creative commons

COMMONS DEED

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

dCollection

교육학석사학위논문

인지진단평가의 문항수준 모형적합도 연구

2019년 2월

서울대학교 대학원

수학교육과 수학교육전공

왕 효 원

인지진단평가의 문항수준 모형적합도 연구

지도교수 유 연 주

이 논문을 교육학석사 학위논문으로 제출함
2018년 12월

서울대학교 대학원
수학교육과 수학교육전공
왕 효 원

왕효원의 교육학석사 학위논문을 인준함
2018년 12월

위 원 장 _____ 이 경 화 _____ (인)

부위원장 _____ 김 서 령 _____ (인)

위 원 _____ 유 연 주 _____ (인)

국문초록

국제 학업성취도 비교 연구의 결과에서 우리나라 학생들의 수학영역 성취도는 높지만, 수학에 대한 정의적 영역 성취도는 낮게 나타났다. 이와 관련하여 성취 중심 평가의 부정적인 측면이 제기되었고, 학교교육의 평가체계가 바뀌어야 한다는 목소리가 높아졌다. 최근 발표된 2015 개정 교육과정 총론에 따르면, 학습의 과정을 중시하는 평가를 강화하고, 평가 결과를 활용하여 교수·학습의 질을 개선하는데 중점을 두어야 한다고 권장한다. 이와 같이 학습내용 성취에 대한 세분화된 정보를 제공할 수 있는 평가를 실시하여 교사와 학생 모두에게 도움을 주어야 할 필요성이 더욱 대두되었다. 그러므로 우리는 학습내용 성취에 대한 자세한 정보를 통해 교사와 학생 모두에게 도움을 줄 수 있는 평가를 지향해야한다.

양적 분석을 이용한 평가의 방법 가운데 인지진단모형(Cognitive Diagnosis Model, CDM)을 바탕으로 한 인지진단평가(Cognitive Diagnosis Assessment, CDA)는 학생들의 인지상태를 통계적으로 추정하여 세분화된 인지요소 벡터로 분석하고 피드백을 제공하는 등 진단적인 역할을 할 수 있다. 학생들이 더 효율적인 학습을 할 수 있게 한다. 인지요소와 문항의 응답 확률과의 관계를 어떻게 가정하고 있느냐에 따라 다양한 인지진단모형이 개발되었다. 여러 가지 인지진단모형 중에는 G-DINA 모형과 같이 문항에서 요구하는 인지요소의 모든 효과를 고려하는 일반적인 모형과 옳은 응답을 위해 인지요소들의 결합적 필요성을 가정하는 DINA 모형과 같은 구체적인 모형이 있다.

인지진단평가는 학생의 인지요소 숙달상태를 통계적으로 추정하기 때문에 그 타당성에 대한 연구가 등장하기 시작했다. 특히 모형적합도는 모형이 실제 관찰된 데이터를 얼마나 잘 나타내는지를 의미하는데, 인지진단평가의 모형-데이터 적합도 연구는 데이터에 적합한 인지진단모형을 찾는데 중점을 둔다. 지금까지의 모형-데이터 적합도 연구는 검사 전

체에 적합한 모형을 찾는 검사 수준의 적합도 연구 위주였다면, 각 문항에 적합한 인지진단모형을 별도로 판정하는 문항수준의 모형-적합도 연구도 등장하였다. 인지진단모형은 인지요소 간의 관계를 모형별로 다양하게 설정하고 있기 때문에, 각 문항에 적합한 인지진단모형은 서로 상이할 수 있다. de la Torre(2011)가 제시한 Wald 검사는 사고 절약의 원리를 바탕으로 하는데, 검사를 구성하는 각 문항에 대해 일반적인 모형과 구체적인 모형의 비교를 통해 구체적인 모형이 일반적인 모형 대신 사용될 수 있는지를 결정하게끔 해준다. Wald 검사는 문항별로 구체적인 모형과 일반적인 모형 사이에 판별을 가능하게 하는데 그쳤지만, Wald 검사에서 계산하는 문항수준 Wald 통계량을 이용하면 후보군의 구체적 인지진단모형 가운데 각 문항에 가장 적합한 인지진단모형을 선택하는 방법을 구성할 수 있다.

본 연구에서는 Wald 통계량을 이용해 각 문항에 적합한 인지진단모형을 선택하는 방법을 개발하고 그 효과를 검증하였다. 이를 위해 모의실험을 실시하였고, 또한 문항에서 요구하는 인지요소의 개수, Q-행렬의 구조, 문항의 질 등의 요인이 개발한 방법의 정확도에 미치는 영향 역시 조사하였다. 이를 위해 위의 조건이 변화할 때 참 모형과 문항수준 모형 선택 방법으로 선택된 모형의 일치율을 구하여 정확도를 평가했다. 그리고 TIMSS 2007 수학 평가 데이터 등을 사용하여, 실제 수학인지진단평가에 Wald 통계량을 이용한 문항수준 모형 선택 방법을 적용한 결과를 분석했다.

분석결과 Wald 통계량을 이용한 문항수준 모형 선택 방법의 정확도는 문항의 질이 높고, 복잡한 Q-행렬이 사용된 상황에서 70% 이상의 정확도를 보였다. 특히 문항에서 요구하는 인지요소의 개수가 많고, 문항의 질이 높을수록 Wald 통계량을 이용한 문항수준 모형 선택 방법의 정확도는 높은 것으로 나타났다. 본 연구에서 개발한 Wald 통계량을 이용한 문항수준 모형 선택 방법을 이용하면, 문항별로 인지진단모형을 다르게 설정하여 좀 더 타당성 있는 인지진단평가를 수행할 수 있을 것으로 기

대된다.

주요어 : 인지진단모형, 인지진단평가, G-DINA 모형, Wald 검사,
모형적합도, 문항수준 모형적합도
학 번 : 2017-29385

목 차

제 1 장 서론 ····· 1

제 1 절 연구의 목적 및 필요성 ····· 1

제 2 절 연구문제와 연구의 중요성 ····· 6

제 2 장 이론적 배경 ····· 9

제 1 절 인지진단모형(Cognitive Diagnosis Model, CDM) ····· 9

1. DINA(Deterministic-Input, Nosiy "And" gate) 모형 ····· 11

2. DINO(Deterministic-Input, Nosiy "Or" gate) 모형 ····· 13

3. ACDM(Additive CDM) ····· 15

4. G-DINA(Generalized DINA) 모형 ····· 18

제 2 절 G-DINA 모형의 모수 추정 ····· 22

1. 주변최대우도추정(Marginalized Maximum Likelihood Estimation, MMLE) ····· 22

2. G-DINA 모형의 설계 행렬 ····· 24

3. 구체적인 모형의 설계 행렬 ····· 26

제 3 절 Wald 검사 ····· 28

제 3 장 연구방법 ····· 31

제 1 절 연구대상 ····· 31

제 2 절 연구절차 ····· 33

1. 모의실험 설계 ····· 33

1.1 간단한 Q-행렬, 문항 당 요구하는 인지요소 2개인 경우 ····· 34

1.2 간단한 Q-행렬, 문항 당 요구하는 인지요소 3개인 경우 ····· 36

1.3 복잡한 구조의 Q-행렬 ····· 38

1.4 복합모형/단일모형 기반 검사 ····· 41

2. Wald 통계량을 이용한 문항수준 모형 선택 ····· 42
3. Wald 통계량을 이용한 문항수준 모형 선택 방법의 정확도 평가 방법 ····· 44
4. TIMSS 2007 응답데이터 분석 ····· 45
5. 윤지영(2016)의 응답데이터 분석 ····· 48

제 4 장 연구결과 ····· 50

제 1 절 모의실험 결과 ····· 50

1. 문항에서 요구하는 인지요소 수의 개수와 모형 선택의 정확도 ····· 50
2. 문항의 질과 모형 선택의 정확도 ····· 57
3. 단일모형 기반 검사의 모형 선택 정확도 ····· 65

제 2 절 실제 데이터 연구 결과 ····· 68

1. TIMSS 2007 응답데이터 분석 결과 ····· 68
2. 윤지영(2016)의 응답데이터 분석 결과 ····· 71

제 5 장 논의 및 제한점 ····· 79

참고문헌 ····· 82

부록 ····· 87

Abstract ····· 93

표 목 차

[표 2-1] $K_j^* = 3$일 때 구체적인 모형과 G-DINA 모형의 문항반응 함수 비교 ····· 21
[표 3-1] 전체 인지요소 5개, 문항에서 요구하는 인지요소 2개인 Q-행렬 ····· 35
[표 3-2] 전체 인지요소 5개, 문항에서 요구하는 인지요소 3개인 Q-행렬 ····· 37
[표 3-3] Lee, et al.(2011)의 Q-행렬의 구조 ····· 38
[표 3-4] 복잡한 구조의 Q-행렬 (Lee, et al.(2011)의 Q-행렬 세 번 반복) ····· 40
[표 3-5] 각 모의실험의 요인 ····· 42
[표 3-6] Wald 통계량을 이용한 모형 선택 비율 예시 ····· 44
[표 3-7] Lee, et al.(2011)의 Q-행렬 ····· 46
[표 3-8] Lee, et al.(2011)이 설정한 인지요소 ····· 47
[표 3-9] 윤지영(2016)이 설정한 인지요소 ····· 48
[표 3-10] 윤지영(2016)이 설정한 Q-행렬 ····· 49
[표 4-1] 모의실험 1과 2의 생성모형별 Wald 통계량을 이용한 모형 선택 비율 ····· 50
[표 4-2] 모의실험 3-3에서 문항에서 요구하는 인지요소 개수에 따른 생성모형별 Wald 통계량을 이용한 모형 선택 비율 ····· 56
[표 4-3] 문항의 질과 문항에서 요구하는 인지요소 개수별 Wald 통계량을 이용한 모형 선택의 TPR ····· 58
[표 4-4] Lee et al.(2011)의 Q-행렬을 사용한 모의실험에서 문항의 질에 따른 Wald 통계량을 이용한 모형 선택의 정확도 ····· 62

[표 4-5] 모의실험 3에서 문항에서 요구하는 인지요소 개수에 따른 생성모형별 Wald 통계량을 이용한 모형 선택의 TPR ····· 63
[표 4-6] 복합모형/단일모형 기반 검사에 대해 생성모형 별 Wald 통계량을 이용한 모형 선택 비율 ····· 66
[표 4-7] TIMSS 2007 응답데이터 분석 결과 ····· 68
[표 4-8] Item ID: M041164가 요구하는 인지요소 (Lee, et al., 2011) ····· 69
[표 4-9] 인지요소 프로파일별 TIMSS 2007 4번 검사지 6번 문항의 정답률 ····· 70
[표 4-10] 인지요소 프로파일별 TIMSS 2007 4번 검사지 1, 21번 문항의 정답률 ····· 71
[표 4-11] 윤지영(2016)의 응답데이터 분석 결과 ····· 72
[표 4-12] 5번 문항이 요구하는 인지요소(윤지영, 2016) ····· 72
[표 4-13] 인지요소 프로파일별 윤지영(2016) 검사지 5번 문항의 정답률 ····· 73
[표 4-14] 7번 문항이 요구하는 인지요소(윤지영, 2016) ····· 74
[표 4-15] 인지요소 프로파일별 윤지영(2016) 검사지 7번 문항의 정답률 ····· 75
[표 4-16] 20번 문항이 요구하는 인지요소(윤지영, 2016) ····· 76
[표 4-17] 인지요소 프로파일별 윤지영(2016) 검사지 20번 문항의 정답률 ····· 76
[표 4-18] 인지요소 프로파일별 윤지영(2016) 검사지 3,22번 문항의 정답률 ····· 77

그 림 목 차

[그림 2-1] 문항-인지요소 간 관계와 그에 대응하는 Q-행렬의 예시 ····· 10
[그림 2-2] G-DINA 모형의 문항반응함수와 DINA 모형, DINO 모형, ACDM의 문항반응함수와의 관계 ····· 21
[그림 3-1] Wald 통계량을 이용한 모형 선택의 예 ····· 43
[그림 4-1] 문항에서 요구하는 인지요소 개수에 대해 생성모형별 Wald 통계량을 이용한 모형 선택의 TPR ····· 57
[그림 4-2] 문항의 질에 따른 Wald 통계량을 이용한 모형 선택의 TPR 변화 - 문항에서 요구하는 인지요소 개수별 변화, 검사 전체 변화 ····· 65
[그림 4-3] TIMSS 2007 Item ID: M041164 ····· 69
[그림 4-4] 윤지영(2016)의 검사지 5번 문항 ····· 72
[그림 4-5] 윤지영(2016)의 검사지 7번 문항 ····· 74
[그림 4-6] 윤지영(2016)의 검사지 20번 문항 ····· 75

제 1 장 서론

제 1 절 연구의 목적 및 필요성

우리나라 학생들이 TIMSS나 PISA와 같은 국제학업성취도평가에서 수학의 인지적 영역 성취도는 최상위권이지만 수학 학습에 대한 정의적 영역 성취도는 하위권에 머물고 있다(조혜정 & 김인수, 2016). 2015년에 실시한 TIMSS 결과보고서를 보면 우리나라 초등학교 4학년과 중학교 2학년 학생들의 수학 성취도는 각각 49개국 중 3위와 39개국 중 2위를 차지하였지만, 수학 학습에 대한 흥미와 자신감 등 정의적 영역의 성취도는 초등학교 4학년 학생의 경우 49개국 중 47-48위, 중학교 2학년 학생의 경우 39개국 중 35-37위를 차지하였다(Mullis et al., 2016). 손원숙 외(2009)는 우리나라의 경우 인지적 영역과 정의적 영역 성취도의 세계 순위는 반대의 양상을 보이지만, 실제 상관관계를 구하면 인지적 영역과 정의적 영역은 높은 상관이 있는 것으로 알려져 있다는 점을 지적하고 이는 인지적 영역에만 집중하고 있는 우리나라 교육에 큰 시사점을 제공한다고 주장했다.

한편 오영수, 전영주(2018)는 실제 학교 현장에서 학생들의 인지적 성취에만 관심을 가지고 평가하기 때문에 학교급과 학년이 올라갈수록 수학학습에 대한 학생들의 감정, 느낌, 흥미 등 정의적 영역은 심각할 정도의 부정적인 편향성을 나타난다고 주장했다. 또한, 김정환(2006)은 평가에 의해 학습동기가 유발되고 평가가 곧 경쟁인 상황에서는 평가는 학생의 불안감 등 정의적 영역에 부정적인 영향을 미치며, 학교교육의 질적 개선을 위해 평가체계가 바뀌어야한다고 주장하는 등 인지적 영역의 성취에만 집중하고 있는 현 평가체계에 대한 부정적인 견해가 제기되고 있다.

교육부(2015)가 발표한 2015 개정 교육과정 총론 중 교육과정 구성의 중점을 보면, 2015 개정 교육과정은 학생 참여형 수업을 활성화하여 자기 주도적 학습 능력을 기르고 학습의 즐거움을 경험하도록 하고, 학습의 과정을 중시하는 평가를 강화하여 학생이 자신의 학습을 성찰하고 평가 결과를 활용하여 교수·학습의 질을 개선하는데 중점을 두었다. Black & William(1998)과 William(2006)에 의하면 평가는 충분히 세분화된 정보를 제공해 학생들의 교수·학습에 적합한 피드백을 제공할 수 있어야 한다. 즉, 우리가 지향해야할 평가는 서열화를 통해 경쟁을 유발하는 방식에서 벗어나 학습내용 성취에 대한 자세한 정보를 통해 교사와 학생 모두에게 도움을 줄 수 있어야 한다.

평가는 사람에 의해서 수행되기 때문에 주관적인 성격을 가지지만, 검사라는 도구를 사용하여 측정치를 제공하고 그를 바탕으로 객관성을 확보한다(성태제, 2018). 교육 현장에서는 학생들이 학습한 내용을 제대로 성취하였는지 측정하기 위해 형성평가, 진단평가 등의 검사를 이용하고, 검사 결과를 통해 학생들이 학습한 내용을 제대로 성취했는지에 대해 평가한다. 오랜 시간동안 고전검사이론(Classical Test Theory, CTT)과 문항반응이론(Item Response Theory, IRT)은 주된 평가도구로 사용되어 왔다(de la Torre & Minchen, 2014). 이 중 최근까지 많이 사용되고 있는 문항반응이론은 학생을 능력 연속체 위의 상대적 위치를 나타내는 점수를 제공하는데, 이 점수는 학생의 능력에 대한 단일차원적인 점수에 불과하다. 예를 들어 수학 검사를 실시했을 때, 문항반응이론이 제공하는 점수는 단일적인 수학 능력에 대한 평가만 가능하게 한다. 이를 개선하기 위해 다차원 문항반응이론(Multi-dimensional IRT, MIRT; Reckase, 1985)이 제시되었지만, 다차원 문항반응이론은 피험자의 능력을 K차원의 벡터로 나누어 측정할 뿐, 세분화된 인지요소로 나누어 인지요소를 습득했는지 습득하지 못했는지에 따라 패턴별로 분류하지는 못한다(Finkelman, Kim, & Roussos, 2009). 즉, 기존 평가도구는 수학의 구체적이고 세부적인 인지 영역에 대한 평가는 불가능하기 때문에 진단적인 용도로는 상당히 제한적이다(de la Torre & Lee, 2013). 기존 이론과 대조되어, 인지 및 심

리 이론이 발전하면서 학습과 관련성 높은 추론을 가능하게 하는 인지진단모형(Cognitive Diagnosis Model, CDM)이 논의되기 시작했다(de la Torre, 2009a). 인지진단모형은 학생의 능력을 세부 인지요소의 다차원적인 결합으로 설정하고 각 인지요소의 숙달/미숙달 여부 등의 정보를 제공할 수 있다. 즉, 세분화된 인지요소에 대한 피드백을 줄 수 있어 기존 문항반응이론이나 고전검사이론보다는 진단적인 성격이 더 강하다는 장점이 있다(de la Torre & Minchen, 2014).

인지진단모형을 기반으로 하는 인지진단평가(Cognitive Diagnosis Assessment, CDA)는 검사에서 각 문항에 대한 학생의 응답을 토대로 문항을 해결하는데 요구되는 인지요소의 숙달 또는 미숙달 여부를 프로파일 형태로 제공하는 통계 추정 모형이다(Tatsuoka, 1995). 세분화된 인지요소에 대한 학생의 숙달/미숙달 프로파일은 학생을 지도하는 교사로 하여금 학습에 대한 피드백을 제공할 수 있고, 학생 스스로도 본인의 부족한 부분을 깨닫고 극복하는 과정을 통해 자기 주도 학습을 경험할 수 있는 등 학습방향 개선의 역할을 할 수 있다. 인지진단모형을 기반으로 한 인지진단평가는 그 등장과 함께 평가의 새로운 가능성을 제시했으며, 인지진단평가를 위해 인지요소들과 문항에 대한 반응 사이의 관계에 대한 가정에 따라 여러 가지 인지진단모형이 개발되었다.

현재 개발된 인지진단모형은 크게 일반적인 모형과 구체적인 모형으로 나눌 수 있다. 일반적인 모형은 인지요소의 주효과와 인지요소 간의 상호작용 효과 모두를 고려하고 있는 모형이다. 일반적인 모형의 예로는 G-DINA(generalized DINA model: de la Torre, 2011), log-linear CDM(Henson, Templin, & Willse, 2009), GDM(generalized diagnostic model: von Davier, 2005) 등이 있다. 구체적인 모형은 일반적인 모형에 상호작용 효과는 없는 것으로 설정하는 등 특정한 가정들이 더해져 만들어진 모형들을 의미한다. 구체적인 모형의 예로는 DINA(deterministic inputs, noisy “and” gate model: Hartel, 1989; Junker & Sijitsma, 2001), DINO(deterministic inputs, noisy “or” gate model: Templin & Henson,

2006), ACDM(additive CDM: de la Torre, 2011) 등이 있는데, DINA 모형은 문항에서 요구하는 인지요소 모두를 숙달해야 문항을 해결 할 수 있다고 가정하고, DINO 모형은 문항에서 요구하는 인지요소 중 하나라도 숙달하면 문항을 해결할 수 있다고 가정한다. 또한, ACDM은 문항에서 요구하는 인지요소 중에서 얼마나 숙달하느냐에 따라 문항을 해결할 수 있는 확률이 달라진다고 가정한다.

인지진단평가에서는 학생들의 응답 자료로부터 인지진단모형의 모수를 추정하고 이를 이용하여 학생의 인지요소 숙달 여부를 판단한다. 이때 이 추정이 타당한지 알아보기 위해 모형-데이터 적합도(model-fit index)를 이용한다(Hu et al., 2016). 여기서 모형-데이터 적합도란 모형이 실제 관찰된 데이터를 얼마나 잘 나타내는지를 의미하는데, 적합도가 좋다는 것은 모형에 의해 예상된 결과가 실제 관찰된 결과와 잘 맞는다는 것을 의미한다. 일반적인 모형은 구체적인 모형에 비해 더 나은 모형-데이터 적합도를 보이지만, 일반적인 모형은 가지는 모수가 많아 추정 과정에서 많은 표본을 요구하고, 반면 구체적인 모형은 특정한 가정들로 인해 일반적인 모형에 비해 간단한 형태를 띠고 의미 있는 해석을 가능하게 한다(de la Torre & Lee, 2013). 각 인지진단모형은 모형이 가지는 모수의 개수도 다르고, 인지요소들 간의 관계를 다르게 가정하고 있다. 특히 표본의 크기가 작을 때, 데이터에 적합한 구체적인 인지진단모형을 사용한 경우가 일반적인 인지진단모형을 사용한 경우보다 인지프로파일 분류의 정확도가 더 높은 것으로 나타났다(Rojas, de la Torre, & Olea, 2012). 그러므로 인지진단평가를 실시함에 있어 적합한 인지진단모형을 활용하는 것이 중요하다.

인지진단모형을 이용한 인지진단평가가 활발히 연구되면서, 동시에 인지진단평가의 타당성에 대한 연구도 많이 진행되고 있다. 학생의 응답 중 부적절한 응답을 찾아내는 피험자 응답 적합도 연구(예: Cui & Leighton, 2009)가 있다. 예를 들어 학생이 검사에 임하는 태도나 동기, 응답 작성, 답안지 작성법 숙지 여부 등의 검사 외적 요인들로 인해 평

가의 타당성을 낮추는 경우가 생길 수 있다(시기자, 2003). 이런 경우, 피험자 응답 적합도 연구를 통해 부적절한 응답을 제거해 평가의 타당성을 향상시킬 수 있다. 또 다른 타당성의 분석으로 인지진단평가의 각 검사에 어떤 인지진단모형이 적합한지를 분석하는 모형적합도 분석이 있다(예: Chen, de la Torre, & Zhang, 2013). 대부분의 모형적합도 연구는 이러한 검사수준 모형적합도를 분석하고 있는데, 검사수준 모형적합도 분석은 모형적합도를 통해 검사 전체에 적합한 인지진단모형을 찾아내는데 집중했다.

다양한 인지진단모형들이 문항을 해결하는데 필요한 인지요소들 간의 관계를 다르게 가정하고 있는 만큼, 검사 내의 각 문항별로 서로 다른 응답모형을 가질 수 있다. 예를 들어, 이차방정식에 대한 검사를 생각해 보자. x^2+2x+1, $x^2+3x+2=0$, $x^2+4x+1=0$, $x^2-2x+3=0$의 네 개의 방정식 중에서 중근을 갖는 방정식을 찾는 문항이 있다고 할 때, 이 문항은 '중근에 대한 이해', '인수분해를 이용하여 이차방정식의 근을 구할 수 있다', '근의 공식을 이용하여 이차방정식의 근을 구할 수 있다' 등 위 세 가지 인지요소 모두를 숙달하고 있어야 문항을 해결할 수 있다. 따라서 이 문항은 DINA 모형이 적합한 모형이라고 판단할 수 있다. 반면, 방정식 $x^2+4x+3=0$의 해를 구하는 문항이 있다고 할 때, '인수분해를 이용하여 이차방정식의 근을 구할 수 있다' 또는 '근의 공식을 이용하여 이차방정식의 근을 구할 수 있다'라는 두 인지요소 중 하나만 숙달하고 있더라도 문항을 해결할 수 있다. 그러므로 이 문항은 DINA 모형 보다는 DINO 모형이 적합하다고 판단할 수 있다. 이처럼 검사 내의 문항마다 적합한 인지진단모형이 다를 수 있기 때문에, 검사수준 모형적합도 분석을 통해 검사 전체에 적합한 인지진단모형을 결정하는 것은 타당하지 않은 평가로 이어질 수 있다. 검사수준 모형적합도 분석에서 더 나아가 문항별로 모형적합도를 따로 분석할 수 있는데, 이를 문항수준 모형적합도(item-level model fit) 분석이라고 한다(de la Torre & Lee, 2013). 검사지 내의 각 문항에 적합한 인지진단모형을 찾아내고,

문항별로 인지진단모형을 달리하여 인지진단평가를 수행할 수 있다면, 평가의 타당성을 더 높일 수 있을 것이다.

de la Torre(2011)는 문항수준에서 모형적합도를 판단할 수 있는 Wald 검사를 제안했다. Wald 검사는 검사내의 각 문항별로 일반적인 인지진단모형과 구체적인 인지진단모형의 적합도를 통계적으로 비교하여 구체적인 인지진단모형이 일반적인 인지진단모형 대신 사용될 수 있는지를 분석할 수 있게 한다. Wald 검사는 각 문항에 대해 구체적인 모형별로 Wald 통계량을 계산할 수 있고, 이를 비교하여 각 문항에 적합한 인지진단모형을 결정할 수 있다. 이에 본 연구에서는 Wald 통계량을 이용하여 문항수준 모형적합도를 분석해 문항수준에서 적합한 인지진단모형을 선택할 수 있는지를 연구하고, Wald 통계량을 바탕으로 문항수준에서 적합한 인지진단모형을 결정하는데 영향을 미치는 요인들을 분석하고자 했다. 그리고 이를 바탕으로 실제 수행된 수학 인지진단평가의 자료를 이용하여 Wald 통계량을 통해 선택된 인지진단모형은 무엇인지 분석하고, 그 의미를 해석해보고자 했다.

제 2 절 연구문제와 연구의 중요성

인지진단모형은 학생들의 능력을 여러 인지요소의 다차원적인 결합으로 가정하고 있다. 인지진단모형은 일종의 응답모형으로 학생들의 인지요소 숙달여부와 각 모형에서 가정하고 있는 인지요소의 관계에 따라 문항에 대해 올바르게 응답할 확률을 나타낸다. 인지진단모형을 바탕으로 하는 인지진단평가는 검사지에 대한 학생들의 응답을 바탕으로 각 인지요소의 숙달여부를 추정하여 학생들의 인지요소 프로파일을 추론할 수 있다. 인지진단평가를 교육현장에 적용한다면 교사들은 학생들의 인지요소 프로파일을 바탕으로 학생들의 학습 상태를 진단하고 피드백을 통해 숙달하지 못한 인지요소를 숙달할 수 있게끔 지도 할 수 있다. 그렇기 때문에 인지진단평가가 적절하게 활용되기 위해서는 인지요소 프로파일

이 적절하게 추정되는 것이 중요하다. 즉, 추정된 인지요소 프로파일이 적합하다고 여겨지려면 인지진단평가의 타당성이 보장되어야 한다. 인지진단평가에서 인지요소 프로파일에 영향을 미칠 수 있는 요소는 Q-행렬이 올바르게 작성되었는지의 여부, 적합한 인지요소가 설정되었는지의 여부, 적합한 인지진단모형이 사용되었는지의 여부 등이 있는데, 본 연구에서는 Q-행렬과 문항에서 요구하는 인지요소가 적합하다고 가정하고 인지요소 프로파일 추정에 있어 적합한 인지진단모형을 결정할 수 있는지에 대해 분석했다.

기존에는 인지진단평가를 수행함에 있어 검사수준에서 적합한 모형을 찾는 것이 주된 연구 내용이었다. 본 연구에서는 검사수준에서 더 나아가 문항수준에서 적합한 인지진단모형을 찾고 이를 바탕으로 인지진단평가의 타당성을 높이는데 도움이 되고자 했다. 이에 본 연구의 연구문제는 다음과 같다.

1. Wald 통계량을 이용하여 문항수준에서 적합한 인지진단모형을 결정할 수 있는가?
2. Wald 통계량을 이용한 문항수준 모형 선택 방법의 정확도에 영향을 미치는 요인은 무엇인가?
3. 수학 인지진단평가에 적합한 인지진단모형은 무엇인가?

이러한 연구문제를 해결하기 위해 먼저 Wald 통계량을 이용해 문항수준에서 적합한 인지진단모형을 선택하는 방법을 개발했다. 그리고 모의실험을 통해 Wald 통계량을 이용하여 문항별로 적합한 인지진단모형을 선택하는 방법의 정확도에 영향을 미치는 요인들을 분석했다. 모의실험에서는 생성모형을 다르게 설정하여 응답데이터를 생성하고, 생성된 응답데이터를 바탕으로 Wald 통계량에 의해 적합하다고 결정된 모형이 생성모형과 일치하는 비율을 조사했다. 그리고 문항에서 요구하는 인지요소 개수, 문항의 질 등 인지진단평가를 이루는 요인들을 달리하여 응답

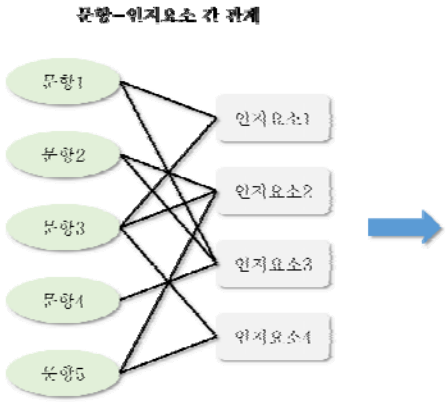
데이터를 생성하고, 각 문항에 대해 생성모형과 Wald 통계량에 의해 적합하다고 결정된 모형이 일치하는 비율을 조사했다. 이를 통해 Wald 통계량을 이용한 모형 선택 방법의 정확도에 영향을 미치는 요인들을 분석해 연구문제 2에 답하고자 했다. 그리고 TIMSS 2007(Martin et al., 2008) 4학년 학생들의 4번 검사지에 대한 응답 데이터와 윤지영(2016)의 일차방정식과 일차부등식 진단평가에 대한 중학교 2,3학년 학생들의 응답데이터 등의 실제 데이터를 분석하여 수학 인지진단평가에서 적합한 인지진단모형은 무엇인지 알아보았다. Wald 통계량을 바탕으로 각 문항에 적합하다고 결정된 인지진단모형, 실제 문항의 내용, 문항에서 요구하는 인지요소 등을 분석하여 Wald 통계량에 의해 결정된 인지진단모형의 의미를 해석하고 연구문제 3에 답하고자 했다.

제 2 장 이론적 배경

제 1 절 인지진단모형(Cognitive Diagnosis Model, CDM)

인지진단모형을 기반으로 하는 인지진단평가는 검사에서 각 문항에 대한 학생의 응답을 바탕으로 문항을 해결하는데 요구되는 인지요소의 숙달/미숙달 여부를 프로파일 형태로 제공하는 통계 추정 모형이다(Tatsuoka, 1995). 인지진단모형을 통한 인지요소 숙달 여부 추론은 문항의 난이도를 고려하여 인지요소 숙달 확률을 추정하는 통계적 기법이다. 기존 검사이론과 달리 문항이 어려워서 문항을 해결하지 못했다는 이유만으로 인지요소를 숙달하지 못한 것으로 판정되거나 문항이 쉬워서 문항을 해결했다는 이유만으로 인지요소를 숙달했다고 판정하지 않으므로 이는 인지진단모형을 기반으로 한 분석방법이 가지는 장점이라고 볼 수 있다(김희경 & 김부미, 2013). 인지진단모형을 기반으로 한 평가 결과는 단순한 능력 추론에 그치는 것이 아니라 검사 이후 학습에 대한 진단적인 정보 등의 피드백을 제공할 수 있다.

인지진단모형은 문항에 대한 학생의 응답을 토대로 인지요소에 대한 숙달 여부를 추론하므로, 문항을 해결하기 위해 요구되는 인지요소 간의 관계가 명확하게 나타나야 한다. 이러한 관계는 Q-행렬을 이용하여 나타낼 수 있는데, Q-행렬은 문항과 인지요소 간의 관계를 $n \times K$행렬(n: 문항의 수, K: 인지요소의 수)로 나타낸다(Tatsuoka, 1983). 문항을 해결하는데 인지요소가 필요하면 1, 필요하지 않으면 0으로 표시한다. 인지요소와 문항 간의 관계를 Q-행렬로 표현한 예는 다음 [그림 2-1]과 같다.



Q-행렬

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

[그림 2-1] 문항-인지요소 간 관계와 그에 대응하는 Q-행렬의 예시

인지진단모형을 실제로 적용할 때에는 참인 Q-행렬은 알 수 없으므로, 실제 데이터를 분석할 때에는 타당한 Q-행렬을 적용해야 한다. Rupp & Templin(2008)은 실제 필요한 인지요소보다 부족하게 설정된 Q-행렬을 사용하면 인지요소에 모두 숙달한 학생들의 정답률이 과 추정되고, 실제 필요한 인지요소보다 많이 설정된 Q-행렬을 사용하면 인지요소에 모두 숙달하지 못한 학생들의 정답률이 과 추정된다는 것을 발견했다. 이렇듯 Q-행렬이 타당하지 않다면, 모수 및 학생의 인지요소 프로파일 추정에 오류가 생길 수 있기 때문에 타당한 Q-행렬을 사용하는 것이 중요하다. 이에 따라 Q-행렬의 불확실성에 관한 연구가 상당히 진행되었지만(예: de la Torre, 2008; Li & Suen, 2013; Wang & Gierl, 2011), 본 연구에서는 Q-행렬이 타당하다고 가정하여 연구를 진행했다.

앞서 언급한 바와 같이 현재까지 다양한 인지진단모형이 개발됐으며, 크게 일반적인 모형과 구체적인 모형으로 구분할 수 있다. 일반적인 모형의 예로는 G-DINA(generalized DINA model: de la Torre, 2011), log-linear CDM(Henson, Templin, & Willse, 2009), GDM(generalized diagnostic model: von Davier, 2005) 등이 있고, 구체적인 모형의 예로는 DINA(deterministic inputs, noisy “and” gate model: Hartel, 1989; Junker & Sijitsma, 2001), DINO(deterministic inputs, noisy “or” gate model:

Templin & Henson, 2006), R-RUM(reduced reparameterized unified model: DiBello, Roussos, & Stout, 2007; Hartz, 2002), ACDM(additive CDM: de la Torre, 2011)등이 있다.

이렇게 인지진단모형이 다양한 만큼, 모형마다 측정하는 인지요소 간의 관계를 다양하게 설정하고 있다. 따라서 인지진단모형을 기반으로 인지진단평가를 수행할 때, 피험자의 인지프로파일을 타당하게 추정하기 위해서는 검사에 적합한 인지진단모형을 사용해야 한다.

1. DINA(Deterministic-Input, Noisy “And” gate) 모형

DINA 모형은 비보상적 모형 중 하나로 분류할 수 있는데, 이는 DINA 모형의 특징을 보면 알 수 있다. DINA 모형은 문항에서 요구하는 인지요소 모두에 숙달해야만 문항을 맞출 수 있다고 가정하고 있다. 즉, 숙달한 인지요소 일부가 숙달하지 못한 다른 인지요소를 보상해주지 못한다고 가정하기 때문에 비보상적인 모형으로 분류된다. DINA 모형에서 문항 j에 대한 문항반응함수(Item Response Function, IRF)는 다음과 같이 나타낼 수 있다.

$$P(\alpha_{lj}^{*}) = \begin{cases} g_j & \text{if } \alpha_{lj}^{*} < 1_{K_j^{*}} \\ 1-s_j & otherwise \end{cases} \tag{2.1}$$

문항반응함수는 피험자 l이 인지요소 벡터 α_{lj}^{*}를 가질 때, 피험자가 문항에 옳게 답할 확률을 나타낸다. 문항반응함수에서 α_{lj}^{*}은 문항 j에서 요구하는 인지요소 벡터 중 피험자 l이 가지는 인지요소 벡터, K_j^{*}는 문항 j에서 요구하는 인지요소의 개수, g_j는 문항 j의 guess 모수, s_j는 문항 j의 slip 모수를 의미한다. 여기서 $\alpha_{lj}^{*} < 1_{K_l^{*}}$는 α_{lj}^{*}가 모두 1이 아닌 경우를 나타내는데, 이는 피험자가 문항에서 요구하는 여러 인지요소 중 하

나라도 숙달하고 있지 않는 경우를 의미한다. 이 경우에 피험자가 문항에 옳게 답할 확률은 g_j, 추측하여 해결하는 확률이 된다. 반대로 피험자가 문항에서 요구하는 인지요소 모두를 숙달하고 있다면, 피험자가 문항에 옳게 답할 확률은 $1-s_j$, 실수하는 경우를 제외한 확률이 된다.

DINA 모형의 문항반응함수는 다음과 같은 형태로도 표현할 수 있다.

$$P(X_{ij}=1|\alpha_i)=(1-s_j)^{\eta_{ij}}g_j^{(1-\eta_{ij})} \tag{2.2}$$

여기서 X_{ij}는 피험자 i가 문항 j를 옳게 답하였는지를 나타내는데, 옳게 답하면 1, 옳게 답하지 않았으면 0으로 나타낼 수 있다. α_i는 피험자 i의 인지요소 벡터를 나타낸다. 즉, 위의 문항반응함수는 인지요소 벡터 α_i를 가지는 피험자 i가 문항 j를 옳게 답할 확률을 나타낸 식이다. s_j와 g_j는 각각 문항 j의 slip 모수와 guess 모수를 나타낸다. η_{ij}는 피험자가 문항에서 요구하는 인지요소 모두를 숙달한 상태인지를 나타낸다. 여기서 η_{ij}는 $\eta_{ij}=\prod_{k=1}^{K_j^*}\alpha_{ik}$과 같은 식으로 나타낼 수 있다. 여기서 K_j^*는 문항 j에서 요구하는 인지요소의 수를 나타내고, α_{ik}는 피험자 i가 k번째 인지요소를 숙달했으면 1, 숙달하지 않았으면 0의 값을 가진다. 즉, DINA 모형의 경우 문항에서 요구하는 인지요소 모두를 숙달한 경우 $\eta_{ij}=1$, 모두 숙달한 경우가 아닌 경우에는 $\eta_{ij}=0$의 값을 가진다.

문항 j에서 요구하는 인지요소가 3개인 경우를 가정해보자. 그러면 3개의 인지요소에 대해 피험자가 가질 수 있는 인지요소의 조합(α_{lj}^*)은 (0,0,0), (0,0,1), (0,1,0), (1,0,0,), (0,1,1), (1,0,1), (1,1,0), (1,1,1)로 총 8가지 경우다. 첫 번째 형태의 문항반응함수에서 각 인지요소 벡터를 대입했을 때 얻어지는 함숫값을 살펴보자. 8가지의 인지요소 벡터 중 (1,1,1)을 제외한 모든 경우에 $\alpha_{lj}^* < 1_3$이 성립하므로 (1,1,1)을 제외한 나머지 7가지 경우에는 피험자가 문항에 옳게 답할 확률이 g_j가 된다. 이는 DINA 모형

의 가정대로 피험자가 문항에서 요구하는 인지요소 중 하나라도 숙달하지 못하면 피험자가 문항에 옳게 답할 확률은 추측하는 확률과 같게 되는 것을 의미한다. 그리고 피험자의 인지요소 벡터가 (1,1,1)인 경우에는 피험자가 문항에 옳게 답할 확률은 $1-s_j$가 되는데, 이는 피험자가 문항에서 요구하는 인지요소 모두를 숙달하였기 때문에 피험자가 문항에 옳게 답할 확률은 실수하는 경우를 제외한 확률과 같다는 것을 의미한다. 그리고 두 번째로 제시된 형태의 문항반응함수에서 살펴보면, (1,1,1)인 경우에는 $\eta_{ij}=1$이므로 피험자가 문항에 옳게 답할 확률은 $1-s_j$가 된다. 그 외의 경우에는 $\eta_{ij}=0$이므로 피험자가 문항에 옳게 답할 확률은 g_j가 된다. 즉, DINA 모형에서 가정하고 있는 바와 같이 피험자는 문항에서 요구하는 인지요소 모두를 숙달하고 있어야 문항에 옳게 답할 수 있고, 그 중 하나라도 숙달하지 못한다면 문항에 옳게 답할 수 없다.

DINA 모형은 어떤 경우에도 각 문항이 두 개의 모수(g_j와 s_j)만 가지기 때문에 간단한 형태를 보이고 있어 지난 십여 년간 많은 관심을 받아왔다(Wang et al., 2015). 하지만 문항에서 요구하는 인지 요소 중 하나만 숙달하지 못하더라도 문항을 맞출 수 없다고 가정하는 것이 지나치게 제한적이기에 극단적인 모형이라는 비판의 시각도 있다(Henson, Templin, & Willse, 2009).

2. DINO(Deterministic-Input, Noisy “Or” gate) 모형

DINO 모형은 보상적 모형 중 하나로 분류할 수 있다. DINO 모형은 문항에서 요구하는 여러 인지요소 중 하나만 숙달해도 문항을 맞출 수 있다고 가정하고 있다. 앞선 DINA 모형의 가정과 달리 숙달한 인지요소가 숙달하지 못한 다른 인지요소를 보상해준다고 가정하고 있기 때문에 DINO 모형은 보상적 모형으로 분류할 수 있다. DINO 모형의 문항반응함수는 다음과 같다.

$$P(\alpha_{lj}^{*}) = \begin{cases} g_j & \text{if } \alpha_{lj}^{*} = 0_{K_j^{*}} \\ 1 - s_j & otherwise \end{cases} \tag{2.3}$$

문항반응함수는 피험자 l이 인지요소 벡터 α_{lj}^{*}를 가질 때, 피험자가 문항 j에 옳게 답할 확률을 나타낸다. 문항반응함수에서 α_{lj}^{*}은 문항 j에서 요구하는 인지요소 벡터 중 피험자 l이 가지는 인지요소 벡터, K_j^{*}는 문항 j에서 요구하는 인지요소의 개수, g_j는 문항 j의 guess 모수, s_j는 문항 j의 slip 모수를 의미한다. 여기서 $\alpha_{lj}^{*} = 0_{K_l^{*}}$는 α_{lj}^{*}의 원소가 모두 0인 경우를 나타낸다. 즉, 피험자가 문항에서 요구하는 인지요소 모두를 숙달하지 않은 상태라면 문항에 옳게 답할 확률은 g_j, 추측하는 확률이 된다. 반대로 학생이 문항에서 요구하는 여러 인지요소 중 하나라도 숙달하고 있다면 문항에 옳게 답할 확률은 $1-s_j$, 실수하는 경우를 제외한 확률이 된다.

DINO 모형의 문항반응함수는 다음과 같은 형태로도 나타낼 수 있다.

$$P(X_{ij} = 1 \mid \alpha_i) = (1 - s_j)^{\omega_{ij}} g_j^{(1-\omega_{ij})} \tag{2.4}$$

여기서 X_{ij}는 피험자 i가 문항 j를 옳게 답하였는지를 나타내는데, 옳게 답하면 1, 옳게 답하지 않았으면 0으로 나타낼 수 있다. α_i는 피험자 i의 인지요소 벡터를 나타내고, 위의 문항반응함수는 인지요소 벡터 α_i를 가지는 피험자 i가 문항 j를 옳게 답할 확률을 의미한다. s_j와 g_j는 각각 문항 j의 slip 모수와 guess 모수를 나타낸다. ω_{ij}는 피험자가 문항에서 요구하는 인지요소 모두를 숙달한 상태인지를 나타내는데, $\omega_{ij} = 1 - \prod_{k=1}^{K_j^{*}} (1 - \alpha_{ik})$과 같이 나타낼 수 있다. 여기서 K_j^{*}는 문항 j에서 요구하는 인지요소의 수를 나타내고, α_{ik}는 피험자 i가 k번째 인지요소를

숙달했으면 1, 숙달하지 않았으면 0의 값을 가진다. DINO 모형의 경우 문항에서 요구하는 인지요소를 적어도 하나 숙달한 경우에는 $\omega_{ij}=1$, 하나의 인지요소라도 숙달하지 못한 경우에는 $\omega_{ij}=0$이다.

마찬가지로 문항 j에서 요구하는 인지요소가 3개인 경우를 가정하자. 그러면 3개의 인지요소에 대해 피험자가 가질 수 있는 인지요소의 조합(α_{lj}^{*})은 (0,0,0), (0,0,1), (0,1,0), (1,0,0,), (0,1,1), (1,0,1), (1,1,0), (1,1,1)로 총 8가지 경우다. 첫 번째로 제시된 DINO 모형의 문항반응함수에서 보면, 8가지의 인지요소 벡터 중 (0,0,0)인 경우만 $a_{lj}^{*}=0_3$이 성립한다. 즉, 피험자의 인지요소 벡터가 (0,0,0)인 경우에 피험자가 문항에 옳게 답할 확률은 g_j가 된다. 이는 피험자가 문항에서 요구하는 인지요소를 전부 숙달하지 못 한 경우에 피험자가 문항에 옳게 답할 확률을 추측하는 확률과 같아진다는 것을 의미한다. (0,0,0)을 제외한 나머지 7가지 경우는 피험자가 문항에서 요구하는 인지요소 중 적어도 하나를 숙달한 경우를 의미하는데, 이 경우에 피험자가 문항에 옳게 답할 확률은 $1-s_j$가 된다. 두 번째로 제시된 DINO 모형의 문항반응함수에서 보면, (0,0,0)인 경우에만 $w_{ij}=0$이 되므로 이 경우에 피험자가 문항에 옳게 답할 확률은 g_j가 된다. 그리고 (0,0,0)을 제외한 나머지 경우에서는 $w_{ij}=1$이 되므로 이 경우 피험자가 문항에 옳게 답할 확률은 $1-s_j$가 된다.

문항에 대해 옳게 답할 확률을 나타낼 때, DINA 모형과 DINO 모형은 인지요소간의 관계를 완전히 다르게 가정하고 있다. 하지만, DINO 모형도 각 문항이 두 개의 모수(g_j와 s_j)를 가지므로 DINA 모형과 마찬가지로 간단하지만 극단적인 특징을 지니고 있다.

3. ACDM(Additive CDM)

ACDM은 앞서 등장한 DINO 모형과 DINA 모형보다는 극단적인 특징이 약하게 나타난다. ACDM의 문항반응함수는 G-DINA 모형의 문항반응함수에서 구할 수 있는데, G-DINA 모형은 일반적인 모형으로 ACDM 이후

에 다루고자 한다. ACDM의 문항반응함수는 다음과 같다.

$$P(\alpha_{lj}^{*}) = \delta_{j0} + \sum_{k=1}^{K_j^{*}} \delta_{jk}\alpha_{lk} \tag{2.5}$$

여기서 α_{lj}^{*}은 문항 j에서 요구하는 인지요소 벡터 중 피험자 l이 가지는 인지요소 벡터, δ_{j0}는 문항 j의 intercept 값, δ_{jk}는 문항 j에서 요구하는 K_j^{*}개의 인지요소 중 k번째 인지요소를 숙달하고 있을 때 k번째 인지요소의 주효과를 의미한다. 그리고 α_{lk}는 α_{lj}^{*}에서 k번째 인지요소가 숙달되었으면 $\alpha_{lk}=1$, 숙달되지 않았으면 $\alpha_{lk}=0$의 값을 가진다. 그러면 ACDM은 각 문항별로 $K_j^{*}+1$개의 모수를 가진다. ACDM의 대표적인 특징은 문항반응함수의 식에서 알 수 있듯이 문항에 옳게 답할 확률은 문항에서 요구하는 인지요소를 얼마나 숙달하고 있느냐에 따라 문항을 해결할 확률이 달라진다는 점이다. 그리고 CDM에서 Additive라는 단어를 붙여 이름을 붙인 것처럼, ACDM은 문항반응함수에서 인지요소들의 주효과를 가산적인 관계로 가정하고 있다.

앞선 DINA 모형과 DINO 모형의 경우와 마찬가지로 문항 j에서 요구하는 인지요소가 3개인 경우의 예를 가정하자. 피험자 l_1이 인지요소를 하나도 숙달하고 있지 않다면, 즉, 피험자 l_1의 인지요소 벡터가 (0,0,0)인 경우 $\alpha_{l1}=\alpha_{l2}=\alpha_{l3}=0$이 된다. 그러면 피험자 l_1이 문항에 옳게 답할 확률은 다음과 같다.

$$P(\alpha_{l_1 j}^{*}) = \delta_{j0} \tag{2.6}$$

다음으로 피험자 l_2가 인지요소 3개 중 첫 번째 인지요소만 숙달하고 있다면, 즉 피험자 l_2의 인지요소 벡터가 (1,0,0)인 경우를 가정하자. 그러면 $\alpha_{l1}=1$, $\alpha_{l2}=\alpha_{l3}=0$이 된다. 그러면 피험자 l_2가 문항 j에 옳게 답할

확률은 다음과 같다.

$$P(\alpha^{*}_{l_2 j}) = \delta_{j0} + \delta_{j1} \tag{2.7}$$

다음으로 피험자 l_3가 문항 j에서 요구하는 인지요소 중 세 번째 인지요소만 숙달하고 있다면, 즉 피험자 l_3의 인지요소 벡터가 (0,0,1)인 경우를 가정하자. 그러면 $\alpha_{l1} = \alpha_{l2} = 0$, $\alpha_{l3} = 1$이 된다. 그러면 피험자 l_3가 문항 j에 옳게 답할 확률은 다음과 같다.

$$P(\alpha^{*}_{l_3 j}) = \delta_{j0} + \delta_{j3} \tag{2.8}$$

마지막으로 피험자 l_4가 문항 j에서 요구하는 인지요수 중 첫 번째와 세 번째 인지요소만 숙달하고 있다고 가정하자. 그러면 피험자 l_4의 인지요소 벡터는 (1,0,1)로 나타낼 수 있고, $\alpha_{l_4 1} = \alpha_{l_4 3} = 1$, $\alpha_{l_4 2} = 0$이 된다. 그러면 피험자 l_4가 문항 j에 옳게 답할 확률은 다음과 같다.

$$P(\alpha^{*}_{l_4 j}) = \delta_{j0} + \delta_{j1} + \delta_{j3} \tag{2.9}$$

위의 4가지 예에서도 알 수 있듯이, ACDM은 문항에서 요구하는 인지요소 중 피험자가 얼마나 숙달하고 있느냐에 따라 피험자가 문항에 옳게 답할 확률이 달라진다. ACDM은 피험자가 문항에서 요구하는 인지요소 모두를 숙달하고 있어야 문항에 옳게 답할 수 있다고 가정하는 DINA 모형과 문항에서 요구하는 인지요소 중 하나만 숙달하고 있더라도 문항에 옳게 답할 수 있다고 가정하는 DINO 모형보다는 많은 모수를 가지고 있고, 모형 자체의 가정이 DINA 모형과 DINO 모형 보다는 극단적인 성격이 약하다.

4. G-DINA(Generalized DINA) 모형

G-DINA 모형은 앞서 언급된 구체적인 인지진단모형(DINA 모형, DINO 모형, ACDM)과는 달리 일반적인 인지진단모형이다. G-DINA 모형은 DINA 모형을 일반화한 모형으로, 피험자가 문항 j에 옳게 답할 확률을 문항 j에서 요구하는 인지요소의 주효과와 인지요소 간의 상호작용 효과로 표현하고 있다. G-DINA 모형의 문항반응함수는 다음과 같이 나타낼 수 있다.

$$P(\alpha_{lj}^{*}) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{lk} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'}\alpha_{lk}\alpha_{lk'} \cdots + \delta_{j12\cdots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk} \quad (2.10)$$

문항반응함수는 피험자 l이 인지요소 벡터 α_{lj}^{*}를 가질 때, 피험자가 문항에 옳게 답할 확률을 나타낸다. 문항반응함수에서 α_{lj}^{*}은 피험자 l의 인지요소벡터 α_l중에서 문항 j에서 요구하는 인지요소로만 이루어진 축약 인지요소 벡터, K_j^*는 문항에서 요구하는 인지요소의 수, δ_{j0}은 intercept 값, δ_{jk}는 인지요소 α_{jk}에 의한 주효과, $\delta_{jkk'}$은 두 인지요소 α_{jk}와 $\alpha_{jk'}$의 상호작용 효과, $\delta_{jkk'k''}$는 세 인지요소 α_{jk}, $\alpha_{jk'}$, $\alpha_{jk''}$의 상호작용 효과, $\cdots$ $\delta_{j12\cdots K_j^*}$는 K_j^*개의 인지요소 $\alpha_1, \alpha_2, \cdots, \alpha_{K_j^*}$의 상호작용 효과를 나타낸다. G-DINA 모형은 문항 j에 대해 $2^{K_j^*}$개의 모수를 가지고 인지요소 프로파일 추정 과정에 있어 이 모수들을 추정해야 하므로, 계산에 상당히 많은 시간이 필요하다. G-DINA 모형이 가지는 가장 큰 특징 중 하나는 G-DINA 모형의 문항반응함수에서 모수를 조정하면 구체적인 인지진단모형을 얻을 수 있다는 점이다.

우선 G-DINA 모형의 문항반응함수에서 δ_{j0}와 모든 인지요소의 상호작용 효과 $\delta_{j12\cdots K_j^*}$를 제외한 모든 모수를 0으로 설정하면 DINA 모형의 문항반응함수를 얻을 수 있다.

$$P(\alpha_{lj}^{*}) = \delta_{j0} + \delta_{j12\cdots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk} \tag{2.11}$$

피험자가 문항 j에서 요구하는 K_j^*개의 인지요소 모두를 숙달했다면 ($\alpha_{lk}=1,\ k=1,2,\cdots,K_j^*$), 피험자가 문항 j를 해결할 확률은 $\delta_{j0}+\delta_{j12\cdots K_j^*}$가 되고, 문항 j에서 요구하는 K_j^*개의 인지요소 중 하나라도 숙달하지 못했다면($\alpha_{lk}=0$인 $k\in\{1,2,\cdots,K_j^*\}$가 적어도 하나 존재), $\prod_{k=1}^{K_j^*}\alpha_{lk}=0$ 이므로, 피험자가 문항 j를 해결할 확률은 δ_{j0}가 된다. 즉, G-DINA 모형의 문항반응함수가 DINA 모형의 문항반응함수와 같아진 것을 알 수 있다.

다음으로, G-DINA 모형의 문항반응함수에서 인지요소에 의한 주효과(δ_{jk})와 인지요소 간의 상호작용에 의한 효과의 절댓값이 모두 동일하다고 가정하고, 모수를 다음과 같이 설정하면 DINO 모형의 문항반응함수를 얻을 수 있다.

$$\delta_{jk} = -\delta_{jk'k''} = \cdots = (-1)^{K_j^*+1}\delta_{j12\cdots K_j^*} \tag{2.12}$$

여기서 $k=1,2,\cdots,K_j^*,\ k'=1,2,\cdots,K_j^*-1,\ k''>k',\ \cdots,K_j^*$다.

DINO 모형의 경우에는 DINA 모형처럼 하나의 식으로 정리되지는 않는다. 주어진 조건에 의해 문항반응함수의 함숫값이 어떻게 나타나는지를 살펴보면 DINO 모형의 문항반응함수와 동일한 것을 알 수 있다. DINO 모형은 문항 j가 요구하는 인지요소 모두에 숙달하고 있는 경우와 일부만 숙달하고 있는 경우에 문항반응함수의 함숫값이 같다. 예를 들어, 문항 j에 옳게 답하는데 필요한 인지요소가 3개(k_{j1}, k_{j2}, k_{j3})라고 하자. 이때, k_{j1}, k_{j2}만 숙달하고 있다면, 인지요소 벡터의 원소는

$\alpha_{jk_1}=1, \alpha_{jk_2}=1, \alpha_{jk_3}=0$과 같고, 문항반응함수의 결과는 다음과 같다.

$$P(\alpha_{lj}^{*})=\delta_{j0}+\delta_{jk_1}+\delta_{jk_2}+\delta_{jk_1k_2} \tag{2.13}$$

이때, 위의 주어진 DINO 모형의 모수 설정에 의하면, $P(\alpha_{lj}^{*})=\delta_{j0}+\delta_{jk_2}$가 된다.

그렇다면, 이번에는 인지요소 모두에 숙달하고 있다고 하자. 그러면 인지요소 벡터의 각 원소는 $\alpha_{jk_1}=1, \alpha_{jk_2}=1, \alpha_{jk_3}=1$이 되고, 문항반응함수의 결과는 다음과 같다.

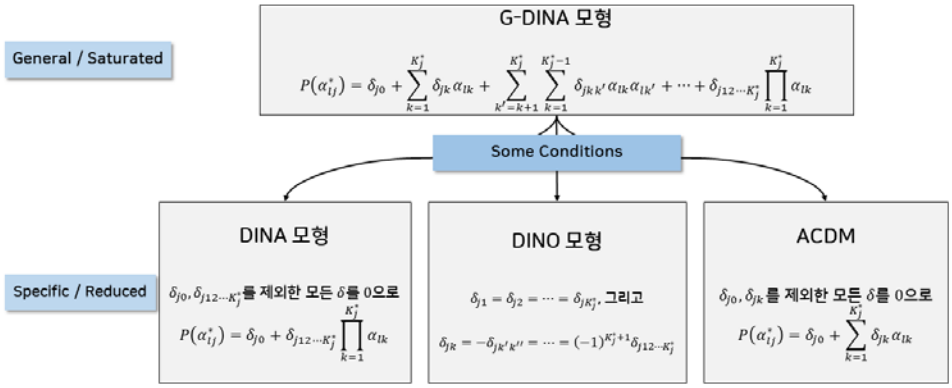
$$P(\alpha_{l'j}^{*})=\delta_{j0}+\delta_{jk_1}+\delta_{jk_2}+\delta_{jk_3}+\delta_{jk_1k_2}+\delta_{jk_1k_3}+\delta_{jk_2k_3}+\delta_{jk_1k_2k_3} \tag{2.14}$$

마찬가지로, DINO 모형의 모수 설정에 의하면 $P(\alpha_{l'j}^{*})=\delta_{j0}+\delta_{jk_3}$가 된다. 앞서 G-DINA 모형의 문항반응함수에서 DINO 모형의 문항반응함수를 구할 때, 인지요소에 의한 주효과는 동일하다고 했으므로, 인지요소를 2개만 숙달한 경우와 인지요소 모두에 숙달한 경우의 문항반응함수의 값은 동일함을 알 수 있다.

다음으로, G-DINA 모형의 문항반응함수에서 상호작용 효과들을 모두 0으로 제한하면 intercept 값과 인지요소의 주효과만 남게 되는데 이는 ACDM의 문항반응함수와 동일한 것을 알 수 있다.

$$P(\alpha_{lj}^{*})=\delta_{j0}+\sum_{k=1}^{K_j^{*}}\delta_{jk}\alpha_{lk} \tag{2.15}$$

G-DINA 모형의 문항반응함수와 DINA 모형, DINO 모형, ACDM의 문항반응함수의 관계를 정리하면 [그림 2-2]와 같이 나타낼 수 있다.



[그림 2-2] G-DINA 모형의 IRF와 DINA 모형, DINO 모형, ACDM의 IRF의 관계

문항에서 요구하는 인지요소가 3개인 경우에 앞서 언급한 세 구체적인 인지진단모형(DINA 모형, DINO 모형, ACDM)과 G-DINA 모형의 문항반응함수의 함숫값을 비교하면 다음 [표 2-1]과 같이 정리할 수 있다.

인지요소 벡터	DINA 모형	DINO 모형	ACDM	G-DINA 모형
(0,0,0)	δ_{j0}	δ_{j0}	δ_{j0}	δ_{j0}
(1,0,0)	δ_{j0}	$\delta_{j0}+\delta_{j1}$	$\delta_{j0}+\delta_{j1}$	$\delta_{j0}+\delta_{j1}$
(0,1,0)	δ_{j0}	$\delta_{j0}+\delta_{j1}$	$\delta_{j0}+\delta_{j2}$	$\delta_{j0}+\delta_{j2}$
(0,0,1)	δ_{j0}	$\delta_{j0}+\delta_{j1}$	$\delta_{j0}+\delta_{j3}$	$\delta_{j0}+\delta_{j3}$
(1,1,0)	δ_{j0}	$\delta_{j0}+\delta_{j1}$	$\delta_{j0}+\delta_{j1}+\delta_{j2}$	$\delta_{j0}+\delta_{j1}+\delta_{j2}+\delta_{j12}$
(1,0,1)	δ_{j0}	$\delta_{j0}+\delta_{j1}$	$\delta_{j0}+\delta_{j1}+\delta_{j3}$	$\delta_{j0}+\delta_{j1}+\delta_{j3}+\delta_{j13}$
(0,1,1)	δ_{j0}	$\delta_{j0}+\delta_{j1}$	$\delta_{j0}+\delta_{j2}+\delta_{j3}$	$\delta_{j0}+\delta_{j2}+\delta_{j3}+\delta_{j23}$
(1,1,1)	$\delta_{j0}+\delta_{j123}$	$\delta_{j0}+\delta_{j1}$	$\delta_{j0}+\delta_{j1}$ $+\delta_{j2}+\delta_{j3}$	$\delta_{j0}+\delta_{j1}+\delta_{j2}+\delta_{j3}$ $+\delta_{j12}+\delta_{j13}+\delta_{j23}+\delta_{j123}$

[표 2-1] $K_j^* = 3$일 때 구체적인 모형과 G-DINA 모형의 문항반응함수 비교

DINA 모형, DINO 모형, ACDM의 문항반응함수의 함숫값을 비교하면 ACDM

의 intercept 값(δ_{j0})은 결국 DINA 모형과 DINO 모형의 guess 모수(g_j)와 같은 것을 알 수 있다. 그리고 문항에서 요구하는 인지요소가 3개인 경우 ACDM이 가지는 모수가 4개이므로 단 2개의 모수(g_j, s_j)를 가지는 DINA 모형과 DINO 모형보다 ACDM이 더 복잡한 구조를 가지는 것을 알 수 있고, ACDM의 구조는 문항에서 요구하는 인지요소의 개수가 증가할수록 더 복잡해질 것이다.

문항에서 요구하는 인지요소가 3개일 때, G-DINA 모형은 $2^3 = 8$개의 모수를 갖는다. 단순히 모수의 개수로만 비교해도 DINA 모형, DINO 모형보다는 4배, ACDM 보다는 2배의 모수를 갖는다. 그리고 G-DINA 모형은 인지요소와 그 상호작용 모두를 고려하고 있기 때문에 구체적인 모형과 비교했을 때 더 복잡하다. 인지진단평가에서 G-DINA 모형을 사용하면 추정하는 모수가 많아 모형이 복잡한 형태를 보이더라도 정확한 추정을 할 수 있다. 하지만 인지요소의 수(K_j^*)가 증가하면 G-DINA 모형이 가지는 모수가 기하급수적으로 증가하기 때문에 많은 표본이 필요하고 소모되는 시간이 많아진다. 반면, 구체적인 모형(DINA 모형, DINO 모형, ACDM)은 일반적인 인지진단모형에 특정한 가정을 더하여 간단한 형태를 나타내고 결과 해석에는 용이하다(de la Torre & Lee, 2013).

제 2 절 G-DINA 모형의 모수 추정

1. 주변최대우도추정(Marginalized Maximum Likelihood Estimation, MMLE)

Neyman & Scott(1948)에 따르면, 문항반응함수 등을 구성하고 있는 구조적인 모수와 인지프로파일 등을 나타내는 부수적인 모수의 결합 최대화를 통해 추정값을 구하면, 일관되지 않은 추정값을 얻을 가능성이 있다. 그러므로 $L(X_i|\alpha_l)$과 같은 조건부우도대신 주변 우도를 사용해야 한다(de la Torre, 2009b). 여기에도 마찬가지로 EM(Expectation Maximization) 알고리즘을 적용하여 추정값을 구할 수 있는데, 이를 주변최대우도추정(Marginalized Maximum Likelihood Estimation, MMLE) 방법

이라고 부를 수 있다.

주변최대우도추정 방법을 사용하기 위해서는 앞서 언급한 주변우도가 정의되어야 한다. 주변우도 $L(X)$는 다음과 같이 정의할 수 있다.

$$L(X) = \prod_{i=1}^{I} L(X_i) = \prod_{i=1}^{I} \sum_{l=1}^{2^K} L(X_i|\alpha_l)p(\alpha_l) \tag{2.16}$$

G-DINA 모형의 모수를 주변최대우도추정 방법을 통해 추정하기 위해서는 응답 데이터의 로그-주변우도를 고려해야 한다. 로그-주변우도 $l(X)$는 다음과 같다.

$$l(X) = \log[L(X)] = \log\left[\prod_{i=1}^{I} \sum_{l=1}^{2^K} L(X_i|\alpha_l)p(\alpha_l)\right] \tag{2.17}$$

여기서 $L(X_i|\alpha_l)$는 인지요소 α_i가 주어졌을 때 피험자 i의 응답에 대한 우도로 $\prod_{j=1}^{J}[P(\alpha_{lj})]^{x_{ij}}[1-P(\alpha_{lj})]^{1-x_{ij}}$로 나타낼 수 있다. G-DINA 모형에서는 문항 j에서 요구하는 인지요소로만 이루어진 축약 인지요소벡터 α_{lj}^*가 사용되므로 $P(\alpha_{lj}^*)$를 사용해야 한다. 이렇게 축약 인지요소벡터를 사용하면 문항 j를 위해 추정해야 하는 모수는 $2^{K_j^*}$개로 줄어든다.

이제 $P(\alpha_{lj}^*)$에 대해 $l(X)$를 미분한 다음, 미분한 값이 0이 되게 하는 $P(\alpha_{lj}^*)$를 구하면 추정값 $\hat{P}(\alpha_{lj}^*)$를 구할 수 있다.

$$\hat{P}(\alpha_{lj}^*) = \frac{R_{\alpha_{lj}^*}}{I_{\alpha_{lj}^*}} \tag{2.18}$$

여기서 $I_{\alpha_{lj}^*} = \sum_{i=1}^{I} p(\alpha_{lj}^*|X_i)$는 인지프로파일 α_{lj}^*를 가질 것으로 예상하는

피험자의 수를 의미하고, $R_{\alpha_{lj}^*} = \sum_{i=1}^{I} p(\alpha_{lj}^* | X_i) X_{ij}$는 인지프로파일 α_{lj}^*를 갖는 피험자 중 문항j에 올바르게 응답할 피험자의 수를 의미한다. 여기서 $p(\alpha_{lj}^* | X_i)$는 피험자i가 인지프로파일 α_{lj}^*를 가질 사후확률을 의미한다.

이제 로그-주변우도($l(X)$)의 $P(\alpha_{lj}^*)$와 $P(\alpha_{l'j}^*)$에 대한 이계도함수는 다음과 같다.

$$\frac{\partial\, l(X)}{\partial\, P(\alpha_{lj}^*)\, \partial\, P(\alpha_{l'j}^*)} = -\sum_{i=1}^{I} \left[L^{-2}(X_i) \frac{\partial\, L(X_i)}{\partial\, P(\alpha_{lj}^*)} \frac{\partial\, L(X_i)}{\partial\, P(\alpha_{l'j}^*)} \right] \qquad (2.19)$$

이를 간단하게 나타내면 다음과 같다.

$$-\sum_{i=1}^{I} \left\{ p(\alpha_{lj}^* | X_i) \frac{X_{ij} - P(\alpha_{lj}^*)}{P(\alpha_{lj}^*[1 - P(\alpha_{lj}^*)])} \right\} \left\{ p(\alpha_{l'j}^* | X_i) \frac{X_{ij} - P(\alpha_{l'j}^*)}{P(\alpha_{l'j}^*[1 - P(\alpha_{l'j}^*)])} \right\} \qquad (2.20)$$

위에서 구한 $\hat{P}(\alpha_{lj}^*)$와 관찰된 X를 이용하면 문항 j의 모수에 대한 근사 정보 행렬 $I(\hat{P}_j^*)$을 구할 수 있다. 여기서 $P_j^* = \{P(\alpha_{lj}^*)\}$를 의미한다. 근사 정보 행렬의 역행렬 $I^{-1}(\hat{P}_j^*)$의 l번째 대각 원소는 $\hat{P}(\alpha_{lj}^*)$의 표준오차 추정치($SE[\hat{P}(\alpha_{lj}^*)]$)다.

2. G-DINA 모형의 설계 행렬

설계 행렬 M_j는 G-DINA 모형의 모수 추정($\hat{P}(\alpha_{lj}^*)$)을 그 구성요소(δ_{j0}, $\delta_{jk}, \delta_{jkk'}, \cdots, \delta_{j12\cdots K_j^*}$)로 변환하는데 중요한 역할을 하고 있다. 설계 행렬의 크기는 $2^{K_j^*} \times P$로 여기서 P는 관심을 두고 있는 모형이 가지는 모수의 수를 나타낸다. 이제 문항 j에서 요구하는 인지요소의 가능한 모든 조합

을 나타내는 $2^{K_j^*} \times K_j^*$행렬 $A_j = \{\alpha_{lk}\}$을 정의하자.

$$A_j = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix} \tag{2.21}$$

일반적인 모형의 설계 행렬 $M_j^{(S)}$의 l번째 행은 A_j의 l번째 행으로부터 얻을 수 있다. 첫 번째 원소는 1, 그 다음 원소는 $\alpha_{lk}(k = 1,2,\cdots,K_j^*)$, 다음 원소는 $\alpha_{lk} \times \alpha_{lk'}(k = 1,2,\cdots,K_j^* - 1,\ k' = k+1,\cdots,K_j^*)$가 되고, 그러면 마지막 원소는 $\prod_{k=1}^{K_j^*} \alpha_{lk}$가 된다. 예를 들어 $K_j^* = 3$인 경우에 일반적인 모형의 설계 행렬 $M_j^{(S)}$는 다음과 같다.

$$M_j^{(S)} = \begin{pmatrix} 1\,0\,0\,0\,0\,0\,0\,0 \\ 1\,1\,0\,0\,0\,0\,0\,0 \\ 1\,0\,1\,0\,0\,0\,0\,0 \\ 1\,0\,0\,1\,0\,0\,0\,0 \\ 1\,1\,1\,0\,1\,0\,0\,0 \\ 1\,1\,0\,1\,0\,1\,0\,0 \\ 1\,0\,1\,1\,0\,0\,1\,0 \\ 1\,1\,1\,1\,1\,1\,1\,1 \end{pmatrix} \tag{2.22}$$

이제 A_j의 각 행에 대하여 문항반응함수의 함숫값, 즉 문항을 해결할 확률 $P(\alpha_{lj}^*)$로 이루어진 $2^{K_j^*}$ 길이의 벡터 $P_j = \{P(\alpha_{lj}^*)\}$를 정의하자. 이제 $\hat{P}_j$가 주어졌을 때, $\delta_j = \{\delta_{j0}, \delta_{j1}, \delta_{j2}, \cdots, \delta_{jK_j^*}, \delta_{j12}, \cdots, \delta_{j12\cdots K_j^*}\}'$의 추정치를 구하려면, 최소 제곱 추정을 이용해 구할 수 있다.

$$\hat{\delta}_j = \left(M_j^{(S)'} M_j^{(S)}\right)^{-1} M_j^{(S)'} \hat{P}_j \tag{2.23}$$

3. 구체적인 모형의 설계 행렬

앞서 살펴 본 설계 행렬은 G-DINA 모형의 모수 추정($\hat{P}_j(\alpha_{lj}^*)$)을 G-DINA 모형의 구성요소들로 변환하는데 사용되었다. 이제 이 설계 행렬을 적절히 사용하면, G-DINA 모형의 모수 추정을 각 구체적인 모형들의 모수들로 변환할 수 있다. $M^{(r)}$을 구체적인 모형 r에 대한 설계 행렬이라고 하자. 그러면, DINA 모형의 설계 행렬은 다음과 같이 나타낼 수 있다.

$$M_j^{(r)}{}_{[2^{K_j^*}\times 2]} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \tag{2.24}$$

위 설계 행렬은 가능한 $2^{K_j^*}$ 가지 경우의 인지요소 벡터 중 모든 인지요소를 숙달한 경우 $(1,1,\cdots,1)$를 제외한 $(0,0,\cdots,0)$부터 $(1,0,\cdots,0)$, $\cdots$, $(0,1,1,\cdots,1)$까지는 모두 같게 만들어주는 역할을 한다.

다음으로, DINO 모형의 설계 행렬은 다음과 같이 나타낼 수 있다.

$$M_j^{(r)}{}_{[2^{K_j^*}\times 2]} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{2.25}$$

위 설계 행렬은 가능한 $2^{K_j^*}$ 가지 경우의 인지요소 벡터 중 인지요소를 하나도 숙달하지 못한 경우 $(0,0,\cdots,0)$를 제외한 $(1,0,\cdots,0)$부터 $(0,1,\cdots,0)$, $\cdots$, $(1,1,1,\cdots,1)$까지는 모두 같게 만들어주는 역할을 한다.

마지막으로, ACDM의 설계 행렬은 G-DINA 모형의 설계 행렬 $M_j^{(S)}$ 중 첫 K_j^*+1개의 열로 이루어진 행렬과 같다. 예를 들어, $K_j^*=3$인 경우, ACDM의

설계 행렬은 다음과 같이 나타낼 수 있다.

$$M_j^{(r)} = \begin{pmatrix} 1\,0\,0\,0 \\ 1\,1\,0\,0 \\ 1\,0\,1\,0 \\ 1\,0\,0\,1 \\ 1\,1\,1\,0 \\ 1\,1\,0\,1 \\ 1\,0\,1\,1 \\ 1\,1\,1\,1 \end{pmatrix} \tag{2.26}$$

위 설계 행렬은 첫 행은 인지요소 벡터가 $(0,0,0)$인 경우, 2~4번째 행은 3개의 인지요소 중 1개의 인지요소만 숙달한 경우($(1,0,0)$, $(0,1,0)$, $(0,0,1)$), 5~7번째 행은 3개의 인지요소 중 2개의 인지요소만 숙달한 경우($(1,1,0)$, $(1,0,1)$, $(0,1,1)$), 마지막 8번째 행은 3개의 인지요소 모두를 숙달한 경우($(1,1,1)$)를 나타낸다.

우리는 문항에서 요구하는 인지요소의 개수에 따라 $2^{K_j^*}$가지의 인지요소 벡터를 만들 수 있고, 이에 속하는 피험자들을 묶어 잠재 집단으로 분류할 수 있다. 그런데 이 집단들이 균등분포를 따르는 것이 아니라면, 모수를 추정함에 있어 집단 간의 상대적인 크기를 고려해야한다. 그래서 $P(\alpha_{lj}^*)$에 가중치를 주어야 하는데, 이 가중치의 역할을 할 수 있는 것이 바로 $I_{\alpha_{lj}^*}$다. 여기서 $I_{\alpha_{lj}^*}$는 식 2.18에서와 마찬가지로 인지요소 벡터 α_{lj}^*를 가질 것으로 예상되는 피험자의 수를 의미한다. $I_{\alpha_{lj}^*}$를 이용하여 가중치의 역할을 할 수 있는 대각 행렬 W_j는 다음과 같이 정의할 수 있다.

$$W_{j[2^{K_j^*} \times 2^{K_j^*}]} = \left\{ I_{\alpha_{lj}^*} \right\} \tag{2.27}$$

이제 앞서 살펴본 G-DINA 모형의 경우와 마찬가지로, 구체적인 모형의 설계행렬과 가중치 행렬을 이용하여 최소제곱추정을 통해 구체적인 모형의 모수 추정 $\hat{\delta}_j$를 구할 수 있다.

$$\hat{\delta}_j = (M_j^{(r)'} W_j M_j^{(r)})^{-1} M_j^{(r)'} W_j \widehat{P}_j \tag{2.28}$$

제 3 절 Wald 검사(Wald test)

인지진단평가를 수행함에 있어 적합한 모형을 선택하는 과정은 모형과 데이터 간의 적합도 검사를 포함한다. 그렇다면 모형과 데이터 간의 적합도를 측정하는 통계량이 필요하다. 지금까지 모형-데이터 적합도에 관한 연구가 많이 진행되었지만, 이는 대부분이 검사수준의 적합도(예: Chen, de la Torre, & Zhang, 2013)에 대한 연구였다. 문항수준 모형적합도를 측정하는 검사 및 그 통계량은 최근 들어서야 제시되기 시작하였는데, 그 중 하나가 de la Torre(2011)가 제시한 Wald 검사다.

Wald 검사를 통해 계산되는 Wald 통계량(W)는 상대 적합도 통계량의 하나로 인지진단모형 간의 비교를 통해 적합도를 판단한다. Wald 통계량을 통한 적합도 비교의 기본 개념은 사고 절약의 원리를 바탕으로 하고 있다. 사고 절약의 원리는 14세기 영국의 논리학자 오컴의 이름을 따 오컴의 면도날이라고도 불리는데, 그가 제시한 원리는 다음과 같은 형태로 실험 과학 방법론에서 언급되었다: 데이터에 대한 두 설명 중, 모든 조건이 동일하다면 간단한 설명을 선택해야한다(Blumer et al., 1987). 즉, Wald 통계량을 바탕으로 한 모형 선택은 인지요소와 Q-행렬이 참이라고 가정할 때 여러 인지진단모형이 데이터에 동등하게 들어맞는다면 그중에서 간단한 모형을 적합한 인지진단모형으로 선택하는 방법이다(Sorrel et al., 2017a).

Wald 검사는 우도 비(likelihood ratio) 검사 등과 같이 문항반응이론 분야에서 연구되던 통계량들이 인지진단모형 분야로 확장되는 것과 다르게 인지진단모형 연구에서 처음으로 제시되었다. Wald 검사의 기본적인 개념은 구체적인 모형이 일반적인 모형 대신에 사용될 수 있는지를 통계적으로 비교한다. 즉, 관찰된 데이터에 대한 구체적인 모형과 일반적인 모

형의 적합도가 통계적으로 비교된다. Wald 통계량은 다음과 같이 구할 수 있다.

$$W = [R_{jp} \times f(P_j)]' \left\{ R_{jp} \times Var[f(P_j)] \times R_{jp}' \right\}^{-1} [R_{jp} \times f(P_j)] \tag{2.29}$$

위의 식에서 R_{jp}는 문항j와 구체적인 모형 p에 대한 $(2^{K_j} - p) \times 2^{K_j}$ 행렬을 나타내고 이때 p는 구체적인 모형이 가지는 모수의 수를 나타낸다. 예를 들어 문항에서 요구하는 인지요소가 3개일 때($K_j = 3$), DINA 모형, DINO 모형, ACDM의 R_{jp}행렬은 각각 다음과 같다.

$$R^{(1)}_{jp6\times 8} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \end{pmatrix} \tag{2.30}$$

$$R^{(2)}_{jp6\times 8} = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \tag{2.31}$$

$$R^{(3)}_{jp4\times 8} = \begin{pmatrix} 1 & -1 & -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 & 0 & 1 & 0 \\ -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 \end{pmatrix} \tag{2.32}$$

DINA 모형의 경우 $R^{(1)}_{jp6\times 8}$행렬은 δ_{j0}과 δ_{j123}을 제외한 다른 모수(δ_{j1}, δ_{j2}, δ_{j3}, δ_{j3}, δ_{j12}, δ_{j13}. δ_{j23})를 모두 0으로 만들어주고, DINO 모형의 경우 $R^{(2)}_{jp6\times 8}$ 행렬은 $\delta_{j1} = \delta_{j2} = \delta_{j3} = -\delta_{j12} = -\delta_{j13} = -\delta_{j23} = \delta_{j123}$의 역할을 하고, ACDM의 경우 $R^{(3)}_{jp4\times 8}$ 행렬은 상호작용 효과 $\delta_{12}, \delta_{13}, \delta_{23}, \delta_{123}$를 모두 0으로 만드는 역할을 한다. $P_j = \left\{ P(\alpha^*_{lj}) \right\}$는 A_j의 각 행에 대하여 문항반응함수의 함숫값, 즉 문항을 해결할 확률 $P(\alpha^*_{lj})$로 이루어진 $2^{K^*_j}$길이의 벡터를

나타낸다. 여기서 A_j는 앞서 언급했듯이 문항j에서 요구되는 인지요소들의 가능한 모든 조합을 나타내는 $2^{K_j^*} \times K_j^*$행렬이다. 다음으로 $f(P_j) = \left(M_j^{(S)'} M_j^{(S)}\right)^{-1} M_j^{(S)'} P_j$는 G-DINA 모형의 모수 $\delta_j = \left\{\delta_{j0}, \delta_{j1}, \cdots, \delta_{jK_j^*}, \delta_{j12}, \cdots, \delta_{j12 \cdots K_j^*}\right\}'$의 최소제곱추정을 나타내고, 여기서 $M_j^{(S)}$는 문항j에 대한 설계행렬이다. 마지막으로 $Var[f(P_j)]$는 다음과 같이 근사될 수 있다. 여기서 $\nabla f(P_j)$는 $f(P_j)$의 gradient, $Var[P_j]$는 P_j의 정보행렬의 역행렬에 -1을 곱한 값을 나타낸다.

$$Var[f(P_j)] \approx \nabla f(P_j)' Var[P_j] \nabla f(P_j) \tag{2.33}$$

Wald 통계량은 귀무가설 하에서 점근적으로 자유도 $2^{K_j^*} - p$인 χ^2분포를 따른다. Wald 검사는 각 문항에 대해 구체적인 모형별로 Wald 통계량을 계산하고, 유의확률을 통해 구체적인 인지진단 모형이 일반적인 인지진단 모형 대신 사용될 수 있는지를 판단할 수 있게 한다. Wald 통계량을 이용해 문항에 적합한 인지진단모형을 결정하는 방법이 가지는 가장 큰 장점은 G-DINA 모형과 같은 일반적인 모형의 모수 추정을 통해 구체적인 모형의 모수 추정을 구할 수 있기 때문에 구체적인 모형의 모수 추정을 구하기 위한 계산과정이 따로 필요하지 않는다는 점이다(Sorrel et al. 2017b). 그리고 Wald 검사는 로그우도 비(log-likelihood ratio) 검사와 같이 검사수준에서 모형적합도를 분석하는 기존 모형적합도 분석과는 달리, 문항수준에서 모형적합도를 분석할 수 있다는 특징을 갖는다.

제 3 장 연구방법

제 1 절 연구대상

본 연구는 모의실험과 실제 데이터 연구를 모두 수행하였다. 모의실험을 통해 응답데이터를 생성할 때 가정한 생성모형과 Wald 통계량에 의해 각 문항에 적합하다고 결정된 인지진단모형 간의 일치율을 조사하였다. 그리고 이를 바탕으로 Wald 통계량을 이용한 문항수준 모형 선택의 정확도에 대해 탐구하고, Q-행렬의 구조, 문항에서 요구하는 인지요소의 개수, 문항의 질 등의 요인을 변화시키며 생성모형과 Wald 통계량을 이용해 결정된 인지진단모형의 비율을 조사했다. 그리고 이를 바탕으로 Wald 통계량을 이용한 모형 선택 방법의 정확도에 영향을 미치는 요인은 무엇인지 분석하고자 했다. 실제 데이터에 적용하기 위해 TIMSS 2007(Martin, Mullis, Foy, & Olson, 2008)의 초등학교 4학년 대상 검사지 4번에 대한 응답데이터와 윤지영(2016)의 일차방정식과 일차부등식 진단평가에 대한 중학교 2,3학년 학생들의 응답데이터 등 실제 데이터에 Wald 통계량을 이용한 문항수준 모형 선택 방법을 적용하고 각 문항에 적합한 인지진단모형은 무엇인지 분석했다. 그리고 Wald 통계량을 이용해 결정한 인지진단모형과 실제 문항, 문항에서 요구하는 인지요소 등을 분석하여 Wald 통계량을 이용한 모형 선택 방법을 수행한 결과를 해석했다.

모의실험에서는 다음과 같은 4가지 요인, 문항에서 요구하는 인지요소의 수, Q-행렬의 구조, 문항의 질, 단일모형 기반 검사/혼합모형 기반 검사 여부 등을 달리하여 Wald 통계량을 이용한 모형 선택 방법의 정확도에 영향을 미치는 요인을 파악하고자 했다. 각 모의실험에서는 가상의 피험자 1,000명에게 검사지가 주어진다고 가정하여 응답 데이터를 산출했다. 모의실험 1-1, 모의실험 1-2, 모의실험 1-3, 모의실험 2-1, 모의실

험 2-2, 모의실험 2-3, 모의실험 3-1, 모의실험 3-2, 모의실험 3-3에서 가정한 검사지는 생성모형을 DINA 모형, DINO 모형, ACDM 순서대로 가정하여 같은 문항군이 세 번 반복되는 혼합모형 기반 검사로 가정하였다. 즉, 한 검사지는 DINA 모형을 생성모형으로 가정하는 문항군, DINO 모형을 생성모형으로 가정하는 문항군, ACDM을 생성모형으로 가정하는 문항군으로 이루어져 있다. 반면 모의실험 4-1에서는 생성모형을 DINA 모형으로 가정한 검사지, 모의실험 4-2에서는 생성모형을 DINO 모형으로 가정한 검사지, 모의실험 4-3에서는 생성모형을 ACDM으로 가정한 검사지와 같이 단일모형 기반 검사를 가정하여 응답데이터를 산출하여 모의실험을 수행했다. 모의실험 설계에 대한 더 자세한 설명은 다음 절 연구절차에서 다루기로 한다.

실제 데이터 연구에서 사용한 실제 데이터는 TIMSS 2007(Martin, Mullis, Foy, & Olson, 2008) 데이터와 윤지영(2016)의 데이터다. 우선, TIMSS는 Trends in International Mathematics and Science Study의 약자로 수학과 과학 두 과목의 성취도를 검사할 수 있는 검사지와 설문지로 이루어져 있다. 이 연구에서는 수학 성취도 검사지에 대한 응답 데이터만 사용한다. TIMSS 2007 수학 성취도 평가는 4학년과 8학년(우리나라의 중학교 2학년) 학생을 대상으로 하는데, 이 연구에서는 4학년 학생들의 4번 검사지에 대한 응답 데이터를 사용했다. 4번 검사지를 사용한 이유는 TIMSS 2007을 연구한 Lee, Park, & Taylan(2011)의 연구에서 기인하는데, 4번 검사지는 총 25개의 문항으로 이루어져 있으며 맞으면 1, 틀리면 0으로 기록되는 이분 문항의 비중이 가장 높다. 25개의 문항 중 문항 ID M041275와 M031247은 다분 문항으로 문항 점수가 2점, 1점, 0점으로 나뉜다. 하지만 Lee, et al.(2011)의 연구에서는 2점을 획득한 경우는 1점으로, 1점과 0점을 획득한 경우는 0점으로 바꾸어 연구를 진행했다.

이 연구 US 표본과 두 벤치마크 주인 Minnesota주와 Massachusetts주의 응답 데이터를 사용했다. 피험자는 US 표본을 구성하는 255개 학교의

564명의 피험자, Minnesota주 48개 학교의 132명의 피험자, 그리고 Massachusetts주 47개 학교의 127명의 피험자로 총 823명의 피험자의 응답데이터를 사용했다.

다음으로 사용한 윤지영(2016)의 데이터는 일차방정식과 일차부등식 진단평가에 대한 응답데이터를 사용했다. 윤지영(2016)의 연구에서는 일차방정식과 일차부등식의 학습이 중학교 1,2학년에 걸쳐서 일어나기 때문에 해당 내용의 숙달 양상을 알아보기 위해 중학교 2,3학년 학생들을 피험자로 선정하여 검사를 실시했다. 본 연구에서는 피험자들의 개인정보는 삭제된 응답데이터만을 사용했다. 즉, 검사지에 대한 피험자들의 응답데이터를 바탕으로 각 문항에 대해 Wald 통계량을 계산하여 이를 이용해 적합한 인지진단모형은 무엇인지 분석했고, 윤지영(2016)이 설정한 인지요소를 바탕으로 그 의미를 알아보았다.

제 2 절 연구절차

1. 모의실험 설계

연구의 목적이 Wald 통계량을 이용하여 문항수준에서 적합한 인지진단모형을 결정할 수 있는지 알아보는 것이므로, 생성모형을 이용해 응답데이터를 산출하고 각 생성모형에 대해 Wald 통계량에 의한 모형 선택 비율을 조사했다. 그리고 각 생성모형에 대해 Wald 통계량에 의한 모형 선택 비율에 영향을 미치는 요인들을 알아보기 위해 다양한 조건에서 모의실험을 수행했다. 각 문항이 요구하는 인지요소의 개수, Q-행렬의 구조, 문항의 질, 단일모형 기반 검사/혼합모형 기반 검사 여부 등을 모의실험 요인으로 설정하였다. 각 문항이 요구하는 인지요소의 개수가 2개인 경우와 3개인 경우를 설정했고, 검사 전체에서 요구하는 인지요소의 개수가 많지 않고 각 문항이 요구하는 인지요소의 개수가 단순해 비교적 간단한 Q-행렬과 검사전체에서 요구하는 인지요소의 개수가 많고, 각 문

항이 요구하는 인지요소의 개수가 다양해 복잡한 Q-행렬이 사용되는 경우를 설정했다. 문항의 질은 guess 모수와 slip 모수의 값을 달리하여 설정했다. guess 모수와 slip 모수의 값이 크면, 인지요소를 숙달하지 않았는데도 문항에 옳게 답할 확률이 커지고, 인지요소를 숙달하였는데도 문항에 옳게 답할 확률이 낮아진다. 즉, guess 모수와 slip 모수의 값이 크면 문항의 질이 낮아지고, 그 값이 작으면 문항의 질이 높아진다고 할 수 있다. 마지막으로 검사지 전체에 세 가지 인지진단모형을 섞어서 응답데이터를 산출한 경우와, 각 인지진단모형을 따로 분리하여 응답데이터를 산출한 경우를 설정했다. 그리고 모의실험에서 조작한 요인들이 변화함에 따라 각 생성모형에 대해 Wald 통계량을 이용한 문항수준 모형 선택 비율을 조사했다.

1.1 간단한 Q-행렬, 문항 당 요구하는 인지요소 2개인 경우

우선, 모의실험 1-1, 모의실험 1-2, 모의실험 1-3에서는 간단한 구조의 Q-행렬을 사용했는데, 검사지 전체에서 요구하는 인지요소는 총 5개로 제한하고, 각 문항이 요구하는 인지요소의 개수는 모두 2개로 동일하게 가정하여 Q-행렬을 구성했다. 그 다음 문항의 질이 미치는 영향을 알아보기 위해 모의실험 1-1, 모의실험 1-2, 모의실험 1-3의 문항의 질을 각각 상, 중, 하로 가정했다. 문항의 질이 상인 경우는 guess 모수와 slip 모수를 각각 0.05로, 문항의 질이 중인 경우는 guess 모수와 slip 모수를 각각 0.1로, 문항의 질이 하인 경우는 guess 모수와 slip 모수를 각각 0.2로 설정했다. 모의실험 1-1, 모의실험 1-2, 모의실험 1-3에서 가정한 검사지는 혼합모형 기반검사로, 생성모형을 달리하여 총 90문항으로 검사지를 구성했다. 1~30번은 DINA 모형, 31~60번은 DINO 모형, 61~90번은 ACDM을 생성모형으로 하여 응답데이터를 산출했다. 각 실험은 1,000명의 응답데이터를 산출하여 문항별 Wald 통계량을 계산하고 문항별로 적합한 인지진단모형을 결정했다. 총 1,000번의 실험을 반복하여 각 생성모형별로 Wlad 통계량에 의한 모형 선택 비율을 조사했다. 모의실험

1-1, 모의실험 1-2, 모의실험 1-3에서 사용한 Q-행렬은 [표 3-1]과 같다.

Item	ATTR1	ATTR2	ATTR3	ATTR4	ATTR5	생성모형
1	1	1	0	0	0	DINA 모형
2	0	1	1	0	0	
3	0	0	1	1	0	
4	0	0	0	1	1	
5	1	0	1	0	0	
6	1	0	0	1	0	
7	1	0	0	0	1	
8	0	1	0	1	0	
9	0	1	0	0	1	
10	0	0	1	0	1	
11	1	1	0	0	0	
12	0	1	1	0	0	
13	0	0	1	1	0	
14	0	0	0	1	1	
15	1	0	1	0	0	
16	1	0	0	1	0	
17	1	0	0	0	1	
18	0	1	0	1	0	
19	0	1	0	0	1	
20	0	0	1	0	1	
21	1	1	0	0	0	
22	0	1	1	0	0	
23	0	0	1	1	0	
24	0	0	0	1	1	
25	1	0	1	0	0	
26	1	0	0	1	0	
27	1	0	0	0	1	
28	0	1	0	1	0	
29	0	1	0	0	1	
30	0	0	1	0	1	
31	1	1	0	0	0	DINO 모형
32	0	1	1	0	0	
33	0	0	1	1	0	
⋮						
58	0	1	0	1	0	
59	0	1	0	0	1	
60	0	0	1	0	1	
61	1	1	0	0	0	ACDM
62	0	1	1	0	0	
63	0	0	1	1	0	
⋮						
88	0	1	0	1	0	
89	0	1	0	0	1	
90	0	0	1	0	1	

[표 3-1] 전체 인지요소 5개, 문항에서 요구하는 인지요소 2개인 Q-행렬

1.2 간단한 Q-행렬, 문항 당 요구하는 인지요소 3개인 경우

다음으로, 모의실험 2-1, 모의실험 2-2, 모의실험 2-3에서는 앞선 세 모의실험과 마찬가지로 간단한 구조의 Q-행렬을 사용하여, 검사지에서 요구하는 전체 인지요소의 개수는 5개로 제한했다. 그리고 각 문항이 요구하는 인지요소의 개수는 모두 3개로 가정하여 Q-행렬을 구성했다. 마찬가지로 문항의 질이 미치는 영향을 알아보기 위해 모의실험 2-1, 모의실험 2-2, 모의실험 2-3의 문항의 질을 각각 상, 중, 하로 가정했다. 문항의 질이 상인 경우는 guess 모수와 slip 모수를 각각 0.05로, 문항의 질이 중인 경우는 guess 모수와 slip 모수를 각각 0.1로, 문항의 질이 하인 경우는 guess 모수와 slip 모수를 각각 0.2로 설정했다. 모의실험 2-1, 모의실험 2-2, 모의실험 2-3도 혼합모형 기반 검사로 생성모형을 달리하여 총 90문항으로 검사지를 구성했다. 1~30번은 DINA 모형, 31~60번은 DINO 모형, 61~90번은 ACDM을 생성모형으로 하여 응답데이터를 산출했다. 각 실험은 1,000명의 응답데이터를 산출하여 문항별 Wald 통계량을 계산하고 문항별로 적합한 인지진단모형을 결정했다. 총 1,000번의 실험을 반복하여 각 생성모형별로 Wlad 통계량에 의한 모형 선택 비율을 조사했다. 모의실험 2-1, 모의실험 2-2, 모의실험 2-3에서 사용한 Q-행렬은 [표 3-2]와 같다.

Item	ATTR1	ATTR2	ATTR3	ATTR4	ATTR5	생성모형
1	1	1	1	0	0	DINA 모형
2	1	1	0	1	0	
3	1	1	0	0	1	
4	1	0	1	1	0	
5	1	0	1	0	1	
6	1	0	0	1	1	
7	0	1	1	1	0	
8	0	1	1	0	1	
9	0	1	0	1	1	
10	0	0	1	1	1	
11	1	1	1	0	0	
12	1	1	0	1	0	
13	1	1	0	0	1	
14	1	0	1	1	0	
15	1	0	1	0	1	
16	1	0	0	1	1	
17	0	1	1	1	0	
18	0	1	1	0	1	
19	0	1	0	1	1	
20	0	0	1	1	1	
21	1	1	1	0	0	
22	1	1	0	1	0	
23	1	1	0	0	1	
24	1	0	1	1	0	
25	1	0	1	0	1	
26	1	0	0	1	1	
27	0	1	1	1	0	
28	0	1	1	0	1	
29	0	1	0	1	1	
30	0	0	1	1	1	
31	1	1	1	0	0	DINO 모형
32	1	1	0	1	0	
33	1	1	0	0	1	
⋮						
58	0	1	1	0	1	
59	0	1	0	1	1	
60	0	0	1	1	1	
61	1	1	1	0	0	ACDM
62	1	1	0	1	0	
63	1	1	0	0	1	
⋮						
88	0	1	1	0	1	
89	0	1	0	1	1	
90	0	0	1	1	1	

[표 3-2] 전체 인지요소 5개, 문항에서 요구하는 인지요소 3개인 Q-행렬

1.3 복잡한 구조의 Q-행렬

앞서 살펴본 총 6개의 모의실험에서는 간단한 구조의 Q-행렬을 사용했다. 하지만 실제 인지진단평가를 수행할 때 사용되는 Q-행렬은 복잡한 구조의 Q-행렬이 사용되는 경우가 빈번하다. 그래서 모의실험 3-1, 모의실험 3-2, 모의실험 3-3에서는 TIMSS 2007(Martin et al., 2008) 데이터를 연구한 Lee, et al.(2011)의 연구에서 사용된 4학년 4번 검사지의 Q-행렬을 사용했다. Lee et al.(2011)의 Q-행렬은 검사에서 요구하는 전체 인지요소는 15개이고, 각 문항이 요구하는 인지요소가 최소 1개에서 최대 6개로 다양하게 이루어져 있어 검사지에서 요구하는 전체 인지요소 5개 중 각 문항이 요구하는 인지요소가 2개 또는 3개로 제한했던 앞선 모의실험들보다는 더 복잡한 구조를 갖는다. TIMSS 2007(Martin et al., 2008) 4학년 4번 검사지는 총 25개의 문항으로 이루어져 있으며, Lee et al.(2011)의 Q-행렬은 문항에서 요구하는 인지요소가 1개인 문항은 3개, 문항에서 요구하는 인지요소가 2개인 문항은 8개, 문항에서 요구하는 인지요소가 3개인 문항은 8개, 문항에서 요구하는 인지요소가 4개인 문항은 4개, 문항에서 요구하는 인지요소가 5개와 6개인 문항은 각각 1개씩으로 이루어져 있다. 이는 다음의 [표 3-3]과 같이 정리할 수 있다.

문항에서 요구하는 인지요소의 수	문항 수
1	3
2	8
3	8
4	4
5	1
6	1

[표 3-3] Lee et al.(2011)의 Q-행렬의 구조

또한, 문항의 질이 미치는 영향을 알아보기 위해 모의실험 3-1, 모의실

험 3-2, 모의실험 3-3의 문항의 질을 각각 상, 중, 하로 가정했다. 문항의 질이 상인 경우는 guess 모수와 slip 모수를 각각 0.05로, 문항의 질이 중인 경우는 guess 모수와 slip 모수를 각각 0.1로, 문항의 질이 하인 경우는 guess 모수와 slip 모수를 각각 0.2로 설정했다. 모의실험 3-1, 모의실험 3-2, 모의실험 3-3도 혼합모형 기반 검사로 생성모형을 달리하여 총 75문항으로 검사지를 구성했다. 1~25번은 DINA 모형, 26~50번은 DINO 모형, 51~75번은 ACDM을 생성모형으로 하여 응답데이터를 산출했다. 각 실험은 1,000명의 응답데이터를 산출하여 이를 바탕으로 Wald 통계량을 계산했다. 모의실험 3-1은 총 187번, 모의실험 3-2는 총 136번, 모의실험 3-3은 총 500번의 실험을 반복하여 각 생성모형별로 Wlad 통계량에 의한 모형 선택 비율을 조사했다. 모의실험 3-1, 모의실험 3-2, 모의실험 3-3에 사용한 Q-행렬은 [표 3-4]와 같다.

Item	ATTR1	ATTR2	ATTR3	ATTR4	ATTR5	ATTR6	ATTR7	ATTR8	ATTR9	ATTR10	ATTR11	ATTR12	ATTR13	ATTR14	ATTR15	생성모형
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	
3	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	
4	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	
5	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	
6	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	
7	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	
8	1	1	1	0	0	0	0	0	0	1	1	0	0	0	0	
9	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
10	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	
11	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	
12	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	
13	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0	DINA 모형
14	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0	
15	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
16	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
17	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	
18	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	
19	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	
20	0	1	1	0	0	0	0	1	0	0	0	0	0	1	0	
21	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	
22	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	
23	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
24	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
25	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	
26	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
27	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	
28	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	
⋮																DINO 모형
48	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
49	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
50	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	
51	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
52	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	
53	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	
⋮																ACDM
73	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
74	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
75	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	

[표 3-4] 복잡한 구조의 Q-행렬

(Lee, et al.(2011)의 Q-행렬 세 번 반복)

1.4 복합모형/단일모형 기반 검사

앞서 제시된 아홉 번의 모의실험의 공통점은 30개 또는 25개의 문항을 생성모형을 달리하여 세 번 반복하여 검사지를 구성했다는 점이다. 즉, 같은 문항군이 세 번 반복되게 검사지를 가정했다. 그러므로 각 모의실험에서 가정한 검사지는 혼합된 모형을 기반으로 한 검사다. 그렇다면 모든 문항을 하나의 생성모형만 가정하여 응답데이터를 산출할 경우, 즉, 혼합모형 기반 검사가 아닌 단일모형 기반 검사의 경우에 각 생성모형별로 Wlad 통계량에 의한 모형 선택 비율이 어떻게 변화하는지를 보기 위한 모의실험을 설계했다. 그래서 모의실험 4-1, 모의실험 4-2, 모의실험 4-3에서는 Lee, et al.(2011)의 연구에서 사용된 Q-행렬을 세 번 반복하는 것이 아니라 한 번만 사용하여 검사지를 구성했다. 그리고 단일모형 기반 검사로 가정하기 위해 각각 모의실험 4-1은 DINA 모형을, 모의실험 4-2는 DINO 모형을, 모의실험 4-3은 ACDM을 생성모형으로 설정하여 응답데이터를 산출했다. 즉, 모의실험 4-1, 모의실험 4-2, 모의실험 4-3에서 가정한 검사지는 모의실험 3-1, 모의실험 3-2, 모의실험 3-3에서 가정한 검사지의 $\frac{1}{3}$이다. 여기서 문항의 질은 하로 설정하여 guess 모수와 slip 모수는 모두 0.2로 설정하고 각 실험은 1,000명의 응답데이터를 산출하여 Wald 통계량을 계산했으며, 총 500번의 실험을 반복하여 각 생성모형별로 Wlad 통계량에 의한 모형 선택 비율을 조사했다. 모의실험 4-1, 모의실험 4-2, 모의실험 4-3에 사용한 Q-행렬은 [표 3-7]과 같다.

지금까지 설계된 총 12개의 모의실험과 모의실험에서 조작한 요인 및 조건은 [표 3-5]와 같이 정리할 수 있다.

	Q-행렬의 구조	간단한 Q-행렬		복잡한 Q-행렬
	문항에서 요구하는 인지요소 수	2개	3개	
혼합모형 기반 검사	문항의 질: 상 guess & slip = 0.05	모의실험 1-1	모의실험 2-1	모의실험 3-1
	문항의 질: 중 guess & slip = 0.1	모의실험 1-2	모의실험 2-2	모의실험 3-2
	문항의 질: 하 guess & slip = 0.2	모의실험 1-3	모의실험 2-3	모의실험 3-3

	Q-행렬의 구조	복잡한 Q-행렬		
단일모형 기반 검사	생성모형	DINA 모형	DINO 모형	ACDM
	문항의 질: 하 guess & slip = 0.2	모의실험 4-1	모의실험 4-2	모의실험 4-3

[표 3-5] 각 모의실험의 요인

2. Wald 통계량을 이용한 문항수준 모형 선택

우리는 모의실험 상에서 생성된 가상의 응답데이터 또는 실제 검사에 대한 피험자들의 실제 응답데이터와 Q-행렬을 바탕으로 각 문항에 Wald 검사를 적용할 수 있다. 검사에 대한 응답데이터를 바탕으로 Wald 검사를 적용하면 각 문항별로 축소모형에 대한 Wald 통계량(W)을 얻을 수 있다. 본 연구에서는 각 문항별로 축소모형에 대한 Wald 통계량(W)을 비교하여 가장 작은 통계량(W)을 나타내는 축소모형을 그 문항에 가장 적합한 모형으로 선택했다. Wald 통계량(W)을 이용해 문항수준에서 적합한 인지진단모형을 선택하는 방법은 다음과 같이 나타낼 수 있다.

$$T = \arg_p \min(W) \tag{3.1}$$

여기서 p는 Wald 검사를 적용할 때 관심을 두고 있는 축소모형을 의미한다. 예를 들어 본 연구에서 관심을 두고 있는 축소모형은 DINA 모형, DINO 모형, ACDM 으로, $p \in \{\mathrm{DINA}, \mathrm{DINO}, \mathrm{ACDM}\}$으로 나타낼 수 있다. 즉, 본 연구에서는 문항별로 DINA 모형, DINO 모형, ACDM의 Wald 통계량(W)을 계산하고 세 개의 Wald 통계량(W) 중 가장 낮은 값을 보이는 모형을 그 문항에 적합한 모형으로 선택했다.

예를 들어 [그림 3-1]과 같이 10개의 문항으로 이루어진 검사에 대한 응답데이터를 바탕으로 Wald 검사를 적용한다고 하자. Wald 검사를 적용하면, 각 문항별로 DINA 모형, DINO 모형, ACDM에 대한 Wald 통계량(W)이 계산된다. 6번 문항을 예시로 들면 DINA 모형에 대한 Wald 통계량은 5.178661, DINO 모형에 대한 Wald 통계량은 11.39908, ACDM에 대한 Wald 통계량은 2.287113이다. 이 중 최솟값($\min(W)$)은 2.287113으로 이에 해당하는 모형($\arg_p \min(W)$)은 ACDM이다. 그러므로 Wlad 통계량을 바탕으로 ACDM을 6번 문항에 가장 적합한 인지진단모형으로 선택할 수 있다.

문항별 Wald 통계량(W) 계산

	DINA	DINO	ACDM
I01	0.085829	2.866193	1.267411
I02	12.28342	62.35792	19.79781
I03	23.46198	271.3139	77.88144
I04	2.145211	2.868761	3.32949
I05	0.954151	18.06085	5.000633
I06	5.178661	11.39908	2.287113
I07	43.01832	43.63892	27.40284
I08	12.23852	15.25459	0.069279
I09	10.83238	10.03871	3.632199
I10	2.803931	4.787763	3.240855

최소값 찾기

$\min(W)$
0.085829
12.28342
23.46198
2.145211
0.954151
2.287113
27.40284
0.069279
3.632199
2.803931

최소값에 해당하는 모형 찾기

	$T = \arg_p \min(W)$
I01	DINA
I02	DINA
I03	DINA
I04	DINA
I05	DINA
I06	ACDM
I07	ACDM
I08	ACDM
I09	ACDM
I10	DINA

[그림 3-1] Wald 통계량을 이용한 모형 선택의 예

3. Wald 통계량을 이용한 문항수준 모형 선택 방법의 정확도 평가 방법

실제 데이터를 분석하는 경우에는 각 문항뿐만 아니라 검사지가 따르는 참 모형을 알 수 없다. 하지만 모의실험에서는 우리가 응답데이터를 생성할 때, 생성모형을 가정하기 때문에 각 문항이 따르는 참 모형을 알 수 있다. 그러므로 생성모형과 Wald 통계량을 이용해 선택된 모형간의 비교를 통해 Wald 통계량을 이용한 문항수준 모형 선택 방법의 정확도를 평가할 수 있다. 그래서 본 연구에서는 [표 3-6]과 같은 표를 통해 각 조건에서 모의실험의 결과, 즉, Wald 통계량을 이용한 문항수준 모형 선택의 비율을 비교하고 이를 바탕으로 Wald 통계량을 이용한 문항수준 모형 선택 방법의 정확도를 분석했다.

Wald / True	DINA	DINO	ACDM
DINA			
DINO		R_{ij}	
ACDM			

[표 3-6] Wald 통계량을 이용한 모형 선택 비율 예시

여기서 각 행은 응답 데이터를 생성할 때 가정한 생성모형, 즉, 문항이 따르는 참 모형을 의미하고, 각 열은 Wald 통계량을 이용해 선택된 모형을 의미한다. 그리고 각 셀에는 R_{ij}의 값이 들어가는데, R_{ij}는 다음과 같이 정의할 수 있다.

$$R_{\mathrm{ij}} = \frac{\arg_p \min(W) = \mathrm{j}\text{인 문항의 수}}{\text{true model이 i인 문항의 수}} \times 100 \qquad (3.2)$$

여기서 $i,j \in \{\mathrm{DINA,DINO,ACDM}\}$로 R_{ij} 참 모형이 i인 문항들 중에서 Wald 통계량에 의해 모형 j가 적합하다고 결정된 문항의 비율을 백분율로 나타낸 값이다. 특히 이 중에서 $i=j$인 경우, 즉, 참 모형과 Wald 통계량에 의해 결정된 모형이 일치하는 경우의 비율을 본 연구에서는 다음과 같이 모형 i가 생성모형인 경우 Wald 통계량을 이용한 문항수준 모형 선택의 True Positive Rate(TPR)이라고 정의했다.

$$TPR = \frac{\arg_p \min(W) = \text{i인 문항의 수}}{\text{true model} = \text{i인 문항의 수}} \times 100 \tag{3.3}$$

본 연구에서는 모의실험에서 문항에서 요구하는 인지요소의 개수, 복잡한 Q-행렬, 문항의 질, 혼합모형/단일모형 기반 검사 여부 등의 요인들이 변화함에 따라 생성모형별로 Wald 통계량을 이용한 모형 선택의 TPR 및 Wlad 통계량에 의한 모형 선택 비율이 어떻게 변하는지 조사했다. 그리고 각 요인들이 각 생성모형별 Wald 통계량을 이용한 모형 선택의 TPR 및 Wald 통계량을 이용한 모형 선택 비율에 영향을 미치는 이유를 분석하여, Wald 통계량을 이용한 모형 선택 방법의 정확도를 평가했다.

4. TIMSS 2007 응답데이터 분석

실제 데이터 연구 1에서는 TIMSS 2007 응답 데이터를 사용했으며 Lee, et al.(2011)의 연구에서 제시한 Q-행렬과 인지요소를 모두 참이라고 가정하고 Wald 통계량을 이용한 문항수준 모형 선택 방법을 수행했다. 그리고 Wald 통계량에 의해 적합하다고 결정된 인지진단모형과 실제 문항, 제시된 인지요소를 바탕으로 Wald 통계량에 의해 결정된 인지진단모형이 적합한지 그 의미를 분석했다. Lee, et al.(2011)은 TIMSS 2007 평가틀(Mullis et al., 2005)을 바탕으로 하여 인지요소를 작성했고, Q-행렬은

해당분야의 전문가들이 작성한 다섯 가지의 Q-행렬을 조합하여 최종 Q-행렬을 확정하여 연구를 진행했다. Lee, et al.(2011)의 연구에서 제시한 Q-행렬과 인지요소는 각각 [표 3-7]과 [표 3-8]과 같다.

Item	ATTR1	ATTR2	ATTR3	ATTR4	ATTR5	ATTR6	ATTR7	ATTR8	ATTR9	ATTR10	ATTR11	ATTR12	ATTR13	ATTR14	ATTR15
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0
4	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
5	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
7	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0
8	1	1	1	0	0	0	0	0	0	1	1	0	0	0	0
9	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
10	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
11	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0
12	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1
13	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0
14	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0
15	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
16	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
17	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
18	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0
19	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0
20	0	1	1	0	0	0	0	1	0	0	0	0	0	1	0
21	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
23	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
25	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1

[표 3-7] Lee, et al.(2011)의 Q-행렬

Number(N) 수	Whole Numbers 정수	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 분수와 소수	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 정수가 포함된 문장	7	Find the missing number or operation and model simple situations involving unknowns in number sentence or expressions.
	Patterns and Relationships 패턴과 관계	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 & Measurement (GM) 기하학적 도형과 측정	Lines and Angles 선과 각	9	Measure, estimate, and understand properties of lines and angles and be able to draw them.
	Two- and Three dimensional Shapes 2차원과 3차원 도형	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 위치와 이동	12	Locate points in an informal coordinate to recognize and draw figures and their movement.
Data & Display(DD) 자료와 표현	Reading and Interpreting 읽기와 해석	13	Read data from tables, pictograph, bar graphs, and pie charts.
		14	Comparing and understanding how to use information from data.
	Organizing and Representing 정리와 나타내기	15	Understanding different representations and organizing data using tables, pictograph, and bar graphs.

[표 3-8] Lee, et al.(2011)이 설정한 인지요소

5. 윤지영(2016)의 응답데이터 분석

실제 데이터 연구2에서는 윤지영(2016)의 연구에서 사용된 응답데이터와 인지요소, Q-행렬을 바탕을 Wald 통계량을 이용한 문항수준 모형 선택 방법을 수행했다. Wald 통계량에 의해 적합하다고 결정된 인지진단모형과 실제 문항, 제시된 인지요소를 바탕으로 Wald 통계량에 의해 결정된 인지진단모형이 적합한지 그 의미를 분석했다. 윤지영(2016)의 연구에서 사용된 인지요소와 Q-행렬은 각각 [표 3-9]와 [표 3-10]과 같다. 윤지영(2016)의 연구에서는 내용과 과정 중심으로 총 10개의 인지요소를 제시하였다. 그리고 Q-행렬은 각 문항이 요구하는 인지요소가 최소 1개에서 최대 6개로 다양하게 구성되어있다.

인지요소	내용
A1	사칙계산 및 식의 계산
A2	등식의 이해
A3	일차방정식의 해의 의미
A4	일차방정식 풀기(등식의 성질 포함)
A5	일차방정식 활용
A6	부등식의 이해
A7	일차부등식의 해의 의미
A8	일차부등식 풀기(부등식의 성질 포함)
A9	일차부등식 활용
A10	대수적 모델링

[표 3-9] 윤지영(2016)이 설정한 인지요소

Item	ATTR1	ATTR2	ATTR3	ATTR4	ATTR5	ATTR6	ATTR7	ATTR8	ATTR9	ATTR10
1	1	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	1	1	0	0	0	0	0	0	0	0
4	1	1	0	0	0	0	0	0	0	0
5	1	1	1	0	0	0	0	0	0	0
6	1	1	1	0	0	0	0	0	0	0
7	1	1	0	1	0	0	0	0	0	0
8	1	1	1	1	0	0	0	0	0	0
9	1	1	1	0	0	0	0	0	0	1
10	1	0	0	0	0	0	0	0	0	1
11	1	1	0	1	1	0	0	0	0	1
12	1	1	0	1	1	0	0	0	0	1
13	1	1	0	0	0	1	0	0	0	0
14	1	1	0	0	0	1	0	0	0	0
15	1	1	0	0	0	1	0	1	0	0
16	1	1	0	0	0	1	1	0	0	0
17	1	1	0	0	0	1	1	1	0	0
18	1	1	0	0	0	1	0	1	0	0
19	1	1	0	0	0	1	0	1	0	0
20	1	1	0	0	0	1	0	1	0	0
21	1	1	0	0	0	1	0	1	1	1
22	1	1	0	0	0	1	0	1	1	1
23	1	1	0	0	0	1	0	1	1	1
24	1	1	0	1	1	0	0	0	0	1
25	1	1	0	0	0	1	0	1	1	1

[표 3-10] 윤지영(2016)이 설정한 Q-행렬

제 4 장 연구결과

제 1 절 모의실험 결과

1. 문항에서 요구하는 인지요소의 개수와 모형 선택의 정확도

다음의 [표 4-1]은 문항의 질과 문항에서 요구하는 인지요소의 개수를 달리한 각 모의실험에서 실제 문항의 생성모형별로 Wald 통계량에 의해 문항수준에서 선택된 모형의 비율을 나타낸 표다.

문항 질	문항에서 요구하는 인지요소 2개				문항에서 요구하는 인지요소 3개			
상	Wald / True	DINA	DINO	ACDM	Wald / True	DINA	DINO	ACDM
	DINA	98.8	0.0	1.2	DINA	96.3	0.0	3.6
	DINO	0.0	98.9	1.1	DINO	0.0	95.9	4.0
	ACDM	20.3	19.9	59.8	ACDM	1.1	1.2	97.7
중	Wald / True	DINA	DINO	ACDM	Wald / True	DINA	DINO	ACDM
	DINA	99.4	0.0	0.6	DINA	94.1	0.0	5.9
	DINO	0.0	99.5	0.5	DINO	0.1	93.6	6.3
	ACDM	18.7	18.7	62.6	ACDM	0.8	0.8	98.4
하	Wald / True	DINA	DINO	ACDM	Wald / True	DINA	DINO	ACDM
	DINA	99.3	0	0.7	DINA	90.1	0	9.9
	DINO	0	99.5	0.5	DINO	0	88.8	11.1
	ACDM	15.4	15.4	69.3	ACDM	0.7	0.9	98.4

단위: %

[표 4-1] 모의실험 1과 2의 생성모형별 Wald 통계량을 이용한 모형 선택 비율

예를 들어, 문항에서 요구하는 인지요소가 2개이고 문항의 질이 '상'인 문항으로 구성된 모의실험 1-1의 결과를 보면, DINA 모형이 생성모형인 문항 중에서 문항의 생성모형과 Wald 통계량에 의해 선택된 모형이 일치한 비율이 98.8%였다. 그리고 DINO 모형이 생성모형인 문항 중에서 문항의 생성모형과 Wald 통계량에 의해 선택된 모형이 일치한 비율은 98.9%였다. 하지만, ACDM의 경우에는 달랐는데, ACDM이 생성모형인 문항 중 문항의 생성모형과 Wald 통계량에 의해 선택된 모형이 일치한 비율은 59.8%에 불과했고, 생성모형이 ACDM임에도 불구하고 ACDM이 생성모형인 문항 중에서 Wald 통계량에 의해 선택된 모형이 DINA 모형인 경우는 20.3%, Wald 통계량에 의해 선택된 모형이 DINO 모형인 경우는 19.9%로 나타났다.

문항에서 요구하는 인지요소의 개수가 2개에서 3개로 증가할 때, DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR의 변화와 ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR의 변화는 서로 다른 양상을 나타냈다. 문항의 질이 상일 때, 문항에서 요구하는 인지요소의 개수가 2개에서 3개로 증가하면 DINA 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 98.8%에서 96.3%로, DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 98.9%에서 95.9%로 하락했다. 이와는 다르게 문항의 질이 상일 때, 문항에서 요구하는 인지요소의 개수가 2개에서 3개로 증가하면 ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 59.8%에서 97.7%로 상승하는 것으로 나타났다. DINA 모형과 DINO 모형이 생성모형인 경우 문항에서 요구하는 인지요소의 개수가 2개에서 3개로 증가하자 DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR이 하락한 폭이 크지는 않지만, ACDM이 생성모형인 경우 문항에서 요구하는 인지요소의 개수가 2개에서 3개로 증가하자, ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 비교적 크게 상승했다.

이러한 경향은 문항의 질이 중인 경우와 하인 경우에서 마찬가지로 나타났다. 문항의 질이 중일 때, 문항에서 요구하는 인지요소의 개수가 2개에서 3개로 증가하면 DINA 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 99.4%에서 94.1%로 감소했고, DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 99.5%에서 93.6%로 하락했다. 반면, ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 62.6%에서 98.4%로 증가했다. 문항의 질이 하일 때는 문항에서 요구하는 인지요소의 개수가 2개에서 3개로 증가하면 DINA 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 99.3%에서 90.1%로 감소했고, DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 99.5%에서 88.8%로 감소했다. 반면, ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 69.3%에서 98.4%로 증가했다.

문항에서 인지요소의 개수가 2개에서 3개로 증가할 때, DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR이 낮아지고, ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR이 높아지는 현상이 나타게 된 이유는 각 인지진단모형의 모수의 개수 차이에 있다고 추측할 수 있다. DINA 모형과 DINO 모형의 모수는 문항의 질, 문항에서 요구하는 인지요소의 개수와 상관없이 문항별로 두 개의 모수, 즉, guess와 slip 모수를 갖는다. 이를 G-DINA 모형의 문항반응함수의 모수 중에서 생각하면 DINA 모형은 intercept 값 δ_{j0}와 상호작용효과 모수 $\delta_{j12\cdots K_j^*}$를 가지고, DINO 모형은 intercept 값 δ_{j0}와 인지요소의 주효과 모수 δ_{j1} 하나만을 가진다. 하지만 ACDM의 경우는 문항이 요구하는 인지요소가 K_j^*개일 때, K_j^*+1개의 모수를 가지므로, 문항이 요구하는 인지요소의 개수가 증가하면 인지진단모형이 가지는 모수의 개수가 커진다.

앞선 언급했듯이 Wald 검사의 기본적인 개념은 사고 절약의 원리로, 인지요소와 Q-행렬이 참이라고 가정할 때 여러 인지진단모형이 데이터

에 동등하게 들어맞는다면 그중에서 간단한 모형을 적합한 인지진단모형으로 선택하는 방법이다(Sorrel et al., 2017a). 인지진단모형이 가지는 모수의 개수가 증가하면, 모형의 적합도도 증가하지만 모형이 더 복잡해진다. 문항에서 요구하는 인지요소의 개수가 2개인 경우에 DINA 모형과 DINO 모형은 2개의 모수를 가지고, ACDM은 intercept 값 δ_{j0}와 인지요소의 주효과 모수 δ_{j1}, δ_{j2} 2개를 포함해 총 3개의 모수를 갖는다. 즉, DINA 모형, DINO 모형과 ACDM이 가지는 모수의 개수 차이가 크지 않기 때문에 적합도의 차이도 크지 않고, 이로 인해 간단한 DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR이 매우 높게 나타난다고 추측할 수 있다.

반면 문항에서 요구하는 인지요소의 개수가 3개인 경우에 ACDM은 intercept 값 δ_{j0}과 인지요소의 주효과 모수 $\delta_{j1}, \delta_{j2}, \delta_{j3}$를 포함해 총 4개의 모수를 갖는다. 문항에서 요구하는 인지요소의 개수가 2개인 경우에는 DINA 모형과 DINO 모형이 가지는 모수의 개수와 ACDM이 가지는 모수의 개수가 1개였지만, 문항에서 요구하는 인지요소의 개수가 3개인 경우에는 ACDM은 DINA 모형과 DINO 모형보다 2개 더 많은 모수를 갖는다. 그렇기 때문에 DINA 모형과 DINO 모형이 ACDM에 비해 간단한 모형이라도 ACDM의 적합도가 월등히 좋다면, DINA 모형과 DINO 모형이 생성모형이더라도 Wald 통계량에 의해 ACDM이 적합한 모형으로 선택될 수 있다. 그래서 문항에서 요구하는 인지요소의 개수가 2개에서 3개로 증가하면 생성모형이 DINA 모형과 DINO 모형인 문항의 경우 Wald 통계량에 의해 ACDM이 적합한 모형이라고 결정되는 비율이 증가해 DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR이 낮아진다고 추측할 수 있다.

다음으로, ACDM이 생성모형인 경우 문항에서 요구하는 인지요소의 개수가 2개에서 3개로 증가하면 Wald 통계량을 이용한 모형 선택의 TPR이 증가하는 이유는 다음과 같이 추측할 수 있다. ACDM(Additive CDM)의 이름과 아래의 문항반응함수에서 알 수 있듯이 ACDM은 인지요소의

주효과(δ_{j1}, $\delta_{j2}\cdots$, $\delta_{jK_j^*}$)와 intercept 값(δ_{j0})의 합으로 피험자가 문항에 옳게 답할 확률이 예측되므로, ACDM의 문항반응함수를 이루는 모수들이 가산적인 관계를 가지고 있다. 이는 ACDM의 문항반응함수를 봐도 알 수 있다.

$$P(\alpha_{lj}^{*}) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{lk} \tag{4.1}$$

그런데 문항에서 요구하는 인지요소의 개수가 2개인 경우에는 ACDM이 가지는 모수가 3개뿐이므로, ACDM 문항반응함수를 이루는 모수들의 가산적인 관계가 두드러지지 않는다. 그렇기 때문에 ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR이 낮게 나타나고, 생성모형이 ACDM인 문항을 간단한 DINA 모형 또는 DINO 모형이 적합하다고 잘못 판단하는 비율이 나타난다고 추측할 수 있다. 실제 실험결과를 보면, 문항에서 요구하는 인지요소의 개수가 2개인 경우, 문항의 질이 상일 때, ACDM을 생성모형으로 하는 문항이 Wald 통계량에 의해 DINA 모형이 적합하다고 결정되는 비율이 20.3%, DINO 모형이 적합하다고 결정되는 비율이 19.9%로 나타났다. 다음으로, 문항의 질이 중일 때, ACDM을 생성모형으로 하는 문항이 Wald 통계량에 의해 DINA 모형 또는 DINO 모형이 적합하다고 결정되는 비율 각각 18.7%씩으로 나타났고, 문항의 질이 하일 때는 ACDM을 생성모형으로 하는 문항이 Wald 통계량에 의해 DINA 모형 또는 DINO 모형이 적합하다고 결정되는 비율이 각각 15.4%씩으로 나타났다. 즉, 문항의 질에 상관없이 생성모형이 ACDM임에도 불구하고 Wald 통계량에 의해 DINA 모형 또는 DINO 모형이 적합하다고 결정된 비율은 모두 15% 이상으로 나타났다.

이제 문항에서 요구하는 인지요소의 개수가 3개인 경우를 보면 ACDM은 guess 모수 δ_{j0}, 인지요소의 주효과 모수 $\delta_{j1},\delta_{j2},\delta_{j3}$ 총 4개의 모수를 갖는다. 이 경우는 DINA 모형, DINO모형과 비교했을 때 앞서 ACDM의

모수가 3개인 경우보다 ACDM의 문항반응함수가 가지는 모수들의 가산적인 관계가 더 드러나게 된다. 그렇기 때문에 DINA 모형과 DINO 모형이 ACDM에 비해 간단하더라도, 문항에서 요구하는 인지요소의 개수가 증가하면 ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR이 증가한다고 추측할 수 있다.

이러한 현상은 복잡한 Q-행렬 디자인을 이용한 모의실험 3-3의 결과를 문항에서 요구하는 인지요소의 개수별로 분석했을 때에도 관찰할 수 있었다. 모의실험 3-3은 TIMSS 2007(Martin et al., 2008)을 연구한 Lee et al.(2011)의 Q-행렬을 사용했는데, Lee et al.(2011)은 검사지 내 각 문항이 요구하는 인지요소의 개수를 최소 1개에서 6개로 다양하게 가정했다. 모의실험 3-3의 결과로부터 문항에서 요구하는 인지요소의 개수 별로 문항을 구분하여 Wald 통계량을 이용한 모형 선택의 정확도를 구하여 [표 4-2]에 나타내었다.

이 경우에도 앞서 살펴본 모의실험 1, 2의 결과와 마찬가지로 문항에서 요구하는 인지요소의 개수가 증가할수록 ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR이 증가하고, DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 하락하고 있음을 알 수 있다. 특히 문항에서 요구하는 인지요소가 5개 이상인 경우에는 문항에 대한 생성모형이 DINA 모형 또는 DINO 모형이더라도 대부분 ACDM이 문항에 적합하다고 결정되었다.

인지요소 개수	Wald / True	DINA	DINO	ACDM
6	DINA	1.6	0.6	97.8
	DINO	0.4	2.2	97.4
	ACDM	0.0	0.2	99.8

문항 수: 1개

인지요소 개수	Wald / True	DINA	DINO	ACDM
3	DINA	48.6	1.3	50.1
	DINO	0.9	45.1	54.0
	ACDM	2.3	2.0	95.7

문항 수: 8개

인지요소 개수	Wald / True	DINA	DINO	ACDM
5	DINA	6.6	1.2	92.2
	DINO	0.8	5.6	93.6
	ACDM	0.4	0.2	99.4

문항 수: 1개

인지요소 개수	Wald / True	DINA	DINO	ACDM
2	DINA	71.4	4.1	24.6
	DINO	6.5	66.0	27.5
	ACDM	6.8	9.1	84.2

문항 수: 8개

인지요소 개수	Wald / True	DINA	DINO	ACDM
4	DINA	19.9	1.3	78.9
	DINO	0.9	18.3	80.9
	ACDM	0.5	0.6	99.0

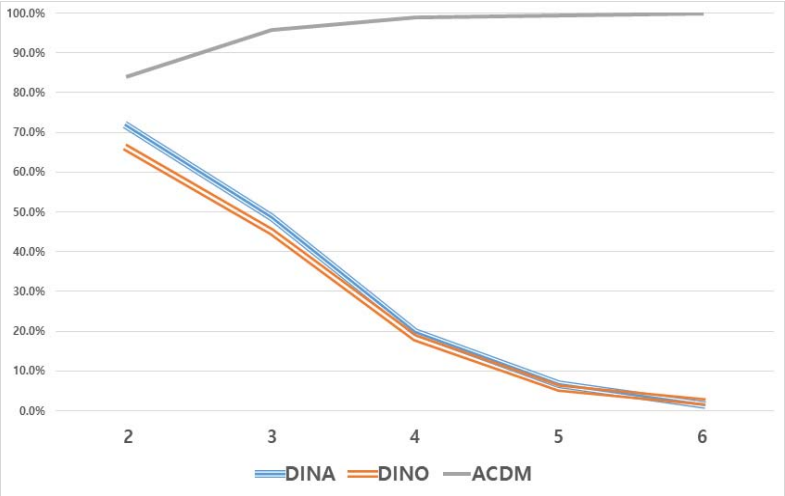
문항 수: 4개

단위: %

*문항의 질: 하

[표 4-2] 모의실험 3-3에서 문항에서 요구하는 인지요소 개수에 따른 생성모형별 Wald 통계량을 이용한 모형 선택 비율

위의 [표 4-2]의 결과를 문항에서 요구하는 인지요소 개수별 각 인지진단모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR을 그래프로 나타내면 [그림 4-1]과 같다. 그래프에서도 알 수 있듯이 문항에서 요구하는 인지요소의 개수가 증가할수록 DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 급격히 하락하고, ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 100%에 계속 가까워지는 것을 알 수 있다.



[그림 4-1] 문항에서 요구하는 인지요소 개수에 대해 생성모형별 Wald 통계량을 이용한 모형 선택의 TPR

위의 결과를 종합하면, DINA 모형, DINO 모형, ACDM이 가지는 모수의 개수의 차이와 ACDM의 문항반응함수에서 모수들의 가산적인 관계로 인해 문항에서 요구하는 인지요소의 개수가 모형 선택의 정확도에 영향을 준다고 추측할 수 있다. 그래서 문항에서 요구하는 인지요소의 개수가 증가하면, DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 낮아지고, ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 증가하는 것으로 나타났다.

2. 문항의 질과 모형 선택의 정확도

다음의 [표 4-3]은 Q-행렬이 비교적 간단한 경우에 문항의 질과 문항에서 요구하는 인지요소의 개수를 달리한 각 모의실험에서 실제 문항의 생성모형별로 Wald 통계량을 이용한 모형 선택의 TPR을 나타낸 표다.

문항의 질	문항에서 요구하는 인지요소 2개			문항에서 요구하는 인지요소 3개		
	DINA	DINO	ACDM	DINA	DINO	ACDM
상	98.8	98.9	59.8	96.3	95.9	97.7
중	99.4	99.5	62.6	94.1	93.6	98.4
하	99.3	99.5	69.3	90.1	88.8	98.4

단위: %

[표 4-3] 문항의 질과 문항에서 요구하는 인지요소 개수별 Wald 통계량을 이용한 모형 선택의 TPR

위의 결과에서 가장 두드러지는 특징은 문항의 질이 증가하면 ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 감소한다는 점이다. 문항에서 요구하는 인지요소의 개수가 3개일 때는 ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR이 거의 변화가 없다고 할 수 있지만, 문항에서 요구하는 인지요소의 개수가 2개일 때는 그 변화의 폭이 비교적 크다.

위 결과에서 알 수 있는 다른 특징은 문항에서 요구하는 인지요소가 3개인 경우에는 문항의 질이 증가하면 DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 증가한다는 점이다. 문항에서 요구하는 인지요소가 2개인 경우에는 DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR 변화는 거의 없다고 볼 수 있다. 문항에서 요구하는 인지요소의 개수가 모형 선택의 정확도에 영향을 미치는 원인이 모수에 있다고 추측했듯이, 문항의 질에 따라 모형 선택의 정확도가 달라지는 이유도 인지진단모형이 가지는 모수의 구조에 있다고 추측할 수 있다.

우선, ACDM의 문항반응함수의 식은 다음과 같다.

$$P(\alpha_{lj}^{*}) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{lk} \tag{4.2}$$

ACDM은 각 인지요소의 주효과들만 존재하고 인지요소들의 상호작용 효과는 모두 0으로 가정하고 있기 때문에 인지요소 개수만큼의 모수 δ_{jk} $(k=1,2,\cdots,K_j^*)$를 갖는다. 그리고 intercept 값에 해당하는 δ_{j0}까지 포함해서 ACDM은 문항 j에 대해 총 K_j^*+1개의 모수를 갖는다. 그리고 DINA 모형과 DINO 모형은 문항에서 요구하는 인지요소 개수와 관계없이 guess 모수(g_j)와 slip 모수(s_j), 단 2개의 모수를 갖는다.

이제 DINO 모형과 DINA 모형이 가지는 guess 모수와 slip 모수, 그리고 ACDM이 갖는 K_j^*+1개의 모수들의 관계를 살펴보면 다음과 같은 관계가 성립한다. (g_j: 문항 j의 guess 모수, s_j: 문항 j의 slip 모수)

$$\begin{gathered} \delta_{j0} = g_j \\ \delta_{j0} + \delta_{j1} + \delta_{j2} + \cdots + \delta_{jK_j^*} = 1 - s_j \end{gathered} \tag{4.3}$$

그런데 문항의 질이 증가하면, 즉, guess 모수와 slip 모수의 값이 낮아지면 intercept 값에 해당하는 δ_{j0}의 크기도 같이 낮아진다. 그러면 ACDM이 가지는 모수들 중에서 δ_{j0}의 영향은 미미해진다. 즉, ACDM의 모수 중 δ_{j0}의 비중이 많이 감소하기 때문에, ACDM은 K_j^*+1개의 모수 중에서 δ_{j0}을 제외한 K_j^*개의 모수만 가지고 있는 것처럼 변하게 된다.

구체적으로 모의실험 결과 중 문항에서 요구하는 인지요소가 2개인 경우를 보면, 문항의 질이 하에서 상으로 증가함에 따라 ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 69.3%에서 59.8%로 낮아진다. 문항에서 요구하는 인지요소가 2개인 경우에 ACDM의 모수는 총 3개다. 그런데 문항의 질이 높아지면, δ_{j0}의 값이 0.2에서 0.05로 낮아지므로 δ_{j0}의 영향이 많이 사라지게 된다. 그러면 ACDM이 가지는 모수는 3개지만 2개의 모수를 가지는 것처럼 여겨지게 된다. 그로 인해

데이터에 대한 ACDM의 적합도와 DINA 모형과 DINO 모형의 적합도의 차이가 미미해지기 때문에 ACDM을 생성모형으로 하는 문항이 ACDM보다 간단한 DINA 모형 또는 DINO 모형이 적합하다고 잘못 결정되는 경우가 발생한다고 추측할 수 있다.

이제 모의실험 결과 중 문항에서 요구하는 인지요소가 3개인 경우를 보자. 이 경우 문항의 질이 하에서 상으로 증가함에 따라 ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 98.4%에서 97.7%로 낮아진다. 하지만 이는 문항에서 요구하는 인지요소가 2개인 경우에 문항의 질이 높아짐에 따라 ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR 변화와 비교했을 때 변화의 폭이 작다. 이러한 차이가 발생하는 이유는 ACDM이 가지는 모수의 개수에 있다고 추측할 수 있다. 앞서 살펴봤듯이 문항의 질이 하에서 상으로 증가함에 따라 δ_{j0}의 영향이 작아진다. ACDM은 δ_{j0}, δ_{j1}, δ_{j2}, δ_{j3}, 4개의 모수를 가지고 있지만 δ_{j0}의 영향이 작아짐에 따라 3개의 모수를 가지는 것처럼 여겨질 수 있다. 그러면 DINA 모형과 DINO 모형이 가지는 모수의 개수와 차이가 줄어들게 되지만 여전히 ACDM은 DINA 모형과 DINO 모형보다는 많은 모수를 가지게 된다. 그래서 문항에서 요구하는 인지요소가 3개인 경우에는 문항의 질이 높아져 δ_{j0}의 영향이 미미해지더라도 ACDM이 DINA 모형과 DINO 모형보다 더 많은 모수를 가지는 데에서 오는 영향이 더 크기 때문에, ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR이 매우 높고 크게 변하지 않는다고 추측할 수 있다.

다음으로 DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR 변화를 살펴보자. 우선 문항에서 요구하는 인지요소가 2개인 경우에는 문항의 질이 높아짐에 따라 DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 거의 변화가 없다. 반면 문항이 요구하는 인지요소가 3개인 경우에는 문항의 질이 높아질수록 DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR이 증가하는 것을 확인할 수 있

다. 이러한 현상도 역시 문항의 질이 높아질수록 δ_{j0}가 작아져 그로인해 나타나는 현상 중 하나로 추측할 수 있다.

문항에서 요구하는 인지요소가 3개인 경우에는 DINA 모형과 DINO 모형이 가지는 모수는 2개고, ACDM이 가지는 모수는 4개다. 그렇기 때문에 문항의 질이 높아져 δ_{j0}의 값이 작아지면 δ_{j0}의 모수의 영향력이 거의 사라진다고 여길 수 있다. 그러면 앞서 언급했듯이, DINA 모형과 DINO 모형이 더 간단하더라도 ACDM이 가지는 모수가 많아 ACDM이 더 나은 적합도를 보이기 때문에 ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 항상 높게 유지된다고 추측할 수 있었다.

하지만, DINA 모형 또는 DINO 모형을 생성모형으로 하는 문항들의 입장에서는 다르게 해석할 수 있다. 문항의 질이 높아지면 DINA 모형과 DINO 모형이 가지는 모수의 개수와 ACDM이 가지는 모수의 개수의 차이가 작아진다. 그렇기 때문에 데이터에 대한 DINA 모형과 DINO 모형의 적합도와 ACDM의 적합도 차이가 작아지게 된다. 그러면 각각 DINA 모형과 DINO 모형이 생성모형인 문항에 대해 Wald 통계량에 의한 모형 선택 방법을 적용하면, ACDM 보다는 비교적 간단한 DINA 모형과 DINO 모형이 적합한 모형으로 각각 결정된다고 추측할 수 있다.

다음으로 Q-행렬이 비교적 복잡한 경우에는 문항의 질이 변함에 따라 모형 선택의 정확도가 어떻게 변화하는지 알아보자. 다음의 [표 4-4]는 Lee et al.(2011)의 Q-행렬을 사용했을 때, 문항의 질을 달리한 각 모의실험에서 문항의 생성모형별로 Wald 통계량을 이용한 모형 선택의 정확도를 나타낸 표다.

문항의 질	각 생성모형별 Wald 통계량을 이용한 모형 선택의 정확도			
상		**DINA**	**DINO**	**ACDM**
	DINA	72.1	0.6	27.3
	DINO	1.0	70.5	28.5
	ACDM	3.0	2.7	94.3
중		**DINA**	**DINO**	**ACDM**
	DINA	66.3	1.1	32.6
	DINO	1.2	65.1	33.7
	ACDM	2.6	2.5	94.9
하		**DINA**	**DINO**	**ACDM**
	DINA	47.6	2.3	50.1
	DINO	2.9	44.1	53
	ACDM	3.4	4.2	92.4

단위: %

[표 4-4] Lee et al.(2011)의 Q-행렬을 사용한
모의실험에서 문항의 질에 따른
Wald 통계량을 이용한 모형 선택의 정확도

위의 표에서 두드러지는 특징은 문항의 질이 하에서 상으로 증가함에 따라 각 생성모형별 Wald 통계량을 이용한 모형 선택의 TPR이 모두 증가한다는 점이다. DINA 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 문항의 질이 낮을 때는 47.6%, 중간일 때는 66.3%, 높을 때는 72.1%로 증가했고, DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 문항의 질이 낮을 때는 44.1%, 중간일 때는 65.1%, 높을 때는 70.5%로 증가했다. 그리고 ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 문항의 질이 낮을 때는 92.4%, 중간일 때는 94.9%, 높을 때는 94.3%로 나타났다. ACDM의 경우

에는 문항의 질이 높을 때 중간일 때에 비해 Wald 통계량을 이용한 모형 선택의 TPR이 0.6% 하락했지만, 문항의 질이 낮을 때 보다는 증가했기 때문에 Q-행렬이 간단할 때 문항의 질이 증가함에 따라 ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR이 감소하던 앞선 모의실험의 결과와는 상반된다.

앞서 살펴봤듯이, Lee et al.(2011)의 Q-행렬은 각 문항에서 요구하는 인지요소의 개수를 다양하게 가정하고 있다. 인지요소 1개를 요구하는 세 문항은 Wald 검사 수행 시 Wald 통계량 계산에서 제외된다. 이 세 문항을 제외하면, 이 Q-행렬에서는 인지요소 2개와 3개를 요구하는 문항은 각각 8문항씩, 인지요소 4개를 요구하는 문항은 4문항, 인지요소 5개와 6개를 요구하는 문항은 각각 1문항씩으로 가정하고 있다. 이제 문항에서 요구하는 인지요소 개수별로 문항의 질에 따라 Wald 통계량을 이용한 모형 선택의 TPR을 비교해보면 [표 4-5]와 같다.

인지요소의 개수: 6개			
문항의 질	DINA	DINO	ACDM
상	17.6	17.1	100.0
중	11.8	8.8	100.0
하	1.6	2.2	99.8

문항 수: 1개

인지요소의 개수: 5개			
문항의 질	DINA	DINO	ACDM
상	38.0	29.9	100.0
중	25.7	22.8	100.0
하	6.6	5.6	99.4

문항 수: 1개

인지요소의 개수: 4개			
문항의 질	DINA	DINO	ACDM
상	58.2	61.5	99.9
중	46.9	48.2	100.0
하	19.9	18.3	99.0

문항 수: 4개

인지요소의 개수: 3개			
문항의 질	DINA	DINO	ACDM
상	81.9	78.3	98.1
중	76.1	74.3	98.1
하	48.6	45.1	95.7

문항 수: 8개

인지요소의 개수: 2개			
문항의 질	DINA	DINO	ACDM
상	80.5	78.9	86.3
중	78.2	76.7	87.8
하	71.4	66	84.2

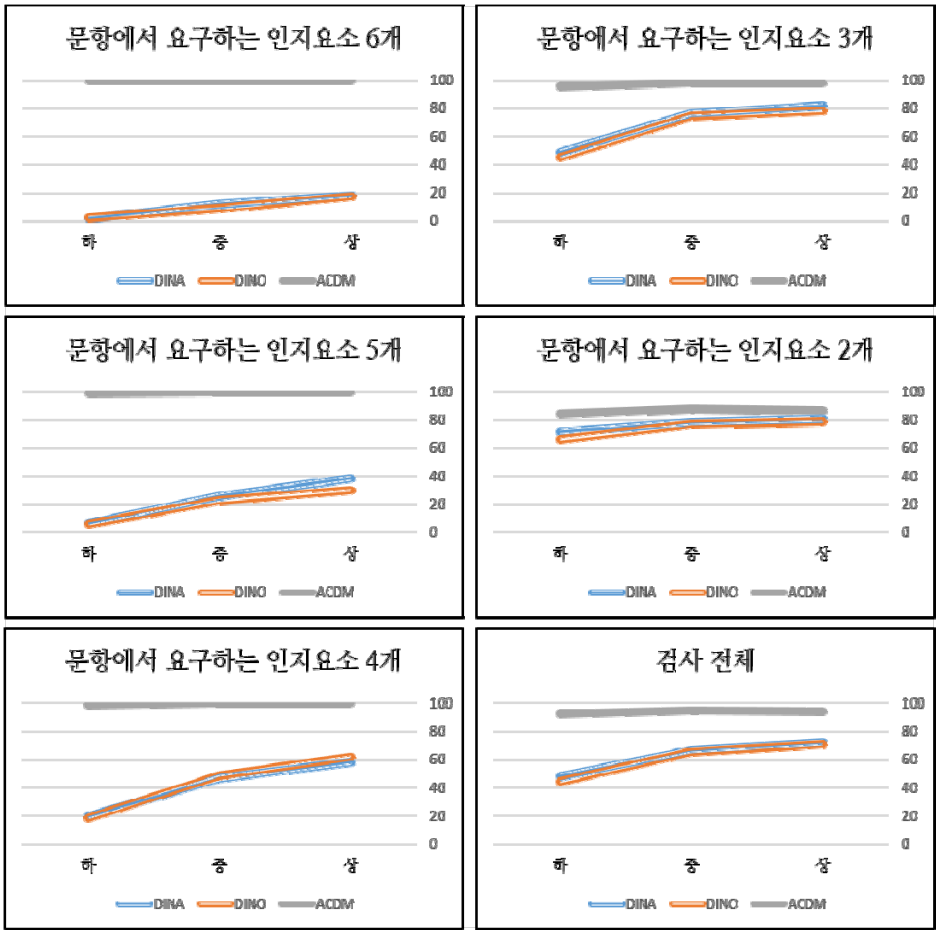
문항 수: 8개

단위: %

[표 4-5] 모의실험 3에서 문항에서 요구하는 인지요소 개수에 따른 생성모형별 Wald 통계량을 이용한 모형 선택의 TPR

DINA 모형과 DINO 모형을 보면 두 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 문항의 질이 증가할수록 같이 커지는 것을 확인할 수 있다. 이는 앞서 살펴본 바와 같이 문항의 질이 증가하면 각 문항이 갖는 intercept 값 δ_{j0}이 작아져 ACDM이 가지는 모수와 DINA 모형과 DINO 모형이 가지는 모수의 차이가 작아져 생기는 변화라고 추측할 수 있다. 즉, 각각 DINA 모형 또는 DINO 모형이 생성모형인 문항에 대해 Wald 통계량을 이용한 모형 선택 방법을 적용하면, ACDM이 가지는 모수의 개수가 더 많아 더 나은 적합도를 보이더라도 DINA 모형과 DINO 모형이 비교적 다 간단하기에 Wald 통계량에 의해 각각 DINA 모형 또는 DINO 모형이 적합한 모형으로 올바르게 결정되는 것이다.

위에서 살펴본 바와 같이 문항의 질이 높아지면, 문항에서 요구하는 인지요소의 개수에 상관없이 DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 눈에 띄게 높아지는 것을 확인할 수 있다. 그로 인해 [표 4-4]에서처럼, 검사 전반에 있어 DINA 모형과 DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR이 높아지는 것이다. 이를 그래프로 표현하면 [그림 4-2]와 같다.



[그림 4-2] 문항의 질에 따른 Wald 통계량을 이용한 모형 선택의 TPR 변화 - 문항에서 요구하는 인지요소 개수별 변화, 검사 전체 변화

3. 단일모형 기반 검사의 모형 선택 정확도

다음의 [표 4-6]은 문항의 질이 낮고, Lee et al.(2011)의 Q-행렬을 사용하여 복합모형 기반 검사와 단일모형 기반 검사를 가정했을 때, 모의실험에서 실제 문항의 생성모형별로 Wald 통계량을 이용한 모형 선택의 비율을 나타낸 표다.

문항의 질: 하 Lee et al.(2011)의 Q-행렬(복잡한 Q-행렬) 사용							
복합모형 기반 검사				단일모형 기반 검사			
Wald / True	DINA	DINO	ACDM	Wald / True	DINA	DINO	ACDM
DINA	47.6	2.3	50.1	**DINA**	84	0.6	15.4
DINO	2.9	44.1	53	**DINO**	0.6	84.1	15.3
ACDM	3.4	4.2	92.4	**ACDM**	0.7	0.9	98.4

단위: %

[표 4-6] 복합모형/단일모형 기반 검사에 대해 생성모형별 Wald 통계량을 이용한 모형 선택 비율

우선 위의 결과에서 Wald 통계량을 이용해 모형 선택을 했을 때, 생성모형을 옳게 찾아내는 비율, 즉, 각 생성모형에 대한 Wald 통계량을 이용한 모형 선택의 TPR을 비교하면 단일모형 기반 검사가 복합모형 기반 검사보다 월등히 높은 것을 확인할 수 있다. DINA 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR을 비교하면, 복합모형 기반 검사에서는 47.6%로 50%에 못 미쳤지만, 단일모형 기반 검사에서는 84%로 많이 상승했다. DINO 모형이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR을 비교하면, 복합모형 기반 검사에서는 44.1%로 세 모형 중 가장 작은 값을 나타냈지만, 단일모형 기반 검사에서는 84.1%로 상당히 많이 상승했다. 마지막으로 ACDM이 생성모형인 경우에 Wald 통계량을 이용한 모형 선택의 TPR을 비교하면, 복합모형 기반 검사에서는 92.4%로 꽤 높은 일치율을 보였고, 단일모형 기반 검사에서는 98.4%로 소폭 상승한 것으로 나타났다.

위의 검사에서 복합모형 기반 검사는 75문항으로 이루어진 전체 검사지의 문항을 3등분하여 각기 DINA 모형, DINO 모형, ACDM을 생성모형으로 하여 응답데이터를 생성했다. 그리고 이 응답데이터를 이용하여 모수들을 추정하고 Wald 통계량을 이용한 모형 선택 방법을 적용하여 각

문항에 적합한 인지진단모형을 결정했다. 반면, 단일모형 기반 검사는 25문항으로 이루어진 검사를 각각 DINA 모형, DINO 모형, ACDM을 생성모형으로하여 응답데이터를 생성하고, 이를 이용하여 모수들을 추정하고 Wald 통계량을 이용한 모형 선택 방법을 적용했다. 여기서 복합모형 기반 검사와 단일모형 기반 검사의 생성모형별 Wald 통계량을 이용한 모형 선택의 TPR이 차이가 나는 이유는 모수 추정과정에 있다고 추측할 수 있다. 복합모형 기반 검사는 G-DINA의 모수 추정 과정에서 다른 문항에 대한 응답이 서로 영향을 미칠 수 있다. 예를 들어, 6번 문항에 대한 모수를 추정한다고 하자. 복합모형 기반 검사는 25개의 문항으로 이루어진 검사를 생성모형을 달리하여 세 번 반복했다. 반면, 단일모형 기반 검사는 25개의 문항으로 이루어진 검사를 반복하지 않고 그대로 사용했다. 그러면 복합모형 기반 검사에서는 6번, 31번, 56번 문항이 똑같은 모수를 가지게 되고, 이 과정에서 6번, 31번, 56번 문항에 대한 응답이 서로에게 영향을 미치게 된다. 즉, 복합모형 기반 검사에서는 단일모형 기반 검사와 달리 G-DINA의 모수 추정 과정에서 다른 문항에 대한 응답이 서로 영향을 미쳐 Wald 통계량을 이용한 모형 선택의 TPR이 단일모형 기반 검사의 결과와 다르다고 추측할 수 있다.

제 2 절 실제 데이터 연구 결과

1. TIMSS 2007 응답데이터 분석 결과

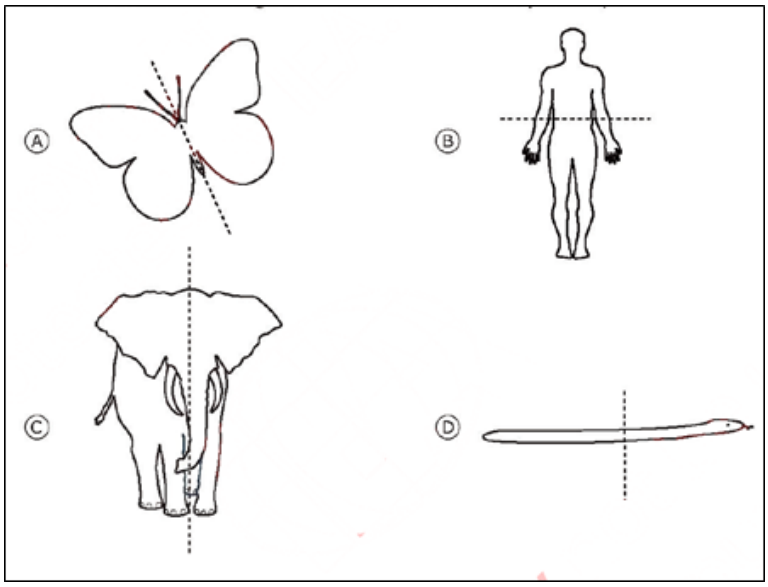
이제 Wald 통계량을 이용한 모형 선택 방법을 실제 데이터에 적용한 결과를 살펴보자. 우선, Wald 통계량을 이용한 문항수준 모형 선택 방법을 TIMSS 2007 응답데이터에 적용하면 [표 4-7]과 같은 결과가 나타난다. 결과를 보면 21문제가 ACDM이 적합하다고 나타났고, 오직 한 문제만 DINO 모형에 적합하다고 나타났다(요구하는 인지요소가 1개뿐인 2,9,24번 문항은 Wald 통계량 계산에서 제외된다).

응답모형	문항 수
DINA	0
DINO	1
ACDM	21

[표 4-7] TIMSS 2007 응답데이터 분석 결과

적합한 인지진단모형이 DINO 모형이라고 결정된 문항은 TIMSS 2007 4번 검사지의 6번 문항(Item ID: M041164)으로 Lee, et al.(2011)의 Q-행렬과 인지요소를 통해 보면 인지요소 10과 인지요소 12를 요구하는 문항이다. 이 문항은 점선을 기준으로 대칭인 도형을 찾는 문제다.

Q. In which of these drawings is the dotted line a line of symmetry?



[그림 4-3] TIMSS 2007 Item ID: M041164

인지요소 10	Classify, compare, and recognize geometric figures and shapes and their relationships and elementary properties. *기하학적 그림과 도형, 그리고 그 관계와 기본적인 성질들을 분류, 비교, 인식함.*
인지요소 12	Locate points in an informal coordinate to recognize and draw figures and their movement. *그림과 그 움직임을 인식하고 그릴 수 있도록 비공식 좌표에서 점들을 찾음.*

[표 4-8] Item ID: M041164가 요구하는 인지요소(Lee, et al., 2011)

이 문항에서 요구하는 인지요소는 위의 [표 4-6]과 같다. 문항에서 요구하는 인지요소들의 내용을 바탕으로 결과를 분석해보면, 도형의 대칭에 대한 성질을 잘 알고 있거나(인지요소 10 숙달), 점선을 기준을 도형을 접어서 포개어지는 사고를 할 수 있다면(인지요소 12 숙달) 위의 문항을 해결할 수 있을 것이다. 즉, 인지요소 10이나 인지요소 12 중 하나만 숙달하고 있어도 위 문항을 해결할 수 있으므로 위 문항에 적합한 인지진단모형이 DINO 모형으로 결정된 것은 적합하다고 해석 할 수 있다.

다만 실제 응답데이터를 바탕으로 각 문항의 모수를 추정한 결과, 가능한 인지요소 프로파일별 6번 문항에 대한 정답률을 조사하면 [표 4-9]와 같다.

	P(00)	**P(10)**	**P(01)**	**P(11)**
Item 6	0.5816	0.9999	0.9999	0.9808

[표 4-9] 인지요소 프로파일별 TIMSS 2007 4번 검사지 6번 문항의 정답률

위의 결과를 보면 $P(10)$, $P(01)$, $P(11)$의 값이 거의 비슷하게 나타난 것을 확인할 수 있다. 즉, 인지요소 10 하나만 숙달한 경우, 인지요소 12 하나만 숙달한 경우, 인지요소 10과 12 모두를 숙달한 경우에 문항을 해결할 확률이 거의 비슷하다. 다만, 여기서 눈여겨 볼 점은 $P(00)$, 인지요소 10과 인지요소 12 모두를 숙달하지 못한 경우의 문항을 해결할 확률이다. 이때의 정답률은 0.5816으로, 6번 문항의 intercept 값 δ_{60}이 0.5816으로 상당히 높다는 것을 의미한다. 우리가 모의실험 시 문항의 질이 낮은 문제는 guess 모수를 0.2로 설정했는데, 이 기준을 통해 보더라도 6번 문항은 문항의 질이 매우 좋지 않은 문항으로 분류 할 수 있다. 또는 Q-행렬 구성이나 인지요소 선정에 오류가 있다고 추측할 수 있다.

ACDM이 적합하다고 결정된 문항들에 대해 각 문항의 모수를 추정한 결과 가능한 인지요소 프로파일별 각 문항에 대한 정답률을 확인하면, 대부분의 문항들이 인지요소를 숙달할수록 문항의 정답률이 선형적으로 증가하는 것을 확인할 수 있다. 다만, 일부 문항의 각 인지요소 프로파일별 정답률을 확인해보면 몇몇 문항은 문항의 정답률이 선형적으로 증가하지 않는 것으로 나타났다. 예를 들어 문항 1과 문항 21의 인지요소 프로파일별 정답률은 다음과 같다.

	P(00)	P(10)	P(01)	P(11)				
Item 1	0.0001	0.9999	0.8051	0.9999				
	P(000)	P(100)	P(010)	P(001)	P(110)	P(101)	P(011)	P(111)
Item 21	0.0001	0.1631	0.0001	0.0001	0.3218	0.0165	0.9999	0.4023

[표 4-10] 인지요소 프로파일별 TIMSS 2007 4번 검사지 1, 21번 문항의 정답률

우선 1번 문항을 살펴보면 1번 문항에서 요구하는 두 개의 인지요소 중 하나만 숙달하더라도 각각 0.9999, 0.8051로 높은 정답률을 보인다. 모의실험에서 문항의 질이 낮은 문항을 $s_j = 0.2$로 설정한 것을 고려하여 $P(01) = 0.8051$의 확률은 문항을 충분히 해결할 수 있는 것으로 보면, 1번 문항은 정답률이 선형적으로 증가하는 경향은 약하다고 볼 수 있다.

다음으로 21번 문항을 보면 문항에서 요구하는 3개의 인지요소 중 2,3번째의 인지요소만 숙달한 경우에는 $P(011) = 0.9999$의 확률을 보이지만, 인지요소 모두를 숙달한 경우에는 $P(111) = 0.4023$의 확률을 보인다. 즉, 인지요소를 더 많이 숙달했음에도 정답률이 눈에 띄게 하락했다. 이러한 현상이 나타나는 이유는 인지요소 선정이나 Q-행렬 구성과정에 오류가 있다고 추측할 수 있다.

검사 전체에서 인지요소 프로파일별 각 문항에 대한 정답률은 부록에 실어두었다.

2. 윤지영(2016)의 응답데이터 분석 결과

Wald 통계량을 이용한 모형 선택 방법을 윤지영(2016)의 응답데이터에 적용하면 [표 4-11]과 같은 결과가 나타난다. 결과를 보면 25문항 중 20문항이 ACDM이 적합하고, 3개의 문항이 DINA 모형이 적합하다고 나타났다(요구하는 인지요소의 개수가 1개인 1,2번 문항은 Wald 통계량 계산에서 제외된다).

응답모형	문항 수
DINA	3
DINO	0
ACDM	20

[표 4-11] 윤지영(2016)의 응답데이터 분석 결과

적합한 인지진단모형이 DINA 모형이라고 결정된 문항은 윤지영(2016)의 연구 속 검사지의 5, 7, 20번 문항이다. 각 문항과 문항이 요구하는 인지요소는 다음과 같다.

5. 다음은 미지수 x의 값에 따른 식 $2x-4$의 값을 나타낸 표이다. 방정식 $x=2x-4$의 해는?

x	-1	0	1	2	3	4	5
$2x-4$	-6	-4	-2	0	2	4	6

① -4 ② -2
③ 0 ④ 2
⑤ 4

[그림 4-4] 윤지영(2016)의 검사지 5번 문항

인지요소 1	사칙계산 및 식의 계산
인지요소 2	등식의 이해
인지요소 3	일차방정식의 해의 의미

[표 4-12] 5번 문항이 요구하는 인지요소(윤지영, 2016)

5번 문항은 방정식의 해를 찾는 문제다. 문항에서 요구하는 인지요소들의 내용을 바탕으로 결과를 분석해보면, 방정식과 관련된 문제를 해결하기 위해서는 기본적으로 사칙계산 및 식의 계산을 숙달해야하고(인지요소 1 숙달), =(등호)가 쓰인 등식의 의미를 알아야(인지요소 2 숙달)

식을 조작하는 등의 풀이가 가능하다. 그리고 위의 5번 문항은 해의 의미, 즉, 일차방정식에서 등식을 만족시키는 값이 곧 해라는 것을 알고 있어야(인지요소 3 숙달) 해결할 수 있다. 그러므로 위 문항에 적합한 인지진단모형이 DINA 모형으로 결정된 것은 적합하다고 판단할 수 있다.

위 문항도 마찬가지로 실제 응답데이터를 바탕으로 각 문항의 모수를 추정한 결과, 가능한 인지요소 프로파일별 5번 문항에 대한 정답률을 조사하면 [표 4-13]과 같다.

	P(000)	P(100)	P(010)	P(001)	P(110)	P(101)	P(011)	P(111)
Item 5	0.1712	0.6205	0.4391	0.3086	0.9999	0.5851	0.7933	0.9730

[표 4-13] 인지요소 프로파일별 윤지영(2016) 검사지 5번 문항의 정답률

위의 결과를 보면 $P(111)$은 0.9730이지만 $P(110)$은 0.9999로 $P(111)$보다 약간 큰 확률을 가지고 있다. 그러므로 위의 정답률을 바탕으로 생각해보면 5번 문항은 DINA 모형의 가정과 완벽히 일치하는 것으로 보기는 어렵다. 하지만 이를 바탕으로 다른 의미 있는 해석을 이끌어낼 수 있다. 문항의 질이 낮은 경우, 즉, slip 모수 s_j가 0.2인 경우, 문항에서 요구하는 인지요소 모두를 숙달했을 때 문항을 해결할 확률은 $1-s_j=0.8$이다. 그렇다면 피험자의 인지요소 프로파일이 $(1,1,1)$, $(1,1,0)$, $(0,1,1)$이라면 5번 문항을 옳게 해결할 수 있을 것이라고 판단할 수 있다. 세 인지요소 프로파일에 공통적으로 포함된 인지요소는 인지요소2다. 그러므로 5번 문항을 해결하는데 중요한 역할을 하는 인지요소는 인지요소 2, 즉, ‘등식의 이해’라고 판단할 수 있다.

7. 일차방정식 $3(2x-1)+2x=21$을 만족하는 x의 값은?

① $-\frac{11}{4}$

② $\frac{11}{4}$

③ -3

④ 3

⑤ $\frac{9}{4}$

[그림 4-4] 윤지영(2016)의 검사지 7번 문항

인지요소 1	사칙계산 및 식의 계산
인지요소 2	등식의 이해
인지요소 4	일차방정식 풀기(등식의 성질)

[표 4-14] 7번 문항이 요구하는 인지요소(윤지영, 2016)

다음으로 7번 문항은 방정식의 해를 구해야하는 문제다. 문항에서 요구하는 인지요소들의 내용을 바탕으로 결과를 분석해보면, 사칙계산 및 식의 계산을 숙달하고 있어야 하고(인지요소 1 숙달), =(등호)가 쓰인 등식이 어떤 의미인지 알아야(인지요소 2 숙달) 주어진 방정식을 정리할 수 있다. 그리고 좌변과 우변의 같은 값을 빼고 더해도 등식은 유지된다는 점을 이용해 방정식의 해를 최종적으로 구해야(인지요소 4 숙달) 7번 문항을 해결할 수 있다. 즉, 인지요소 1,2,4 모두를 숙달해야 7번 문항을 해결할 수 있다. 그러므로 위 문항에 적합한 인지진단모형이 DINA 모형으로 결정된 것은 적합하다고 볼 수 있다.

7번 문항도 마찬가지로 실제 응답데이터를 바탕으로 각 문항의 모수를 추정한 결과, 가능한 인지요소 프로파일별 7번 문항에 대한 정답률을 조

사하면 [표 4-15]와 같다.

	P(000)	P(100)	P(010)	P(001)	P(110)	P(101)	P(011)	P(111)
Item 7	0.0382	0.2748	0.0001	0.9999	0.9999	0.3574	0.9999	0.9999

[표 4-15] 인지요소 프로파일별 윤지영(2016) 검사지 7번 문항의 정답률

위의 결과를 보면 $P(001)=P(110)=P(011)=P(111)=0.9999$로 모두 같다. 그러므로 위의 정답률을 바탕으로 생각해보면 7번 문항은 DINA 모형의 가정과 완벽하게 일치하지 않는 것으로 볼 수 있다. 덧붙여 위의 [표 4-13]을 바탕으로 7번 문항을 분석해보면, 피험자는 인지요소 4만 숙달해도 7번 문항을 해결할 수 있을 것으로 기대된다. 즉, 피험자가 등식의 성질을 바탕으로 일차방정식을 풀 수 있다면, 7번 문항은 해결할 수 있을 것으로 기대된다. 하지만 등식의 이해가 선행되어야 등식의 성질을 바탕으로 일차방정식을 풀 수 있기 때문에 위의 결과는 올바른 결과가 아니라고 해석할 수 있다. 올바르지 않은 결과가 나온 원인은 역시 인지요소가 불명확하게 선정되었거나, Q-행렬 구성에 있어 오류가 있다고 추측할 수 있다.

20. $a<0$일 때, 일차부등식 $ax+2>1$를 풀면?

① $x<\frac{1}{a}$

② $x>\frac{1}{a}$

③ $x<\frac{5}{a}$

④ $x>-\frac{1}{a}$

⑤ $x<-\frac{1}{a}$

[그림 4-6] 윤지영(2016)의 검사지 20번 문항

인지요소 1	사칙계산 및 식의 계산
인지요소 2	등식의 이해
인지요소 6	부등식의 이해
인지요소 8	일차부등식 풀기(부등식의 성질)

[표 4-16] 20번 문항이 요구하는 인지요소(윤지영, 2016)

마지막으로, 20번 문항은 부등식의 해를 구하는 문항이다. 문항에서 요구하는 인지요소들의 내용을 바탕으로 결과를 분석해보면, 앞선 문제들과 마찬가지로 기본적으로 사칙계산(인지요소 1) 및 등식의 이해(인지요소 2)가 숙달되어야 문제를 해결할 수 있다. 그리고 이 문항은 부등식과 관련된 문제이므로 부등호($<,>$)가 포함된 부등식에 대한 이해를 요구한다(인지요소 6 숙달). 이제 피험자가 $ax>-1$을 거쳐 $x<-\frac{1}{a}$라는 해를 구하기 위해서는 부등식의 성질을 숙달하고 있어야(인지요소 8 숙달) 한다. 그러므로 이 문항도 DINA 모형이 적합한 인지진단모형이라고 결정된 것은 적합하다고 볼 수 있다.

20번 문항도 마찬가지로 실제 응답데이터를 바탕으로 각 문항의 모수를 추정한 결과, 가능한 인지요소 프로파일별 20번 문항에 대한 정답률을 조사하면 [표 4-17]과 같다.

	P(0000)	P(1000)	P(0100)	P(0010)	P(0001)	P(1100)	P(1010)	P(1001)
Item 20	0.0001	0.1917	0.1364	0.0322	0.3045	0.0683	0.0917	0.3187
	P(0110)	P(0101)	P(0011)	P(1110)	P(1101)	P(1011)	P(0111)	P(1111)
	0.0001	0.1567	0.2802	0.1968	0.2252	0.0001	0.7423	0.7546

[표 4-17] 인지요소 프로파일별 윤지영(2016) 검사지 20번 문항의 정답률

위의 결과를 보면 인지요소 모두를 숙달했을 때((1111)) 문항을 해결할 확률이 가장 높다. 그리고 인지요소 프로파일이 (0111)인 경우를 제외하면 문항을 해결할 확률이 0.3과 비슷하거나 낮으므로 20번 문항은 어느 정도 DINA

모형의 가정과는 일치한다고 해석할 수 있다. 그리고 20번 문항이 DINA 모형을 따른다고 했을 때, 20번 문항의 slip 모수는 약 0.25가 되고 guess 모수는 0.0001이 된다.

윤지영(2016)의 응답데이터를 분석한 결과 역시 ACDM이 적합하다고 결정된 문항들에 대해 각 문항의 모수를 추정한 결과 가능한 인지요소 프로파일별 각 문항에 대한 정답률을 확인하면, 대부분의 문항들이 인지요소를 숙달할수록 문항의 정답률이 선형적으로 증가하는 것을 확인할 수 있다. 다만, 일부 문항의 각 인지요소 프로파일별 정답률을 확인해보면 몇몇 문항은 정답률이 선형적으로 증가하지 않는 경우도 있었다. 예를 들어 문항 3과 문항 22의 인지요소 프로파일별 정답률은 다음과 같다.

Item 3	P(00)	P(10)	P(01)	P(11)												
	0.5628	0.9313	0.8649	0.9999												
Item 22	P(000000)	P(100000)	P(010000)	P(001000)	P(000100)	P(000010)	P(000001)	P(110000)	P(101000)	P(100100)	P(100010)	P(100001)	P(011000)	P(010100)	P(010010)	P(010001)
	0.0001	0.3861	0.0001	0.0001	0.9999	0.3326	0.9999	0.0001	0.0001	0.3052	0.5510	0.0001	0.0001	0.0001	0.2011	0.9999
	P(001100)	P(001010)	P(001001)	P(000110)	P(000101)	P(000011)	P(111000)	P(110100)	P(110010)	P(110001)	P(101100)	P(101010)	P(101001)	P(100110)	P(100101)	P(100011)
	0.0001	0.0001	0.4103	0.0001	0.0001	0.0001	0.9999	0.3961	0.7700	0.0001	0.0001	0.3477	0.0001	0.0001	0.0001	0.0001
	P(011100)	P(011010)	P(011001)	P(010110)	P(010101)	P(010011)	P(001110)	P(001101)	P(001011)	P(000111)	P(111100)	P(111010)	P(111001)	P(110110)	P(110101)	P(110011)
	0.0001	0.0001	0.0001	0.2262	0.2557	0.0001	0.9999	0.9999	0.9999	0.9999	0.6211	0.0001	0.0001	0.0001	0.7966	0.6173
	P(101110)	P(101101)	P(101011)	P(100111)	P(011110)	P(011101)	P(011011)	P(010111)	P(001111)	P(111110)	P(111101)	P(111011)	P(110111)	P(101111)	P(011111)	P(111111)
	0.9999	0.0001	0.0831	0.0001	0.0001	0.0001	0.5685	0.9999	0.0001	0.0001	0.4048	0.6422	0.3628	0.9999	0.9999	0.7863

[표 4-18] 인지요소 프로파일별 윤지영(2016) 검사지 3번, 22번 문항의 정답률

우선, 3번 문항의 결과를 살펴보면 문항에서 요구하는 2개의 인지요소 중 1개만 숙달한 경우의 정답률은 각각 0.9313, 0.8649로 상당히 높다. 즉, 모의실험에서 문항의 질이 낮은 문항의 slip 모수를 $s_j = 0.2$로 설정한 점을 고려

하면, $P(01)=0.8649$인 경우에는 문항을 해결할 수 있다고 볼 수 있다. 그리고 또 하나 주목할 점은 문항에서 요구하는 인지요소를 하나도 숙달하지 못한 경우의 정답률은 $P(00)=0.5628$로 0.5를 넘는 값이다. 그러므로 5번 문항은 guess 모수가 매우 커 문항의 질이 낮다고 볼 수 있다.

다음으로 22번 문항의 결과를 보면 문항에서 요구하는 인지요소 모두를 숙달하지 못했음에도 불구하고 정답률이 0.9999인 경우가 있다. 하지만 인지요소 모두를 숙달한 경우의 정답률은 $P(111111)=0.7863$으로 0.8에 못 미치는 수준이다. 즉, 이 경우에는 문항에서 요구하는 인지요소 모두를 숙달하고 있음에도 정답률이 0.8 아래로 하락하는 현상이 나타났다. 이 경우 역시 인지요소 선정이나 Q-행렬 구성에 오류가 있다고 추측할 수 있다.

윤지영(2016)의 검사 전체에서 인지요소 프로파일별 각 문항에 대한 정답률은 부록에 실어두었다.

제 5 장 논의 및 제한점

연구 결과를 종합했을 때, Wald 통계량을 이용한 모형 선택 방법은 어느 정도 효율이 있다고 볼 수 있다. 문항에서 요구하는 인지요소의 개수가 많을수록, ACDM이 생성모형인 경우 Wald 통계량을 이용한 모형 선택의 TPR은 높아지는 것을 확인할 수 있었다. 특히, 문항에서 요구하는 인지요소의 개수가 1개부터 6개까지 다양하게 설정한 Lee et al.(2011)의 Q-행렬을 사용한 모의실험에서 확인할 수 있듯이, 문항의 질이 높은 경우 DINA 모형, DINO 모형, ACDM을 생성모형으로 하는 경우 Wald 통계량을 이용한 모형 선택 방법의 TPR은 70% 이상으로 높은 일치율을 보이는 것을 확인할 수 있었다.

하지만 Wald 통계량을 이용한 모형 선택 방법의 몇 가지 한계를 확인할 수 있었다. 모의실험 결과에서, Wald 통계량을 이용한 모형 선택의 정확도는 문항에서 요구하는 인지요소의 수, Q-행렬의 복잡성, 문항의 질, 혼합모형/단일모형 기반 검사 여부 등에 영향을 받는 것으로 나타났다. 그러나 이 연구에서 수행한 모의실험에는 제시된 문항수준 모형적합도에 영향을 주는 요인들이 상호작용하였을 때 어느 정도의 정확도를 보장하는지에 대해 명확한 가이드라인을 제시할 만큼 다양한 실험조건을 시도하지 못하였다. 따라서 Wald 통계량을 이용한 모형 선택의 정확도에 영향을 주는 요인들을 더 찾아보고 더 광범위한 모의실험을 통해 Wald 검사의 성질을 더 확실하게 규명할 필요가 있다.

구체적으로, 모의실험 3에서 ACDM의 경우 Wald 통계량을 이용한 모형 선택의 TPR은 문항에서 요구하는 인지요소의 개수가 4개 이상인 경우 문항의 질과 상관없이 거의 100%를 유지하고 있고, 문항에서 요구하는 인지요소의 개수가 3개인 경우에도 상당히 높은 값을 보이고 있다. 다만, 문항에서 요구하는 인지요소의 개수가 2개인 경우에는 모의실험 1과 2처럼 Q-행렬이 간단한 경우와 다소 차이가 나타난다.

우선, Q-행렬이 간단하고 문항에서 요구하는 인지요소의 개수가 2개인 문

항들로만 이루어진 검사를 사용한 모의실험 1의 경우 ACDM이 생성모형일 때 Wald 통계량을 이용한 모형 선택의 TPR은 문항의 질이 낮을 때는 69.3%, 중간일 때는 62.6%, 높을 때는 59.8%로 점점 낮아지고 70%에 미치지 못하는 수준이었다. 그래서 이러한 현상의 이유를 모수의 변화에 있다고 추측했었다. 즉, 문항의 질이 높아짐에 따라 각 문항이 갖는 intercept 값 δ_{j0}이 작아지고, 그로 인해 ACDM이 가지는 모수와 DINA 모형, DINO 모형이 가지는 모수의 차이가 작아져 ACDM이 생성모형임에도 불구하고 Wald 통계량에 의해 DINA 모형 또는 DINO 모형이 적합한 모형이라고 결정되는 경우가 많아 ACDM이 생성모형일 때 Wald 통계량을 이용한 모형 선택의 TPR이 작아지는 것이다.

이와는 다르게 복잡한 Q-행렬을 사용한 모의실험 3에서는 문항에서 요구하는 인지요소의 개수가 2개인 문항들의 경우만 보면, ACDM이 생성모형일 때 Wald 통계량을 이용한 모형 선택의 TPR은 문항의 질이 낮을 때는 84.2%, 중간일 때는 87.8%, 높을 때는 86.4%로 나타난다. 전반적으로 80%이상의 일치율을 보이며 문항의 질이 중간일 때는 낮을 때보다 그 폭은 작지만 증가하는 모습도 나타났다. 이러한 현상이 나타난 이유는 복잡한 Q-행렬을 사용한 경우에 검사 전체에서 요구하는 인지요소도 더 다양하고, 인지요소를 많이 요구하는 문항에 영향을 받는다고 추측할 수 있기 때문에 이에 대한 추가적인 모의실험을 실시해볼 필요가 있다.

또 다른 측면의 한계점은 Wald 통계량은 상대 적합도 통계량이라는 것이다. 그러므로 인지진단평가를 실시 할 때에는 모형적합도를 판단할 수 있는 절대 적합도 통계량이나 다른 상대 적합도 통계량을 함께 사용하여 Wald 통계량을 이용한 모형 선택 방법을 보완할 필요가 있다. 절대 적합도 통계량의 한 예로 $S-X^2$ 통계량이 있다. Sorrel et al.(2017a)은 $S-X^2$을 인지진단모형의 문항수준 절대 적합도 통계량으로 이용한다면, 제1종 오류율은 유지되만, 검정력은 받아들일 수 없는 수준이라고 결론지었다. 그러므로 $S-X^2$을 바로 문항수준 적합도 통계량으로 사용하기에는 무리가 있다. 하지만 Sorrel et al.(2017a)은 Wang et al.(2015)의 연구를 언급하며 $S-X^2$의 문제점 개선의 실마리를 언급하고 있어 $S-X^2$ 통계량을 이용하여 Wald 통계량을 이용한 모형 선택 방법의 정확도를

개선할 수 있는 방법을 고려해볼 수 있다.

추가적으로 우도 비 검사는 대부분 문항반응이론 프레임워크에서 많이 연구된 내용으로 아직 인지진단모형 프레임워크에서 많이 연구되지 않았다. Sorrel et al.(2017a)에 의하면 우도 비 검사는 Wald 검사에 비해 계산과정이 복잡하고 오래 걸린다는 단점이 있지만, 우도 비 검사가 Wald 검사에 비해 강력하다고 주장했다. 그런데 우도 비 검사를 수행함에 있어 우도는 검사수준에서 계산되기 때문에 이를 문항수준 모형적합도로 사용하는 것이 타당한지에 대해서는 추가적인 연구가 필요하다.

참 고 문 헌

교육부 (2015). 수학과 교육과정. 교육부 고시 제2015-74호 [별책 8].

김정환. (2006). 학습자 성취도 평가의 문제와 발전방향 탐색. **교육문화연구**, **12**, 199-218.

김희경, & 김부미. (2013). 인지진단모형을 활용한 수학 학업성취 결과 분석. **학교수학**, **15**(2), 289-314.

성태제. (2001). **문항반응이론의 이해와 적용**. 서울: 교육과학사.

성태제. (2018). 인지진단모형에 의한 중 1 수학 '수와 연산'의 형성평가 도구 개발과 분석. **교육평가연구**, **31**, 29-53.

손원숙, 김경희, 박정, & 박효희. (2009). 한국, 핀란드, 홍콩-중국의 과학 성취 모형 비교. **교육평가연구**, **22**, 129-149.

시기자. (2003). 검사매체와 검사시행 모형에 따른 피험자 능력모수와 피험자 응답 적합도비교. 이화여자대학교 박사학위논문.

오영수, & 전영주. (2018). 문제제기 활동이 수학에 대한 정의적 영역에 미치는 영향. **한국콘텐츠학회논문지**, **18**(2), 541-552.

윤지영. (2016). 요소 간 위계 방식과 인공신경망을 적용한 수학 인지진단 평가 연구, 서울대학교 대학원 석사학위논문.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74.

Blumer, A., Ehrenfeucht, A., Haussler, D., &Warmuth, M. K. (1987). Occam's razor. *Information processing letters, 24*(6), 377-380.

Chen, J., Torre, J., & Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, *50*(2), 123-140.

Cui, Y., & Leighton, J. P. (2009). The hierarchy consistency index: Evaluating person fit for cognitive diagnostic assessment. *Journal of Educational Measurement*, 46(4), 429-449.

Cui, Y., & Li, J. (2015). Evaluating person fit for cognitive diagnostic assessment. *Applied Psychological Measurement*, 39(3), 223-238.

de la Torre, J. (2008). An empirically based method of Q matrix validation for the DINA model: Development and applications. *Journal of educational measurement*, 45(4), 343-362.

de la Torre, J. (2009a). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement*, *33*(3), 163-183.

de la Torre, J. (2009b). DINA model and parameter estimation: A didactic. *Journal of educational and behavioral statistics*, 34(1), 115-130.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76(2), 179-199.

de la Torre, J., & Lee, Y. S. (2013). Evaluating the Wald test for item level comparison of saturated and reduced models in cognitive diagnosis. *Journal of Educational Measurement*, 50(4), 355-373.

de la Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicología Educativa*, *20*(2), 89-97.

DiBello, L. V., Roussos, L. A., & Stout, W. (2007). *Review of cognitively diagnostic assessment and a summary of psychometric models*. CR Rao, & S. Sinharay (Eds.), Handbook of statistics, Vol. 26: Psychometrics (pp. 970 - 1030).

Finkelman, M., Kim, W., &Roussos, L. A. (2009). Automated test assembly for cognitive diagnosis models using a genetic algorithm. *Journal of educational measurement*, *46*(3), 273-292.

Haertel, E. H. (1989). Using restricted latent class models to map the

skill structure of achievement items. *Journal of Educational Measurement*, 26(4), 301-321.

Hartz, S. M. (2002). A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality (Doctoral dissertation, ProQuest Information & Learning).

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74(2), 191.

Hong, C. Y., Chang, Y. W., & Tsai, R. C. (2016). Estimation of Generalized DINA Model with Order Restrictions. *Journal of Classification*, 33(3), 460-484.

Hu, J., Miller, M. D., Huggins-Manley, A. C., &Chen, Y. H. (2016). Evaluation of model fit in cognitive diagnosis models. *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(3), 258-272.

Lee, Y. S., Park, Y. S., & Taylan, D. (2011). A cognitive diagnostic modeling of attribute mastery in Massachusetts, Minnesota, and the US national sample using the TIMSS 2007. *International Journal of Testing*, 11(2), 144-177.

Li, H., & Suen, H. K. (2013). Constructing and validating a Q-matrix for cognitive diagnostic analyses of a reading test. *Educational Assessment*, 18(1), 1-25.

Martin, M. O., Mullis, I. V., Foy, P., & Olson, J. F. (2008). TIMSS 2007 international mathematics report. *Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*, 319-323.

Mullis, I. V. S., Martin, M. O., Foy, P., &Hooper, M.(2016). *TIMSS 2015 InternationalResults in Mathematics.* Retrieved from BostonCollege, TIMSS & PIRLS International StudentCenter. Retrieved from http://timssandpirls.bc.edu/timss2015/international-results/

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.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, 1-32.

Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24(1), 50-64.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied psychological measurement*, *9*(4), 401-412.

Rojas, G., de la Torre, J., &Olea, J. (2012). Choosing between general and specific cognitive diagnosis models when the sample size is small. In *annual meeting of the National Council of Measurement in Education, Vancouver, British Columbia, Canada.*

Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement*, 6(4), 219-262.

Sorrel, M. A., Abad, F. J., Olea, J., de la Torre, J., & Barrada, J. R. (2017a). Inferential Item-Fit Evaluation in Cognitive Diagnosis Modeling. *Applied Psychological Measurement*, 41(8), 614-631.

Sorrel, M. A., de la Torre, J., Abad, F. J., & Olea, J. (2017b).

Two-step likelihood ratio test for item-level model comparison in cognitive diagnosis models. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 13(S1), 39.

Stone, C. A. (2000). Monte Carlo Based Null Distribution for an Alternative Goodness of Fit Test Statistic in IRT Models. *Journal of Educational Measurement*, 37(1), 58-75.

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

Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. *Cognitively diagnostic assessment*, 327-359.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological methods*, 11(3), 287.

von Davier, M. (2005). A general diagnostic model applied to language testing data. *ETS Research Report Series*, 2005(2).

Wang, C., & Gierl, M. J. (2011). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in critical reading. *Journal of Educational Measurement*, 48(2), 165-187.

Wang, C., Shu, Z., Shang, Z., & Xu, G. (2015). Assessing item-level fit for the DINA model. *Applied Psychological Measurement*, 39(7), 525-538.

Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational assessment*, 11(3-4), 283-289.

<부 록>

[부록 1] TIMSS 2007 응답데이터 분석 결과 - 인지요소 프로파일별 각 문항에 대한 정답률

	P(0)	P(1)
Item 2	0.0001	0.9999
Item 9	0.5396	0.9999
Item 24	0.1919	0.6877

	P(00)	P(10)	P(01)	P(11)
Item 1	0.0001	0.9999	0.8051	0.9999
Item 4	0.0001	0.1050	0.9999	0.9999
Item 6	0.5816	0.9999	0.9999	0.9808
Item 10	0.0001	0.9999	0.4594	0.9999
Item 15	0.2909	0.6836	0.9339	0.9696
Item 16	0.0724	0.4953	0.2651	0.9618
Item 17	0.0001	0.0001	0.3735	0.9999
Item 23	0.1214	0.5488	0.5650	0.9771

	P(000)	P(100)	P(010)	P(001)	P(110)	P(101)	P(011)	P(111)
Item 3	0.0001	0.0105	0.4877	0.0001	0.9999	0.0001	0.9999	0.9999
Item 5	0.1712	0.6205	0.4391	0.3086	0.9999	0.5851	0.7933	0.9730
Item 7	0.0382	0.2748	0.0001	0.9999	0.9999	0.3574	0.9999	0.9999
Item 12	0.0001	0.0001	0.0001	0.9999	0.0001	0.9999	0.9999	0.9999
Item 18	0.0001	0.2055	0.0001	0.9999	0.4941	0.9999	0.9999	0.9999
Item 19	0.0001	0.0001	0.0001	0.2133	0.0001	0.9881	0.9999	0.9999
Item 21	0.0001	0.1631	0.0001	0.0001	0.3218	0.0165	0.9999	0.4023
Item 22	0.0001	0.0001	0.9999	0.0001	0.8328	0.3850	0.9999	0.9999

	P(0000)	P(1000)	P(0100)	P(0010)	P(0001)	P(1100)	P(1010)	P(1001)	P(0110)	P(0101)	P(0011)	P(1110)	P(1101)	P(1011)	P(0111)	P(1111)
Item 11	0.2656	0.0629	0.2222	0.3402	0.0001	0.9002	0.2084	0.2574	0.9714	0.9606	0.2031	0.4027	0.0001	0.2101	0.0728	0.4788
Item 13	0.0001	0.0001	0.9999	0.5801	0.5492	0.9737	0.7888	0.9999	0.6450	0.0872	0.9999	0.8341	0.9999	0.5001	0.9999	0.8938
Item 20	0.0488	0.3807	0.3853	0.9999	0.9697	0.3665	0.5669	0.0001	0.9999	0.7775	0.1052	0.9999	0.4035	0.9999	0.6358	0.9999
Item 25	0.2297	0.1222	0.9999	0.8013	0.3165	0.0001	0.9999	0.1576	0.9999	0.9999	0.5210	0.9999	0.4553	0.9611	0.9999	0.9999

Item 8	P(00000)	P(10000)	P(01000)	P(00100)	P(00010)	P(00001)	P(11000)	P(10100)	P(10010)	P(10001)	P(01100)	P(01010)	P(01001)	P(00110)	P(00101)	P(00011)
	0.1661	0.2963	0.0001	0.0001	0.0001	0.3340	0.9999	0.2937	0.1462	0.1444	0.0001	0.9999	0.2540	0.9999	0.4671	0.0001
	P(11100)	P(11010)	P(11001)	P(10110)	P(10101)	P(10011)	P(01110)	P(01101)	P(01011)	P(00111)	P(11110)	P(11101)	P(11011)	P(10111)	P(01111)	P(11111)
	0.5112	0.0001	0.5147	0.4078	0.2294	0.9999	0.9999	0.0001	0.4085	0.0001	0.9999	0.9999	0.9999	0.0001	0.0001	0.7488
Item 14	P(000000)	P(100000)	P(010000)	P(001000)	P(000100)	P(000010)	P(000001)	P(110000)	P(101000)	P(100100)	P(100010)	P(100001)	P(011000)	P(010100)	P(010010)	P(010001)
	0.0001	0.0001	0.0001	0.0001	0.0001	0.5861	0.0001	0.0001	0.2010	0.6262	0.3397	0.4257	0.0001	0.0001	0.9999	0.0001
	P(001100)	P(001010)	P(001001)	P(000110)	P(000101)	P(000011)	P(111000)	P(110100)	P(110010)	P(110001)	P(101100)	P(101010)	P(101001)	P(100110)	P(100101)	P(100011)
	0.4961	0.0001	0.0001	0.3574	0.0001	0.9999	0.0001	0.0419	0.9999	0.0001	0.0001	0.0001	0.9999	0.0001	0.9999	0.0001
	P(011100)	P(011010)	P(011001)	P(010110)	P(010101)	P(010011)	P(001110)	P(001101)	P(001011)	P(000111)	P(111100)	P(111010)	P(111001)	P(110110)	P(110101)	P(110011)
	0.0001	0.7442	0.0001	0.0001	0.0001	0.3586	0.0001	0.0001	0.9999	0.9999	0.0001	0.9999	0.0001	0.9999	0.0001	0.9999
	P(101110)	P(101101)	P(101011)	P(100111)	P(011110)	P(011101)	P(011011)	P(010111)	P(001111)	P(111110)	P(111101)	P(111011)	P(110111)	P(101111)	P(011111)	P(111111)
	0.0001	0.0001	0.9999	0.0001	0.0001	0.0001	0.1315	0.0001	0.9999	0.9999	0.2111	0.9999	0.9999	0.0001	0.9999	0.8779

[부록 2] 윤지영(2016)의 응답데이터 분석 결과 - 인지요소 프로파일별 각 문항에 대한 정답률

	P(0)	P(1)
Item 1	0.2969	0.9649
Item 2	0.4125	0.9999

	P(00)	P(10)	P(01)	P(11)
Item 3	0.5628	0.9313	0.8649	0.9999
Item 4	0.3472	0.5430	0.5075	0.8773
Item 10	0.2310	0.7991	0.8517	0.9999

	P(000)	P(100)	P(010)	P(001)	P(110)	P(101)	P(011)	P(111)
Item 5	0.0001	0.7125	0.1359	0.8687	0.6805	0.7077	0.8480	0.9999
Item 6	0.4255	0.9023	0.4940	0.3701	0.9604	0.9872	0.9999	0.9858
Item 7	0.0001	0.9195	0.0001	0.7622	0.8997	0.8525	0.9999	0.9716
Item 13	0.0001	0.7272	0.8916	0.9628	0.9860	0.9303	0.9999	0.9789
Item 14	0.0704	0.2715	0.3148	0.3456	0.4001	0.5424	0.4359	0.8882

	P(0000)	P(1000)	P(0100)	P(0010)	P(0001)	P(1100)	P(1010)	P(1001)	P(0110)	P(0101)	P(0011)	P(1110)	P(1101)	P(1011)	P(0111)	P(1111)
Item 8	0.0001	0.0001	0.1948	0.2512	0.0001	0.0001	0.1359	0.4158	0.0001	0.2431	0.0001	0.3037	0.9999	0.9999	0.2864	0.9999
Item 9	0.0870	0.0001	0.3292	0.4692	0.0001	0.9999	0.9346	0.0001	0.0001	0.8749	0.2302	0.5164	0.8069	0.9999	0.3728	0.9926
Item 15	0.0754	0.0001	0.0001	0.1479	0.2708	0.6297	0.4228	0.9999	0.3224	0.9461	0.4073	0.8102	0.7497	0.7934	0.4366	0.9879
Item 16	0.0001	0.0001	0.0001	0.0001	0.6320	0.9302	0.0001	0.9999	0.8510	0.8728	0.9999	0.5794	0.9999	0.9999	0.9999	0.9999
Item 18	0.2243	0.1783	0.2361	0.2030	0.5841	0.4085	0.9999	0.3811	0.9999	0.5110	0.9999	0.9004	0.9999	0.7855	0.3093	0.9999
Item 19	0.1131	0.1712	0.0728	0.1164	0.1363	0.2702	0.6823	0.3160	0.6618	0.2576	0.0001	0.9676	0.8950	0.9999	0.7655	0.9897
Item 20	0.0001	0.1917	0.1364	0.0322	0.3045	0.0683	0.0917	0.3187	0.0001	0.1567	0.2802	0.1968	0.2252	0.0001	0.7423	0.7546

Item 11	P(00000)	P(10000)	P(01000)	P(00100)	P(00010)	P(00001)	P(11000)	P(10100)	P(10010)	P(10001)	P(01100)	P(01010)	P(01001)	P(00110)	P(00101)	P(00011)
	0.0001	0.2355	0.0001	0.0001	0.0001	0.9999	0.7055	0.0001	0.4555	0.0001	0.0001	0.4392	0.0001	0.0001	0.2983	0.0001
	P(11100)	P(11010)	P(11001)	P(10110)	P(10101)	P(10011)	P(01110)	P(01101)	P(01011)	P(00111)	P(11110)	P(11101)	P(11011)	P(10111)	P(01111)	P(11111)
	0.9999	0.8809	0.7609	0.0001	0.8597	0.9999	0.2161	0.0001	0.5503	0.9999	0.7360	0.0013	0.9349	0.9999	0.8079	0.9999
Item 12	P(00000)	P(10000)	P(01000)	P(00100)	P(00010)	P(00001)	P(11000)	P(10100)	P(10010)	P(10001)	P(01100)	P(01010)	P(01001)	P(00110)	P(00101)	P(00011)
	0.0001	0.1771	0.0001	0.0001	0.9999	0.0001	0.0001	0.9999	0.3832	0.9999	0.3938	0.0001	0.2350	0.9999	0.0001	0.3895
	P(11100)	P(11010)	P(11001)	P(10110)	P(10101)	P(10011)	P(01110)	P(01101)	P(01011)	P(00111)	P(11110)	P(11101)	P(11011)	P(10111)	P(01111)	P(11111)
	0.7803	0.9999	0.4802	0.9999	0.6338	0.9999	0.9999	0.7632	0.9999	0.9999	0.6086	0.9999	0.9999	0.8145	0.8228	0.9999
Item 17	P(00000)	P(10000)	P(01000)	P(00100)	P(00010)	P(00001)	P(11000)	P(10100)	P(10010)	P(10001)	P(01100)	P(01010)	P(01001)	P(00110)	P(00101)	P(00011)
	0.0001	0.4714	0.0001	0.0001	0.9999	0.0001	0.0001	0.6305	0.3308	0.0001	0.1514	0.0001	0.0001	0.1633	0.9999	0.4931
	P(11100)	P(11010)	P(11001)	P(10110)	P(10101)	P(10011)	P(01110)	P(01101)	P(01011)	P(00111)	P(11110)	P(11101)	P(11011)	P(10111)	P(01111)	P(11111)
	0.7258	0.9999	0.5127	0.0001	0.5866	0.7164	0.9999	0.0001	0.8351	0.2532	0.7297	0.4390	0.9999	0.8287	0.9999	0.9489
Item 24	P(00000)	P(10000)	P(01000)	P(00100)	P(00010)	P(00001)	P(11000)	P(10100)	P(10010)	P(10001)	P(01100)	P(01010)	P(01001)	P(00110)	P(00101)	P(00011)
	0.0001	0.0001	0.0001	0.4490	0.4838	0.3970	0.0001	0.0001	0.9999	0.0001	0.0001	0.6291	0.9999	0.3192	0.0001	0.1975
	P(11100)	P(11010)	P(11001)	P(10110)	P(10101)	P(10011)	P(01110)	P(01101)	P(01011)	P(00111)	P(11110)	P(11101)	P(11011)	P(10111)	P(01111)	P(11111)
	0.0001	0.6886	0.9999	0.9999	0.0001	0.0001	0.9999	0.9999	0.5335	0.0001	0.9999	0.9602	0.8083	0.9999	0.4185	0.9999

Item 21	P(000000)	P(100000)	P(010000)	P(001000)	P(000100)	P(000010)	P(000001)	P(110000)	P(101000)	P(100100)	P(100010)	P(100001)	P(011000)	P(010100)	P(010010)	P(010001)
	0.1106	0.4390	0.0001	0.2107	0.0001	0.0001	0.0001	0.0001	0.1583	0.1123	0.5477	0.0001	0.0001	0.2442	0.9999	0.1363
	P(001100)	P(001010)	P(001001)	P(000110)	P(000101)	P(000011)	P(111000)	P(110100)	P(110010)	P(110001)	P(101100)	P(101010)	P(101001)	P(100110)	P(100101)	P(100011)
	0.0001	0.9999	0.0001	0.7419	0.4659	0.9999	0.0001	0.5355	0.9999	0.9472	0.8024	0.6126	0.6682	0.3474	0.6442	0.9999
	P(011100)	P(011010)	P(011001)	P(010110)	P(010101)	P(010011)	P(001110)	P(001101)	P(001011)	P(000111)	P(111100)	P(111010)	P(111001)	P(110110)	P(110101)	P(110011)
	0.0001	0.9999	0.5345	0.7773	0.9999	0.4838	0.9999	0.0001	0.9999	0.9999	0.9999	0.7248	0.3851	0.9999	0.9999	0.9999
	P(101110)	P(101101)	P(101011)	P(100111)	P(011110)	P(011101)	P(011011)	P(010111)	P(001111)	P(111110)	P(111101)	P(111011)	P(110111)	P(101111)	P(011111)	P(111111)
	0.7396	0.7459	0.6627	0.9999	0.0001	0.9999	0.9999	0.9999	0.9999	0.9999	0.8393	0.9086	0.9063	0.9999	0.4488	0.9999
Item 22	P(000000)	P(100000)	P(010000)	P(001000)	P(000100)	P(000010)	P(000001)	P(110000)	P(101000)	P(100100)	P(100010)	P(100001)	P(011000)	P(010100)	P(010010)	P(010001)
	0.0001	0.3861	0.0001	0.0001	0.9999	0.3326	0.9999	0.0001	0.0001	0.3052	0.5510	0.0001	0.0001	0.0001	0.2011	0.9999
	P(001100)	P(001010)	P(001001)	P(000110)	P(000101)	P(000011)	P(111000)	P(110100)	P(110010)	P(110001)	P(101100)	P(101010)	P(101001)	P(100110)	P(100101)	P(100011)
	0.0001	0.0001	0.4103	0.0001	0.0001	0.0001	0.9999	0.3961	0.7700	0.0001	0.0001	0.3477	0.0001	0.0001	0.0001	0.0001
	P(011100)	P(011010)	P(011001)	P(010110)	P(010101)	P(010011)	P(001110)	P(001101)	P(001011)	P(000111)	P(111100)	P(111010)	P(111001)	P(110110)	P(110101)	P(110011)
	0.0001	0.0001	0.0001	0.2262	0.2557	0.0001	0.9999	0.9999	0.9999	0.9999	0.6211	0.0001	0.0001	0.0001	0.7966	0.6173
	P(101110)	P(101101)	P(101011)	P(100111)	P(011110)	P(011101)	P(011011)	P(010111)	P(001111)	P(111110)	P(111101)	P(111011)	P(110111)	P(101111)	P(011111)	P(111111)
	0.9999	0.0001	0.0831	0.0001	0.0001	0.0001	0.5685	0.9999	0.0001	0.0001	0.4048	0.6422	0.3628	0.9999	0.9999	0.7863

Item 23	P(000000)	P(100000)	P(010000)	P(001000)	P(000100)	P(000010)	P(000001)	P(110000)	P(101000)	P(100100)	P(100010)	P(100001)	P(011000)	P(010100)	P(010010)	P(010001)
	0.0001	0.0001	0.1207	0.4894	0.0001	0.9999	0.3521	0.0717	0.0001	0.5108	0.5504	0.4218	0.0001	0.0001	0.1997	0.1395
	P(001100)	P(001010)	P(001001)	P(000110)	P(000101)	P(000011)	P(111000)	P(110100)	P(110010)	P(110001)	P(101100)	P(101010)	P(101001)	P(100110)	P(100101)	P(100011)
	0.4326	0.9999	0.0001	0.8830	0.0001	0.0001	0.3054	0.0001	0.0001	0.0001	0.0001	0.9999	0.0001	0.0001	0.2315	0.6496
	P(011100)	P(011010)	P(011001)	P(010110)	P(010101)	P(010011)	P(001110)	P(001101)	P(001011)	P(000111)	P(111100)	P(111010)	P(111001)	P(110110)	P(110101)	P(110011)
	0.0001	0.0001	0.0001	0.0001	0.6927	0.0001	0.9999	0.0001	0.0001	0.9999	0.0001	0.1083	0.0001	0.5179	0.1036	0.7751
	P(101110)	P(101101)	P(101011)	P(100111)	P(011110)	P(011101)	P(011011)	P(010111)	P(001111)	P(111110)	P(111101)	P(111011)	P(110111)	P(101111)	P(011111)	P(111111)
	0.3565	0.0001	0.5431	0.0001	0.9999	0.0001	0.1260	0.9999	0.9999	0.0001	0.9999	0.3494	0.4542	0.3382	0.0001	0.6309
Item 25	P(000000)	P(100000)	P(010000)	P(001000)	P(000100)	P(000010)	P(000001)	P(110000)	P(101000)	P(100100)	P(100010)	P(100001)	P(011000)	P(010100)	P(010010)	P(010001)
	0.0001	0.0001	0.1271	0.0001	0.2461	0.0001	0.0001	0.0001	0.0001	0.0001	0.9999	0.1359	0.0001	0.9999	0.0001	0.0001
	P(001100)	P(001010)	P(001001)	P(000110)	P(000101)	P(000011)	P(111000)	P(110100)	P(110010)	P(110001)	P(101100)	P(101010)	P(101001)	P(100110)	P(100101)	P(100011)
	0.2147	0.9999	0.0001	0.0766	0.0001	0.9999	0.0001	0.7312	0.0001	0.4955	0.4372	0.3758	0.7150	0.9999	0.9999	0.6323
	P(011100)	P(011010)	P(011001)	P(010110)	P(010101)	P(010011)	P(001110)	P(001101)	P(001011)	P(000111)	P(111100)	P(111010)	P(111001)	P(110110)	P(110101)	P(110011)
	0.0001	0.0001	0.1013	0.0001	0.7958	0.9999	0.3291	0.9999	0.1270	0.9999	0.0001	0.0001	0.2603	0.0001	0.0001	0.3108
	P(101110)	P(101101)	P(101011)	P(100111)	P(011110)	P(011101)	P(011011)	P(010111)	P(001111)	P(111110)	P(111101)	P(111011)	P(110111)	P(101111)	P(011111)	P(111111)
	0.9999	0.9999	0.0001	0.0001	0.9999	0.3128	0.3387	0.0001	0.9999	0.9999	0.9311	0.8821	0.7112	0.0001	0.7008	0.9195

Abstract

Evaluation of Item-level Model Fit in Cognitive Diagnosis Assessment

Wang, Hyo Won
Department of Mathematics Education
The Graduate School
Seoul National University

The recent international achievement comparison studies show that Republic of Korea was ranked high place in performance of mathematics. However, Korean students exhibit significantly low marks in the attitude toward mathematics. Some educators claim that assessment systems in classrooms which focus only on final achievement adversely affect students' attitude toward mathematics. In this matter, assessment systems in classroom need to be changed for improved quality of education. The revised 2015 Mathematics curriculum in Republic of Korea promotes self-directed learning and informative use of assessment results. Therefore, assessment in school should aim for diagnostic purposes, rather than putting students in order.

Cognitive Diagnosis Assessment (CDA) using Cognitive Diagnosis

Model (CDM) is a form of assessment to provide the examinees' cognitive profiles consisting of mastery status of multiple attributes. CDA can offer information to feedback on students' learning and help improve teaching and learning methods for teachers. Many CDMs assuming various attribute-response relations have been developed. CDMs can be divided into two types: general models and specific models. General models such as generalized-DINA (G-DINA) model (de la torre, 2011) include all parameters associated with main effects and attribute-to-attribute interaction effects. In contrast, specific models consider special constraints, such as the Deterministic Inputs, Noise "And" Gate (DINA) model (Hartel, 1989; Junker & Sijitsma, 2001) which assumes that all attributes required for an item need to be mastered to solve the item.

Considering the fact that CDA statistically estimates each student's attributes profile, many researchers have become interested in the validity of CDA. Among a variety of CDMs, it is important to choose an appropriate model for a specific assessment. For this reason, methods for model-data fit have been suggested by many researchers. As of now, however, analyses of the model fit of CDMs are mostly limited to test-level. Since the appropriate model may be different for each item, deciding a single model for an entire test might not be appropriate for certain CDAs. Wald test proposed by de la Torre(2011) is a new method to compare the model fit between a general model and a specific model at an item-level. Wald test can only help deciding appropriate model between general and specific models at an item-level, but using the Wald statistics computed during Wald tests, it is possible to develop an item-level model selection method.

In this study, an item-level model selection method is developed

based on Wald statistics. Simulation studies are carried out to verify that the model selection method using Wald statistics can identify an appropriate model for each item using item response data. Moreover, factors that might affect the performance of the proposed item-level model selection method have been examined, including complexity of Q-matrix, the number of attributes per item, and item quality. Also, the proposed method is applied to real data sets and the results of chosen CDMs are analyzed regarding the nature of items.

The results of simulation studies indicate that the model selection method using Wald statistics yields the accuracy of at least 70% in the conditions such that item quality is high and complex Q-matrix is used. Increasement in the number of attributes per item and item quality increases the accuracy of the proposed method. We expect the proposed methods can be applied to CDAs consisting of heterogeneous items with different response models and contribute to improve validity of the CDAs.

keywords : Cognitive Diagnosis Model, Cognitive Diagnosis Assessment, G-DINA Model, Wald test, Model Fit, Item-level Model Fit

Student Number **: 2017-29385**