답데이터를 생성하도록 할 수 있다. 이로써 한 검사의 응답데이터를 생성하는데 여러 인지진단모형을 혼합할 수 있게 된다. 1000명의 피험자 수, 1000세트의 인지프로파일 참값 생성, 1000세트의 응답데이터 생성, 그리고 높은 문항수준의 검사지와 낮은 문항수준의 검사지를 가정하는 등의 설계구조는 앞 절의 내용과 동일하다. 하나의 생성 방식에 대해 세 가지 인지진단모형 각각의 추정

을 진행하는 방식 또한 동일하다. 2절의 모의실험 설계구조는 <표 III-5>와 같다.

<표 III-5> 인지요소 수가 3개, 5개, 8개일 때 모의실험 설계구조

인지요소 수	생성모형 (1000번반복)	문항수준 (slip, guess 모수로 결정)	추정 (1000번반복)
3	DINA DINO ACDM	high	DINA
			DINO
			ACDM
		low	DINA
			DINO
			ACDM
5	DINA DINO ACDM	high	DINA
			DINO
			ACDM
		low	DINA
			DINO
			ACDM
8	DINA DINO ACDM	high	DINA
			DINO
			ACDM
		low	DINA
			DINO
			ACDM

## 1) 인지요소가 3개인 경우

앞 절에서 인지요소가 3개일 때의 Q행렬은  $(0, 0, 0)$ 을 제외하고 가능한 인지요소 패턴을 세 개씩 나열하는 구조로 제시하였다. 이 절에서는 하나의 검사지에 대한 응답데이터를 세 가지 인지진단모형으로 생성하기 때문에 q벡터의 적절한 분배가 요구된다. 인지요소가 3개인 경우의 기존 Q행렬이 동일한 q벡터를 세 번씩 반복적으로 사용하는 구조이므로 균등한 분배를 위해 각각의 q벡터를 삼등분 하는 방식을 채택한다. 즉, 각각의 인지진단모형에  $(0, 0, 0)$ 을 제외한 가능한 모든 인지요소패턴이 하나씩 해당되도록 한다. 하나의 인지진단모형이 응답데이터를 생성하기 위해 사용하는 q벡터 7가지는 다음과 같다.

$$(0, 0, 1), (0, 1, 0), (1, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0), (1, 1, 1)$$

이러한 q벡터 집합을 하나의 세트라 하면 DINA 모형, DINO 모형, 그리고 ACDM 모형 각각이 각 세트에 대해 높은 문항수준에서와 낮은 문항수준에서의 응답데이터를 생성하게 된다. 한 세트에 7개의 q벡터가 있으므로 각 모형별로 7문항씩, 총 21문항에 대한 응답데이터를 생성하는 것이다. 앞 절에서 인지요소 수가 3개인 경우와 문항의 순서만 바뀌었을 뿐 기본적인 Q행렬의 구성요소는 동일하다. 이 절의 모의실험을 통해 보고자 한 것은 문항별로도 생성모형과 생성모형이 아닌 모형 간 인지요소 관계 차이에 따른 추정치가 달라지는지에 관한 것이므로, 한 검사지 안에서의 문항 수준은 높든지 낮든지 한 가지 성질로 통일한다. 혼합검사지의 Q행렬은 다음의 <표 III-6>로 나타내었다.

<표 III-6> 인지요소 3개를 측정하는 혼합 검사지의 Q행렬

생성모형	문항	인지요소1	인지요소2	인지요소3	합계
DINA	1	1	0	0	1
	2	0	1	0	1
	3	0	0	1	1
	4	1	1	0	2
	5	1	0	1	2
	6	0	1	1	2
	7	1	1	1	3
DINO	8	1	0	0	1
	9	0	1	0	1
	10	0	0	1	1
	11	1	1	0	2
	12	1	0	1	2
	13	0	1	1	2
	14	1	1	1	3
ACDM	15	1	0	0	1
	16	0	1	0	1
	17	0	0	1	1
	18	1	1	0	2
	19	1	0	1	2
	20	0	1	1	2
	21	1	1	1	3

## 2) 인지요소가 5개인 경우

인지요소가 5개일 때 Q행렬은 앞 절에서 인지요소 한 개짜리 10문제, 두 개짜리 10문제, 3개짜리 10문제로 구성하였다. 10문제는 3개의 인지진단모형으로 나누어 분배하기 부자연스러운 측면이 있다. 단, 모의실험에서는 실제 문항을 사용하는 것이 아니기 때문에 일정 범위 내로 문항 수준의 통일이 가능하고 인지요소 또한 그 순서를 바꿔 재배치한다 해도

아무런 문제가 없이 대칭적이다. 따라서 각 인지진단모형이 반드시 동일한 q벡터를 모두 사용해야만 하는 것은 아니다. 이 절의 의의가 인지요소 간 관계성, 문항해결을 위해서는 인지요소를 다 알아야 하는지, 일부만 알아도 되는지, 선형적으로 통합되는지를 보고자 하는 것이기 때문에 문항에서 측정하는 인지요소가 1개인 경우보다는 2개, 3개인 경우에 보다 초점을 두고자 Q행렬을 재편성하였다. 문항에서 측정하는 인지요소가 1개인 경우는 생성모형과 추정모형의 적합성이 반드시 일치하지 않을 수도 있고 인지요소가 여러 개 있어 그 관계성이 필연적으로 작용하는 경우보다 불분명 할 수도 있지만, 인지요소가 더 적은 경우와 더 많은 경우인 3개, 8개의 모의실험에서 확인 가능하기 때문에 인지요소가 5개인 경우에 한 해 측정하는 인지요소 수가 2개, 3개인 문항에 더욱 초점을 맞추었다. 이에 모든 문항이 인지요소 두 개를 측정하는 검사도구 하나와 인지요소 세 개를 측정하는 검사도구 하나를 제작하여 인지요소가 5개일 때에 한해 두 검사에 대한 추정을 시행하였다.

이에 따라 Q행렬 또한 두 가지로 존재한다. 첫 번째 Q행렬은 각 인지진단모형이  $5C_2=10$ 개의 동일한 q벡터를 통해 응답데이터를 생성하게끔 한다. 동일한 q벡터는 세 인지진단모형을 통해 응답데이터 생성에 사용되는데, 높은 문항수준과 낮은 문항수준의 두 가지 응답데이터로 생성한다. 인지요소가 3개인 경우와 마찬가지로, 인지진단모형만 혼합할 뿐 문항수준은 혼합하지 않는다. 두 검사 도구를 통해 계산하는 추정치는 문항수준에서의 값이기 때문에 문항수준을 혼합한 검사를 따로 마련해야 할 필요는 없다고 결정하였다. 문항마다 두 개의 인지요소를 측정하는 검사지의 Q행렬은 <표 III-7>에서 확인 가능하다. 두 번째 Q행렬에서도 각 인지진단모형이 동일한 q벡터를 사용하여 응답데이터를 생성하도록 설정하였다. 단, 각 문항에서 인지요소 두 개가 아닌 세 개를 측정하도록 하였다. 인지요소를 하나만 측정할 때보다 두 개를 측정할 때가 각 인지진단모형이 가정하는 관계성이 더 잘 드러날 것이라는 예상과 함께, 측정하는 인지요소 수가 세 개로 두 개보다 많다면 이 또한 추정치에 차이가 있을 것이라는 예상 또한 확인하고자 한다. 문항에서 측정하는 인

지요소가 3개인 경우 역시  $5C3 = 10$ 개로 q벡터의 수가 동일하고, 따라서 Q행렬의 수, 검사지의 문항 수도 동일하다. 두 번째 검사지의 Q행렬은 <표 III-8>에서 확인 가능하다.

<표 III-7> 인지요소 5개를 측정하는 혼합 검사지의 Q행렬-1

생성모형	문항	인지1	인지2	인지3	인지4	인지5	합계
DINA	1	1	1	0	0	0	2
	2	1	0	1	0	0	2
	3	1	0	0	1	0	2
	4	1	0	0	0	1	2
	5	0	1	1	0	0	2
	6	0	1	0	1	0	2
	7	0	1	0	0	1	2
	8	0	0	1	1	0	2
	9	0	0	1	0	1	2
	10	0	0	0	1	1	2
DINO	11	1	1	0	0	0	2
	12	1	0	1	0	0	2
	13	1	0	0	1	0	2
	14	1	0	0	0	1	2
	15	0	1	1	0	0	2
	16	0	1	0	1	0	2
	17	0	1	0	0	1	2
	18	0	0	1	1	0	2
	19	0	0	1	0	1	2
	20	0	0	0	1	1	2
ACDM	21	1	1	0	0	0	2
	22	1	0	1	0	0	2
	23	1	0	0	1	0	2
	24	1	0	0	0	1	2
	25	0	1	1	0	0	2
	26	0	1	0	1	0	2
	27	0	1	0	0	1	2
	28	0	0	1	1	0	2
	29	0	0	1	0	1	2
	30	0	0	0	1	1	2

<표 III-8> 인지요소 5개를 측정하는 혼합 검사지의 Q행렬-2

생성모형	문항	인지1	인지2	인지3	인지4	인지5	합계
DINA	1	1	1	1	0	0	3
	2	1	1	0	1	0	3
	3	1	1	0	0	1	3
	4	1	0	1	1	0	3
	5	1	0	1	0	1	3
	6	1	0	0	1	1	3
	7	0	1	1	1	0	3
	8	0	1	1	0	1	3
	9	0	1	0	1	1	3
	10	0	0	1	1	1	3
DINO	11	1	1	1	0	0	3
	12	1	1	0	1	0	3
	13	1	1	0	0	1	3
	14	1	0	1	1	0	3
	15	1	0	1	0	1	3
	16	1	0	0	1	1	3
	17	0	1	1	1	0	3
	18	0	1	1	0	1	3
	19	0	1	0	1	1	3
	20	0	0	1	1	1	3
ACDM	21	1	1	1	0	0	3
	22	1	1	0	1	0	3
	23	1	1	0	0	1	3
	24	1	0	1	1	0	3
	25	1	0	1	0	1	3
	26	1	0	0	1	1	3
	27	0	1	1	1	0	3
	28	0	1	1	0	1	3
	29	0	1	0	1	1	3
	30	0	0	1	1	1	3

### 3) 인지요소가 8개인 경우

검사에서 측정하는 인지요소가 8개인 경우의 Q행렬은 앞 절에서  $8C1 + 8C2 = 8 + 28 = 36$ 개로 제시하였다. 인지요소가 8개인 경우는  $8C1 = 8$ 개를 제외하면  $8C2 = 28$ 개,  $8C3 = 56$ 개로  $28 \times 3 = 84$ 개,  $56 \times 3 = 168$ 개여서 동일한 q벡터를 모든 인지진단모형에 적용할 수도 없고, q벡터를 나누어 사용하려고 해도 문항에서 측정하는 인지요소 수가 3개인 경우는 검사지 길이가  $8C3 = 56$ 개, 4개인 경우는  $8C4 = 70$ 개가 되어 현실성이 떨어진다는 문제점이 있다. 따라서 인지요소가 8개인 경우는 기존의 Q행렬을 그대로 유지하여 인지요소가 1개, 2개인 경우에 대해 측정하고 각 q벡터를 각 인지진단모형별로 분배한다. 문항에서 측정하는 인지요소가 1개인 경우의 8문항을 3, 3, 2개로, 측정하는 인지요소가 2개인 경우의 28문항을 9, 9, 10개로 나눈다. 앞서 말한 것과 같이 인지요소의 번호는 순열을 통해 재배치하여도 전혀 상관이 없기 때문에 동일한 q벡터를 사용하지 못하는 것에 대해 문제가 없을 것으로 가정하였다. 혼합 모형에 대한 인지요소 8개짜리 Q행렬은 <표 III-9>로 나타났다. 단, 앞 절에서 Q행렬과 동일한 q벡터를 사용하지만 문항의 순서는 다르다.

혼합 생성모형은 한 검사지에 세 인지진단모형을 모두 적용하여 응답 데이터를 생성하기 때문에, 앞 절과는 검사지의 수가 다르고 따라서 추정의 수도 다르다. 앞 절에서는 하나의 인지요소 수, 그리고 Q행렬에 대해 각각의 생성모형이 있고 문항 수준에 따라  $3 \times 2 = 6$ 개의 검사지를 설정하였다. 각 검사지는 1000명의 피험자가 응시한다고 가정하여 1000명에 대한 응답데이터를 얻고, 이를 1000번 반복한다. 이렇게 얻은 1000명의 1000세트 응답데이터는 세 인지진단모형에 의해 추정되고, 하나의 인지요소 수에 대해 총 18000세트의 추정 데이터를 얻는다.

혼합 생성모형은 각 인지요소 수와 Q행렬에 대해 문항수준만 두 가지를 고려하므로 총  $1 \times 2 = 2$ 의 검사지를 구성한다. 피험자의 수, 검사의

반복은 동일하다. 즉, 각 검사지에 대해 1000명의 피험자로 1000세트 응답데이터를 생성하며 생성된 응답데이터가 각 인지진단모형에 의해 추정되는 방식은 같으므로 총  $2 \times 1000 \times 3 = 6000$ 세트의 추정 데이터를 얻는다. 이는 앞 절과 비교하면  $\frac{1}{3}$ 에 해당하는 데이터 수를 얻게끔 한다. 두 모의실험의 목적 및 성격차이로 인해 설계 또한 그 규모를 달리한다. 이 절의 모의실험은 검사지 내 각 문항에 대한 인지진단모형 적합성을 판정하는데 그 의의가 있다.

<표 III-9> 인지요소 8개를 측정하는 혼합 검사지의 Q행렬

생성 모형	문항	인지1	인지2	인지3	인지4	인지5	인지6	인지7	인지8	합계
DINA	1	0	0	0	0	0	0	0	1	1
	2	0	0	0	0	0	0	1	0	1
	3	0	0	0	0	0	1	0	0	1
	4	0	0	0	0	0	0	1	1	2
	5	0	0	0	0	0	1	0	1	2
	6	0	0	0	0	0	1	1	0	2
	7	0	0	0	0	1	0	0	1	2
	8	0	0	0	0	1	0	1	0	2
	9	0	0	0	0	1	1	0	0	2
	10	0	0	0	1	0	0	0	1	2
	11	0	0	0	1	0	0	1	0	2
	12	0	0	0	1	0	1	0	0	2
DINO	13	0	0	0	0	1	0	0	0	1
	14	0	0	0	1	0	0	0	0	1
	15	0	0	1	0	0	0	0	0	1
	16	0	0	0	1	1	0	0	0	2
	17	0	0	1	0	0	0	0	1	2
	18	0	0	1	0	0	0	1	0	2
	19	0	0	1	0	0	1	0	0	2
	20	0	0	1	0	1	0	0	0	2
	21	0	0	1	1	0	0	0	0	2
	22	0	1	0	0	0	0	0	1	2
	23	0	1	0	0	0	0	1	0	2
	24	0	1	0	0	0	1	0	0	2
ACDM	25	0	1	0	0	0	0	0	0	1
	26	1	0	0	0	0	0	0	0	1
	27	0	1	0	0	1	0	0	0	2
	28	0	1	0	1	0	0	0	0	2
	29	0	1	1	0	0	0	0	0	2
	30	1	0	0	0	0	0	0	1	2
	31	1	0	0	0	0	0	1	0	2
	32	1	0	0	0	0	1	0	0	2
	33	1	0	0	0	1	0	0	0	2
	34	1	0	0	1	0	0	0	0	2
	35	1	0	1	0	0	0	0	0	2
	36	1	1	0	0	0	0	0	0	2

## 2.3 TIMSS 2007 데이터

TIMSS 2007은 기본적으로 의도된 교육과정이 교사를 통해 수업에서 잘 실현되고 있는지, 학생의 성취로 이어졌는지를 단계적으로 고려하는 평가이다. 즉 TIMSS 2007의 평가는 학생의 현재 수준에 대한 판단의 근거로써의 정보를 줄뿐만 아니라 교육과정의 실현 과정 전반을 살펴보고 피드백 한다는 데 그 의의가 있다. 이는 역량평가가 교육과정의 설계와 수업을 실행하는 교사의 진행방식, 학생의 주도적인 학업 계획 등에 영향을 주는 데 목적이 있다는 점과 일맥상통한다. TIMSS 2007 검사가 내용 영역 뿐 아니라 인지영역을 구체적으로 평가하는 사실은 TIMSS 2007을 역량평가의 실례로 고려할만한 이론적 근거를 마련해 준다. 또한 내용 영역과 인지 영역으로의 구분은 측정하는 인지요소가 갖는 특징의 차이를 큰 틀에서 파악하기 쉽도록 한다. 개별 인지요소의 성격 및 특징은 측정하는 인지요소 사이의 작용 방식, 비보상적인지 보상적인지를 가정하는데 유용한 정보를 제공할 수 있다.

TIMSS 2007 4학년 수학 평가들의 내용 영역은 수(the Number), 도형과 측정(Geometric Shapes and Measures), 자료 표현(Data Display)으로 구성되며, 인지 영역은 사고 과정인 알기(the Knowing), 적용하기(Applying), 추론하기(Reasoning)로 구성된다. Lee, Park & Tylan(2011)에 따른 TIMSS 2007 4학년 수학의 영역별 평가 비율은 검사 문항 전체와 4번 검사지 각각에서 다음의 <표 III-10>와 같이 나타난다. <표 III-10>에서도 알 수 있듯이, 4번 검사지는 인지진단모형으로 분석하기 쉬운 응답데이터를 제공해준다는 강점 외에도 TIMSS 2007 검사지 전반이 평가하는 영역 별 분포에 대한 대표성을 가진다는 점에서 좋은 표본이라 할 수 있다. 연구대상에서 밝힌 바와 같이, 우리나라의 4학년 학생들은 TIMSS 2007에 참여하지 않았고, 이에 미국 학생들의 응답데이터 자료를 분석한다.

〈표 III-10〉 영역에 따른 문항 분포

과목	평가영역	인지요소	문항전체	4번검사지
수학	내용영역	수	50%	44%
		도형과 측정	35%	32%
		자료 표현	15%	24%
	인지영역	알기	40%	40%
		적용하기	40%	40%
		추론하기	20%	20%

### 1) Q행렬

인지진단평가를 실제데이터에 적용함에 있어서 타당성이 떨어지는 Q행렬의 사용으로 추정된 결과해석은 타당하지 못할 수 있다(Rupp & Templin, 2008). 실제 데이터에는 모의실험과 같이 생성 모형이나 생성 Q행렬이 따로 존재하지 않기 때문에 올바른 Q행렬이라는 개념은 없지만 타당한 Q행렬을 적용해야 한다. 이에 이 연구는 검증된 Q행렬을 사용하기 위해 선행연구인 Lee, Park & Tylan(2011)의 Q행렬을 그대로 적용한다. 이 연구에서는 Lee, Park & Tylan(2011)의 Q행렬 뿐 아니라 그 대상이 되는 4번 검사지, 그리고 같은 국가 학생들의 응답데이터를 사용하기 때문에 그 타당성을 보장받는다고 할 수 있다.

실제 데이터에 대한 결과를 심층적으로 해석하기 위해서는 Q행렬의 구조와 더불어 인지요소 각각의 내용을 알아야 한다. Lee, Park & Tylan(2011)은 내용 영역에 따른 평가 요소(Mullis et al., 2005)를 참고하여 TIMSS 2007이 수학과목에서 4학년용 대상으로 측정하는 인지요소 기준틀을 〈표 III-11〉와 같이 작성하였다. 또한 인지요소 기준을 따라 4번

검사지에 대한 Q 행렬을 작성하였는데, 다섯 가지 가능한 Q행렬 후보들을 합의하는 과정을 거쳤다고 설명하고 있다. 또한 만일 하나의 문항을 해결하는데 다양한 전략의 사용이 가능하다면 가장 주된, 문제풀이에 자주 이용되는 전략을 기준으로 Q행렬을 작성하였다고 밝히고 있다. 사용한 Q 행렬은 <표 III-12>와 같다. ‘문항’ 오른쪽에 표시된 1, 2, ..., 15는 인지요소 1, 인지요소 2, ..., 인지요소 15를 의미한다. 문항 마지막 행의 합계는 검사지 내에서 각 인지요소가 측정되는 횟수를, 인지요소 마지막 열의 합계는 각 문항이 측정하는 인지요소의 수를 나타낸다.

〈표 III-11〉 Lee, Park & Tylan(2011)에서 사용한 인지요소 틀

Number (N)	Whole Numbers (4)	1. Representing, comparing, and ordering whole numbers as well as demonstrating knowledge of place value.
		2. Recognize multiples, computing with whole numbers using the four operations, and estimating computations.
		3. Solve problems, including those set in real life contexts (for example, measurement and money problems).
		4. Solve problems involving proportions.
	Fractions and Decimals (2)	5. Recognize, represent, and understand fractions and decimals as parts of a whole and their equivalents.
		6. Solve problems involving simple fractions and decimals including their addition and subtraction.
	Number Sentences with Whole Numbers (1)	7. Find the missing number or operation and model simple situations involving unknowns in number sentence or expressions.
	Patterns and Relationships (1)	8. Describe relationships in patterns and their extensions; generate pairs of whole numbers by a given rule and identify a rule for every relationship given pairs of whole numbers.
Geometric Shapes & Measurem ent (GM)	Lines and Angles (1)	9. Measure, estimate, and understand properties of lines and angles and be able to draw them.
	Two- and Three- dimensional Shapes (2)	10. Classify, compare, and recognize geometric figures and shapes and their relationships and elementary properties.
		11. Calculate and estimate perimeters, area, and volume.
	Location and Movement (1)	12. Locate points in an informal coordinate to recognize and draw figures and their movement.
Data & Display (DD)	Reading and Interpreting (2)	13. Read data from tables, pictographs, bar graphs, and pie charts.
		14. Comparing and understanding how to use information from data.
	Organizing and Representing (1)	15. Understanding different representations and organizing data using tables, pictographs, and bar graphs.

<표 III-12> Lee, Park & Tylan(2011)에서 사용한 Q행렬

문항	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	합계
M041052	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2
M041056	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
M041069	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	3
M041076	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	2
M041281	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	3
M041164	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	2
M041146	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	3
M041152	1	1	1	0	0	0	0	0	0	1	1	0	0	0	0	5
M041258A	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
M041258B	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	2
M041131	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	4
M041275	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	3
M041186	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0	4
M041336	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0	6
M031303	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2
M031309	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2
M031245	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	2
M031242A	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	3
M031242B	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	3
M031242C	0	1	1	0	0	0	0	1	0	0	0	0	0	1	0	4
M031247	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	3
M031219	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	3
M031173	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2
M031085	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
M031172	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	4
합계	6	16	11	3	3	2	2	3	3	7	2	3	4	3	2	70

## 2.4 TIMSS 2007 데이터에 기반한 모의실험 설계

처음 두 절에서 소개한 모의실험은 단일 모형을 생성모형으로 하는 검사와 여러 모형을 혼합하여 생성하는 검사의 두 단계로 구성되었다. 특히 혼합하여 생성하는 경우 각 모형이 균등한 문항수의 응답데이터를 생성할 뿐 아니라 문항 당 측정하는 인지요소의 수도 동일하게 배분되었다. 이러한 설계는 특정 모형에 비중을 두지 않기 위함이나 다소 현실성이 떨어진다는 문제점이 있다. 이에 이 절에서는 TIMSS 2007 데이터의 추정치를 바탕으로 보다 현실적인 모의검사를 추가적으로 구성한다.

각 문항의 생성모형 결정은 TIMSS 2007 데이터의 추정치 중 RMSEA 값을 기준으로 한다. 이는 다른 적합도 지수와 달리 RMSEA만이 문항별 값을 계산해주기 때문이다. 즉, 각 문항별 RMSEA의 값을 비교하여 가장 작은 RMSEA 값을 갖도록 하는 추정모형을 모의검사의 응답데이터 생성모형으로 설정한다. 각 문항의 slip, guess 모수 역시 실제 데이터의 추정치를 그대로 사용한다. 모의검사의 생성모형 및 slip, guess 모수 선택은 <표 III-13>과 같다. 이 표에서 문항수준은  $1 - s_i - g_i$  값으로, 필요한 인지요소를 모두 숙달한 학생이 문항을 맞출 확률과 필요한 인지요소를 하나도 숙달하지 못한 학생이 문항을 맞출 확률의 차이를 나타낸다.

RMSEA 값이 문항별로 계산되기는 하지만 앞서 이론적 배경에서 설명한 바와 같이 특정 인지진단모형의 추정 내에서 계산되는 값이고 특히 정답률의 추정치와 관측치의 차이를 계산하는 데 각 인지진단모형의 문항반응함수를 사용한다. 결국 각 문항의 RMSEA 값을 구하더라도 특정 문항반응함수를 통해 하나의 검사 전체를 해석하여 얻어지는 값이기 때문에 하나의 추정모형 내에서 계산하는 문항 간 추정치는 종속될 수밖에 없다. 따라서 추정모형 간 적합성을 비교하기 위해 검사 전체의 RMSEA가 아닌 문항별 RMSEA를 비교하는 것은 유용하지 않다. 이 절에서 RMSEA 값을 이용해 생성모형을 선택함은 문항별 추정치를 제공하기 때문이다. 이 실험의 결과는 100번 반복시행의 평균값을 따른다.

<표 III-13> TIMSS 2007 데이터의 RMSEA 값에 따른 모의실험 설계구조

문항	추정모형			인지요소	문항수준	생성모형
	DINA	DINO	ACDM			
1	0.058	0.192	0.101	2	0.730	DINA
2	0.05	0.03	0.077	1	0.965	DINO
3	0.157	0.212	0.133	3	0.820	ACDM
4	0.086	0.053	0.105	2	0.909	DINO
5	0.202	0.186	0.115	3	0.731	ACDM
6	0.078	0.12	0.091	2	0.179	DINA
7	0.105	0.161	0.111	3	0.554	DINA
8	0.096	0.133	0.061	5	0.824	ACDM
9	0.067	0.222	0.11	1	0.600	DINA
10	0.119	0.173	0.142	2	0.418	DINA
11	0.135	0.216	0.111	4	0.366	ACDM
12	0.139	0.219	0.092	3	0.601	ACDM
13	0.106	0.117	0.096	4	0.754	ACDM
14	0.132	0.314	0.069	6	0.839	ACDM
15	0.121	0.223	0.125	2	0.390	DINA
16	0.134	0.189	0.109	2	0.892	ACDM
17	0.042	0.008	0.081	2	0.701	DINA
18	0.121	0.199	0.091	3	0.869	ACDM
19	0.1	0.086	0.01	3	0.964	ACDM
20	0.149	0.231	0.125	4	0.600	ACDM
21	0.126	0.17	0.098	3	0.369	ACDM
22	0.143	0.193	0.131	3	0.353	ACDM
23	0.156	0.229	0.119	2	0.802	ACDM
24	0.195	0.192	0.177	1	0.380	ACDM
25	0.075	0.012	0.109	4	0.984	DINO

## IV. 연구결과

### 1. 각각의 인지진단모형에 대한 모의실험 결과분석

#### 1.1 DINA 모형으로 생성한 응답데이터의 적합도 지수

##### 1) 적합도 지수

DINA 모형으로 생성한 응답데이터에 대한 추정결과는 인지요소 수에 관계없이 대체로 비슷한 경향을 보였다. 우선 절대 적합도 지수에 해당하는 MADcor, SRMSR은 문항 수준에 관계없이 생성모형에서 가장 좋은 값을 보였다. 그리고 극단적인 비보상적 모형인 DINA 모형이 생성모형인 만큼, 정반대의 방향으로 극단적인 보상적 모형 DINO에서 가장 안 좋은 추정치를 보였다. 추정치는 DINA, ACDM, 그리고 DINO 모형 순서로 좋게 나타났다. 또한 측정하는 인지요소의 수가 많아질수록, 문항 수준이 좋지 않을수록 모형 간 추정치 차이가 좁혀지는 양상을 보였다.

상대적합도 지수는 문항 수준에 더 영향을 받는 것으로 나타났다. 생성모형이어도 문항 수준이 낮으면 높은 문항수준의 검사지를 추정한 비생성모형, 심지어는 정반대의 가정을 갖는 DINO 모형보다도 좋지 못한 추정치를 갖는 것으로 드러났다. 문항 수준 내에서 상대적합도 지수 또한 DINA, ACDM, 그리고 DINO 모형 순서로 좋게 나타났다. 모든 인지요소를 알아야 한다고 가정하는 DINA 모형으로 응답데이터를 생성했기 때문에, 극단적 반대인 DINO 모형의 추정은 가장 안 좋게 나타났고, 두 모형 사이의 구조를 취하는 ACDM 모형이 중간 추정치를 나타냈다. 세 모형 간 추정하는 모수의 수가 동일하지는 않지만, 페널티에 의해 순위가 뒤바뀌기에는 deviance 값과 비교했을 때 모수의 수가 상대적으로 작기 때문에 순서를 바꿀만한 비중을 갖지 않는 것으로 해석된다.

특이한 점은 문항관련 절대적합도 지수인 RMSEA 값에 관한 것인데, 인지요소가 3개, 5개일 때는 다른 절대적합도 지수와 마찬가지로 문항 수준에 관계없이 생성모형에서 가장 좋게 나타났지만, 인지요소가 8개인 경우에는 문항수준이 낮은 경우에서 가장 좋은 두 값이 나왔고, 심지어 ACDM 모형 대신 DINO 모형의 추정치가 좋게 나타난 점이다. 이는 여러 인지진단모형을 혼합한 검사 결과의 문항별 추정치 비교에서 좀 더 이해할 수 있을 것이다. 간단히 살펴보면 문항에서 측정하는 인지요소의 수가 하나인 경우 생성모형과 추정모형 간 일관성이 없었고, 오히려 생성하지 않은 모형에서의 추정치가 더 좋게 나오는 경우가 많았다. 인지요소 수가 하나 보다 많은 경우에는 생성모형의 RMSEA 값이 가장 좋은 것으로 일관성 있게 나타났지만 인지요소 수가 증가하면서 RMSEA 값들의 차이가 매우 작아졌기 때문에 인지요소 수가 하나인 경우의 추정치와 합쳐지면서 순위의 구분이 희미해진 것으로 이해할 수 있다.

각 인지요소 수에서 적합도 지수는 3개, 5개, 8개 순서로 <표 IV-1>, <표 IV-2>, <표 IV-3>로 정리하였다.

<표 IV-1> 인지요소 3개일 때 DINA 모형으로 생성한 응답데이터의 적합도 지수

		MADcor	SRMSR	RMSEA	deviance	AIC	BIC	#parameters
high quality	DINA	0.017	0.022	0.020	16215.0	16894.2	17134.7	49
	DINO	0.146	0.193	0.103	21347.1	21852.6	22093.0	49
	ACDM	0.121	0.151	0.099	19354.5	19936.7	20250.8	64
low quality	DINA	0.022	0.027	0.023	23737.8	23213.8	23454.3	49
	DINO	0.063	0.079	0.061	25520.5	25246.6	25487.1	49
	ACDM	0.047	0.059	0.093	24769.2	24455.1	24769.2	64

<표 IV-2> 인지요소 5개일 때 DINA 모형으로 생성한 응답데이터의 적합도 지수

		MADcor	SRMSR	RMSEA	deviance	AIC	BIC	#parameters
high quality	DINA	0.019	0.024	0.045	24403.4	24585.4	25032.0	91
	DINO	0.099	0.129	0.117	31567.8	31749.8	32196.4	91
	ACDM	0.061	0.084	0.117	28828.4	29040.4	29560.6	106
low quality	DINA	0.023	0.028	0.049	34399.1	34581.1	35027.7	91
	DINO	0.045	0.057	0.060	36641.5	36823.5	37270.1	91
	ACDM	0.031	0.039	0.103	35836.9	36048.9	36569.1	106

<표 IV-3> 인지요소 8개일 때 DINA 모형으로 생성한 응답데이터의 적합도 지수

		MADcor	SRMSR	RMSEA	deviance	AIC	BIC	#parameters
high quality	DINA	0.019	0.024	0.177	32050.2	32704.2	34309.0	327
	DINO	0.067	0.092	0.173	41375.4	42029.4	43634.2	327
	ACDM	0.048	0.064	0.195	38086.3	38360.3	39032.7	137
low quality	DINA	0.022	0.028	0.155	42341.7	42995.7	44600.5	327
	DINO	0.037	0.046	0.103	44894.1	45548.1	47153.0	327
	ACDM	0.027	0.033	0.162	44040.2	44314.2	44986.5	137

## 2) 인지프로파일 분류 정확도

검사자가 측정하는 각 인지요소에 대한 marginal rate과 피험자에 대한 인지프로파일 분류 정확도, marginal rate 모두 생성 모형인 DINA모형, 그리고 ACDM모형과 DINO 모형 순의 정확도를 나타내었다. 인지요소 수가 적을 때는 문항 수준과 관계없이 생성모형의 정확도가 가장 좋았고, 인지요소 수가 많아지면 높은 문항수준에서 ACDM 모형 추정치가 낮은 문항수준에서 생성모형인 DINA의 추정치보다 좋게 나타나기도 했다. 그러나 인지요소의 구조를 정반대로 설정하는 DINO 모형의 추정치는 항상 가장 안 좋게 나타났다. 특히 인지프로파일 전체에 대한 정확도는 생성 모형이 항상 가장 좋은 추정치를 보였다. 측정하는 인지요소의 수가 8개

로 가장 많을 때 피험자의 프로파일 전체를 맞게 추정하는 비율은 절반 정도로 나타났다. 그러나 marginal rate은 90%에 가까운 수준으로 나타났다.

<표 IV-4> 인지요소 3개일 때 DINA로 생성한 응답데이터 인지요소 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
attribute 1	mle	0.983	0.770	0.880	0.949	0.798	0.883
	map	0.983	0.762	0.879	0.949	0.764	0.881
attribute 2	mle	0.990	0.738	0.939	0.958	0.841	0.909
	map	0.990	0.741	0.941	0.958	0.819	0.908
attribute 3	mle	0.993	0.917	0.948	0.923	0.727	0.808
	map	0.993	0.920	0.948	0.922	0.654	0.797

<표 IV-5> 인지요소 3개일 때 DINA로 생성한 응답데이터 프로파일 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
the attribute profile rate	mle	0.975	0.402	0.807	0.859	0.456	0.705
	map	0.975	0.423	0.809	0.859	0.366	0.693
the marginal profile rate	mle	0.989	0.808	0.922	0.944	0.789	0.867
	map	0.989	0.807	0.923	0.943	0.746	0.862

<표 IV-6> 인지요소 5개일 때 DINA로 생성한 응답데이터 인지요소 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
attribute 1	mle	0.986	0.896	0.936	0.937	0.839	0.887
	map	0.986	0.870	0.934	0.936	0.802	0.882
attribute 2	mle	0.977	0.797	0.902	0.896	0.715	0.810
	map	0.977	0.726	0.902	0.894	0.626	0.797
attribute 3	mle	0.981	0.820	0.924	0.928	0.794	0.870
	map	0.981	0.762	0.924	0.927	0.740	0.863
attribute 4	mle	0.975	0.822	0.893	0.926	0.794	0.867
	map	0.974	0.748	0.892	0.924	0.729	0.861
attribute 5	mle	0.990	0.927	0.963	0.908	0.752	0.830
	map	0.990	0.929	0.963	0.906	0.659	0.815

<표 IV-7> 인지요소 5개일 때 DINA로 생성한 응답데이터 프로파일 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
the attribute profile rate	mle	0.923	0.382	0.682	0.711	0.213	0.457
	map	0.922	0.251	0.681	0.706	0.104	0.425
the marginal profile rate	mle	0.982	0.853	0.923	0.919	0.779	0.853
	map	0.981	0.807	0.923	0.917	0.711	0.844

<표 IV-8> 인지요소 8개일 때 DINA로 생성한 응답데이터 인지요소 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
attribute 1	mle	0.983	0.851	0.916	0.913	0.819	0.865
	map	0.982	0.835	0.919	0.905	0.777	0.859
attribute 2	mle	0.975	0.715	0.879	0.871	0.706	0.800
	map	0.974	0.635	0.884	0.860	0.650	0.788
attribute 3	mle	0.988	0.936	0.946	0.904	0.764	0.846
	map	0.987	0.937	0.947	0.897	0.725	0.841
attribute 4	mle	0.968	0.721	0.836	0.905	0.784	0.852
	map	0.966	0.642	0.833	0.897	0.741	0.846
attribute 5	mle	0.977	0.709	0.885	0.906	0.732	0.841
	map	0.976	0.641	0.890	0.899	0.699	0.837
attribute 6	mle	0.986	0.900	0.952	0.902	0.740	0.836
	map	0.985	0.905	0.953	0.895	0.694	0.828
attribute 7	mle	0.973	0.769	0.869	0.890	0.706	0.817
	map	0.971	0.705	0.867	0.882	0.662	0.809
attribute 8	mle	0.979	0.799	0.903	0.892	0.731	0.825
	map	0.977	0.757	0.900	0.884	0.685	0.817

<표 IV-9> 인지요소 8개일 때 DINA로 생성한 응답데이터 프로파일 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
the attribute profile rate	mle	0.880	0.024	0.453	0.529	0.020	0.249
	map	0.880	0.560	0.453	0.491	0.011	0.121
the marginal profile rate	mle	0.979	0.800	0.898	0.898	0.748	0.835
	map	0.977	0.757	0.899	0.890	0.704	0.828

## 1.2 DINO 모형으로 생성한 응답데이터의 적합도 지수

### 1) 적합도 지수

생성모형이 DINO인 경우의 적합도 지수는 DINA 모형으로 생성했을 때와 모든 점에서 유사한 경향을 보인다. DINA 모형으로 응답데이터를 생성했을 때 인지요소의 관계성 설정이 정반대되는 DINO 모형의 추정이 가장 안 좋게 나타난 것과 같이, DINO 모형으로 생성한 응답데이터의 추정은 DINA 모형에서 가장 안 좋게 나타났다. DINO 모형은 응답데이터를 해석할 때 하나의 인지요소만 있어도 문항을 해결 했을 것으로 가정하는 것과 달리 DINA 모형은 모든 인지요소를 숙달했을 것으로 가정하기 때문에 완전히 반대된다.

모든 추정은 DINO, ACDM, DINA 모형 순으로 좋게 나타났고, 절대 적합도 지수는 문항 수준보다 생성모형과의 일치에, 상대 적합도 지수는 생성모형보다 문항 수준에 더 영향을 받는 것으로 나타났다. 검사에서 측정하는 인지요소 수가 많아지거나 문항수준이 좋지 않을수록 절대적합도 지수의 차이가 줄어드는 경향도 동일하게 나타났다. 또한 검사에서 추정하는 모수의 수가 모형별로 다르고, 인지요소 수가 증가할수록 그 차이가 커짐에도 기본적인 deviance 차이 때문에 패널티를 적용한 AIC, BIC에서도 그 순위가 바뀌지 않았다. 또한 DINA 모형에서처럼 인지요소가 8개인 경우의 RMSEA 값은 생성모형과 관계없는 결과를 나타냈다.

<표 IV-10> 인지요소 3개일 때 DINO 모형으로 생성한 응답데이터의 적합도 지수

		MADcor	SRMSR	RMSEA	deviance	AIC	BIC	#parameters
high quality	DINA	0.158	0.206	0.164	20878.6	20976.6	21217.1	49
	DINO	0.016	0.021	0.026	15320.7	15418.7	15659.1	49
	ACDM	0.090	0.127	0.194	19774.8	19902.8	20216.9	64
low quality	DINA	0.069	0.087	0.089	25467.1	25565.1	25805.5	49
	DINO	0.021	0.027	0.034	23363.2	23461.2	23701.7	49
	ACDM	0.038	0.051	0.132	24696.0	24824.0	25138.1	64

<표 IV-11> 인지요소 5개일 때 DINO 모형으로 생성한 응답데이터의 적합도 지수

		MADcor	SRMSR	RMSEA	deviance	AIC	BIC	#parameters
high quality	DINA	0.096	0.125	0.164	32184.1	32366.1	32812.7	91
	DINO	0.019	0.025	0.065	25264.1	25446.1	25892.7	91
	ACDM	0.053	0.074	0.180	30498.8	30710.8	31231.0	106
low quality	DINA	0.046	0.058	0.095	36747.7	36929.7	37376.3	91
	DINO	0.023	0.028	0.071	34486.8	34668.8	35115.4	91
	ACDM	0.027	0.035	0.139	35877.3	36089.3	36609.5	106

<표 IV-12> 인지요소 8개일 때 DINO 모형으로 생성한 응답데이터의 적합도 지수

		MADcor	SRMSR	RMSEA	deviance	AIC	BIC	#parameters
high quality	DINA	0.063	0.085	0.174	41791.8	42445.8	44050.7	327
	DINO	0.019	0.025	0.181	33267.8	33921.8	35526.6	327
	ACDM	0.036	0.046	0.210	39260.6	39534.6	40207.0	137
low quality	DINA	0.036	0.045	0.103	45606.4	46260.4	47865.2	327
	DINO	0.022	0.028	0.156	43161.2	43815.2	45420.0	327
	ACDM	0.025	0.031	0.165	44783.1	45057.1	45729.5	137

## 2) 인지프로파일 분류 정확도

DINO 모형으로 생성한 응답데이터의 인지요소별, 그리고 인지프로파일별 정확도는 모두 높은 문항 수준에서 생성모형으로 추정된 경우가 가장 높게 나타났다. DINO 모형과 인지요소 간 관계성을 가장 다르게 정의하는 DINA 모형의 정확도가 가장 낮았고, ACDM 모형은 DINO모형의 정확도와 DINA 모형의 정확도 사이의 값을 나타냈다. 높은 문항 수준에서 ACDM 모형의 분류 정확도와 낮은 문항 수준에서 DINO 모형의 분류 정확도는 비슷한 수준으로 나타났다. 인지요소의 수가 많아질수록 낮은 문항에서의 인지 프로파일 정확도는 생성모형 조차 매우 낮은 정확도를 보였다. DINA 모형의 경우와 비슷하게 절반 정도의 정확도를 보였다. 한편 marginal rate은 DINO 모형 역시 90%에 가까운 정확도를 보였다. 즉, 피험자의 인지프로파일 전체를 맞춘 비율은 인지요소 수가 많아질수

록 낮은 수준으로 떨어지지만, 각 피험자의 특정 인지요소 숙달여부를 맞추는 비율은 높게 유지된다고 할 수 있다.

<표 IV-13> 인지요소 3개일 때 DINO로 생성한 응답데이터 인지요소 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
attribute 1	mle	0.842	0.995	0.949	0.834	0.957	0.907
	map	0.843	0.995	0.947	0.812	0.956	0.899
attribute 2	mle	0.913	0.993	0.965	0.787	0.941	0.887
	map	0.911	0.993	0.964	0.745	0.941	0.876
attribute 3	mle	0.779	0.991	0.915	0.770	0.951	0.905
	map	0.779	0.991	0.909	0.735	0.951	0.893

<표 IV-14> 인지요소 3개일 때 DINO로 생성한 응답데이터 프로파일 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
the attribute profile rate	mle	0.537	0.980	0.838	0.421	0.871	0.730
	map	0.539	0.980	0.831	0.344	0.871	0.700
the marginal profile rate	mle	0.845	0.993	0.943	0.797	0.950	0.900
	map	0.845	0.993	0.940	0.764	0.949	0.889

<표 IV-15> 인지요소 5개일 때 DINO로 생성한 응답데이터 인지요소 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
attribute 1	mle	0.858	0.987	0.928	0.794	0.916	0.866
	map	0.792	0.987	0.903	0.719	0.914	0.857
attribute 2	mle	0.873	0.984	0.914	0.842	0.928	0.887
	map	0.842	0.984	0.899	0.802	0.926	0.882
attribute 3	mle	0.857	0.985	0.899	0.850	0.936	0.895
	map	0.805	0.985	0.871	0.815	0.935	0.887
attribute 4	mle	0.873	0.978	0.913	0.818	0.930	0.879
	map	0.811	0.977	0.890	0.772	0.929	0.864
attribute 5	mle	0.888	0.985	0.918	0.795	0.928	0.867
	map	0.849	0.985	0.908	0.718	0.927	0.850

<표 IV-16> 인지요소 5개일 때 DINO로 생성한 응답데이터 프로파일 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
the attribute profile rate	mle	0.450	0.929	0.632	0.328	0.731	0.514
	map	0.298	0.928	0.543	0.195	0.727	0.469
the marginal profile rate	mle	0.870	0.984	0.879	0.820	0.928	0.879
	map	0.820	0.984	0.894	0.765	0.926	0.868

<표 IV-17> 인지요소 8개일 때 DINO로 생성한 응답데이터 인지요소 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
attribute 1	mle	0.892	0.987	0.927	0.721	0.879	0.824
	map	0.863	0.986	0.914	0.665	0.869	0.814
attribute 2	mle	0.767	0.970	0.895	0.772	0.910	0.861
	map	0.721	0.968	0.879	0.730	0.903	0.853
attribute 3	mle	0.724	0.966	0.879	0.731	0.898	0.837
	map	0.663	0.964	0.859	0.687	0.890	0.824
attribute 4	mle	0.734	0.970	0.888	0.739	0.895	0.840
	map	0.662	0.967	0.861	0.695	0.887	0.827
attribute 5	mle	0.908	0.986	0.931	0.839	0.921	0.883
	map	0.895	0.985	0.915	0.800	0.914	0.873
attribute 6	mle	0.804	0.974	0.912	0.754	0.886	0.832
	map	0.755	0.971	0.897	0.698	0.876	0.808
attribute 7	mle	0.852	0.980	0.912	0.767	0.910	0.857
	map	0.830	0.979	0.882	0.739	0.903	0.848
attribute 8	mle	0.797	0.979	0.925	0.702	0.881	0.816
	map	0.763	0.978	0.911	0.651	0.872	0.808

<표 IV-18> 인지요소 3개일 때 DINO로 생성한 응답데이터 프로파일 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
the attribute profile rate	mle	0.033	0.859	0.467	0.025	0.521	0.249
	map	0.021	0.848	0.362	0.014	0.483	0.202
the marginal profile rate	mle	0.810	0.977	0.909	0.753	0.897	0.844
	map	0.769	0.975	0.890	0.708	0.889	0.832

### 1.3 ACDM 모형으로 생성한 응답데이터의 적합도 지수

#### 1) 적합도 지수

ACDM 모형의 경우도 앞서 살펴 본 두 모형의 추정치와 유사한 결과를 나타냈다. 절대 적합도 지수는 문항 수준과 관계없이 생성모형인 ACDM 모형으로 추정할 때가 가장 좋은 것으로 나타났다. 마찬가지로, 인지요소의 수가 많아지거나 문항의 수준이 좋지 않을 경우에는 생성모형과 다른 모형 사이의 추정치 차이가 줄어드는 경향이 나타났다. 상대 적합도 지수는 낮은 문항수준에서 생성모형의 추정정보는 높은 문항수준에서 다른 모형의 추정이 더 좋게 나타나, 좋은 수준의 문항을 사용하는 것이 중요하게 나타났다. 그러나 같은 검사지에 대해서는 여전히 생성모형의 추정치가 가장 좋기 때문에 검사 내 적합성 순위는 바뀌지 않는다. 인지요소가 8개일 때 RMSEA 값은 그 차이가 점점 줄어들어 순위가 역전되기도 하는 모습을 동일하게 보였다. 한 가지 눈여겨볼 점은, 생성모형인 ACDM를 제외한 두 모형, DINA, DINO 모형의 추정치가 전반적으로 매우 유사하게 나타났다는 사실이다. 두 모형의 추정치는 인지요소의수에 관계없이 일관성 있게 유사하여 거의 구별이 안 될 정도로 나타났다. 이를 통해 인지요소 간의 관계가 극단적인 두 모형 사이에서 어느 한 쪽과 더 가깝기보다 거의 정중앙에 위치해 있다고 볼 수 있다.

<표 IV-19> 인지요소 3개일 때 ACDM모형으로 생성한 응답데이터의 적합도 지수

		MADcor	SRMSR	RMSEA	deviance	AIC	BIC	#parameters
high quality	DINA	0.101	0.146	0.114	23605.6	23703.6	23944.1	49
	DINO	0.101	0.145	0.111	23684.0	23782.0	24022.5	49
	ACDM	0.016	0.021	0.023	20951.6	21079.6	21393.7	64
low quality	DINA	0.069	0.097	0.076	26105.3	26203.3	26443.8	49
	DINO	0.069	0.097	0.075	26126.1	26224.1	26464.6	49
	ACDM	0.019	0.024	0.024	24704.9	24832.9	25147.0	64

<표 IV-20> 인지요소 5개일 때 ACDM모형으로 생성한 응답데이터의 적합도 지수

		MADcor	SRMSR	RMSEA	deviance	AIC	BIC	#parameters
high quality	DINA	0.085	0.136	0.128	34903.6	35085.6	35532.2	91
	DINO	0.085	0.136	0.129	34851.8	35033.8	35480.4	91
	ACDM	0.017	0.023	0.054	29874.8	30086.8	30607.0	106
low quality	DINA	0.046	0.066	0.065	39127.9	39309.9	39756.5	91
	DINO	0.046	0.066	0.092	39116.9	39298.9	39745.5	91
	ACDM	0.021	0.026	0.047	37805.4	38017.4	38537.6	106

<표 IV-21> 인지요소 8개일 때 ACDM모형으로 생성한 응답데이터의 적합도 지수

		MADcor	SRMSR	RMSEA	deviance	AIC	BIC	#parameters
high quality	DINA	0.052	0.088	0.177	43541.9	44195.9	45800.7	327
	DINO	0.052	0.088	0.179	43520.2	44174.2	45779.1	327
	ACDM	0.019	0.025	0.198	39708.8	39982.8	40655.1	137
low quality	DINA	0.039	0.059	0.130	46643.9	47297.9	48902.7	327
	DINO	0.039	0.058	0.129	46656.0	47310.0	48914.8	327
	ACDM	0.021	0.027	0.146	45141.2	45415.2	46087.5	137

## 2) 인지프로파일 분류 정확도

ACDM 모형으로 응답데이터를 생성한 경우의 인지요소 및 프로파일 분류 정확도 역시 높은 문항수준에서 생성모형으로 추정했을 때 가장 좋게 나타났다. 낮은 문항 수준에서 생성모형의 추정치는 높은 문항 수준에서 DINA, DINO 모형의 추정치와 큰 차이 없이 비슷하게 나타났으나 전반적으로는 더 좋은 추정치를 보였다. DINA, DINO 모형의 여러 적합도 지수가 거의 동일한 수준으로 나타난 것과 같이 두 모형의 인지요소 및 프로파일 정확도 역시 매우 유사한 수준으로 나타났다. 인지요소의 수가 많아질수록 전반적인 분류 정확도가 떨어지나 프로파일 분류 정확도 외에는 근소한 차이이며, 프로파일에 관한 marginal rate은 ACDM에서도 90%정도의 수준으로 나타났다.

<표 IV-22> 인지요소 3개일 때 ACDM으로 생성한 응답데이터 인지요소 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
attribute 1	mle	0.970	0.968	0.992	0.915	0.912	0.964
	map	0.968	0.965	0.992	0.905	0.902	0.964
attribute 2	mle	0.933	0.919	0.986	0.870	0.867	0.950
	map	0.924	0.906	0.986	0.845	0.841	0.949
attribute 3	mle	0.952	0.952	0.990	0.911	0.910	0.962
	map	0.947	0.946	0.990	0.898	0.897	0.962

<표 IV-23> 인지요소 3개일 때 ACDM으로 생성한 응답데이터 프로파일 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
the attribute profile rate	mle	0.859	0.842	0.968	0.715	0.708	0.883
	map	0.842	0.821	0.968	0.670	0.662	0.883
the marginal profile rate	mle	0.952	0.946	0.989	0.899	0.897	0.959
	map	0.946	0.939	0.989	0.882	0.880	0.958

<표 IV-24> 인지요소 5개일 때 ACDM으로 생성한 응답데이터 인지요소 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
attribute 1	mle	0.948	0.946	0.992	0.887	0.888	0.939
	map	0.943	0.942	0.992	0.885	0.885	0.939
attribute 2	mle	0.890	0.912	0.982	0.845	0.838	0.915
	map	0.869	0.892	0.982	0.819	0.814	0.914
attribute 3	mle	0.909	0.888	0.985	0.788	0.784	0.902
	map	0.886	0.865	0.985	0.732	0.731	0.901
attribute 4	mle	0.922	0.924	0.990	0.777	0.777	0.885
	map	0.905	0.907	0.990	0.712	0.714	0.883
attribute 5	mle	0.862	0.859	0.979	0.801	0.806	0.903
	map	0.831	0.829	0.979	0.748	0.754	0.902

<표 IV-25> 인지요소 5개일 때 ACDM으로 생성한 응답데이터 프로파일 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
the attribute profile rate	mle	0.580	0.579	0.932	0.335	0.333	0.623
	map	0.508	0.509	0.932	0.242	0.243	0.619
the marginal profile rate	mle	0.906	0.906	0.986	0.819	0.819	0.909
	map	0.887	0.887	0.986	0.779	0.780	0.908

<표 IV-26> 인지요소 8개일 때 ACDM으로 생성한 응답데이터 인지요소 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
attribute 1	mle	0.852	0.848	0.956	0.737	0.739	0.856
	map	0.828	0.822	0.955	0.697	0.698	0.853
attribute 2	mle	0.867	0.873	0.962	0.828	0.829	0.908
	map	0.844	0.852	0.962	0.801	0.803	0.907
attribute 3	mle	0.825	0.833	0.944	0.815	0.816	0.909
	map	0.794	0.803	0.944	0.787	0.787	0.908
attribute 4	mle	0.820	0.821	0.954	0.820	0.819	0.914
	map	0.791	0.793	0.954	0.795	0.794	0.913
attribute 5	mle	0.831	0.820	0.952	0.818	0.817	0.900
	map	0.804	0.792	0.952	0.787	0.785	0.898
attribute 6	mle	0.869	0.875	0.971	0.791	0.792	0.893
	map	0.849	0.855	0.971	0.759	0.760	0.892
attribute 7	mle	0.835	0.841	0.953	0.806	0.806	0.894
	map	0.807	0.814	0.952	0.770	0.771	0.892
attribute 8	mle	0.914	0.912	0.983	0.840	0.838	0.921
	map	0.898	0.896	0.983	0.818	0.815	0.920

<표 IV-27> 인지요소 8개일 때 ACDM으로 생성한 응답데이터 프로파일 분류 정확도

		estimating model					
		high quality			low quality		
		DINA	DINO	ACDM	DINA	DINO	ACDM
the attribute profile rate	mle	0.206	0.210	0.720	0.139	0.138	0.430
	map	0.156	0.159	0.718	0.096	0.096	0.423
the marginal profile rate	mle	0.852	0.853	0.959	0.807	0.807	0.900
	map	0.827	0.828	0.959	0.777	0.777	0.898

## 2. 여러 인지진단모형을 혼합한 모의실험 결과분석

세 가지 인지진단모형을 혼합한 검사는 문항별로 응답데이터를 생성한 모형이 다르기 때문에, 문항별 RMSEA 값을 계산하여 그 차이가 어떻게 나타나는지 살펴보았다. 문항에서 측정하는 인지요소가 한 개인 경우는 인지요소 간 관계랄 게 없어서인지 생성모형과 좋은 추정을 보이는 모형 사이의 일관성이 없었다. 주목할 만 한 점은, 인지요소를 하나만 측정하는 문항의 경우 생성모형에 관계없이 DINA, DINO 모형의 RMSEA 값이 ACDM의 RMSEA 값보다 대체로 더 좋게 나타났다는 점이다.

검사에서 측정하는 인지요소 수가 3개인 경우, 문항이 인지요소를 1개만 측정할 때는 생성모형과 추정치 사이에 별 관계가 없었으나 문항이 2개 또는 3개의 인지요소를 측정할 때 문항수준에 관계없이 생성모형이 추정모형일 때의 RMSEA 값이 가장 좋은 것으로 나타났다.

검사에서 측정하는 인지요소의 수가 5개인 경우, 문항이 인지요소를 2개 측정할 때 생성모형이나 문항 수준과 상관없이 생성모형과 추정모형이 일치하면 RMSEA 값이 가장 좋은 것으로 나타났다. 문항이 인지요소를 3개 측정할 때, DINA 모형과 DINO 모형은 각각이 생성모형일 때 추추정치 또한 생성모형에서 가장 좋게 나타났지만, ACDM 모형은 순위가 뒤바뀌기도 하는 모습을 보였다. 문항에서 측정하는 인지요소가 1개인 경우에서 극단적으로 드러나듯이, 기본적으로 DINA, DINO 모형의 RMSEA 값이 ACDM 보다 더 좋게 나오는 경향이 있다. 이러한 상황에서 DINA 또는 DINO 모형이 생성모형인 경우, 예를 들어 DINA 모형이 생성모형이면 DINO 모형과의 추정 차이가 생기면서 생성모형으로써 독보적으로 좋은 RMSEA 값을 얻게 되지만, ACDM 모형은 생성모형이 되어 더 좋은 RMSEA 값을 얻어도 DINA, DINO 모형과 그 값의 차이가 적어서 순위가 바뀌기도 하는 것으로 보인다. 이러한 경향은 인지요소의 수가 8개인 경우의 결과표에서도 동일하게 나타난다. 이러한 맥락에서 실제 데이터의 세 모형 간 RMSEA 값에서 DINA, DINO 모형 간 값 차이가 거의

없는 상태에서 세 모형 간 값이 비슷하다면 ACDM이 생성모형인 경우로 짐작해볼 수 있다.

단, 연구절차에서 밝힌 바와 같이 기본적으로 RMSEA의 문항별 추정치는 검사 내에서 주어지는 다른 문항의 응답데이터를 함께 고려하여 얻어지는 검사 전반에 대한 추정치이다. 따라서 문항에 대한 모형 간 적합성을 비교하기 위한 일반적 지표로 적용하기에는 한계가 있음을 주의해야 한다.

<표 IV-28> 인지요소 3개일 때 여러 모형으로 생성한 응답데이터의 RMSEA 값

generat- ing model	item number	estimating model						#attrib- utes
		high quality			low quality			
		DINA	DINO	ACDM	DINA	DINO	ACDM	
DINA	1	0.025	0.024	0.038	0.027	0.028	0.049	1
	2	0.035	0.035	0.041	0.034	0.034	0.049	1
	3	0.035	0.034	0.066	0.029	0.028	0.048	1
	4	0.026	0.414	0.167	0.034	0.208	0.108	2
	5	0.030	0.423	0.135	0.038	0.202	0.118	2
	6	0.028	0.401	0.156	0.036	0.213	0.115	2
	7	0.027	0.348	0.200	0.035	0.153	0.108	3
DINO	8	0.020	0.020	0.037	0.033	0.033	0.048	1
	9	0.012	0.014	0.034	0.030	0.029	0.052	1
	10	0.036	0.036	0.066	0.030	0.030	0.048	1
	11	0.495	0.026	0.280	0.191	0.036	0.118	2
	12	0.438	0.032	0.195	0.224	0.032	0.126	2
	13	0.417	0.021	0.249	0.228	0.030	0.124	2
	14	0.317	0.041	0.207	0.164	0.031	0.137	3
ACDM	15	0.029	0.030	0.039	0.030	0.029	0.048	1
	16	0.027	0.026	0.038	0.035	0.035	0.049	1
	17	0.024	0.025	0.071	0.029	0.030	0.049	1
	18	0.319	0.321	0.034	0.182	0.190	0.045	2
	19	0.300	0.304	0.041	0.184	0.187	0.043	2
	20	0.341	0.343	0.049	0.167	0.171	0.042	2
	21	0.323	0.325	0.034	0.194	0.195	0.040	3

<표 IV-29> 인지요소 5개일 때 여러 모형으로 생성한 응답데이터의 RMSEA 값-1

generat- ing model	item number	estimating model						#attrib- utes
		high quality			low quality			
		DINA	DINO	ACDM	DINA	DINO	ACDM	
DINA	1	0.05	0.25	0.17	0.04	0.10	0.09	2
	2	0.06	0.22	0.14	0.05	0.11	0.09	2
	3	0.05	0.23	0.14	0.05	0.11	0.08	2
	4	0.06	0.20	0.12	0.05	0.11	0.10	2
	5	0.07	0.23	0.11	0.04	0.10	0.08	2
	6	0.05	0.26	0.17	0.05	0.09	0.08	2
	7	0.04	0.28	0.18	0.05	0.10	0.09	2
	8	0.05	0.23	0.14	0.05	0.12	0.09	2
	9	0.07	0.18	0.12	0.04	0.10	0.07	2
	10	0.06	0.20	0.12	0.05	0.10	0.09	2
DINO	11	0.27	0.06	0.17	0.10	0.05	0.09	2
	12	0.21	0.06	0.18	0.11	0.05	0.10	2
	13	0.23	0.06	0.13	0.12	0.05	0.08	2
	14	0.25	0.06	0.14	0.10	0.05	0.09	2
	15	0.26	0.05	0.21	0.11	0.05	0.10	2
	16	0.31	0.05	0.24	0.11	0.05	0.09	2
	17	0.24	0.06	0.17	0.09	0.05	0.08	2
	18	0.20	0.06	0.17	0.11	0.05	0.10	2
	19	0.19	0.06	0.14	0.11	0.04	0.09	2
	20	0.25	0.06	0.15	0.10	0.04	0.09	2
ACDM	21	0.19	0.15	0.08	0.09	0.07	0.07	2
	22	0.15	0.14	0.08	0.10	0.08	0.07	2
	23	0.19	0.13	0.08	0.09	0.08	0.07	2
	24	0.17	0.13	0.08	0.10	0.08	0.07	2
	25	0.14	0.18	0.07	0.09	0.08	0.07	2
	26	0.19	0.16	0.07	0.09	0.08	0.07	2
	27	0.18	0.15	0.08	0.08	0.08	0.07	2
	28	0.14	0.15	0.08	0.09	0.09	0.07	2
	29	0.14	0.15	0.08	0.09	0.08	0.07	2
	30	0.17	0.15	0.08	0.08	0.08	0.06	2

<표 IV-30> 인지요소 5개일 때 여러 모형으로 생성한 응답데이터의 RMSEA 값-2

generat- ing model	item number	estimating model						#attrib- utes
		high quality			low quality			
		DINA	DINO	ACDM	DINA	DINO	ACDM	
DINA	1	0.04	0.11	0.12	0.03	0.07	0.07	3
	2	0.04	0.11	0.12	0.02	0.06	0.06	3
	3	0.03	0.14	0.13	0.03	0.06	0.06	3
	4	0.03	0.12	0.13	0.02	0.05	0.06	3
	5	0.03	0.13	0.12	0.03	0.04	0.06	3
	6	0.02	0.13	0.12	0.02	0.06	0.06	3
	7	0.03	0.10	0.13	0.03	0.04	0.06	3
	8	0.03	0.09	0.10	0.03	0.04	0.06	3
	9	0.03	0.08	0.11	0.03	0.04	0.06	3
	10	0.03	0.07	0.11	0.03	0.04	0.06	3
DINO	11	0.11	0.03	0.12	0.07	0.03	0.08	3
	12	0.11	0.03	0.13	0.06	0.03	0.07	3
	13	0.10	0.03	0.16	0.08	0.02	0.07	3
	14	0.11	0.03	0.15	0.05	0.03	0.07	3
	15	0.10	0.04	0.13	0.06	0.02	0.07	3
	16	0.08	0.04	0.10	0.05	0.03	0.07	3
	17	0.09	0.03	0.13	0.04	0.03	0.07	3
	18	0.08	0.03	0.10	0.05	0.03	0.07	3
	19	0.08	0.03	0.11	0.04	0.03	0.06	3
	20	0.07	0.04	0.10	0.04	0.03	0.06	3
ACDM	21	0.16	0.11	0.06	0.06	0.07	0.05	3
	22	0.16	0.10	0.07	0.05	0.06	0.05	3
	23	0.16	0.10	0.06	0.06	0.07	0.05	3
	24	0.15	0.09	0.07	0.04	0.05	0.06	3
	25	0.15	0.08	0.06	0.05	0.06	0.05	3
	26	0.14	0.07	0.07	0.04	0.05	0.05	3
	27	0.11	0.06	0.08	0.04	0.05	0.06	3
	28	0.13	0.07	0.07	0.04	0.05	0.06	3
	29	0.12	0.05	0.08	0.04	0.05	0.06	3
	30	0.14	0.04	0.08	0.04	0.04	0.06	3

<표 IV-31> 인지요소 8개일 때 여러 모형으로 생성한 응답데이터의 RMSEA 값

generat- ing model	item number	estimating model						#attrib- utes
		high quality			low quality			
		DINA	DINO	ACDM	DINA	DINO	ACDM	
DINA	1	0.04	0.06	0.14	0.08	0.08	0.12	1
	2	0.09	0.13	0.16	0.08	0.07	0.12	1
	3	0.13	0.15	0.19	0.11	0.10	0.15	1
	4	0.15	0.33	0.22	0.11	0.17	0.16	2
	5	0.14	0.33	0.22	0.09	0.17	0.15	2
	6	0.11	0.32	0.22	0.10	0.16	0.14	2
	7	0.10	0.39	0.22	0.11	0.15	0.15	2
	8	0.08	0.32	0.23	0.08	0.17	0.14	2
	9	0.05	0.32	0.24	0.10	0.14	0.15	2
	10	0.08	0.43	0.24	0.08	0.17	0.13	2
	11	0.08	0.35	0.24	0.12	0.15	0.15	2
	12	0.06	0.35	0.24	0.09	0.15	0.14	2
DINO	13	0.13	0.15	0.18	0.10	0.10	0.14	1
	14	0.07	0.11	0.11	0.11	0.11	0.14	1
	15	0.17	0.10	0.18	0.11	0.10	0.14	1
	16	0.44	0.12	0.29	0.19	0.11	0.16	2
	17	0.22	0.10	0.27	0.16	0.09	0.14	2
	18	0.22	0.13	0.25	0.16	0.10	0.14	2
	19	0.22	0.10	0.28	0.16	0.08	0.14	2
	20	0.22	0.11	0.28	0.15	0.12	0.15	2
	21	0.21	0.12	0.27	0.15	0.09	0.14	2
	22	0.25	0.10	0.27	0.17	0.11	0.16	2
	23	0.23	0.11	0.27	0.17	0.10	0.15	2
	24	0.24	0.13	0.25	0.17	0.10	0.16	2
ACDM	25	0.16	0.13	0.18	0.11	0.09	0.12	1
	26	0.09	0.11	0.07	0.10	0.10	0.11	1
	27	0.19	0.25	0.20	0.14	0.14	0.13	2
	28	0.20	0.26	0.20	0.13	0.13	0.12	2
	29	0.15	0.32	0.20	0.12	0.17	0.13	2
	30	0.28	0.27	0.22	0.16	0.14	0.14	2
	31	0.29	0.21	0.18	0.16	0.14	0.13	2
	32	0.28	0.21	0.19	0.15	0.14	0.15	2
	33	0.26	0.24	0.23	0.15	0.12	0.13	2
	34	0.26	0.25	0.23	0.15	0.13	0.14	2
	35	0.16	0.29	0.18	0.12	0.15	0.13	2
	36	0.18	0.26	0.21	0.13	0.14	0.12	2

### 3. TIMSS 2007 데이터 결과분석

<표 IV-32> TIMSS 2007 4학년 미국 응답데이터의 적합도 지수

		MADcor	SRMSR	deviance	AIC	BIC	Npars
estimating model	DINA	0.047	0.059	22365.8	87999.8	242664.9	32817
	DINO	0.058	0.073	22423.6	88057.6	242722.7	32817
	ACDM	0.026	0.035	21919.3	22351.3	23369.3	216

DINA, DINO, ACDM 모형 중 절대 적합도 지수인 MADcor, SRMSR 모두에서 0.05미만의 값을 갖는 모형은 ACDM 모형 뿐인 것으로 나타났다. RMSEA에 대해 0.05미만의 값을 가지는 모형은 없지만 ACDM 모형의 값이 가장 작게 나타났다. 절대 적합도 지수 전반에서 ACDM 모형 다음으로 DINA 모형의 값이 작게 나타나 DINO 모형 보다는 나은 것처럼 보인다. 그러나 절대 적합도 지수는 모형 간 적합성 비교를 위한 값이 아니라 추정의 유효성을 판별하기 위한 값이기 때문에 다른 두 모형 간 비교 보다는 ACDM 모형의 추정 타당성에 주목하는 것이 합리적인 것으로 판단하였다.

상대 적합도 지수인 deviance, AIC, BIC에서도 ACDM 모형의 값이 가장 작았다. 특히 AIC, BIC 값에서 그 차이가 더욱 크게 벌어지는 양상을 보였다. 이론적 배경에서 살펴 본 AIC, BIC의 식을 고려하여 각 모형이 추정하는 모수의 수를 나타내는 Npars(the number of parameters) 항목과 deviance 값을 함께 보면 이러한 차이를 쉽게 이해할 수 있다. 우선 DINA, DINO 모형은 ACDM보다 그 가정이 극단적인 만큼 구조 또한 간단하여 각 문항에 대해 slip, guess 모수만 추정하면 된다. 따라서 문항 모수의 측면에서는 ACDM 보다 추정할 모수의 수가 더 적다. 그러나 인지 프로파일의 수를 추정하는 과정에서 DINA, DINO 모형은 프로파일 각각을 별개로 추정한다. 피험자 수가 정해져 있어 모수 하나를 제외한다고 해도 여전히  $32767(=2^{15} - 1)$ 개의 skill class 모수가 필요하다. 그러나

이와 다르게 ACDM 모형은 훨씬 적은 수인 121개만 추정한다. 각 검사에서 추정하는 DINA, DINO 모형의 skill class 모수는 기하급수적으로 증가하는데 비해 ACDM 모형의 문항 모수는 선형적으로 증가하여 그 차이 또한 (문항에서 측정하는 인지요소 수-1)에 불과하다. 특히 수학교과에서 한 검사에 해당하는 문항의 수는 현실적으로 30개를 넘기 어렵기 때문에 검사에서 측정하는 인지요소의 수가 많을수록 AIC, BIC에서 ACDM 모형의 패널티는 DINA, DINO 모형의 패널티에 비해 상대적으로 매우 작은 수준으로 부여되는 것으로 이해할 수 있다. 정리하자면 ACDM 모형이 받은 패널티가 눈에 띄게 작긴 하지만, deviance 값에서도 가장 좋게 나타났기 때문에 세 모형 중 추정의 효율이 가장 뛰어났을 뿐 아니라 추정치 자체의 적합성도 가장 좋은 것으로 판단할 수 있다.

적합도 지수를 살펴본 결과, TIMSS 2007 4번 검사지는 세 모형 중 ACDM 모형으로 추정하는 것이 여러 면에서 가장 적합한 것으로 나타났다. 그러나 이 4번 검사지는 여러 단원에 걸친 내용을 평가하고 있고 그만큼 다양한 내용 영역의 인지요소 숙달여부를 진단하기 때문에 모든 문항이 ACDM 모형으로 추정하기 적합한 것은 아닐 수 있다. 이를 확인하기 위해 문항별 RMSEA 값을 확인하였다. 그 값은 <표 IV-33>과 같다. 결과에 따르면 DINA에서 가장 좋은 추정을 보인 문항이 7개, DINO에서 3개, ACDM에서 15개로 나타난다. 문항에서 측정하는 인지요소의 수가 하나인 경우 인지요소의 관계성이 없음을 고려해 두 문항을 제외한 경우의 분포는 동일한 모형의 순서에서 6개, 3개, 14개로 유사하게 나타난다. 즉, 모든 문항에서 ACDM의 추정이 가장 좋은 것은 아니지만 그 비중이 60% 가량으로 높아 검사에 대한 인지진단모형 적합성 결과와 일맥상통함을 확인하였다. 한 편으로는 ACDM의 추정이 적합한 문항이 많기 때문에 검사에 대한 적합성 또한 ACDM에서 가장 좋게 나온 것으로 이해할 수 있다.

추정 적합성을 나타내는 인지진단모형이 가정하는 인지요소 관계성과 문항이 실제로 측정하는 인지요소 관계성의 연관성을 확인하기 위해 몇 가지 문항 예시를 살펴보겠다. 우선 <그림 IV-1>은 DINA 모형의 추정이

적합한 것으로 나타난 1번 문항이다(TIMSS 2007 문항 번호 M041052).

<p>Which number equals 3 ones + 2 tens + 4 hundreds?</p> <p>Ⓐ 432 Ⓑ 423 Ⓒ 324 Ⓓ 234</p> <p style="text-align: right;">M041052</p>
---

<그림 IV-1> TIMSS 2007 4번 검사지 1번 문항

Copyright © 2008 International Association for the Evaluation of Educational Achievement (IEA). All rights reserved.

이 문항에서 측정하는 인지요소는 1, 2번으로 인지요소 1번은 정수의 표현을 이해하는 것을 포함하고 2번은 사칙연산을 포함한다. 지문에서 나타내는 ones, tens, hundred의 표현을 이해하고, 연산을 사용하는 것 두 가지 모두를 성공하지 못한 사람은, 가령 지문의 표현을 정확히 이해하지 못하거나 연산을 이용해 제대로 계산하지 못한 사람은 문항의 답을 맞히기 어려움을 쉽게 이해할 수 있다. 이러한 연산 문제는 DINA 모형이 적합한 전형적 예시라 할 수 있다.

다음으로, DINO 모형의 추정이 적합한 것으로 나타난 25번 문항을 아래 <그림 IV-2>로 살펴보겠다(TIMSS 2007 문항 번호 M031172).

Street	Number of houses
Main	
Center	
First	
Hill	

Mary is making a chart to show the number of houses on some streets. Every  stands for 5 houses. There are 20 houses on Hill Street. How many  should Mary put in the chart beside Hill Street?

- Ⓐ 4
- Ⓑ 5
- Ⓒ 15
- Ⓓ 20

M031172

<그림 IV-2> TIMSS 2007 4번 검사지 25번 문항

Copyright © 2008 International Association for the Evaluation of Educational Achievement (IEA). All rights reserved.

이 문항에서 측정하는 인지요소는 1, 2, 13, 15번이다. 어떤 학생은 지문의 5 그리고 20을 보고  $20 \div 5 = 4$ 의 계산을 통해 정답이 4를 얻는 데 성공할 수 있다. 이 학생은 기호의 의미를 해석하는 과정을 거치지 않거나 (인지요소 1의 미숙달) 다른 거리(street)에 대한 정보를 이해하지 않거나 (인지요소 13의 미숙달), 4개로 계산한 결과를 그래프와 연결 지어 생각하지 않더라도(인지요소 15의 미숙달) 나름의 풀이 전략을 통해 하나의 인지요소(인지요소 2의 숙달)만 활용하여도 문항을 해결할 수 있다. 즉, 인지요소 1, 2, 13, 15번 중 하나만 숙달해도 좋은 전략을 이용하여 답에도달할 수 있으며 따라서 DINO 모형으로의 추정이 가장 적합하다.

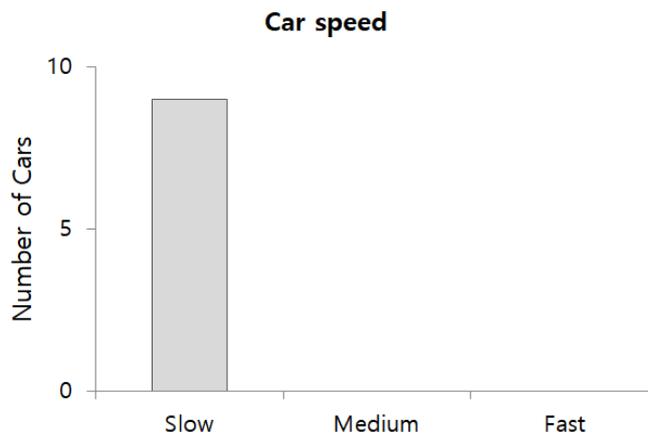
마지막으로 아래 <그림 IV-3>은 ACDM 모형의 추정이 적합한 것으로 나타난 12번 문항이다(TIMSS 2007 문항 번호 M041275).

Several students were collecting information about how fast cars were driving by their school. The table below shows the results for 20 cars.

Car	Slow	Medium	Fast
1		X	
2	X		
3	X		
4			X
5			X
6	X		
7		X	
8		X	
9	X		
10	X		
11	X		
12		X	
13	X		
14			X
15			X
16	X		
17		X	
18	X		
19		X	
20			X

To make the results easier to read, the students started to put the information into the bar graph.

Complete the bar graph.



M041275

<그림 IV-3> TIMSS 2007 4번 검사지 12번 문항

Copyright © 2008 International Association for the Evaluation of Educational Achievement (IEA). All rights reserved.

이 문항에서 측정하는 인지요소는 1, 13, 15번이다. 문항을 해결하기 위해서는 주어진 표를 읽고 (인지요소 13의 숙달) 구성을 이해하여 (인지요소 15의 숙달) 나타낼 수 있어야 한다. 어떤 피험자는 인지요소 13의 숙달을 통해 위의 표에서 느린, 중간 속도의, 빠른 차의 수를 계산하는데 성공할 수 있다. 그러나 인지요소 15를 숙달하지 못했다면 이 수치를 아래 막대그래프로 나타내기 위해서는 막대를 얼마나 그릴 것인지 잘 이해하지 못할 수 있다. 피험자는 그래프를 충분히 이해하지 못한 채로, 표를 읽은 값을 토대로 어느 정도의 추측을 사용해 문항을 맞힐 수도 있고 실패할 수도 있다. 그러나 인지요소 13과 함께 인지요소 15를 함께 숙달하면 더 높은 확률로 정답에 도달할 수 있다. 즉, 일부를 숙달해도 일정 수준의 정답률을 가질 수 있지만 필요한 인지요소를 모두 숙달하면 더 높은 정답률을 얻을 수 있어 ACDM 모형으로 추정하기에 적합하다.

<표 IV -33> TIMSS 2007 4학년 미국 응답데이터의 문항별 RMSEA 값

item number	estimating model			합계
	DINA	DINO	ACDM	
1	0.058	0.192	0.101	2
2	0.05	0.03	0.077	1
3	0.157	0.212	0.133	3
4	0.086	0.053	0.105	2
5	0.202	0.186	0.115	3
6	0.078	0.12	0.091	2
7	0.105	0.161	0.111	3
8	0.096	0.133	0.061	5
9	0.067	0.222	0.11	1
10	0.119	0.173	0.142	2
11	0.135	0.216	0.111	4
12	0.139	0.219	0.092	3
13	0.106	0.117	0.096	4
14	0.132	0.314	0.069	6
15	0.121	0.223	0.125	2
16	0.134	0.189	0.109	2
17	0.042	0.008	0.081	2
18	0.121	0.199	0.091	3
19	0.1	0.086	0.01	3
20	0.149	0.231	0.125	4
21	0.126	0.17	0.098	3
22	0.143	0.193	0.131	3
23	0.156	0.229	0.119	2
24	0.195	0.192	0.177	1
25	0.075	0.012	0.109	4

#### 4. TIMSS 2007 데이터에 기반한 모의실험 결과분석

TIMSS 2007의 미국학생 4학년 응답데이터를 분석한 결과 세 가지 인지진단모형 중 가장 좋은 추정을 보인 모형은 ACDM 모형인 것으로 나타났다. 특히 문항별 RMSEA 비교에서도 ACDM의 추정치가 가장 좋은 문항이 검사 전체 문항의 60% 정도를 차지하였다. 앞서 각 인지진단모형의 생성 비중을 균등하게 가정한 실험에서는 문항별 RMSEA에서도 생

성모형과 추정모형 간 일치도가 어느 정도 유지되는 것으로 나타났다. 그러나 이 실험에서는 실제 데이터인 TIMSS 4번 검사지의 추정치를 토대로 ACDM 모형으로 15문항, DINA 모형으로 7문항, DINO 모형으로 3문항의 응답데이터를 생성하였고 이로 인해 각 인지진단모형의 응답데이터 생성이 균등하지 않고 ACDM 모형의 응답데이터 생성비율이 가장 높고 DINO 모형의 생성비율은 가장 적다. 따라서 생성모형을 균등하게 배분한 검사와 같이 문항별 RMSEA 추정 비교가 생성모형과 강한 연관을 갖기 어렵다. 이러한 점을 고려하여 실험의 결과는 검사별 추정치로 비교하였고 특히 100번의 반복시행에서 각 시행마다 가장 좋은 RMSEA 추정치를 나타낸 추정모형이 무엇인지 그 횟수를 확인하였다. 먼저 적합도 지수는 <표 IV-34>와 같다.

<표 IV-34> TIMSS 2007 데이터에 기반한 모의검사 적합도 지수

		MADcor	SRMSR	deviance	AIC	BIC	Npars
estimating model	DINA	0.041	0.059	23795.1	89429.1	244094.2	32817
	DINO	0.038	0.055	23762.6	89396.6	244061.7	32817
	ACDM	0.021	0.027	23407.4	23839.3	24857.3	216

절대적합도인 MADcor, SRMSR 그리고 상대적합도인 deviance, AIC, BIC 모두에서 ACDM 모형의 추정이 가장 적합한 것으로 나타났다. 특히 ACDM 모형의 추정이 가장 좋고 DINA, DINO 모형의 추정이 비슷한 양상은 실제 데이터의 결과와 같다. RMSEA는 아래 <표 IV-34>와 같다.

<표 IV-34> TIMSS 2007 데이터에 기반한 모의검사 RMSEA 값

	DINA	DINO	ACDM
MEAN	0.176	0.163	0.118
#MIN	6	10	84

RMSEA역시 ACDM에서 가장 작은 값을 나타냈다. #MIN 항목은각 추정모형이 100번의 시행 중 가장 작은 RMSEA 값을 나타낸 시행의 횟수를 의미한다. 결과에 따르면 문항 응답데이터의 생성비율이 높은 ACDM의 추정이 좋은 경우가 대부분이었으며 DINA, DINO 모형도 가장 작은 값을 가질 때가 있는 것으로 나타났다. 문항 응답데이터 생성비율이 60%인 것보다 더 높은 확률인 85% 정도의 비율로 좋은 추정을 나타내었다.

## V. 논의 및 제한점

### 1. 논의 및 결론

인지진단모형은 최근 30여 년간 선진적 평가도구로 여겨지며 활발한 연구를 통해 주목받아왔다. 인지진단모형은 특히 인지요소의 숙달여부를 진단한다는 점에서 학생 개인에게 뿐만 아니라 수업을 이끄는 교사 그리고 교육과정에 핵심적인 정보를 제공할 수 있다. 그러나 인지요소의 관계성 설정에 따라 인지진단모형의 종류가 매우 다양하고 수학교과 내용 체계를 구성하는 인지 및 내용 영역의 차원성 때문에 어떤 인지진단모형을 사용하는 것이 적합한가를 결정하는 것은 어려운 문제다. 따라서 이 연구는 검사의 문항 구성에 따라 어떤 인지진단모형으로 추정하는 것이 가장 적합한지, 또 적합하지 않은 모형의 추정치는 어떻게 나타나는지를 연구하였다.

우선 검사에서 측정하는 인지요소 간 관계성이 극단적으로 비보상적이거나 보상적이거나, 또는 선형적인 결합방식 중 어느 하나만을 갖는 경우에는 생성모형과 추정모형이 일치할 때 가장 좋은 적합성을 나타냄을 확인했다. 검사지의 문항 수준이 높든지 낮든지 간에, 검사지 내 적합성은 생성모형으로 추정한 경우가 가장 높다는 일관된 결과를 얻었다. 특히 절대 적합도 지수는 문항 수준에 관계없이 생성모형으로 추정해야만 좋은 수준을 보이는 것으로 나타났다. 생성모형이 아닌 모형으로 추정할 때는 생성모형과 인지요소의 관계 구조 차이가 클수록 더 낮은 수준의 추정치를 나타냈다. 이러한 결과는 실제 수학교과 시험에 인지진단모형을 적용하기에 앞서 검사하는 인지요소의 성격과 그로 인해 인지요소 간 관계성이 어떠한지에 대한 점검을 거쳐 적합한 인지진단모형을 선정하는 과정이 필요함을 시사한다.

인지요소 간 결합방식이 문항별로 다양한 검사에서는 가장 큰 비중을 차지하는 인지요소 관계와 일치하는 인지진단모형의 추정이 가장 적합한

것으로 나타났다. TIMSS 2007에서 문항별 RMSEA 값은 어느 특정모형에서 항상 가장 작은 값을 갖지 않고 다양하게 분포하였다. 비중이 가장 높은 것은 ACDM 모형으로, 과반 이상인 60%의 문항에서 ACDM 모형의 RMSEA 값이 가장 작았다. 절대 적합도 지수와 상대 적합도 지수를 포함하는 검사 전반의 추정치 역시 ACDM 모형에서 가장 적합한 것으로 나타나 문항의 비중과 일치하는 모형을 선택하는 것이 적합함을 확인하였다.

전통적으로 수학교과 의 인지진단평가는 비보상적 모형에 적합한 것으로 여겨져 왔으나 기존의 견해는 선행연구에서 적용한 검사지의 특성으로 인한 결과일 수 있다. 최근 확률 문제에 관한 인지진단평가에서는 보상적 모형인 ACDM이 비보상적 모형보다 더 적합한 추정을 나타내었다 (Philipp et al., 2017). 따라서 인지진단모형의 평가 적합성을 수학교과 전체에 대해 판단하는 것은 지나치게 포괄적인 측면이 있으며 검사에서 측정하는 인지요소의 성격과 그에 따른 결합방식, 특정 결합방식을 가지는 문항의 비중을 파악하여 적합한 인지진단모형을 선택할 수 있어야 한다.

## 2. 제한점 및 제언

이 연구는 인지진단모형에 적합하고 또 검증된 검사지를 사용하고자 TIMSS 2007 데이터와 선행연구에 따른 Q행렬을 사용하였다. 그러나 DINA, DINO, ACDM 모형을 통한 모수 추정에서 몇몇 문항의 guess 또는 slip 모수는 매우 큰 수준으로 나타나 상당히 낮은 수준의 문항이 포함되어 있음을 확인하였다. 이로 인해 몇 문항의 응답데이터 타당도는 낮은 수준일 수 있다. 또한 검증된 Q행렬을 사용하기 위해 선행연구를 따랐는데, 25문항과 823명의 피험자로 15개의 인지요소를 추정하여 문항수와 피험자수 대비 많은 수의 인지요소 숙달여부를 추정해야 하는 어려움이 있다. 더 정확한 추정을 위해서는 인지요소 숙달여부 추정을 위한 문항

및 검사를 개발하여 사용할 필요가 있다.

또한 수학은 교과 특성상 인지요소 간 위계성이, 즉 어떤 인지요소를 숙달하지 않고서는 다음 인지요소를 숙달하기 어려운 관계가 존재할 가능성이 크지만 이 연구의 모의실험이나 실제 데이터를 분석함에 있어 그러한 점을 고려하지는 않았다. 뿐만 아니라 위계성이 없기 때문에 검사지가 측정하는 인지요소 수가 증가할수록, 모형이 추정해야하는 인지프로파일의 수 또한 기하급수적으로 증가하는 어려움이 있다.

마지막으로 실제 검사에서 측정하는 인지요소의 관계성은 문항별로 상이할 가능성이 높지만 인지진단모형을 통한 추정은 검사 전반의 데이터를 통해 이루어지기 때문에 문항별로 나누어 추정할 수 없다는 어려움이 있다. 즉, 많은 비중을 차지하는 관계성을 알고 적절한 인지진단모형을 선택함이 중요할 뿐 아니라 유용하지만 더 나아가 문항별로 적절한 인지진단모형을 적용하여 추정하는 것은 지금 시점에서 어렵다는 한계가 있다.

## 참 고 문 헌

- 교육부 (2015). 수학과 교육과정. 교육부 고시 제2015-74호 [별책 8].
- 김명연, 강태훈(2016). 보상성 및 비보상성 가정을 평가하기 위한 모형선택지수의 활용가능성 탐색: DINA와 DINO 모형 비교를 중심으로. **교육평가연구**, 제29권 제3호, 383-404.
- 김희경, 한정아, 최숙기, 김부미(2013). 인지진단모형을 적용한 학업성취 프로파일 분석 및 결과 보고 방안. **한국교육과정평가원**.
- 반재천, 김선(2016). 인지진단모형, 인지요소패턴 추정 방법, 사례수에 따른 인지요소패턴 추정의 분류 일관성 및 분류 정확성 비교. **교육평가연구**, 29(3), 405-431.
- 이규진(2015). DINA 모형 기반의 초등학교 수학 진단평가 개발과 활용. **교육평가연구**, 28(4), 1129-1151.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Birgit Harks, Eckhard Klieme, Johannes Hartig & Dominik Leiss (2014). Separating Cognitive and Content Domains in Mathematical Competence. *Educational Assessment*, 19(4), 243-266.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7-74.
- Bloom, B. S. (Ed.). (1956). Taxonomy of educational objectives: *The classification of educational goals. Handbook I: Cognitive domain*. New York, NY: David McKay.
- Borg, I. (1986). Facettentheorie: Prinzipien und Beispiele [Facet theory: Principles and examples]. *Psychologische Rundschau*, 37, 121-37.
- Chen J., de la Torre J., Zhang Z. (2013). Relative and absolute fit evaluation in cognitive diagnostic modeling. *Journal of Educational Measurement*, 50, 123-140.
- Csapó, B. (2010). Goals of learning and the organization of knowledge.

*Zeitschrift für Pädagogik*, 56, 12-7.

- DiBello, L., Roussos, L. A., & Stout, W. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. V. Rao & S. Sinharay (Eds.), *Handbook of statistics*: Vol. 26. Psychometrics (pp. 979-1027). Amsterdam, the Netherlands: Elsevier.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461-464.
- Ravand, H. (2015). Application of a Cognitive Diagnostic Model to a High-Stakes Reading Comprehension Test. *SAGE*, 34(8), 782-799.
- Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74, 191-210.
- Huebner, A., & Wang, B. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement*, 71(2), 407-419.
- Jimmy de la Torre (2011). The Generalized DINA Model Framework. *Psychometrika*, 76(2), 179-199.
- Jimmy de la Torre, Guadalupe Carmona, Vincent Kieftenbeld, Hartono Tjoe & Cynthia Lima (2016). Diagnostic Classification Models and Mathematics Education Research: Opportunities and Challenges. *JRME Monograph*, 15, 53-72.
- Hu, J., Miller, M. D., Huggins-Manley, A. C., & Chen, Y. H. (2016). Evaluation of Model Fit in Cognitive Diagnosis. *International Journal of Testing*, 16(2), 119-141.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.
- Klieme, E., Hartig, J., & Rauch, D. P. (2008). The concept of

- competence in educational contexts. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of Competencies in Educational Contexts* (pp. 3-22). Göttingen, Germany: Hogrefe & Huber.
- Koeppen, K., Hartig, J., Klieme, E., & Leutner, E. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie/Journal of Psychology*, 216(2), 61-73.
- Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2009). A practical illustration of multidimensional diagnostic skills profiling: Comparing results from confirmatory factor analysis and diagnostic classification models. *Studies in Educational Evaluation*, 35, 64-70.
- Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2012). The Impact of Model Misspecification on Parameter Estimation and Item-Fit Assessment in Log-Linear Diagnostic Classification Models. *Journal of Educational Measurement*, 49, 59-81.
- Lee, Y-S., Park, Y-S., & Taylan, D. (2011). A cognitive Diagnostic Modeling of Attribute Mastery in Massachusetts, Minnesota, and the U.S. National Sample using the TIMSS 2007. *International Journal of Testing*, 11, 144-177.
- Lee, Y-W., & Sawaki, Y. (2009). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly*, 6(3), 239-263.
- Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models. *Measurement: Interdisciplinary Research and Perspective*, 11(3), 71-101.
- Mogens, N. (2003). Mathematical competencies and the learning of mathematics: The danish kom project. *Mediterranean Conference On Mathematical Education*, 3, 115-126.
- Mullis, I. V. S., Martin, M. O., & Foy, P. (with Olson, J. F., Preuschoff,

- C., Erberber, E., Arora, A., & Galia, J.). (2008). *TIMSS 2007 International Mathematics Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., Trong, K. L., & Sainsbury, M. (2009). *PIRLS 2011 assessment framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O' Sullivan, C. Y., Arora, A., & Erberber, E. (2005). *TIMSS 2007 assessment frameworks*. Chestnut Hill, MA: IEA.
- Niss, M. (2003). Mathematical competencies and the learning of mathematics: The Danish KOM project. In A. Gagatsis & S. Papastavridis (Eds.), *Mediterranean Conference on Mathematical Education* (pp. 115–124). Athens, Greece: Hellenic Mathematical Society and Cyprus Mathematical Society.
- 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.
- Rupp, A. A., & Templin, J. L. (2008). Unique Characteristics of Diagnostic Classification Models: A Comprehensive Review of the Current State-of-the-Art. *Measurement: Interdisciplinary Research and Perspectives*, 6(4), 219–262.
- Rupp, A. A., Templin, J. L., & Henson, R. A. (2010). Diagnostic Measurement: Theory, Methods, and Applications. *Psychometrika*, 76(4), 715–716.
- Tatsuoka, K. K. (1983). Rule Space: An Approach for Dealing with Misconceptions Based on Item Response Theory. *Journal of Educational Measurement*, 20, 345–354.
- Tatsuoka, K. K. (1995). *Cognitive Diagnostic Assessment for Education: Theory and Applications*. Jacqueline Leighton, Mark Gierl.

- Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11(3), 287-305.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45-65). Bern, Switzerland: Hogrefe & Huber.
- Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, 11, 283-289.

## Abstract

# A Study on Model Fit of Cognitive Diagnosis Models of the Cognitive and Content Domains in Mathematical Competence

You Min-kyung

Department of Mathematics Education

The Graduate School

Seoul National University

In mathematics, traditional teaching has emphasized the acquisition concepts, principles, and laws. However, as the limit of content-based learning has been pointed out, developing students' competency in cognitive domain has become a goal of the curriculum and its importance is continuously emphasized. Cognitive Diagnosis Assessment(CDA) can provide important information for competence evaluation by inferring the profile about mastering the content and cognitive elements based on student's item response. However, in spite of the diversity of the Cognitive Diagnosis Models(CDMs) according to the models about the relationship between item response and mastery

of the attributes, there is little research on the model fit of the CDMs considering the characteristics of the attributes to be estimated. Therefore, this study compared and analyzed the model fit of evaluation models according to the characteristics of the attributes in each simulation and actual data TIMSS 2007 by applying three representative CDMs (DINA, DINO and ACDM).

According to the simulation results, the CDM that assumes the same structure as the one used to generate the data was found most appropriate showing best fit, and the more different the relationship between the generation and the estimation model, the poorer the model fit results. In the verification of fitness through the absolute model index, applying the correct model regardless of the level of item qualities on the test was decisive factor for the appropriateness of the estimation. In the relative fitness index, it was found that the use of the high quality items can improve the model fit than to select the model properly. Therefore, in order to obtain useful information through the CDA, it is necessary to select the appropriate CDM and to use items appropriate for the purpose of the test.

Since the attributes measured for each test item are different and the characteristics for each item are different, the relationship between attributes and item responses can also vary even in a same test. As a result of applying the CDM to the response data of the US students for TIMSS 2007 test number 4, the model with the lowest RMSEA value was different for each item among the results using various CDMs. The RMSEA values of the 15 items, which accounted for more than half of the 25 items, were lowest in the ACDM model. These results are consistent with the fact that ACDM has been shown to be best model to the TIMSS 2007 mathematics test number 4 data of US students in our analysis and various previous studies. In addition, when we analyzed

the items qualitatively regarding the cognitive model to solve that item, the results matched with the per-item best fit model.

**Keywords** : competence, cognitive domain, content domain, Cognitive Diagnosis Model, model fit

***Student Number*** : 2016-21567