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

인지진단 검사의 Q행렬 타당성 분석 방법 연구

2016년 8월

서울대학교 대학원
수학교육과
안 희 상

인지진단 검사의 Q행렬 타당성 분석 방법 연구

지도교수 유 연 주

이 논문을 교육학석사학위논문으로 제출함

2016년 8월

서울대학교 대학원
수학교육과
안 희 상

안희상의 석사학위논문을 인준함
2016년 8월

위 원 장 _____ (인)
부 위 원 장 _____ (인)
위 원 _____ (인)

국문초록

인지진단 검사의 Q행렬 타당성 분석 방법 연구

오랜 시간동안 평가는 교육 활동의 과정이 종료된 후에 학생들에게 제대로 학습이 이루어졌는지를 확인하기 위한 수단으로 인식되어왔다. 그러나 최근 들어 교육 측정 분야에서는 평가의 역할을 단순히 학생들의 능력을 측정하여 수량화, 서열화 하는 것이 아니라 학습 과정 중에 학습 목표를 이끄는 활동으로 이해하는 경향이 나타나고 있다. 이에 따라 학생들의 인지적 강점과 약점을 결과로서 제공함으로써 향후 학습을 위한 구체적인 피드백이 가능한 인지진단평가에 대한 관심 또한 커지고 있다.

인지진단평가에서는 지식의 상태가 여러 인지요소의 다차원적 결합으로 이루어짐을 가정하고, 학생들이 습득해야 하는 세부적 인지요소의 숙달 여부 혹은 정도를 평가 문항에 대한 응답을 통해 측정한다. 이때 각각의 문항과 인지요소의 관계를 나타내는 행렬이 Q행렬이다. 인지진단 검사의 결과는 Q행렬과 학생의 문항응답으로 결정이 되기 때문에 정확하지 못한 Q행렬이 사용될 경우 검사 결과의 해석이 타당하지 않을 수 있다. 따라서 어떤 인지진단 검사의 결과 분석이 신뢰성을 확보하기 위해서는 Q행렬의 타당성을 입증하는 절차가 필요하다.

본 연구에서는 Q행렬의 타당성을 분석하는 기존의 여러 가지 방법들 중에서 Chia-Yi Chiu의 Q행렬 정련방법에 주목하여 문제점을 분석하고 이를 개선한 $AVERSS$ 타당성 분석 방법을 제시한다. $AVERSS$ 타당성 분석 방법은 Q행렬의 타당성을 판정하는 양적인 방법으로 $AVERSS$라는 타당성 지수를 사용한다. Q행렬 타당성 분석 방법의 적용 연구를 위해서는 윤지영(2015)에서 이루어진 일차방정식과 일차부등식 평가의 Q행렬과 평가 결과가 활용되었으며, 효율성 및 정확도 분석에는 가상데이터를 이용하였다.

Chia-Yi Chiu의 Q행렬 정련방법은 RSS라는 타당성 지수를 활용하여 기

존의 Q행렬을 정련해 나가는 방법으로 이를 실제 인지진단 검사에 적용하여 본 결과 이론에 맞지 않는 정련이 일어났음을 확인할 수 있었다. 이와 같은 문제점은 RSS가 최소일 때가 정확한 Q행렬이라는 Q행렬 정련방법의 가정에서 비롯된 것일 수 있으며, 유사한 가정을 가지는 Hamming모형 역시 그 실효성에 문제가 있음을 확인하였다. 이는 실제 인지진단 검사의 Q행렬을 분석함에 있어서 탐색적 접근이 가지는 문제점을 드러내는 결과이다.

따라서 본 연구에서는 확인적 접근의 방법인 $AVERSS$ 타당성 분석 방법을 개발하였으며 Q행렬 전체의 타당성을 대표할만한 지수로 $AVERSS$를 정의하였다. $AVERSS$ 타당성 분석 방법은 검사 데이터를 통하여 검사에 사용된 Q행렬의 $AVERSS$를 얻은 후 이를 컷오프 값과 비교하여 Q행렬의 타당성을 판정한다. 이때 컷오프 값으로는 모의실험 데이터를 통하여 얻은 $AVERSS$ 분포의 95 백분위수를 사용하였다. 또한 잘못된 Q행렬들에 대하여 $AVERSS$ 타당성 분석 방법을 적용한 결과 Q행렬의 3%만 잘못되어도 93%이상 타당하지 않다고 판정하였고, 이로써 $AVERSS$ 타당성 분석 방법이 높은 효율성으로 Q행렬의 타당성을 판정함을 확인하였다.

이 연구는 탐색적 접근의 기존의 Q행렬 타당화 방법들이 가지는 문제점을 보완하여 실제 인지진단 검사에 적용 가능한 확인적 접근의 Q행렬 타당성 분석 방법을 개발하였다는 것에 그 의의가 있다. 또한 타당성지수인 $AVERSS$를 활용하여 검사를 개선하는 프로세스를 제시함으로써 $AVERSS$가 수학 인지진단 검사에 활용될 수 있는 또 다른 방향성을 제시하였다는 점에서도 의의를 찾을 수 있다.

주요어 : 인지진단평가, Q행렬 타당화, 확인적

학 번 : 2014-20937

목 차

국문초록 ····· i
목차 ····· iii
표 목차 ····· v
그림 목차 ····· vi

Ⅰ. 서론 ····· 1
1. 연구의 필요성 ····· 1
2. 연구의 목적 및 질문 ····· 5

Ⅱ. 이론적 배경 ····· 6
1. 고전 검사 이론과 문항 반응 이론 ····· 6
2. 인지진단평가 ····· 10
2.1. 인지진단평가의 개념 및 용어 설명 ····· 10
2.2. RSM모형과 Hamming모형 ····· 14
2.3. DINA모형과 Fusion모형 ····· 17
3. Q행렬 타당화 ····· 21
4. Chia-Yi Chiu의 Q행렬 정련방법 ····· 24

Ⅲ. 연구 방법 ····· 26
1. 적용연구 ····· 26
2. 모의실험 ····· 30
3. 타당성 분석 방법 ····· 32
3.1. Chia-Yi Chiu의 Q행렬 정련 방법 ····· 32
2.2. *AVERSS*를 활용한 Q행렬 타당성 분석 방법 ····· 32

4. 연구절차 ····· 35

Ⅳ. 연구결과 ····· 37

1. Chia-Yi Chiu의 Q행렬 정련 방법의 문제점 분석 ····· 37

1.1. Chia-Yi Chiu의 Q행렬 정련 방법 적용 결과 분석 ····· 37

1.2. Chia-Yi Chiu의 Q행렬 정련 방법의 문제점 분석 ····· 40

2. *AVERSS* 타당성 분석 방법 개발 및 적용 결과 ····· 43

2.1. 모의실험을 통한 *AVERSS* 분포 및 컷오프 값 분석 ··· 43

2.2. *AVERSS* 타당성 분석 방법 적용 결과 ····· 47

3. *AVERSS*를 이용한 인지진단 검사의 개선방법 ····· 49

3.1. $AVERSS_j$값이 큰 문항들에 대한 질적 분석 ····· 54

3.2. 방정식과 부등식 검사의 개선 ····· 61

4. 두 타당화 방법 비교 ····· 66

5. 이탈모수가 *AVERSS*에 미치는 영향 ····· 69

Ⅴ. 논의 및 결론 ····· 71

1. 연구결과 요약 및 논의 ····· 71

2. 연구의 제한점 및 제안 ····· 74

참고문헌 ····· 76

[부록] 잘못된 Q행렬에 대한 정보 ····· 81

ABSTRACT ····· 89

표 목 차

<표 Ⅲ-1> 학교별 연구 참여 대상 분포 ····· 26
<표 Ⅲ-2> 일차방정식과 일차부등식 영역의 인지요소 ····· 27
<표 Ⅲ-3> 방정식과 부등식 검사의 Q행렬 ····· 28
<표 Ⅲ-4> 모의실험에 사용된 문항 모수 ····· 31
<표 Ⅳ-1> Q행렬 정련방법 적용 후 정련된 Q행렬 ····· 37
<표 Ⅳ-2> Q행렬 정련방법 적용 전과 후 비교 행렬 ····· 38
<표 Ⅳ-3> ANN모형과 Hamming모형 적용 결과 비교 ····· 41
<표 Ⅳ-4> 검사 문항의 이탈모수에 따른 $AVERSS$의 평균 및 표준편차 ···· 45
<표 Ⅳ-5> 검사 문항의 이탈모수에 따른 $AVERSS$ 컷오프 값 ····· 45
<표 Ⅳ-6> 잘못된 Q행렬에 $AVERSS$ 타당성 분석 방법 적용 결과 ····· 46
<표 Ⅳ-7> 방정식과 부등식 검사 Q행렬의 $AVERSS$와 문항별 $AVERSS_j$ ·· 47
<표 Ⅳ-8> 문항 질적 분석을 통해 발견할 수 있는 2가지 오류 예시 ····· 49
<표 Ⅳ-9> 예시 문항A와 B의 Q벡터 ····· 49
<표 Ⅳ-10> $AVERSS_j$를 기준으로 문항 분류 ····· 54
<표 Ⅳ-11> 방정식과 부등식 평가의 Q행렬에서 인지요소3, 인지요소4와 관련된 부분 ····· 57
<표 Ⅳ-12> 문항 제거 후 검사의 재구성 ····· 63
<표 Ⅳ-13> 검사 1~6의 $AVERSS$ ····· 64
<표 Ⅳ-14> 이탈-추측 상황과 RSS의 관계 ····· 69
<표 Ⅳ-15> 검사 문항의 이탈모수에 따른 $AVERSS$의 평균 및 컷오프 값 · 69

그 림 목 차

[그림 Ⅱ-1] 문항특성곡선 ····· 8
[그림 Ⅱ-2] Q행렬 제작 ····· 11
[그림 Ⅱ-3] 인지진단모형을 적용한 평가 결과 분석 및 활용 ····· 13
[그림 Ⅲ-1] 인지요소 간 위계 구조 ····· 27
[그림 Ⅲ-2] 연구 절차 ····· 36
[그림 Ⅳ-1] Set1의 $AVERSS$ 분포 ····· 43
[그림 Ⅳ-2] Set2의 $AVERSS$ 분포 ····· 44
[그림 Ⅳ-3] Set3의 $AVERSS$ 분포 ····· 44
[그림 Ⅳ-4] $AVERSS$를 이용한 검사의 개선 프로세스 ····· 54
[그림 Ⅳ-5] Chia-Yi Chiu의 Q행렬 정련방법과 $AVERSS$ 타당성 분석 방법의 프로세스 ····· 66
[그림 Ⅳ-6] 검사 문항의 이탈모수에 따른 $AVERSS$의 평균 및 표준편차 ·· 69

Ⅰ. 서론

1. 연구의 필요성

우리나라는 국제 수학-과학 성취도 변화 추이 비교연구(Trends in International Mathematics and Science Study : TIMESS)와 국제 학업 성취도 평가 (Programme for International Student Assessment : PISA)에 참여하여 읽기, 수학, 과학 분야에서 높은 학업 성취도를 보인 반면, 수학에 대한 흥미, 자신감, 가치인식에 관한 정의적 영역의 평가에서는 매우 낮은 수준의 결과를 보이고 있다(최승현 외, 2014). 정의적 특성은 학교 교육의 중요한 목표 중 하나로 인간의 삶의 질과도 매우 밀접한 관계가 있으며, 특히 학습자의 수학에 대한 흥미, 자신감 가치인식 등의 성향은 개인의 사고에 장시간 영향을 끼칠 수 있다(Eggen & Kauchak, 2010; 한혜숙, 최계연, 2011). 이에 따라 우리나라 학생들의 수학에 대한 부정적 인식은 학교를 넘어서 사회적 이슈로 확대되었고, 이를 개선하기 위한 구체적인 교육 방안을 마련하려는 노력이 진행되고 있다.

최승현, 황혜정(2014)은 학생들의 수학에 대한 부정적 인식에 대한 원인 중 하나로 평가에 의한 불안감을 제시하고 있다. 학생들은 점수로 제공되는 평가 결과에 의해 자신감과 유능감을 잃게 되고, 수학이나 평가에 대해 가지는 불안감이 심화된다. 그리고 대다수의 학생들에게 수학의 가치는 대학 진학에 필요한 과목으로 인식되며, 평가는 학생들을 서열화 시키기 위한 수단으로 사용된다. 장상호(2004)는 오늘날 학교 현장에서 이루어지고 있는 평가가 교육 본래의 면모 보다는 더 나은 혜택을 얻기 위한 점수 경쟁을 보여 주고 있다고 말한다. 또한, 김동화(2003)는 우리나라 수학 평가의 문제점으로 객관식 선다형과 단답형 문제풀이에 의한 지필 고사라는 점, 수학과정의 내용보다 회상, 모방과 기계적 암기에 의존하는 것, 학습내용을 이해했는가에 대한결정을 성적순에 따라 정하는 것, 학생들의 인지적 개념구조에 대한 획일적 해석만 허용는 것 등을 제시한다.

우리나라의 2015 개정 교육과정에서는 평가의 목적에 대하여 “평가는

학생의 교육 목표 도달도를 확인하고, 교수·학습의 질을 개을 개선하는데 목적을 둔다”라고 명시되어 있다. 이와 유사한 맥락에서 평가를 교육에 의한 결과물을 측정하고 수량화 하는 것이 아닌 학습 과정 중에 학습 목표로 이끄는 활동으로 이해하는 경향이 생겼는데, Dylan William(2011)은 이러한 평가에 대한 관점의 변화를 검토하고 ‘학습을 위한 평가’의 의미를 명확히 하였다. 그의 논문에서 다루고 있는 학습을 위한 평가에 관한 선행연구들은 학습에 있어서 평가로부터 제공되는 피드백의 영향에 대하여 다루고 있다. Nyquist(2003)는 이 피드백의 정도에 따라서 ‘단지 약한 피드백(Weaker feedback only)’, ‘단지 피드백(Feedback only)’, ‘약한 형성평가(Weak formative assessment)’, ‘보통의 형성평가(Moderate formative assessment)’, ‘강한 형성평가(Strong formative assessment))’ 총 5단계로 형성평가를 분류하였는데, 이러한 분류체계는 평가에 있어서 피드백의 역할의 중요성을 보여준다. Nyquist의 분류 체계에 따르면 통상적인 피드백이 포함된 평가(결과정보 및 정답에 대한 피드백)는 두 번째 단계인 ‘단지 피드백’으로 분류되며, 마지막 단계인 ‘강한 형성평가’단계의 평가는 결과 정보 및 정답에 대한 피드백 뿐 아니라 적당한 설명과 향상을 위해 착수할 구체적인 활동까지 제시하여야 한다.

기존의 평가 체제인 고전검사이론, 문항반응이론 등은 피험자의 현재 능력 수준에 대한 상대적 서열 정보 제공에는 효과적이지만, 교수·학습의 개선을 위한 구체적인 정보 제공이라는 측면에서는 부족하다. 이에 따라 피험자의 강약점을 파악함으로써 구체적인 피드백을 제공할 수 있는 인지진단 모형이 최근 주목받고 있다(이영주, 2014).

인지진단평가는 학생들의 지식구조의 상태와 과정능력을 측정하기 위해 만들어진 평가 방식으로 학생들의 인지적 강점과 약점에 대한 정보를 제공한다(Leighton & Gierl, 2007). 인지진단평가에서는 학습에 필요한 사전지식들과 기능들을 몇 가지 인지요소의 조합으로 보는데, 학생들은 인지진단평가를 통하여 자신이 어떤 인지요소를 숙달하였고, 어떤 인지요소를 숙달하지 못했는지에 관하여 피드백을 받을 수 있다(Jang, 2008). 인지진단평가는 각 문항들과 문항이 측정하고자 하는 인지요소 간의 관계를 토대로

학생들의 각 문항에 대한 응답을 분석하여 개별학생들의 인지상태 프로파일을 제공한다(김희경 외, 2012). 따라서 목적에 부합하는 인지진단 검사가 이루어지기 위해서는 각 인지요소의 숙달을 측정할 수 있는 적절한 문항들이 사용되어야 할 뿐 아니라 문항과 요소의 관계가 올바르게 가정되어야 한다.

인지진단 검사의 Q행렬이란 검사의 문항과 요소간의 관계를 나타내는 행렬로 검사 내용지식을 아는 전문가들이 이론과 그 외 실질적인 고려사항들을 기반으로 문항과 인지요소 사이의 관계를 가정하여 제작한다(Dibello et al., 2007). 그러나 문항과 요소간의 관계성이 분명하게 드러나지만은 않기 때문에 이러한 절차로 결정된 Q행렬이 적절하거나 일관된다고 보장할 수 없으며, 실제 검사의 참된 Q행렬과 다를 수 있다. 그럼에도 불구하고 교육평가에서 이루어지고 있는 많은 인지진단평가에서는 Q행렬의 타당성을 전문가의 검토를 통해 확보하는 것에 그치고 있다.

이에 따라 데이터를 통하여 기존에 제작된 Q행렬의 타당성을 분석하는 방법들에 관한 연구들이 진행되었으나 아직까지 적절한 방법이 충분히 마련되지는 않았다(Feng, 2013). de la Torre(2008)는 베이지안 방법론에 근거하여 MCMC(markov Chia-Yin monte carlo) 추정 알고리즘에 의한 ‘Sequential EM-based δ방법’ 을 제안하고 DINA모형에서 Q행렬의 잘못된 부분을 수정하는데 성공하였으며, de la Torre & Chiu(2015)에서는 이 방법이 G-DINA모형에서 적용 가능함을 증명하였다. 또한, Chiu(2013)는 Residual Sum을 사용하여 데이터를 통해 Q행렬의 항을 개선해나가는 Q행렬 정련 방법을 제안하였다. 그 외에도 De Corlo(2012)의 베이지안 기법을 사용한 방법, Barnes(2010)의 hill-climbing 방법, Liu et al.(2012)의 ‘loss function’ 을 최소화 시키는 방법, Close(2012)의 주성분분석을 통한 타당화 방법 등이 개발되었다.

정확하지 않은 Q행렬은 잘못된 피험자의 능력측정, 잘못된 모델 모수 측정, 좋지 못한 모델 적합성 등의 원치 않는 결과를 초래할 수 있으므로 인지진단평가가 성공적으로 활용되기 위해서는 Q행렬의 타당성을 검증할 수 있는 적절한 방법의 제작이 매우 중요하다(Rupp & Templin, 2008;

Feng, 2013). 본 연구에서는 전문가가 제시한 Q행렬의 타당성을 검사하는 방법과 더 나아가 정확한 Q행렬을 제작하기 위한 방법에 대하여 연구한다.

2. 연구의 목적 및 질문

본 연구의 목적은 수학 인지진단 검사를 제작하는데 필요한 Q행렬의 타당성을 판정하는 양적인 방법을 개발하는 것에 있다. 기존에 연구된 Q행렬 정련방법을 변형하여 Q행렬의 타당도를 나타내는 수치를 개발하고 이를 실제 수학 인지진단 검사를 수행한 자료에 적용하여 볼 것이다. 또한 가상 데이터를 이용하여 개발한 방법의 효율성과 정확도를 평가할 것이다.

구체적인 연구 질문은 다음과 같다.

ⅰ. 기존의 Q행렬 정련방법의 문제점은 무엇이며 이를 보완한 새로운 Q행렬 타당성 분석 방법은 무엇인가.

ⅱ. 새로운 Q행렬 타당성 분석 방법은 기존의 방법과 비교하여 어떤 점에서 향상되었는가.

ⅲ. 수학교육의 인지진단 검사에 새로운 Q행렬 타당성 분석 방법을 어떻게 활용할 수 있는가.

Ⅱ. 이론적 배경

1. 고전검사이론과 문항반응이론

검사이론(test theory)에 따르면 검사(test)란 학생들의 어떤 정신적, 심리적, 인지적 구성체(psychological construct)를 측정하는 것을 의미한다(Crocker & Aljina, 1986). 고전검사이론(Classical test theory, CTT)은 총점에 근거하여 학생들의 능력을 추정하는 이론으로 현재까지도 가장 대중적으로 사용된다. Spearman(1907)은 고전검사이론에 기반을 둔 고전 진점수 모형(Classical true score model)을 고안하였는데, 이 모형에서 검사의 결과로부터 얻은 학생의 관찰점수(X)는 측정 오류가 없을 때의 학생의 점수, 즉 학생의 실제 능력치를 의미하는 진점수(T)와 오차점수(E)의 합으로 표현된다. 여기서 관찰 점수는 하나의 확률 변수로 볼 수 있고, 피험자에 따라 관찰점수의 분포가 달라진다.

$$X = T + E$$

다음은 고전 진점수 모형이 가지고 있는 세 가지의 기본적인 가정이다.

1. 전체 피험자에 대하여 오차점수 평균은 0이다. (즉, $\mu_E = 0$)
2. 전체 피험자에 대하여 진점수와 오차점수의 상관계수는 0이다. (즉, $\rho_{TE} = 0$)
3. 만약 피험자들이 두 개의 서로 다른 시험을 보고, 각각의 시험에 대하여 관찰점수가 서로 다른 분포로부터 임의로 결정된다면, 두 오차점수 사이의 상관계수는 0이다. (즉, $\rho_{E_1E_2} = 0$)

또한, 고전 진점수 모형은 검사 결과의 일관된 정도를 나타내는 신뢰도(reliability)를 설명하였다는 것에도 큰 의의를 가진다(Crocker & Aljina,

1986). 고전 진점수 모형에서 검사의 신뢰도 지수(reliability index)와 신뢰도 계수(reliability coefficient)는 다음과 같이 정의되며, 일반적으로 검사의 신뢰도를 말할 때에는 신뢰도 계수를 사용한다.

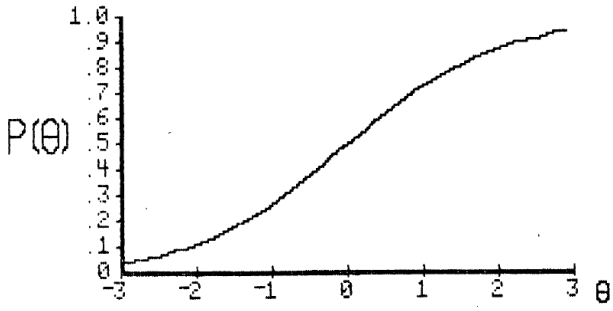
$$\text{신뢰도 지수 : } \rho_{XT} = \frac{\sigma_T}{\sigma_X}$$

$$\text{신뢰도 계수 : } \rho_{X_1X_2} = \frac{\sigma_T^2}{\sigma_X^2}$$

더불어 문항난이도, 문항변별도 등의 개념도 고전검사이론에서 출발하였다(Devellis, 2006). 문항난이도는 어떤 문항이 어려운 정도를 뜻하므로 전체 피험자 수에 대한 해당문항을 맞춘 피험자의 수의 비율로 구할 수 있다. 그리고 어떤 문항이 검사를 잘 표현하는지를 나타내는 문항변별도는 해당 문항과 전체 문항들 사이의 상관계수로 구할 수 있다.

위와 같이 고전검사이론에서는 검사를 시행한 후 얻은 결과를 통하여 문항난이도와 문항변별도를 측정하는 것이 가능하다. 그러나 문항난이도와 문항변별도가 다른 어떤 두 문항에 대하여 능력치가 다른 피험자들이 어떻게 반응할지 예측하는 것은 불가능하다. 평가 제작에 있어서 각 문항에 대하여 피험자들이 어떻게 반응할지를 예측하는 것 매우 중요한 정보이다. 이러한 관점에서 문항 자체의 특성을 고려하여 피험자들의 문항에 대한 반응을 예측할 수 있는 문항반응이론(Item response theory, IRT)은 하나의 해결책이 될 수 있다(Crocker & Aljina, 1986).

문항반응이론 역시 고전검사이론과 마찬가지로 피험자의 능력치(latent trait)를 어떤 측정 스케일(scale of measurement)를 사용하여 측정하는 것에 그 목적이 있으며, 그 핵심은 문항마다 고유하게 가지는 문항특성곡선(Item characteristic curve)이라 할 수 있다(Baker, 2001). 문항특성곡선은 피험자의 능력에 따라 문항에 어떻게 반응할지를 나타내는 곡선으로 대개 [그림 Ⅱ-1]과 같은 형태를 가지며, 2차원 좌표공간에서 x축은 피험자의 능력치(θ)를 y축은 문항을 맞힐 확률($p(\theta)$)을 나타낸다.



[그림 Ⅱ-1] 문항특성곡선 출처 : Baker(2001)

문항반응이론에서는 확률함수 $p(\theta)$의 구조와 문항모수의 개수에 따라 여러 가지 모형이 존재하는데, 가장 보편적으로는 로지스틱 모형이 사용된다(Baker, 2001). 1모수 로지스틱 모형은 문항 난이도(β)만을 모수로 가지는 모형이고, 2모수 로지스틱 모형은 문항난이도와 문항 변별도(α)를 모수로 가지며, 3모수 로지스틱 모형은 위 두 개의 문항모수 외에도 문항 추측도(γ)를 사용한다. 각각의 로지스틱 모형을 나타내는 식은 다음과 같다.

1모수 로지스틱 모형 : $p(\theta) = \dfrac{1}{1+e^{-(\theta-\beta)}}$

2모수 로지스틱 모형 : $p(\theta) = \dfrac{1}{1+e^{-\alpha(\theta-b)}}$

3모수 로지스틱 모형 : $p(\theta) = \gamma + (1-\gamma)\dfrac{1}{1+e^{-a(\theta-b)}}$

문항반응이론의 문항모수들은 피험자의 집단과는 독립적인 문항 고유의 값이며, 문항난이도는 문항특성곡선에서 문항을 맞힐 확률이 0.5가 될 때의 능력치를, 문항변별도는 이 지점에서의 곡선의 기울기를, 문항추측도는 능력이 전혀 없는 학생($\theta \to -\infty$)이 정답을 맞힐 확률을 뜻한다.

문항반응이론에서 검사 문항들에 대한 피험자의 응답을 통하여 피험자의 능력을 추정할 때에는 최대우도 측정(Maximum likelihood estimation)이 주로 사용된다(Baker & Kim, 2004). 우도(likelihood)란 피험자의 능력(θ)에 따라 실제 피험자의 응답처럼 응답할 확률을 의미하는데, 이는 문항별 확

률함수 $p(\theta)$를 활용하면 θ에 대한 식으로 표현할 수 있다. 최대우도 측정에서는 우도가 최대가 되게 하는 값을 피험자의 능력 θ로 택하며, 이 과정에서 뉴튼방법(Newton’s method)등의 추정 방법이 사용될 수 있다.

2. 인지진단평가 (Cognitive Diagnostic Assessment)

평가의 목적이 피험자의 능력을 점수화하는 것에서 피험자의 현 상태를 진단하여 더 나은 교수·학습으로 나아갈 수 있도록 돕는 것으로 확장됨에 따라 개별 피험자들의 인지적 강약점에 관한 정보를 제공하는 인지진단모형이 주목받고 있다(Dylan Wiliam, 2007; 이영주, 2014). 인지진단평가는 기존의 측정 이론에 인지심리 이론을 도입한 것으로, 개별학생의 학업성취 수준을 정교한 프로파일 형태로 제공하여 학생들이 습득해야 하는 세부적 인지요소의 숙달 정도 및 학생의 학습 개선을 위한 유용한 정보를 제공하는 방향으로 분석이 이루어진다(Leighton & Griel, 2008; 김희경, 김부미, 2013).

2.1. 인지진단평가의 개념 및 용어 설명

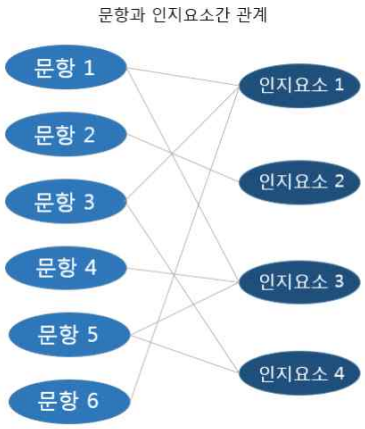
• **인지요소 (attribute, skill)** : 인지요소란 학생이 문항을 해결하는데 필요한 능력, 기능, 지식, 인지과정 등을 통합하여 의미하는 말이다(Tatsuoka, 1983). 각각의 문항들은 한 가지 혹은 여러 가지 인지요소를 필요로 하는데, 예를 들어 '$2+3\times5=?$'이라는 문제는 '덧셈을 할 수 있느냐', '곱셈을 할 수 있느냐', '덧셈과 곱셈의 순서를 알고 있느냐' 등의 인지요소와 관련되어 있다. 또한 위의 예시와 같은 어떤 수학적 내용을 알고 있는지에 관한 지식 외에도 문제를 푸는 과정에서 필요한 기능을 인지요소로 볼 수도 있다. 김희경 외(2012)는 2011년 학업성취도 수학 평가에서 적용 가능한 인지요소로 '직관적으로 추론하기', '문제 상황 해석하기' 등 총 14가지의 인지요소를 제시하였는데, 이것들은 모두 과정중심의 인지요소라 할 수 있다.

• **Q행렬 (Q-matrix)** : Q행렬은 인지요소와 문항이 주어졌을 때, 각 문항과 각 요소간의 관계를 나타내는 행렬로 Tatsuoka(1983)에 의해 처음으로 제

시되었다. 일반적으로 Q행렬에서 각 행은 서로 다른 문항을 나타내며, 각 열은 서로 다른 인지요소를 나타낸다. 이때, i번째 행에 해당하는 문항을 해결하는데 j번째 열에 해당하는 인지요소가 요구되는 경우 Q행렬의 (i,j)-원소의 값은 1로 나타내고, 이 인지요소가 요구되지 않는 경우는 0으로 나타낸다. 즉, Q행렬은 다음과 같이 나타내어진다.

$$Q=(q_{ij}).\ \text{단,}\ \begin{cases} q_{ij}=1 & (\text{문항 } i\text{를 맞히는데 인지요소 } j\text{가 요구될 때}) \\ q_{ij}=0 & (\text{문항 } i\text{를 맞히는데 인지요소 } j\text{가 요구되지 않을 때}) \end{cases}$$

예를 들어 [그림 Ⅱ-2]에 나타나 있듯이 여섯 개의 문항이 네 개의 인지요소와 관련되며 그 관계가 왼쪽의 다이어그램과 같은 경우, 오른쪽 도표와 같이 Q행렬을 제작할 수 있다.



Q행렬

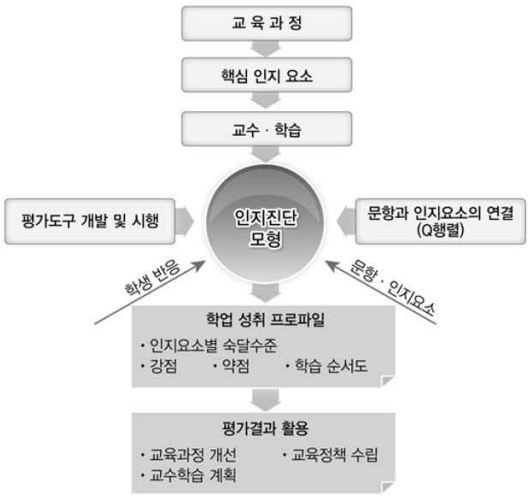
	인지1	인지2	인지3	인지4
문항1	1	0	1	0
문항2	0	1	0	0
문항3	1	0	0	1
문항4	0	0	1	0
문항5	0	0	1	1
문항6	1	0	0	0

[그림 Ⅱ-2] Q행렬 제작 출처: 김희경 외(2012)

• **응답벡터 (response pattern)** : 응답벡터란 피험자의 각 문항에 대한 응답을 하나의 벡터로 표현한 것으로 응답벡터의 길이는 문항의 개수와 같고 각각의 원소는 0과 1로 구성된다. 이때 응답벡터의 i번째 원소가 1일 경우 피험자가 i번 문항의 정답을 맞혔음을, 0일 경우 정답을 맞히지 못했음을 의미한다.

• **피험자의 인지요소벡터 분류** : 인지요소벡터(attribute vector)란 인지요소의 조합으로 가능한 것들을 벡터형태로 나타낸 것으로, 그 길이는 인지요소의 개수와 같다. 예를 들어 인지요소의 개수가 K개라고 하면, 이 K개의 인지요소를 조합한 인지요소벡터는 $(0,0,...,0)$부터 $(1,1,...,1)$까지 최대 2^K개이고, 위계관계에 따라 개수가 줄어들 수도 있다. 인지진단검사에서 피험자의 능력은 피험자가 숙달한 인지요소의 조합으로 표현되기 때문에, 피험자의 능력 추정을 또 다른 말로 피험자의 인지요소벡터 분류라고 한다. 즉, 피험자의 능력을 추정한다는 것은 이 2^K개의 인지요소벡터 중에 피험자가 숙달한 인지요소의 조합이 어느 인지요소벡터에 해당하는 지를 찾는 과정이라 할 수 있다.

• **기대반응벡터 (expected response pattern)** : 기대반응이란 각각의 인지요소벡터에 속한 피험자가 해당 인지요소벡터의 인지요소의 조합에 따라 각 문항에 대하여 어떻게 이상적으로 반응할지를 나타낸 것으로 만약 피험자가 해당 문제를 푸는데 필요한 모든 인지요소를 습득했다면 1의 값을, 그렇지 못하다면 0의 값을 가진다. 이때 한 피험자의 모든 문항에 대한 기대반응을 벡터 형태로 나열한 것을 기대반응벡터라 한다. 예를 들어 6개의 문항으로 이루어진 검사를 통해 4개의 인지요소에 대한 피험자들의 능력을 측정하는 상황에서 Q행렬이 [그림 Ⅱ-2]와 같다고 가정하자. 이때 피험자가 속할 수 있는 인지요소벡터로는 $(0,0,0,0)$, $(1,0,0,0)$, $(0,1,0,0)$, $\cdots$, $(1,1,1,1)$의 총 16개의 인지요소벡터가 존재 한다. 대부분의 인지진단모형에서는 한 문항이 측정하고자 하는 인지요소들을 모두 숙달하였을 때 해당 문항을 풀 수 있다고 가정한다. 예를 들어, 인지요소벡터 $(1,1,0,0)$에 속한 피험자들은 이상적으로 인지요소1과 인지요소2만을 필요로 하는 문항2나 문항6은 맞힐 수 있는 반면, 인지요소3이나 인지요소4를 필요로 하는 문항1, 문항3, 문항4, 문항5는 맞힐 수 없다. 이럴 경우 인지요소벡터 $(1,1,0,0)$의 기대반응벡터는 $(0,1,0,0,0,1)$이라 할 수 있다.



[그림 Ⅱ-3] 인지진단모형을 적용한 평가 결과 분석 및 활용

출처 : 김희경 외 (2012)

[그림 Ⅱ-3]은 인지진단 평가의 절차를 나타낸다. 먼저 교육과정에 따라 학생들이 습득해야 하는 핵심 인지요소가 결정되면, 이에 맞는 교수학습이 이루어지고 이를 평가할만한 평가도구가 개발된다. 평가는 각 인지요소의 숙달 정도를 파악할 만한 문항들로 이루어지며, 인지요소와 문항이 모두 준비되면 이들의 관계를 나타내는 Q행렬이 제작된다. 검사가 시행된 후 피험자들은 문항에 대한 응답을 토대로 각 인지요소의 숙달 프로파일을 결과로서 제공받게 되고, 이는 학생의 학습과 교사의 교수법 개선에 활용할 수 있다.

이때 피험자들의 각 문항에 대한응답을 토대로 인지요소벡터를 분류하는 과정에서 인지진단모형이 필요하다. 인지진단평가의 모형으로는 대표적으로 RSM모형(Tatsuoka,1985), DINA모형(Junker & Sijtsma, 2001), Fusion모형(Hartz, 2002) 등이 있고, 그 외에도 여러 가지 모형들이 존재하며, 연구자는 검사 상황에 맞는 모형을 사용하여 검사 결과를 분석할 수 있다. 모든 인지진단모형은 피험자의 각 문항에 대한 정답여부와, 문항-요소간의 관계를 통하여 피험자의 각 인지요소별 숙달여부를 추정하지만, 모형에 따라 그 추정과정에 차이가 존재한다.

2.2 RSM모형과 Hamming모형

RSM모형(Rule Space Model)은 학생들의 오류 혹은 오해를 분석하는 규칙-접근 방법(rule-assessing method)과 개개인의 특성 접근(personal-index approach)을 결합한 모델로 K.Tatsuoka에 의해 개발된 초창기 인지진단모형이다(Tatsuoka, 1983). 초기의 RSM모형은 학생들이 문제를 풀 때 발생할 수 있는 여러 가지 오류 규칙(rule)에 대하여 분석하는 형태였으나, 인지요소나 Q행렬 등의 개념이 발전함에 따라 다음과 같이 정립되었다.

RSM모형의 첫 번째 과정은 피험자들의 인지요소벡터를 규명하고, 각 인지요소벡터의 기대반응벡터를 구하는 것이다. Tatsuoka는 어떤 문제를 맞히기 위해서는 이 문제가 포함하는 인지요소들을 모두 숙달해야 하고, 반대로 피험자가 숙달한 인지요소들만으로 이루어진 문제는 맞힐 수 있다고 가정하였다. 그는 인지요소벡터와 기대반응벡터 사이의 위와 같은 관계를 나타내는 함수를 Boolean Descriptive Function(BDF)이라 명명하였다(Tatsuoka,1991).

피험자들의 응답벡터와 기대반응벡터는 일치하지 않을 수 있다. 따라서 응답벡터와 가까운 기대반응벡터를 찾아주는 과정이 필요하다. 이를 위해 RSM모형에서는 먼저 피험자들의 응답벡터와 기대반응벡터를 규칙공간(rule space)이라는 공간에 매핑(mapping)한다. 이 Rule space는 IRT를 통해 얻을 수 있는 피험자들의 능력 변수 θ와 기대반응벡터의 예외성(unusualness)을 나타내는 변수 ζ, 그리고 문항의 부분집합 상에서 기대반응벡터의 예외성을 다루는 변수 ζs, 총 3개의 변수를 축으로 하는 3차원 공간이다. (θ와 ζ만을 축으로 하는 2차원 공간에 매핑하는 것도 가능하지만, 3차원 공간을 사용하는 것이 피험자들의 인지요소벡터 분류에 더 이 용이하다; Tatsuoka, 1996) Tatsuoka는 피험자들의 응답벡터가 기대반응벡터와 같은 경우는 매우 드물다고 생각하였고, 응답벡터를 'fuzzy response pattern'으로 취급하였다(Tatsuoka & Tatsuoka, 1987). 그는 규칙공간상에서 fuzzy response pattern에 대응되는 점들이 기대반응벡터에 대응되는 점을 중심으로 하는 확률타원(probability ellipse) 주변에 떼 지어 몰려다닌다고(swarm around)

표현하였다.

마지막 과정은 피험자의 응답벡터와 각 인지요소벡터의 기대반응벡터의 Rule space상에서 Mahalanobis Distances(D^2)을 측정하여 피험자의 응답벡터와 가장 가까운 확률타원을 찾는 것이다. 그리고 가장 가까운 확률 타원에 해당되는 인지요소벡터를 피험자의 인지요소벡터로 택한다. 이 과정에서 오류를 최소화하기 위해 베이즈 의사결정 규칙(Bayes' decision rule)이 사용된다.

RSM모형은 2001년 10월부터 PSAT/NMSQT의 분석에 활용되었고 (Milewski & Baron, 2002), 그 외에도 SAT (Tatsuoka, 1993; Guerrero, 2001), TOEFL (Scott, 1998), TOEIC (Buck & Tatsuoka, 1998), NAEP Science Assessment (Yepes-Baraya et al., 1998), GRE-Q (Tatsuoka & Gallagher, 1998; Tatsuoka & Boodoo, 2000), TIMSS 및 TIMSS-R (Tatsuoka et al., 2004; Xin et al., 2004) 등의 몇 가지 주요 검사의 분석에 활용되었다(Dogan & Tatsuoka, 2007).

RSM모형과 같이 인지요소 벡터의 분류에 거리 개념을 사용하는 모형으로는 Hamming모형이 있다. Hamming모형은 Chiu & Douglas(2013)에 의해 개발된 비모수 모형으로피험자의 문항에 대한 실제 응답벡터와 기대반응벡터 사이의 Hamming distance를 최소로 만드는 인지요소벡터를 피험자의 능력으로 택한다. 문항의 개수가 총 J개이고, 피험자i의 문항 j에 대한 응답을 Y_{ij}, 기대반응을 n_{ij}라 할 때, Hamming distance는 다음과 같다.

$$d_H = \sum_{j=1}^{J} |Y_{ij} - n_{ij}|$$

Hamming모형에서는 모든 인지요소벡터에 대하여 기대반응벡터와 피험자의 응답벡터사이의 거리인 d_H를 구하고, d_H값이 가장 작을 때의 인지요소벡터를 해당 피험자의 능력으로 결정한다. 따라서 문항과 인지요소의 관계를 나타내는 Q행렬만 주어지면 문항 모수의 추정과정 없이 학생들의 인지요소벡터를 분류할 수 있다. 비모수 모형은 데이터의 규모가 작을 때,

반복적인 추정이 일어나야 하는데 컴퓨터의 사양이 이를 받쳐주기 어려울 때 등 문항모수의 측정이 어려울 경우에도 사용이 가능하다는 강점을 가진다. 또한, 굳이 문항모수의 측정이 어려운 경우가 아니라 하더라도 가장 간단하게 피험자의 능력을 유추해보기 위해 사용되는 방법이 Hamming모형이라 할 수 있다.

Chiu & Douglas(2013)에서는 Hamming 모형을 두 가지 방법으로 확장하였는데, 첫 번째는 Hamming distance(d_H) 대신 문항별 가치를 고려한 weighted Hamming distance (d_{wH})를 사용하는 방법이다. 아래 식에서 $\overline{p_j}$는 피험자들 중에서 문항 j를 맞힌 피험자의 비율을 나타낸다.

$$d_{wH} = \sum_{j=1}^{J} \frac{1}{\overline{p_j}(1-\overline{p_j})} |Y_{ij} - n_{ij}|$$

두 번째 방법은 다음절에서 설명할 DINA모형의 아이디어를 적용하여 해당 문항을 풀 수 있는 인지요소를 모두 가지고 있지만 틀린 경우인 slip 상황과, 부족한 인지요소가 있음에도 불구하고 맞힌 경우인 guess상황을 고려한 penalized Hamming방법이다. 이 방법에서는 guessing에 해당하는 가중치 w_g와 slip에 해당하는 가중치 w_s를 사용하여 거리를 나타낸다.

$$d_{pH} = \sum_{j=1}^{J} w_g I[Y_{ij} = 1] |Y_{ij} - n_{ij}| + \sum_{j=1}^{J} w_s I[Y_{ij} = 0] |Y_{ij} - n_{ij}|$$

이때, w_g와 w_s를 문항별로 다르게 하여 문항별 w_{gj}와 w_{sj}를 사용하면 penalized Hamming 방법을 더욱 일반화 수 있다.

이와 같이 확장된 Hamming 모형들 역시 기존 Hamming 모형과 마찬가지로 기대반응벡터와 피험자의 응답벡터 사이의 거리인 d_{wH} 혹은 d_{pH}를 최소화 시키는 방법으로 피험자의 인지요소벡터를 분류한다.

2.3 DINA모형과 Fusion모형

위에서 소개한 모형들이 인지진단모형의 초기형태라면, 현재 가장 널리 활용되는 인지진단모형으로는 DINA모형과 Fusion모형을 들 수 있다. DINA모형은 Haertel(1984)에 의해 RLCM이라는 이름으로 처음 제시가 되었고, Junker & Sijtsma(2001)에 의해 이탈(slip)모수와 추측(guess)모수를 가지는 DINA모형(Deterministic Inputs, Noisy-And-Gate)으로 정립되었다. 여기서 이탈모수는 문항을 맞히는데 필요한 모든 인지요소를 습득한 피험자가 문항을 맞히지 못할 확률을 나타내고, 추측모수는 문항을 맞히는데 필요한 인지요소들 중에서 하나라도 습득하지 않은 피험자가 문항을 맞힐 확률을 나타낸다. 문항 j의 이탈모수를 s_j, 추측모수를 g_j라 할 때(단, $1-s_j > g_j$), s_j와 g_j는 다음과 같이 정의할 수 있다.

$$s_j = P(X_{ij} = 0 \mid \eta_{ij} = 1)$$
$$g_j = P(X_{ij} = 1 \mid \eta_{ij} = 0)$$

여기서 X_{ij}는 피험자i의 문항j에 대한 가상응답을, η_{ij}는 피험자 i가 문항 j를 푸는데 필요한 모든 인지요소를 숙달했는지 여부를 나타낸다. 만약 α_{ik}를 피험자 i의 인지요소 k 숙달 여부로 정의하고, Q행렬의 (j,k)항을 q_{jk}로 표기한다면, η_{ij}는 $\prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$ (K:인지요소 개수)라고 표현할 수 있다. 또한 이렇게 정의한 α_{ik}를 벡터형태로 나열한 $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iK})$는 피험자$i$의 인지요소벡터를 나타내는 형식으로 사용가능하다. 이때, DINA모형은 피험자 i가 문항j를 맞힐 확률을 다음과 같이 정의한다.

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

피험자의 능력 추정방법으로는 최대우도측정(Maximum likelihood estimation)이 가장 보편적으로 사용된다. 먼저 피험자의 응답벡터가 주

어졌을 때, 각 인지요소벡터에 대하여 해당 응답벡터가 나올 우도(likelihood)를 계산한다. 우도는 문항별 독립이 가정된 상황에서, 각 문항에 대해 실제 응답같이 반응할 확률을 모두 곱하여 나타낼 수 있다. 최대우도측정은 우도가 가장 높은 인지요소벡터를 피험자의 인지요소벡터로 채택하는 방법이다. 문항의 개수를 J개라 하고, 피험자 i의 실제 응답벡터를 Y_i라 할 때, 우도를 수식으로 나타내면 다음과 같다.

$$P(Y_i|\theta) = \prod_{j=1}^{J} [\{(1-s_j)^{\eta_{ij}} g_j^{(1-\eta_{ij})}\}^{y_{ij}} \{1-(1-s_j)^{\eta_{ij}} g_j^{(1-\eta_{ij})}\}^{1-y_{ij}}]$$

DINA모델명의 And-Gate는 한 문항을 풀기 위해서는 문항과 관련된 모든(and) 인지요소를 숙달해야 함을 의미한다. DINA모형과 유사한 DINA모형은(Deterministic Inputs, Noisy-Or-Gate) 문항별로 이탈모수, 추측모수를 가진다는 점에서 DINA모형과 흡사하나 각 문항에 관련된 인지요소들 중에서 하나만 숙달하면 해당 문항을 풀기에 충분하다고 가정한다(Templin & Henson, 2006). 즉, DINO모형에서 피험자 i가 문항j를 맞힐 확률은 DINA모형과 동일한 식으로 표현되지만, η_{ij}의 값이 $1-\prod_{k=1}^{K}(1-\alpha_{ik})^{q_{jk}}$ (K:인지요소 개수)으로 달라진다.

DINA모형이나 DINA모형에서는 만약 어떤 문항을 푸는데 5개의 인지요소가 필요하다면, 이 중에 1개의 인지요소만을 숙달한 피험자와 4개의 인지요소를 숙달한 피험자가 이 문항을 맞힐 확률이 동일한다. 실제 상황에서는 그렇지 않을 경우가 종종 있는데, 이럴 경우 문항별 이탈모수, 추측모수가 아닌 요소별 이탈모수와 추측모수를 가지는 NIDA모형(Noisy Input, Deterministic-And-Gate)이나 NIDO모형(Noisy Input, Deterministic-Or-Gate)을 사용할 수 있다. NIDA모형은 어떤 문제를 풀기 위해서는 관련된 모든 인지요소를 습득해야 함을 가정한다는 점에서 DINA모형과 공통점을 가진다. NIDA모형에서 인지요소 k의 이탈모수를 s_k, 추측모수를 g_k라 할 때 피험자 i가 문항 j를 맞힐 확률은 다음과 같다(Junker & Sijtsma, 2001).

$$P(X_{ij}=1|\alpha_i)=\prod_{k=1}^{K}[(1-s_k)^{\alpha_{ik}}g_k^{(1-\alpha_{ik})}]^{q_{jk}}$$

NIDO모형은 어떤 문항을 푸는데 관련된 인지요소 중에서 하나만을 숙달해도 문항을 풀기에 충분한 상황에서 사용되며, 위의 세 모형과는 다르게 이탈모수, 추측모수가 아닌 요소별 intercept모수($\lambda_{0,k}$), slope모수($\lambda_{1,k}$)를 사용한다(Templin, 2006). NIDO모형에서 피험자 i가 문항 j를 맞힐 확률은 다음과 같다.

$$P(X_{ij}=1|\alpha_i)=\frac{\exp(\sum_{k=1}^{K}(\lambda_{0,(k)}+\lambda_{1,(k)}\alpha_{ik})q_{jk})}{1+\exp(\sum_{k=1}^{K}(\lambda_{0,(k)}+\lambda_{1,(k)}\alpha_{ik})q_{jk})}$$

Fusion모형은 NIDA모형의 확장된 형태로 피험자의 고유능력이 고려된 모델이다. NIDA모형은 요소별로 이탈모수(s_k)와 추측모수(g_k)를 가지는데 이때 문항에 의한 차이는 고려되지 않았다. 즉, 두 문항이 같은 인지요소 조합으로 이루어져 있다면, 이 문항들은 정확히 같은 확률함수를 가진다. Dibello et al.(1995)와 Maris(1999)는 이를 확장하여 문항별로 각각 다른 이탈모수(s_{jk})와 추측모수(g_{jk})를 가지는 모형을 만들었고, 이것이 최종적으로 재모수화된 것이 Hartz(2002)의 RUM모형(Fusion모형)이다(Dibello et al., 2007).

π_{jk}를 문항 j를 풀 때 인지요소 k를 숙달한 학생이 이를 정확히 사용할 확률이라고 정의하자. (즉, $\pi_{jk}=1-s_{jk}$) 이 때 $\pi_j{}^*=\prod_{k=1}^{K}\pi_{jk}^{q_{jk}}$를 Q행렬에 기반을 둔 문항 j의 난이도 모수라 할 수 있는데, 이는 문항 j를 푸는데 요구되는 모든 인지요소를 숙달한 학생이 이를 정확히 사용하여 문항 j의 정답을 맞힐 확률을 뜻한다. 또, r_{jk}를 인지요소 k를 숙달하지 못한 학생이 문항 j에 이를 우연히 옳게 적용할 확률(즉, $r_{jk}=g_{jk}$)로 정의하면, $r_{jk}{}^*=\frac{r_{jk}}{\pi_{jk}}$는 Q행렬에 기반을 둔 문항 j의 변별도 모수라고 볼 수 있

다. 이때 인지요소벡터 α_i에 속하는 피험자 i가 문항 j를 맞힐 확률은 다음과 같다.

$$P(X_{ij}=1|\alpha_i,\theta_i)=\pi_j{}^*\prod_{k=1}^{K}(r_{jk}{}^*)^{(1-\alpha_{ik})\times q_{jk}}P_{c_j}(\theta_i)$$

여기서 θ_i는 Q행렬에 명시된 인지요소를 제외한 피험자 i의 기타 능력을 뜻하는 연속적 모수이고, c_j는 문항 j가 이 잔여능력에 얼마나 의존하고 있는지를 나타내는 수치이다. $P_{c_j}(\theta_i)$는 피험자 i가 Q행렬에 명시되지 않은 기타능력을 문항 j에 올바르게 적용할 확률을 의미하는데 로지스틱 모형을 사용하여 $P_{c_i}(\theta_j)=\dfrac{1}{1+\exp(-1.7(\theta_j+c_i))}$ 으로 정의한다.

Fusion 모형의 장점은 문제해결을 위해 필요한 요소들이 모두 Q행렬에 명시되지 않을 수 있다는 가능성을 고려하였다는 것이다(Dibello et al., 2007). 즉, Fusion모형에서는 Q행렬의 불완전성(incompleteness)을 해결하기 위해 피험자의 기타능력을 나타내는 연속적 모수 θ_i를 도입하였다. 또한, 문항 • 요소별 모수 π_{jk}, r_{jk}를 사용함으로써 기존 DINA모형, NIDA모형의 확장에 성공하였다.

3. Q행렬 타당화(Q-matrix Validation)

정확한 Q 행렬을 얻지 못한다면 인지진단 검사를 수행한 결과의 해석이 타당하지 않을 수 있다(Rupp & Templin, 2008). Q행렬의 오류(misspecification)는 정확하지 않은 문항모수의 측정과 좋지 않은 모델핏을 발생시키고, 이는 피험자들의 인지요소벡터가 잘못 분류되는 것으로 이어질 수 있기 때문이다(Henson & Templin, 2009; Rupp & Templin, 2008; Liu et al., 2012). 따라서 Q행렬이 얼마나 정확하게 만들어졌는지는 검사의 타당도와 직결되는 부분이라 할 수 있다. 이와 관련하여 정확한 Q행렬을 만들기 위한 연구가 최근 들어 이루어지기 시작했다.

검사의 타당도(validity)란 검사의 목적에 비추어 보았을 때 검사 결과의 해석을 증거 혹은 이론들이 지지하는 정도를 의미한다(Crocker & Algina, 1986; Wainer & Braun, 1988). 따라서 어떠한 검사가 타당하다는 것은 그 검사의 결과가 보다 유의미하고 적절하며, 의사결정에 사용될 수 있음을 뜻한다. 검사의 타당도를 평가하는 대표적인 방법으로는 요인분석(factor analysis)이 있다. 요인분석이란 결과 혹은 관심이 있는 대상 $Y_1, Y_2, \ldots, Y_l$들이 보이지 않는 요인 $F_1, F_2, \ldots, F_k$ $(k < l)$들과 어떠한 일차적 관계가 있는지를 조사하는 분석법이다. 즉, 요인분석에서는 결과에 영향을 주는 요인들이 얼마나 있는지를 찾아내고, 그 요인들이 어떻게 그룹 지어져 있는지를 찾아낸다. 요인분석은 확인적 요인분석(Confirmatory Factor Analysis)과 탐색적 요인분석(Exploratory Factor Analysis), 크게 두 가지 방법으로 나뉜다(Thompson, 2004). 탐색적 요인분석이란 보이는 결과들과 보이지 않는 요인들 사이의 구조를 온전히 데이터를 통하여 얻어내는 방법이다. 반면, 확인적 요인분석에서는 검증된 이론들을 사용하여 요인들의 구조를 사람이 설정한 뒤, 이 구조를 뒷받침할만한 증거들을 조사하여 설정된 구조의 정당성을 확인한다.

Q행렬을 구축하는 가장 일반적인 방법은 해당 교과의 내용 전문가들이 문항과 인지요소 간의 관계를 제시하는 것이다(김희경 외, 2012). 그러나 이렇게 제안된 Q행렬은 제한된 인원의 전문가에 의해 제작되므로 절대적

일 수 없으며, 이론적 관점에 따라서 어떤 문항이 어떤 요소를 포함하고 있는지, 그렇지 않은지가 다를 수 있기 때문에 일관성을 보장할 수 없다. 그러므로 연구자들이 제시한 Q행렬이 반드시 적절한 Q행렬이라고 주장하기 어렵다. 이때 Q행렬이 가질 수 있는 오류(misspecification)로는 문항과 관련된 인지요소를 관련되지 않다고 표기한 경우(underspecified), 관련이 없는 인지요소를 관련된다고 표기한 경우(overspecified), 그리고 이 두 가지가 복합적으로 발생한 경우가 있다(Feng 2013). 이러한 오류를 개선하기 위해 Q행렬은 전문가의 판단과 통계적 분석에 따른 수정과정을 거치게 된다(김희경 외, 2012).

이에 따라 양적 분석을 통해 정확한 Q행렬을 찾아나가는 방법들이 제안되기 시작했고, 이는 Q행렬의 타당도를 수치화시켜서 기존의 Q행렬을 개선해나가는 Q행렬 타당성 분석 방법(혹은, Q행렬 타당화 방법: Q-matrix validation method)에 관한 연구들로 이어졌다(Feng, 2013; Torre, 2008; 2009; Torre & Chiu 2010; 2015; De carlo, 2011; 2012; Liu et al, 2012; Chiu, 2013; Desmarais & Naceur, 2013). Q행렬의 타당성을 분석함에 있어서 필수적인 것은 타당성지수(validation index)로 이는 Q행렬이 타당한 정도를 나타내는 어떤 기준이 되는 수치를 나타내며, 보통 피험자의 응답 자료를 토대로 양적분석을 통해 얻어진다.

de la Torre는 베이지안 방법론에 근거하여 MCMC(markov Chain monte carlo) 추정 알고리즘을 사용하는 'Sequential EM-based δ방법'을 제안하였는데, 이 방법은 DINA모형과 G-DINA모형에서 Q행렬의 잘못된 부분을 수정하는데 성공하였다(Torre, 2008; Torre & Chiu, 2015). 또한, Chiu(2013)에서는 피험자의 기대응답벡터와 실제 응답벡터 사이의 RSS(residual sum of square)를 최소화 시키는 방향으로 Q행렬을 개선하는 Q행렬정련방법(Q-matrix refinement method)를 소개하고, 이론적으로 증명하였다. De corlo(2012)는 베이지안 기법을 사용하여 Q행렬의 불명확한 부분을 찾아내는 방법을 고안하였으며 Liu et al.(2012)는 'objective(loss) function'을 최소화 시키도록 Q행렬을 개선하는 방법을 제안하였다. 또한 Romero et al.(2014)에서는 LSDM모형(Least Squares Distance Model)에서 사용가능한

두 개의 타당성지수(MAD, LSD)를 개발하였고, Desmarais & Naceur(2013)에서는 응답벡터행렬(각 열이 한명의 피험자의 응답벡터)을 Q행렬과 피험자들의 인지요소벡터행렬(각 열이 한명의 피험자의 인지요소벡터)로 분해하여 분석하는 방법인 ALS(Alternative Least-Square Factorization)방법을 소개하였다.

이 Q행렬 타당화 방법들은 모두 기존에 제작된 Q행렬을 검사 데이터의 통계적 분석을 통하여 정련(refinement)해나가는 방법으로 그 과정은 다음과 같다. 먼저 전문가가 문항-요소 관계를 가정하여 제작한 Q행렬을 초기 Q행렬로 설정하고, 검사 데이터로부터 이 Q행렬의 타당성 지수를 계산한다. 그 다음 Q행렬의 일부를 변형시켜가면서 타당성 지수가 더 좋아지는 방향으로 Q행렬의 원소들을 변형하는 정련과정을 거친다. 만약 타당성 지수가 작을수록 Q행렬이 타당함을 의미하면 타당성 지수가 작아지는 방향으로, 타당성 지수가 클수록 Q행렬이 타당함을 의미하면 타당성 지수가 커지는 방향으로 Q행렬을 정련하여 최종적으로 가장 좋은 타당성 지수를 가지는 Q행렬을 얻는다. 그러므로 Q행렬 정련방법들은 검사 데이터에 의존하여 Q행렬의 원소를 찾아나가는 탐색적 접근의 방법이라 할 수 있다.

위와 같이 다양한 논문에서 여러 가지 타당성 분석 방법들이 소개된 가운데 본 연구에서는 Chiu(2013)의 Q행렬 정련 방법(Q-matrix refinement method)에 초점을 두고 연구를 진행하였다.

4. Chia-Yi Chiu의 Q행렬 정련방법(Q-matrix refinement method)

Chia-Yi Chiu에 의해 개발된 Q행렬 정련방법은 RSS(Residul Sum of Square)라는 타당성지수를 활용하여 Q행렬을 개선해나가는 방법이다(Chiu, 2013). Q행렬 정련방법은 알고리즘이 이론적으로 증명이 되었을 뿐 아니라, 비모수모형을 사용하기 때문에 컴퓨터의 사양에 상관없이 빠르게 결과를 얻을 수 있으며, 모델에 대한 특별한 가정이 필요하지 않다는 장점도 가지고 있다.

Q행렬 정련방법의 타당성지수인 RSS값을 구하는 방법은 다음과 같다. 피험자 i의 문항 j에 대한 실제 응답을 Y_{ij}라하고, 기대반응을 η_{ij}라 하자. 이때 피험자 i의 문항 j에 대한 RSS_{ij}값은 다음과 같다.

$$RSS_{ij} = (Y_{ij} - \eta_{ij})^2$$

따라서 전체 피험자 수를 N명이라 할 때, 모든 피험자의 문항 j에 대한 RSS_j값은 다음과 같다.

$$RSS_j = \sum_{i=1}^{N} RSS_{ij} = \sum_{i=1}^{N} (Y_{ij} - \eta_{ij})^2$$

Q행렬 정련 방법에서는 문항별 RSS_j값들을 최소화 시키는 방향으로 Q행렬을 개선해나간다. 한번의 iteration은 RSS_j값이 큰 문항 순서대로 Q벡터를 개선하여 모든 문항을 거치는 것으로 구성되는데, Q벡터를 개선할 때에는 기존 Q벡터를 해당 문항의 RSS_j값을 최소로 만드는 벡터로 대체한다. 이를 더 이상 Q행렬이 변하지 않을 때까지 반복하여 최종적으로 RSS값이 최소가 되는 Q행렬을 얻는다. 따라서 한 번의 iteration동안 피험자의 능력추정은 문항의 개수만큼 진행되고, 전체 프로세스동안 문항의 개수와 iteration 횟수를 곱한 만큼 이루어진다. 그러므로 많은 횟수의

추정을 감당하기 위해서 비모수 추정 방법인 Hamming모형이 사용된다. 구체적인 알고리즘은 다음과 같다(Chiu, 2013).

Q행렬 정련방법 알고리즘

1. $S_0 = \{1,2,...,J\}$을 문항의 전체 집합이라 하고, $Q^{(0)}$를 초기 Q행렬이라 하자.
2. $Q^{(0)}$와 피험자의 문항응답 데이터로부터 Hamming 추정을 사용하여 피험자들의 인지요소벡터를 분류한다.
3. 피험자의 문항응답 데이터와 (2에서 얻은) 피험자의 인지요소 숙달 프로파일을 가지고 S_0의 문항별 RSS_j값을 측정한다. S_0의 문항들 중에서 문항별 RSS_j값이 가장 큰 문항을 찾아내어 문항j라 하자.
4. $Q^{(0)}$에서 문항j에 해당하는 Q벡터 q_j를 다른 2^K-2개의(K는 요소개수) 0-1벡터로 바꾸어 보면서 RSS_j값을 계산하고, 이 값이 최소가 되는 벡터로 q_j를 대체한다. 이렇게 바뀐 Q를 $Q^{(1)}$이라 하고, $S_1 = S_0/\{j\}$라 하자.
5. $Q^{(0)}$에 $Q^{(1)}$을, $S^{(0)}$에 $S^{(1)}$을 넣어서 2-4 과정을 모든 문항을 거칠 때까지($S^{(1)}$이 공집합이 될 때까지) 반복한다.
6. 위 2-5과정을 더 이상 Q가 변하지 않을 때까지, 혹은 maximum iteration까지 반복한다.

Ⅲ. 연구방법

1. 적용연구

본 연구에서는 중학교 학생들 936명의 방정식과 부등식 영역의 인지요소의 숙달 여부를 평가한 이전 연구의 인지진단 평가에 쓰인 Q행렬의 타당성을 검증하였다(윤지영, 2015). 이 검사는 일차방정식과 일차부등식 문항을 합하여 총 25문항으로 구성되었고, 중학교 2학년과 3학년 학생을 대상으로 하였다. 방정식과 부등식 영역에 필요한 인지요소로는 <표 Ⅲ-2>의 10개의 인지요소를 다루었으며, 이들의 위계구조는 [그림 Ⅲ-1]에서 확인할 수 있다. 학생들의 숙달 인지요소를 추정하는 과정에서는 요소들 간의 위계 구조를 반영한 인지진단모형인 요소 간 위계 방식(Attribute Hierarchy Method; AHM)을 사용하였다.

평가는 서울과 경기, 인천 지역의 총 6개 중학교 40학급의 학생들을 대상으로 하였으며, 2학년 학생 580명과 3학년 학생 356명이 참여하였다. 아래 <표 Ⅲ-1>은 각 학교당 몇 명의 학생들이 참여하였는지를 나타내고 있다.

학교	sample 수		성별		
	2학년	3학년	남	여	미기재
A학교	106	-	53	46	7
B학교	72	-	33	34	5
C학교	49	-	25	24	-
D학교	96	-	38	55	3
E학교	76	-	76	-	-
	-	117	117	-	-
F학교	181	-	-	181	-
	-	239	-	239	-
	580 (61.97%)	356 (38.03%)	342 (36.54%)	579 (61.86%)	15 (1.60%)

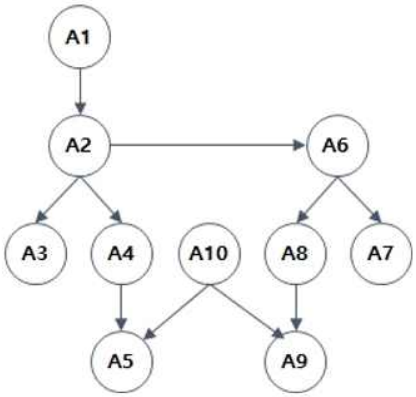
<표 Ⅲ-1> 학교별 연구 참여 대상 분포 출처 : 윤지영(2015)

윤지영(2015)에서는 일차방정식과 일차부등식 영역의 문항들을 해결하는데 필요한 내용과 기능을 아래 <표 Ⅲ-2>에서 설명하는 10개의 인지요소의 조합으로 보고, 학생들의 일차방정식과 일차부등식에 관련된 이해도를 측정할 수 있는 25개의 문항으로 검사를 구성하였다.

인지요소	요소 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	일차부등식 해의 의미
A8	일차부등식 풀기(부등식의 성질 포함)
A9	일차부등식 활용
A10	대수적 모델링

<표 Ⅲ-2> 일차방정식과 일차부등식 영역의 인지요소

위 10개의 인지요소는 다음 [그림 Ⅲ-1]과 같은 위계구조를 가진다. 이 위계구조는 내부 전문가들이 중학교 수학 교육과정을 참고하여 여러 차례의 논의를 통해 결정하였다(윤지영, 2015).



[그림Ⅲ-1] 인지요소 간 위계 구조

출처 : 윤지영(2015)

각각의 검사 문항과 인지요소들 사이의 관계는 <표 Ⅲ-3>에서 확인할 수 있으며 이 Q행렬에는 10개의 인지요소사이의 위계구조가 반영되어있다.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	1	0	0	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0	0	0
문항3	1	1	0	0	0	0	0	0	0	0
문항4	1	1	0	0	0	0	0	0	0	0
문항5	1	1	1	0	0	0	0	0	0	0
문항6	1	1	1	0	0	0	0	0	0	0
문항7	1	1	0	1	0	0	0	0	0	0
문항8	1	1	1	1	0	0	0	0	0	0
문항9	1	1	1	0	0	0	0	0	0	1
문항10	1	0	0	0	0	0	0	0	0	1
문항11	1	1	0	1	1	0	0	0	0	1
문항12	1	1	0	1	1	0	0	0	0	1
문항13	1	1	0	0	0	1	0	0	0	0
문항14	1	1	0	0	0	1	0	0	0	0
문항15	1	1	0	0	0	1	0	1	0	0
문항16	1	1	0	0	0	1	1	0	0	0
문항17	1	1	0	0	0	1	1	1	0	0
문항18	1	1	0	0	0	1	0	1	0	0
문항19	1	1	0	0	0	1	0	1	0	0
문항20	1	1	0	0	0	1	0	1	0	0
문항21	1	1	0	0	0	1	0	1	1	1
문항22	1	1	0	0	0	1	0	1	1	1
문항23	1	1	0	0	0	1	0	1	1	1
문항24	1	1	0	1	1	0	0	0	0	1
문항25	1	1	0	0	0	1	0	1	1	1

<표 Ⅲ-3> 방정식과 부등식 검사의 Q행렬

윤지영(2015)에서는 각 학생들의 인지요소 숙달 프로파일을 얻기 위해서는 요소 간 위계 방식(AHM) 모형에(Leighton et al., 2004), 인공신경망을

적용하여 학생들의 문항 응답 데이터를 분석하였다. 이를 위해 먼저 위계 구조를 따르는 모든 가능한 인지상태와 그에 따른 기대반응벡터로 학습데이터(training set)를 구성한 후, 이를 이용하여 인공신경망방법을 이용한 기계 학습(machine learning)을 진행하였다. 이렇게 학습된 모델에 학생들의 문항에 대한 실제 응답을 입력하면, 각 학생들의 인지요소 숙달 프로파일을 얻을 수 있다.

2. 모의실험

본 연구에서는 R프로그래밍을 활용해 모의실험을 진행하였다. 한 번의 모의실험마다 1000명의 가상 피험자의 데이터를 생성하였고, 한명의 데이터는 인지요소 10개에 대한 숙달 여부 및 25개 문항에 대한 가상 응답 결과를 포함하였다. 가상 피험자의 인지요소벡터를 설정할 때, 실제 상황과 유사한 데이터를 얻기 위하여 윤지영(2016)의 연구에 사용된 936명의 학생들의 인지요소벡터의 분포를을 이용하였다. 이 분포를 반영한 R의 sample 함수를 사용하여 각 가상 피험자에 대한 인지요소벡터를 임의 추출하였다.[1] 이렇게 각 가상피험자의 인지요소벡터가 설정되면, DINA모형을 통해 25개의 문항에 대한 가상 응답을 결정하였다.
예를 들어 문항j의 이탈모수가 s_j, 추측모수가 g_j인 경우, 가상피험자 i의 문항 j에 대한 응답Y_{ij}는 다음과 같은 확률로 1의 값을 가지도록 설정하여 0과 1의 값 중에 임의로 택한다.

$$P(Y_{ij}=1|\alpha_i)=(1-s_j)^{n_{ij}}g_j^{(1-n_{ij})}$$

여기서 α_i는 피험자의 인지요소벡터를 의미하고, n_{ij}는 피험자 i가 문항 j를 푸는데 필요한 모든 인지요소를 숙달하였을 때 1의 값을, 그렇지 않았을 때 0의 값을 가진다.

본 연구에서는 DINA모형에 사용된 문항의 이탈모수에 따라 총 3개의 실험 변수 세트를 설계하였다. 이때 문항의 추측모수는 0.2로 모두 고정하였다. 각 세트 별로 1000번의 모의실험을 반복하였고, 그로부터얻은 1000개의 실험결과(각 실험결과는 1000명의 가상피험자에 대한 데이터에 대해 얻어진 것임)를 의 평균값과 분포 등을 분석하였다. 각 세트에서 문항의 이탈모수는 모든 문항에 대해 통일하였고, 그 값은 <표 Ⅲ-4>에서 확인할 수 있다.

---

1) R의 sample함수에서는 각각의 인지요소벡터가 추출될 확률을 다르게 설정하는 것이 가능하다.

Set 1	Set 2	Set 3
이탈=0.05	이탈=0.1	이탈=0.2
추측=0.2	추측=0.2	추측=0.2

<표 Ⅲ-4> 모의실험에 사용된 문항 모수

이와 같은 방법으로 제작된 가상데이터는 문항의 이탈모수가 0.05, 0.1, 0.2인 3가지 상황에 대하여 각각 1000명의 피험자를 대상으로 하는 검사를 1000번 진행한 효과를 가진다.

3. 타당성 분석 방법

본 연구에서는 Chia-Yi Chiu의 Q행렬 정련방법에서 사용된 아이디어로부터 $AVERSS$라는 타당성지수를 구성하여 이를 이용한 Q행렬 타당성 분석방법을 개발하였다. Chia-Yi Chiu의 Q행렬 정련방법과 $AVERSS$를 활용한 Q행렬 타당성 분석 방법은 다음과 같다.

3.1. Chia-Yi Chiu의 Q행렬 정련 방법

Chia-Yi Chiu가 개발한 Q행렬 정련방법은 피험자의 기대반응과 실제 응답사이의 거리를 좁혀주는 방향으로 Q행렬을 개선해나가는 방법이다 (Ⅱ-4절 참조). 전체 피험자 수를 N명이라 할 때, 모든 피험자의 문항 j에 대한 RSS_j값은

$$RSS_j = \sum_{i=1}^{N} RSS_{ij} = \sum_{i=1}^{N} (Y_{ij} - n_{ij})^2$$

이라 할 수 있다. Q행렬 정련 방법에서는 이 문항별 RSS_j값들을 최소화 시키는 방향으로 Q행렬을 개선해나간다.

3.2 $AVERSS$를 활용한 타당성 분석 방법

본 연구에서 소개하는 $AVERSS$ 타당성 분석 방법은 RSS값의 어떤 컷오프(cut-off)값을 제시하여 데이터로부터 얻은 Q행렬의 RSS값이 컷오프 값 이하일 경우 전문가에 의해 제작된 Q행렬을 타당하다고 판정하고, 컷오프 값 보다 클 경우 Q행렬이 타당하지 않다고 판정한다. 이를 위해 Q행렬 전체의 타당성을 대표할만한 타당성지수로 문항별 RSS_j값의 평균

을 의미하는 $AVERSS$를 사용하였다.

기존의 Q행렬 정련방법에서 사용된 RSS는 수치는 문항별 값으로 이는 각 문항에 대하여 피험자들의 응답과 기대반응이 다를 경우를 세어준 값이다. 본 연구에서 개발하고자 한 것은 문항별 Q벡터가 아닌 Q행렬 전체의 타당성이므로 문항별 RSS_j값들을 하나의 개념으로 통합할 필요가 있었다. 이를 위해서 먼저 RSS_j값을 피험자의 규모와 무관한 문항 고유의 값으로 바꾸어주는 과정이 필요하다. RSS_j값은 피험자의 수가 많을수록 값이 커지므로 RSS_j값을 피험자의 수로 나누었다. 이를 문항별 $AVERSS_j$라 정의하였고, 이는 문항j의 기대반응과 실제응답이 다른 경우(피험자)의 비율을 나타내게 된다. 즉,

$$AVERSS_j = \frac{RSS_j}{N} = \frac{\sum_{i=1}^{N} RSS_{ij}}{N} \qquad (N\text{: 피험자의 수})$$

이렇게 정의한 문항별 $AVERSS_j$를 Q행렬 전체를 대표하는 하나의 개념으로 통합하기 위해서는 총합(sum)과 평균(average) 두 가지를 생각할 수 있다. $SUMRSS$를 각각의 $AVERSS_j$값을 더한 값으로, $AVERSS$를 $AVERSS_j$의 평균값으로 정의한다.

$$SUMRSS = \sum_{j=1}^{J} AVERSS_j = \frac{\sum_{j=1}^{J} RSS_j}{N} = \frac{\sum_{j=1}^{J}\sum_{i=1}^{N} RSS_{ij}}{N}$$

$$AVERSS = \frac{SUMRSS}{J} = \frac{\sum_{j=1}^{J} AVERSS_j}{J} = \frac{\sum_{j=1}^{J} RSS_j}{NJ} = \frac{\sum_{j=1}^{J}\sum_{i=1}^{N} RSS_{ij}}{NJ}$$

Q행렬의 타당성지표는 검사의 규모와 상관없이 Q행렬의 타당도를 판정할 수 있어야하므로 피험자의 규모 뿐 아니라 문항의 개수에도 영향을

받지 않아야 한다. 따라서 문항이 많을수록 커지는 $SUMRSS$와 문항의 개수에 영향을 받지 않는 $AVERSS$ 중에서 $AVERSS$를 타당성지수로 선택하였다.

$AVERSS$는 모든 피험자와 모든 문항에 대한 RSS_{ij}의 평균값이라 할 수 있으며 각각의 RSS_{ij}가 0 또는 1의 값을 가지므로 $AVERSS$는 0부터 1까지의 값을 가진다. 그리고 $RSS_{ij}=1$인 경우는 피험자 i의 문항 j에 대한 기대반응과 실제응답이 다를 경우를 나타내므로 $AVERSS$는 전체 경우 중에서 기대반응과 실제응답이 다른 경우의 비율을 의미한다. 따라서 $AVERSS$값이 0에 가까울수록 기대반응과 실제응답이 일치할 확률이 높다는 의미이고, 이는 연구에 사용된 Q행렬이 데이터를 잘 설명한다고 해석할 수 있다. 그러므로 $AVERSS$를 Q행렬 전체를 나타내는 타당성지수로 사용하는 것이 가능하다.

$AVERSS$ 타당성 분석방법은 어떤 검사에 사용된 Q행렬의 타당성을 판정하는 방법으로 검사 데이터로부터 얻은 $AVERSS$값과 컷오프 값을 비교하여 $AVERSS$값이 컷오프 값 이하이면 Q행렬이 타당하다고, 크다면 타당하지 않다고 판정한다. 이때 사용되는 컷오프 값은 Ⅳ-2절에서 제시한다.

기존 Chia-Yi Chiu의 Q행렬 정련방법에서는 RSS값이 작아지는 방향으로 Q행렬을 정련하였다면, $AVERSS$ 타당성 분석 방법에서는 정련과정 없이 Q행렬의 타당성을 판정하는데 그치므로 기존 Q행렬 정련방법을 축약한 형태라 볼 수 있다. 그러나 기존의 Q행렬 정련방법은 데이터를 통해 Q행렬을 결정하는 탐색적 접근의 방법인 반면, 이 새로운 Q행렬 타당화 방법은 전문가가 제시한 Q행렬의 타당성을 확인하는 차원의 확인적 접근의 방법이라는 것에서 큰 차이가 존재한다.

4. 연구 절차

본 연구는 수학 인지진단 검사에 적용 가능한 Q행렬 타당화 방법 제작에 그 목표가 있다. 이를 위해 Ⅲ-3.2절에서 제시한 $AVERSS$를 이용한 타당성 분석 방법이 기존 방법에 비해서 얼마나 수학 인지진단 검사에 적합한지에 관한 연구가 이루어졌다.

먼저 기존 Chia-Yi Chiu의 Q행렬 정련방법이 어떻게 Q행렬을 개선하는지 알아보기 위해 Ⅲ-1절에서 소개한 윤지영(2016)의 방정식과 부등식 인지진단평가에 사용된 Q행렬에 대하여, 기존의 Q행렬 정련방법을 적용하여 그 결과를 분석하였다. 알고리즘 구현은 R 프로그램을 이용하였고, NPCD 패키지에 내제되어있는 Q-refine 함수를 사용하였다. 정련된 Q행렬을 얻은 후에는 검사 문항과 요소를 직접 대조해보며 실질적인 Q행렬 개선이 이루어졌는지를 확인하였다. 또한, Q행렬 정련방법의 알고리즘을 살펴보며 이 방법이 가지고 있는 문제점의 원인을 찾아보았다.

다음으로는 $AVERSS$ 타당성 분석 방법의 구축을 위해 Ⅲ-2절의 가상 데이터를 활용하여 $AVERSS$의 분포를 얻고, Q행렬의 타당도를 판정할만한 $AVERSS$의 컷오프 값을 결정하였다. 그리고 이렇게 결정된 컷오프 값이 잘못된 Q행렬들에 대하여 얼마나 정확히 진단할 수 있는지를 분석하여 $AVERSS$ 타당성 분석 방법의 효율성을 측정하였다. 그리고 $AVERSS$ 타당성 분석 방법을 적용하여 Ⅲ-1절에서 소개한 윤지영(2016)의 방정식과 부등식 인지진단평가에 사용된 Q행렬의 타당성을 분석하였다.

추가적으로 $AVERSS$를 활용하여 검사를 개선시키는 방법을 연구하였다. 기존의 방정식과 부등식 검사를 개선하기 위해 먼저 각 문항별 $AVERSS_j$값을 관찰하여 $AVERSS_j$값이 어떤 컷오프 값(위의 컷오프 값과 다른 의미)보다 큰 문항들의 특성을 질적으로 분석하였다. 그리고 문항들의 특성을 고려하여 적합한 방향으로 검사를 개선하였다. 이렇게 검사를 개선한 후 AHM모형을 적용하여 피험자의 능력을 다시 추정하였고, 개선된 검사의 Q행렬에 다시 $AVERSS$ 타당성 분석 방법을 적용시켜 타

당도의 변화를 관찰하였다.

다음은 본 연구의 절차를 나타낸다.

[그림 Ⅲ-2] 연구 절차

Ⅳ. 연구 결과

1. Chia-Yi Chiu의 Q행렬 정련방법의 문제점 분석

1.1. Chia-Yi Chiu의 Q행렬 정련방법 적용 결과 분석

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	1	0	0	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0	0	0
문항3	1	0	0	0	0	0	0	0	0	0
문항4	1	1	0	0	0	0	0	0	0	1
문항5	1	0	1	0	0	0	0	0	0	0
문항6	1	1	0	0	0	0	0	0	0	0
문항7	1	1	0	1	0	0	0	0	0	0
문항8	1	1	1	1	0	1	0	1	0	0
문항9	1	1	1	0	0	0	0	0	0	1
문항10	1	0	0	0	0	0	0	0	0	1
문항11	1	1	0	1	1	0	0	0	0	1
문항12	0	0	0	1	1	0	0	0	0	0
문항13	1	0	0	0	0	1	0	0	0	0
문항14	1	1	1	0	0	1	0	0	0	1
문항15	1	1	0	0	0	1	0	1	0	0
문항16	0	1	0	0	0	0	1	0	0	0
문항17	1	0	0	0	0	1	1	1	0	0
문항18	0	1	0	0	0	1	0	1	0	0
문항19	1	1	0	0	0	1	0	1	0	0
문항20	0	0	1	1	1	0	1	1	1	1
문항21	1	0	0	0	0	0	0	1	0	1
문항22	1	1	0	0	0	1	0	1	1	1
문항23	0	0	0	0	0	1	0	1	1	1
문항24	1	0	0	1	1	0	0	0	0	1
문항25	1	0	0	0	0	0	0	1	1	1

<표 Ⅳ-1> Q행렬 정련방법 적용 후 정련된 Q행렬

Ⅲ-1의 방정식과 부등식 평가에서 사용된 Q행렬을 Chia-Yi Chiu의 Q 행렬 정련방법을 통해 정련한 결과 위와 같은 새로운 Q행렬을 얻었다. 이를 <표 Ⅲ-3>의 기존 Q행렬과 비교한 결과는 <표 Ⅳ-2>와 같다. 여기서 1은 기존 Q행렬에서는 0이었으나 Q행렬 정련방법 실행 후 1로 바뀐 항을 나타내고, -1은 기존 Q행렬에서는 1이었으나 Q행렬 정련방법 실행 후에는 0으로 바뀌게 된 항을 나타낸다.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	0	0	0	0	0	0	0	0	0	0
문항2	0	0	0	0	0	0	0	0	0	0
문항3	0	-1	0	0	0	0	0	0	0	0
문항4	0	0	0	0	0	0	0	0	0	1
문항5	0	-1	0	0	0	0	0	0	0	0
문항6	0	0	-1	0	0	0	0	0	0	0
문항7	0	0	0	0	0	0	0	0	0	0
문항8	0	0	0	0	0	1	0	1	0	0
문항9	0	0	0	0	0	0	0	0	0	0
문항10	0	0	0	0	0	0	0	0	0	0
문항11	0	0	0	0	0	0	0	0	0	0
문항12	-1	-1	0	0	0	0	0	0	0	-1
문항13	0	-1	0	0	0	0	0	0	0	0
문항14	0	0	1	0	0	0	0	0	0	1
문항15	0	0	0	0	0	0	0	0	0	0
문항16	-1	0	0	0	0	-1	0	0	0	0
문항17	0	-1	0	0	0	0	0	0	0	0
문항18	-1	0	0	0	0	0	0	0	0	0
문항19	0	0	0	0	0	0	0	0	0	0
문항20	-1	-1	1	1	1	-1	1	0	1	1
문항21	0	-1	0	0	0	-1	0	0	-1	0
문항22	0	0	0	0	0	0	0	0	0	0
문항23	-1	-1	0	0	0	0	0	0	0	0
문항24	0	-1	0	0	0	0	0	0	0	0
문항25	0	-1	0	0	0	-1	0	0	0	0

<표 Ⅳ-2> Q행렬 정련방법 적용 전과 후 비교 행렬

<표 Ⅳ-2>에서 알 수 있듯 25개의 문항 중 9개의 문항을 제외하고는 모두 변화가 있었다. 그런데 이를 검사문항과 대조해본 결과 적절한 해석이 가능한 개선을 제시하지 않았음을 알 수 있었다. 예를 들어 두 개의 인지요소가 추가된 문항8과, 거의 모든 인지요소의 유무가 뒤바뀌게 된 문항20을 관찰해보도록 하자. 문항8과 문항20은 다음과 같다.

8. 다음 두 방정식의 해가 같을 때, 상수 a의 값을 구하여라. $3(\frac{2}{5}-0.1x)=0.2(2x+\frac{5}{2})$ $\frac{x+3}{6}=\frac{a-2x}{3}-1$ ① 3 ② 4 ③ 5 ④ 6 ⑤ 7	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 $x<-\frac{1}{a}$

정련된 Q행렬에서는 문항8을 풀기 위한 세부능력으로 인지요소6과 인지요소8이 추가되었다. 그런데 '부등식의 이해'를 나타내는 인지요소6과 '일차부등식 풀기(부등식의 성질 포함)'를 나타내는 인지요소8은 문항8을 푸는데 필요한 세부능력이 아님을 확인할 수 있다. 또한, 문항20에서 필수적인 인지요소6('부등식의 이해')은 제거되었고, 문항20과 전혀 관련이 없어 보이는 인지요소10('대수적 모델링')이 추가된 것 역시 제대로 개선되었다고 볼 수 없다. 다른 문항들 관찰해보아도 마찬가지로 Q행렬 정련방법을 통하여 Q행렬의 실질적인 개선이 이루어지지 않음을 확인해볼 수 있었다. 이는 기존 Chia-Yi Chiu의 Q행렬 정련방법이 실제 인지진단 검사의 Q행렬에 적용하는 것에 어려움이 있음을 의미한다.

1.2. Chia-Yi Chiu의 Q행렬 정련방법의 문제점 분석

그렇다면 Chia-Yi Chiu의 Q행렬 정련방법이 위와 같은 문제점을 가지게 된 근본적인 원인은 무엇일까? 그 원인은 Q행렬 정련방법의 가정과 관련되었을 수 있다. Q행렬 정련방법은 문항별 RSS_j값들이 가장 작을 때, 다시 말해 기대반응과 피험자의 실제 응답 사이의 거리들의 합이 가장 작을 때가 실제 Q벡터가 된다는 가정을 토대로 정련이 진행된다. 최종적으로 정련된 Q행렬을 얻기 위해서는 문항별 RSS_j값들이 작아지는 방향으로 Q벡터를 변경하고, Hamming모형으로 피험자의 인지요소벡터를 다시 분류하는 과정을 여러 차례 반복한다. 여기서 RSS_j값은 Hamming distance의 합이라 할 수 있으므로 Q행렬 정련방법은 Hamming distance가 가장 짧을 경우가 실제 Q행렬과 올바른 인지요소벡터 분류라는 가정을 수반한다. 이는 Hamming distance가 가장 짧을 때가 올바른 인지요소벡터 분류라는 Hamming모형의 가정과 유사하다. 다시 말해 Q행렬 정련방법은 Hamming모형의 아이디어를 Q행렬 타당화에 적용시킨 방법이라 할 수 있다. 그러므로 Q행렬 정련방법의 실효성은 Hamming 추정이 얼마나 올바르게 피험자의 인지요소벡터를 분류해주는가와 직결되어 있다.

<표 Ⅳ-3>은 3.1의 방정식과 부등식 평가에 참여한 학생들의 검사지 응답을 토대로 인지진단을 한 결과의 일부이다. 방법1의 결과는 인지요소의 위계구조를 반영한 AHM모형을 적용하여 얻은 결과이고, 방법2의 결과는 Hamming모형으로 얻은 결과이다. <표 Ⅳ-3>에서 알 수 있듯 Hamming모형을 적용했을 때와 ANN모형을 적용했을 때 피험자의 인지요소벡터는 매우 다르다. 20명 중에서 회색으로 음영이 들어간 9명의 진단이 다르게 나왔고, 두 모형을 적용했을 때 결과가 같은 11명 중 10명은 모든 인지요소를 숙달한 학생들이었다. 이렇듯 두 모형을 적용한 결과가 다른 이유는 Hamming 모형에서는 인지요소의 위계구조를 고려하지 않았다는 점에 있다. 인지진단모형과 Q행렬의 형태에 따라 정도의 차이가 존재하지만, 정해진 개수의 문항들을 통해 위계구조 없이 검사할

수 있는 인지요소의 개수는 제한적이다. 위계구조가 없다면 10개의 인지요소는 총 2^{10}개의 인지요소벡터를 생성하므로 25개의 문항은 피험자의 인지요소벡터를 올바로 분류하기에 현저히 적다. 따라서 인지요소 10개를 검사하기 위해서는 가능한 인지요소벡터의 개수를 줄여줄 수 있는 위계구조가 필요하다. 그러나 Hamming모형에서는 위계구조를 고려하지 못하므로 피험자의 인지요소벡터 분류가 제대로 이루어지지 않았을 것이라 예상된다.

	방법1 (ANN모형)										방법2 (Hamming모형)									
인지요소	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
피험자1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
피험자2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
피험자3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
피험자4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
피험자5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
피험자6	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
피험자7	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
피험자8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
피험자9	1	1	1	1	0	1	1	1	0	1	1	1	0	1	0	1	1	0	0	1
피험자10	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
피험자11	1	1	1	0	0	1	0	0	0	1	1	1	1	0	1	1	0	0	1	1
피험자12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
피험자13	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	0	0	0	0	1
피험자14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
피험자15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
피험자16	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1
피험자17	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	0
피험자18	1	1	1	1	0	1	1	1	0	1	1	1	1	0	0	1	1	1	0	1
피험자19	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
피험자20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

<표 Ⅳ-3> ANN모형과 Hamming 모형 적용 결과 비교

이와 같은 결과는 Hamming distance가 짧은 것을 선택하는 것이 항상 올바른 선택이 아닐 수도 있다는 것을 보여준다. 따라서 같은 아이디어를 사용하는 Q행렬 정련방법 역시 올바른 방향으로 개선이 일어나지 않았을 것이라 예상된다.

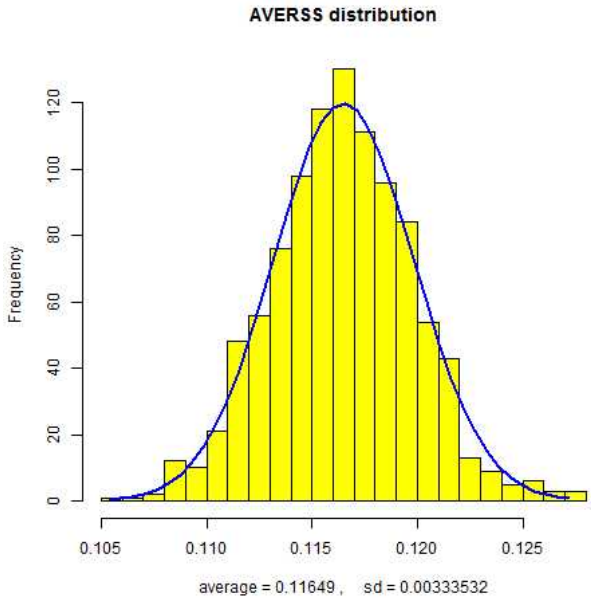
Chia-Yi Chiu의 Q행렬 정련 방법을 적용한 결과는 탐색적 접근의 한계를 잘 드러낸다. 탐색적 접근에서는 각각의 요인에 대해 의미나 해석을 부여하지 않기 때문에 그 결과가 이론과 전혀 맞지 않을 수 있다(Thompson, 2004). Chia-Yi Chiu의 Q행렬 정련방법은 연구자가 이론을 토대로 제시한 Q행렬을 초기 Q행렬로 잡고 정련을 시작하지만, 이후의 정련과정에서는 온전히 검사 데이터에 의존하여 문항별 RSS_j값이 가장 작을 때를 타당한 Q행렬로 택하고 있다. 따라서 Q행렬 정련 방법은 데이터를 측정할 때 생긴 여러 가지 오류로 인해 잘못된 Q행렬을 개선된 Q행렬로 제시할 수도 있으며 특히 오류에 대한 방어능력(robustness)이 약할 수 있다. 그런데 실제 인지진단평가에서는 학생의 컨디션, 검사 환경 등 다양한 변인이 존재하고, 이에 따라 다양한 오류가 발생할 수 있으므로 검사 데이터에 의존하는 탐색적 접근의 결과가 전문가가 이론적 근거에 기반을 두어 제시한 것보다 더 적절하다고 보장하기 어렵다. 그리하여 본 연구에서는 Q행렬의 타당성의 분석에 있어서 탐색적 접근이 아닌 전문가가 이론을 기반으로 제시한 Q행렬의 타당성을 확인하는 확인적 접근의 분석 방법 개발에 초점을 두게 되었다.

2. *AVERSS* 타당성 분석 방법 개발 및 적용 결과

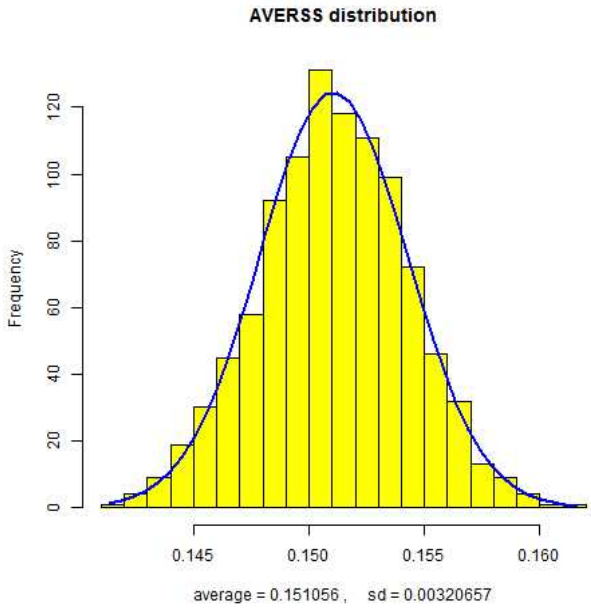
AVERSS 타당성 분석 방법이 효율적으로 Q행렬의 타당도를 판정하기 위해서는 적절한 컷오프 값이 사용되어야 한다. 이 타당성 분석 방법은 컷오프 값을 기점으로 Q행렬의 타당성 여부를 결정하므로 컷오프 값의 역할이 매우 중요하다. 본 연구에서는 가상데이터를 활용하여 *AVERSS* 값의 분포를 분석하고, 이를 기반으로 적절한 컷오프 값을 제시한다.

2.1. 모의실험을 통한 *AVERSS* 분포 및 컷오프 값 분석

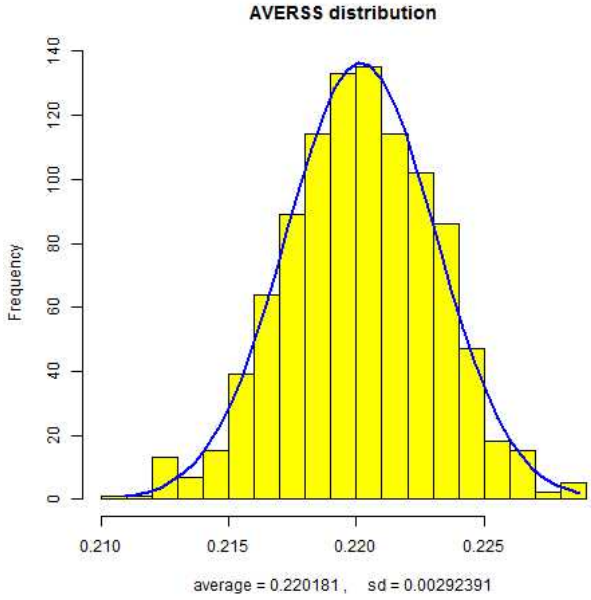
본 연구에서는 *AVERSS*의 분포를 얻기 위해 연구방법의 2절에서 소개된 모의실험 데이터를 활용하였다. [그림 Ⅳ-1] ~ [그림 Ⅳ-3]은 각각 문항 별 이탈모수를 0.05, 0.1, 0.2로 설정해준 3가지 시뮬레이션 데이터 세트로부터 얻은 *AVERSS*의 분포를 보여준다.



[그림 Ⅳ-1] Set1의 *AVERSS*분포 (이탈=0.05, 추측=0.2)



[그림 Ⅳ-2] Set2의 $AVERSS$분포 (이탈=0.1, 추측=0.2)



[그림 Ⅳ-3] Set3의 $AVERSS$분포 (이탈=0.2, 추측=0.2)

각각의모의실험 데이터 세트는 1000번의 검사 결과를 포함하고 있으므로 이를 통해 얻어지는 1000개의 *AVERSS*값으로부터 각 문항모수 상황에 대해 대략적인 *AVERSS* 분포를 얻어낼 수 있다. 다음의 <표 Ⅳ-4>는 검사 문항의 이탈 모수에 따른 *AVERSS*의 평균과 표준편차를 보여주는데, 이탈모수가 큰 상황에서 *AVERSS* 값의 평균 또한 커진다는 것을 확인할 수 있다.

	*AVERSS*의 평균	*AVERSS*의 표준편차
Set1 (이탈=0.05)	0.116	0.003
Set2 (이탈=0.1)	0.151	0.003
Set3 (이탈=0.2)	0.220	0.003

<표 Ⅳ-4> 검사 문항의 이탈모수에 따른 *AVERSS*값의 평균 및 표준편차

*AVERSS*의 분포는 검사를 구성하고 있는 문항의 모수에 영향을 받기 때문에(Ⅳ-5절에서 분석) 타당성 판정의 기준이 될 *AVERSS*의 컷오프 값은 검사의 문항 모수에 따라 다르게 제시되어야 한다. 본 연구에서는 *AVERSS*의 컷오프 값으로 가상데이터를 통해 얻은 *AVERSS* 분포 상에서 95 백분위수의 값을 제시한다. 이 컷오프 값(<표 Ⅳ-5>의 굵은 글씨)을 사용할 경우 모의실험 상에서 정확한 Q행렬에 대하여 95%의 확률로 타당하다고 판정하게 된다.

	이탈=0.05	이탈=0.1	이탈=0.2
90 백분위수	0.121	0.155	0.224
95 백분위수	**0.122**	**0.156**	**0.225**
99 백분위수	0.125	0.159	0.227

<표 Ⅳ-5> 검사 문항의 이탈모수에 따른 *AVERSS* 컷오프 값

AVERSS 타당성 분석 방법이 유효하기 위해서는 정확한 Q행렬을 타당하다고 판정할 뿐 아니라 잘못된 Q행렬이 타당하지 않다는 판정 역시 가능하여야 한다. 이를 알아보기 위하여 기존 Q행렬을 변형하여 잘못된

Q행렬들을 제작 하였다. 잘못된 Q행렬로는 정확한 Q행렬의 1%, 3%, 5%가 변형된 Q행렬을 각각 5개씩 사용하였는데, Q행렬을 변형할 때에는 기본적으로 변형할 항을 임의로 선택하였으나 인지요소의 위계구조에 어긋나지 않도록 조정해가며 변형을 진행하였다. 그리고 한 인지요소를 측정하는 문항의 개수가 3개 미만이 되면 인지요소의 숙달 여부를 측정하기 어려울 수 있으므로 변형된 Q행렬에서 각각의 인지요소를 측정하는 문항이 적어도 3개 이상이 되도록 조정하였다. Q1~Q5는 기존의 정확한 Q행렬에서 1%(2~3개의 항)가 변형된 Q행렬들을, Q6~Q10은 3%(7~8개의 항)가 변형된 Q행렬을, Q11~Q15는 5%(12~13개의 항)가 변형된 Q행렬을 나타내며 이 15개의 Q행렬에 대한 자세한 정보는 부록에 첨부하였다.

Q1~Q15를 문항 모수가 다른 3가지 상황에 대하여 각각 *AVERSS* 타당성 분석 방법을 적용한 결과는 다음과 같다. 아래 <표 Ⅳ-6>에서 나타내는 수치는 서로 다른 문항 모수 상황의 1000개의 모의실험 데이터에 대하여 *AVERSS* 타당성 분석 방법을 적용한 결과 잘못된 Q행렬이 타당하지 않다고 판정된 경우의 비율을 의미한다.

(1%변형)	Q1	Q2	Q3	Q4	Q5
이탈=0.05 (c=0.122)	0.319	0.334	0.844	0.232	0.913
이탈=0.1 (c=0.156)	0.268	0.327	0.834	0.202	0.911
이탈=0.2 (c=0.225)	0.224	0.342	0.893	0.189	0.945
(3%변형)	Q6	Q7	Q8	Q9	Q10
이탈=0.05	1	0.982	0.972	1	1
이탈=0.1	1	0.981	0.960	0.999	1
이탈=0.2	1	0.983	0.939	1	1
(5%변형)	Q11	Q12	Q13	Q14	Q15
이탈=0.05	1	1	1	1	1
이탈=0.1	1	1	1	1	1
이탈=0.2	1	1	1	1	1

<표 Ⅳ-6> 잘못된 Q행렬에 *AVERSS* 타당성 분석 방법 적용 결과

<표 Ⅳ-6>에 비추어볼 때, 이탈모수가 다른 세 가지 상황에서 모두 1%가 변형된 Q행렬들의 타당성에 대하여는 불안정하게 분석하였지만, 3%가 변형된 Q행렬들의 경우 안정적으로 90%이상의 Q행렬이 타당하지 않다고 판정하였고, 5%가 변형된 Q행렬들은 거의 대부분 타당하지 않다고 판정하는 것에 성공하였다. 즉, 위에서 제시한 컷오프 값은 정확한 Q행렬과 잘못된 Q행렬을 판가름할 수 있는 적절한 값임을 알 수 있다. 이는 $AVERSS$ 타당성 분석 방법을 통해 타당한 Q행렬에 대한 검증을 할 뿐 아니라 잘못된 Q행렬에 대한 판정 또한 제대로 이루어질 수 있음을 의미한다.

2.2. $AVERSS$ 타당성 분석 방법 적용 결과

다음은 Ⅲ-1절의 방정식과 부등식 인지진단 평가의 결과를 토대로 검사에 사용된 Q행렬의 $AVERSS$를 구한 결과이다.

	문항1	문항2	문항3	문항4	문항5	문항6	문항7	문항8	문항9
RSS_j	135	86	52	241	128	75	16	305	181
$AVERSS_j$	0.1442	0.0919	0.0556	0.2575	0.1368	0.0801	0.0171	0.3259	0.1934
	문항10	문항11	문항12	문항13	문항14	문항15	문항16	문항17	문항18
RSS_j	50	167	179	42	320	170	41	249	164
$AVERSS_j$	0.0534	0.1784	0.1912	0.0449	0.3419	0.1816	0.0438	0.2660	0.1752
	문항19	문항20	문항21	문항22	문항23	문항24	문항25	SUM RSS	AVE RSS
RSS_j	121	434	369	245	223	207	212		
$AVERSS_j$	0.1293	0.4637	0.3942	0.2618	0.2382	0.2212	0.2265	4.717	0.1885

<표 Ⅳ-7> 방정식과 부등식 검사 Q행렬의 $AVERSS$와 문항별 $AVERSS_j$

방정식과 부등식 검사 데이터를 통해 얻은 $AVERSS$값은 0.1885가량으로 전체 경우 중에서 기대반응과 실제응답이 다른 경우가 19%정도였음

을 알 수 있다. 이를 Ⅳ-2.1절에서 제시한 컷오프 값에 비추어 보았을 때, 검사 문항들의 이탈모수가 0.2이상 혹은 그에 가까운 값이라면 본 Q 행렬이 타당하다고 말할 수 있으나, 문항들의 이탈모수가 0.1이하 혹은 그에 가까운 값이라면 본 Q행렬은 타당하다고 할 수 없다. 즉, 이 검사에 사용된 문항들이 어느 정도의 이탈모수를 가지는지에 따라 *AVERSS* 타당성 분석 방법을 적용한 결과 역시 달라진다. 따라서 Q행렬의 타당성 판정을 위해서는 검사 문항 모수의 측정이 선행되어야 한다. 검사의 문항 모수를 측정하는 과정은 *AVERSS* 측정에 사용된 검사 데이터가 아닌 다른 데이터나 이론에 근거한 분석을 통해 이루어져야 한다.

3. $AVERSS$를 이용한 인지진단 검사의 개선방법

Ⅳ-3절에서는 $AVERSS$를 Q행렬의 타당성을 확인하는 것에서 그치지 않고 이를 활용하여 인지진단 검사를 개선하는 방법을 제시한다. 어떤 검사의 데이터가 컷오프 값에 비추어보았을 때 큰 $AVERSS$값을 가질 경우, 검사의 Q행렬에 전체적인 문제가 있을 수도 있지만 일부 문항들에 의해서 $AVERSS$값이 크게 증폭되었을 수도 있다. 이럴 경우 해당 문항들을 개선함으로써 더욱 타당한 검사를 구성하는 것이 가능하다. 본 연구에서는 각 문항별 $AVERSS_j$값을 관찰하여 유의미하게 큰 값을 가지는 문항들의 질적 분석을 통해 검사를 개선하는 방법을 제시한다.

먼저 각 문항별 $AVERSS_j$값이 어떤 컷오프 값($AVERSS$ 타당성 분석 방법의 컷오프 값과는 다른 의미)보다 큰 문항들에 대하여 질적 분석을 진행한다. 문항별 $AVERSS_j$은 해당 문항과 인지요소들의 관계가 타당하게 구성되어있는지를 드러내므로 $AVERSS_j$값이 큰 문항일수록 인지요소와의 관계가 잘못 설정되었을 가능성이 크기 때문이다.

$AVERSS_j$값이 큰 문항들을 질적 분석하는 과정에서 검사 내용지식 전문가가 필요하다. 문항 분석에서는 Q행렬에서 해당 문항의 세부 능력으로 설정되어있는 인지요소들이 문항을 잘 설명하고 있는지를 중점적으로 다룬다.

문항A. 다음 식을 계산 하여라.

$$(5-3)\times 0+7$$

문항B. 다음 식을 계산 하여라.

$$8\times 2+3\times 4$$

인지요소 목록
V1. 덧셈을 할 수 있다.
V2. 뺄셈을 할 수 있다.
V3. 곱셈을 할 수 있다.
V4. 나눗셈을 할 수 있다.
V5. 괄호의 의미를 알고 있다.

	V1	V2	V3	V4	V5
문항A	1	1	1	0	1
문항B	1	0	1	0	0

<표 Ⅳ-9> 예시 문항A와 B의 Q벡터

<표 Ⅳ-8> 문항 질적 분석을 통해 발견할 수 있는 2가지 오류 예시

질적 분석을 통해서 파악할 수 있는 문제 상황은 크게 두 가지가 있는데 위의 예시상황의 문항A와 문항B는 이 2가지 경우를 잘 드러내고 있다. 첫 번째 경우는 문항-요소 관계가 잘못 설정된 경우이다. 해당 문항을 풀기위해 필요한 인지요소가 Q행렬 상에서는 필요 없다고 설정되었거나, 거꾸로 필요 없는 인지요소가 Q행렬 상에서는 필요 하다고 설정된 경우가 이에 해당된다. 위 예시상황의 문항A도 이러한 오류를 포함하고 있다. 문항A는 '-' 기호가 등장하므로 인지요소 V2('뺄셈을 할 수 있다')를 필요로 하는 것처럼 보이지만, 그 뒤에 0과 곱하는 과정이 들어있으므로 실제로는 문항의 답이 뺄셈의 결과에 의존하지 않는다. 이러한 분석을 통하여 위의 Q행렬에서 문항A와 인지요소 V2관계가 잘못 설정되었음을 알 수 있다. 내용중심의 인지요소들은 문항-요소 관계가 보다 명확하므로 문항-요소 관계에 문제가 있는 경우가 드물지만, 문제풀이 과정중심의 인지요소들은 보는 관점에 따라 문항과의 관계가 모호할 수 있기 때문에 문항-요소관계가 잘못 설정될 가능성이 있다.

두 번째는 해당 문항을 풀기 위해서 기존에 설정된 인지요소들로는 설명할 수 없는 새로운 능력이 필요할 경우이다. 인지진단 검사를 제작하는 과정에서 인지요소들이 결정된 후 각각의 문항에 필요한 인지요소들을 대응시킬 때 문항을 풀기 위해 필요한 어떤 세부능력이 빠지는 경우가 발생할 수 있는데, 위 예시상황의 문항B가 이러한 오류를 나타내고 있다. 문항B는 인지요소 V1 ('덧셈을 할 수 있다')과 인지요소 V3 ('곱셈을 할 수 있다')를 필요로 하는 문항이다. 그런데 과연 학생이 덧셈과 곱셈을 할 수 있다고 해서 문항B를 풀 수 있을까? 사실 문항B를 풀기 위해서는 덧셈과 곱셈을 할 수 있을 뿐 아니라 이들의 계산 순서도 알아야 한다. 즉, 덧셈과 곱셈을 할 수 있지만 이들의 계산 순서를 모르는 학생은 올바른 답을 얻을 수 없다. 따라서 문항B에는 기존 5개의 인지요소로는 설명할 수 없는 새로운 인지요소 V6('덧셈과 곱셈의 계산 순서를 알고 있다')를 필요로 한다. 이렇듯 문항이 Q행렬에 명시되지 않은 새로운 인지요소를 필요로 하는 경우는 질적 분석을 통해서만 문제점을 명확히 파악할 수 있다.

$AVERSS_j$값이 큰 문항에 대한 문항들에 대한 질적 분석을 마치면 분석 결과에 따라 적합한 방향으로 검사를 개선한다. 본 연구에서는 문항의 질적 분석 후 취할 수 있는 개선 방안으로 다음의 다섯 가지를 제시한다.

개선방안1. 문항-요소관계가 잘못된 경우 Q행렬 개선
개선방안2. 문항-요소관계가 잘못된 경우 문항 개선
개선방안3. 문항들에게 새로운 세부능력이 필요할 경우 이를 배재하여 문항을 개선
개선방안4. 문항들에게 새로운 세부능력이 필요할 경우 이를 새로운 요소로 추가
(필요할 경우 새로운 요소를 담고 있는 문항 추가)
개선방안5. 문항들을 제거한 검사를 분석

*개선방안1*은 문항-요소관계가 잘못 설정되었을 때 Q행렬 개선하는 방안이다. 이때 문항-요소의 관계를 재결정함에 있어서는 반드시 이론적 근거가 필요하다. 또한 이 방안을 사용하기 위해서는 바뀐 Q행렬 역시 각각의 인지요소를 측정하기에 적합한 형태여야 한다. 즉, Q행렬의 원소를 재결정함으로써 어떤 인지요소를 다루는 문항의 개수가 매우 적은 상황이나, 필요한 인지요소벡터에 해당하는 문항이 모자란 상황이 발생한다면 이 방안을 적용하는 것을 재고해볼 필요가 있다. 위의 예시상황에서 문항A에 대하여 Q행렬을 다음과 같이 변형해 주는 것이 *개선방안1*에 해당한다.

	V1	V2	V3	V4	V5
문항A	1	0	1	0	1

*개선방안1*은 Q행렬을 개선한다는 측면에서 기존의 Q행렬 정련방법들과 유사하지만, 데이터가 아닌 이론을 근거로 개선이 이루어진다는 것에서 차이점이 존재한다.

그러나 때때로 피험자들의 인지요소벡터를 분류하기 위해서 해당 문항이 가지고 있는 인지요소의 조합을 요구하는 문항이 꼭 필요할 경우가 있다. 이럴 경우 Q행렬을 변경하지 않고 문항을 개선함으로써 문항-요소관계를 바로잡을 수 있다. *개선방안2*는 문항-요소관계가 잘못 설정된 경우 문항을 개선하는 방안이다. 위에 예시상황에서 다음과 같이 문항A를 변형해주는 것이 *개선방안2*에 해당한다.

문항A. 다음 식을 계산 하여라. $(5-3)\times 2+7$

*개선방안2*는 Q행렬의 구조에 문제가 생기지 않도록 문항을 개선하기 때문에 검사 자체가 변형된다. 따라서 이렇게 재구성한 검사를 실시하여 다시 타당도를 분석할 수 있는 경우에만 사용 가능하다.

문항들에 새로운 세부능력이 필요할 경우 가장 쉽게 이를 개선할 수 있는 방안은 추가적으로 필요한 세부 능력을 배재하여 문항을 변형하는 *개선방안3*이다. *개선방안3*에서도 *개선방안2*와 마찬가지로 인지요소의 구성이나 Q행렬이 변하지 않기 때문에 기존 검사의 추정과정에 문제가 없었다면 개선방안 적용 후에도 추정과정에 문제가 생길 위험성이 없다. 위의 예시상황의 문항B에 *개선방안3*을 적용한다면 다음과 같이 문항을 변경할 수 있다.

문항B. 다음 식을 계산 하여라. $8\times 2+3$

그러나 *개선방안3* 역시 *개선방안2*와 마찬가지로 검사자체에 변형이 일어나기 때문에 이렇게 바뀐 검사의 타당도를 다시 분석할 수 있을 경우에만 사용가능하다.

만약 검사의 변형을 피해야 하는 상황이라면 문항들이 새로운 세부능력을 필요로 할 경우 이를 새로운 요소로 추가하는 *개선방안4*를 적용할

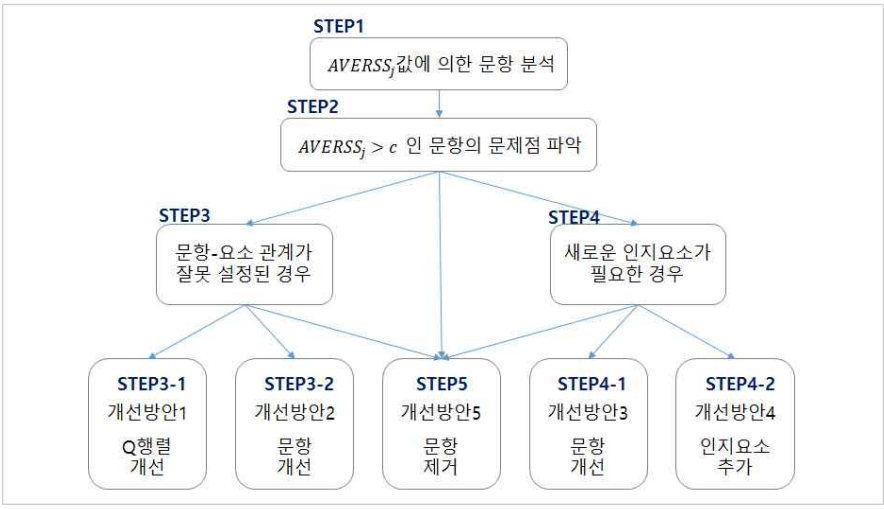
수 있다. *개선방안4*는 문항들이 공통적인 추가 세부능력을 필요로 할 때 사용할 수 있다. 예를 들어 위의 예시상황에서 덧셈과 곱셈의 계산 순서를 알아야 풀 수 있는 다른 문항들이 존재한다면 '덧셈과 곱셈의 계산 순서를 알고 있다'라는 새로운 인지요소 V6를 추가하는 것이 가능하다. 이때 문항B에 해당되는 Q벡터는 다음과 같이 변경 된다.

	V1	V2	V3	V4	V5	V6
문항B	1	0	1	0	0	1

*개선방안4*의 경우 인지요소의 개수가 증가하므로 각 인지요소 숙달 여부를 추정하는 과정에 문제가 생길 수 있다. 따라서 *개선방안4*를 적용할 때에는 Q행렬이 각각의 인지요소의 숙달여부를 추정하는 것이 가능한 구조인지를 꼭 확인해주어야 한다. 만약 연구자가 *개선방안4*를 적용하여 새로운 인지요소를 추가하고 싶으나 이 새로운 인지요소를 필요로 하는 문항의 개수가 너무 적어서 추정과정에 문제가 생길 수 있다면, 새로운 인지요소를 담고 있는 문항을 추가할 수 있다. 그런데 문항을 추가할 경우 검사의 변형이 발생하므로 재검사를 실시하여 그 타당도를 분석할 수 있을 경우에만 사용가능하다.

마지막으로 위의 개선 방향을 모두 적용하기 어려울 경우 $AVERSS_j$값이 큰 문항들을 제거함으로써 문제점을 없애버리는 *개선방안5*를 택할 수 있다. 예를 들어 어떤 $AVERSS_j$값이 큰 문항이 기존의 인지요소로는 설명될 수 없고, 해당 문항을 푸는데 필요한 새로운 인지요소를 추가하기에는 이 새로운 인지요소와 다른 검사문항들 간의 연관성이 떨어질 때, 해당 문항을 제거한 검사를 분석하는 것이 가장 합리적이다. *개선방안5*를 적용할 때의 유의점은 문항의 개수 감소로 인해 추정이 제대로 일어나지 않을 가능성을 고려해야 한다는 것이다. 즉, 해당 문항을 제거하여도 여전히 Q행렬이 각각의 인지요소의 숙달 여부를 측정하기에 적합한 구조인지를 확인해 주어야 한다.

Ⅳ-3.1, Ⅳ-3.2절에서는 위에서 제시한 절차를 통하여 3.1의 방정식과 부등식 평가를 개선하는 과정을 설명하였다.



[그림 Ⅳ-4] $AVERSS$를 이용한 검사의 개선 프로세스

3.1. $AVERSS_j$값이 큰 문항들에 대한 질적 분석

<표 Ⅳ-7>에서 방정식과 부등식의 평가에 사용된 일부 문항의 $AVERSS_j$값이 다른 문항보다 유달리 높다는 것을 알 수 있다. 다음 <표 Ⅳ-10>은 문항별 $AVERSS_j$값을 기준으로 검사 문항들을 분류한 것이다. 본 연구에서는 0.3을 문항분석대상의 컷오프 값으로 잡고, 0.3보다 큰 $AVERSS_j$값을 가지는 문항8, 문항14, 문항20, 문항21에 대하여 질적 분석을 진행하였다.

$AVERSS_j$의 범위	문 항
0.2미만	1,2,3,5,6,7,9,10,11,12,13,15,16,18,19
0.2이상 0.3미만	4,17,22,23,24,25
0.3이상 0.4미만	8,14,21
0.4이상	20

<표 Ⅳ-10> $AVERSS_j$기준으로 문항 분류

먼저 가장 큰 $AVERSS_j$값을 가지는 문항20을 살펴보면, 이 문항은 인지요소1(사칙계산 및 식의 계산), 인지요소2(등식의 이해), 인지요소6(부등식의 이해), 인지요소8(일차부등식 풀기(부등식의 성질 포함))을 필요로 한다. 이 4가지의 인지요소만으로 이루어진 문항에는 문항20외에도 문항18과 문항19가 있다. 그런데 문항18과 문항19에 비하여 문항20의 정답률은 현저히 떨어진다. 이 세 문항은 같은 인지요소조합을 가지므로 이상적으로는 모든 피험자에 대하여 세 문항의 응답이 같아야 한다. 따라서 정답률의 차이는 $AVERSS_j$값의 차이로 이어진다.

그렇다면 문항20이 문항18, 문항19와 다른 점은 무엇일까? 문항18, 문항19는 기본적인 일차 부등식을 푸는 문제로 미지수 x를 제외하고는 숫자상수들로 부등식이 구성되어있다. 반면, 문항20에 등장하는 부등식은 미지수 x이외에도 문자 상수 a를 포함하고 있다. 따라서 학생들은 단순히 식을 정리하고 양변을 숫자상수로 나누어주면 풀리는 문항18, 문항19에 비해 문항20을 어렵게 느꼈을 것이다.

18. 일차부등식 $3x-2<1$을 풀면? ① $x<4$ ② $x<1$ ③ $x>1$ ④ $x<0$ ⑤ $x>0$	19. 다음 부등식을 풀면? $0.2x+1<\frac{1}{5}(2x-1)$ ① $x=6$ ② $x>6$ ③ $x\geq 6$ ④ $x\leq -6$ ⑤ $x>-6$	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 $x<-\frac{1}{a}$

문항20의 풀이과정을 살펴보면 주어진 부등식의 좌변의 2를 이항시킨 후 a로 양변을 나누어준다. 이때 a가 음수이기 때문에 부등호의 방향을 바꾸어주어야 하므로 정답은 5번이다. 많은 학생들이 부등호의 방향을

바꾸는 것을 실패하여 4번으로 답할 것이라는 예상과 달리, 오답 중에서 1번의 비중이 월등히 높았다. 즉, 학생들은 음수로 나눌 때 부등호의 방향을 바꾸어야 한다는 것은 알고 있으나 a라는 문자 상수가 '-' 부호를 담고 있다는 개념이 잘 적립되어있지 않은 것이다. 이는 부등식에 관한 인지요소라기 보다는 문자 상수에 대한 이해에 관련된 부분이다. 따라서 문항20은 연구자가 잡은 10개의 인지요소로는 설명할 수 없는 기타 능력을 필요로 한다. 지난 연구에서는 이 능력에 해당하는 요소가 잡혀있지 않았기 때문에 문항20이 문항18, 문항19와 같은 유형의 문제로 분류된 것이다. 문항20외에 문자 상수를 사용하는 문항에는 문항8과 문항17이 있는데, 문항8의 경우 $AVERSS_j$값이 컷오프 값보다 큰 문항으로 분류되었고 문항17 역시 5번째로 $AVERSS_j$값이 큰 문항임을 <표 Ⅳ-7>에서 확인할 수 있다. 이로부터 문자 상수 존재가 $AVERSS$값을 증가시키는 하나의 요인으로 작용한 것이라 예상할 수 있다.

8. 다음 두 방정식의 해가 같을 때, 상수 a의 값을 구하여라. $3(\frac{2}{5}-0.1x)=0.2(2x+\frac{5}{2})$ $\frac{x+3}{6}=\frac{a-2x}{3}-1$ ① 3 ② 4 ③ 5 ④ 6 ⑤ 7	17. $2x-a \geq 3$의 해가 1을 포함할 때, 상수 a의 값의 범위는? ① $a<0$ ② $a \leq 0$ ③ $a<-1$ ④ $a \leq -1$ ⑤ $a<-3$

문항8도 문항20과 마찬가지로 문자 상수를 사용한다. 문항8의 경우 문자 상수를 조작하는 과정은 필요로 하지 않고, 단순히 대입을 통하여 문자 상수의 값을 구해주면 된다. 그리고 문항8은 문자 상수가 존재한다는 것 이외에도 다른 문제들과의 차이점이 존재하는데, 문제에 두개의 식이 주어진다는 것이다. 사실 이 두 식은 함께 조작될 필요가 없고, 첫 번째

식의 해를 구하여 두 번째 식에 대입해주면 문제를 풀 수 있다. 그러나 아직 연립방정식을 배우지 않은 학생들은 식이 두개 주어졌다는 것만으로도 다른 문제들보다 어렵게 느낄 수 있을 것이다.

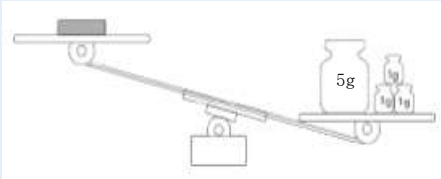
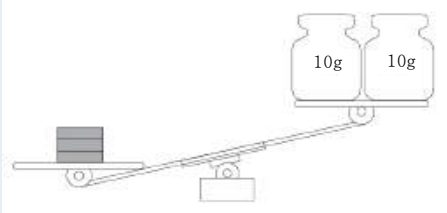
Item No.	A3	A4	Item No.	A3	A4
1			14		
2			15		
3			16		
4			17		
5	1		18		
6	1		19		
7		1	20		
8	1	1	21		
9	1		22		
10			23		
11		1	24		1
12		1	25		
13					

<표 Ⅳ-11> 방정식과 부등식 평가의 Q행렬에서 인지요소3, 인지요소4와 관련된 부분

문항8은 인지요소1(사칙계산 및 식의계산), 인지요소2(등식의 이해), 인지요소3(일차방정식의 해의 의미), 인지요소4(일차방정식 풀기), 총 4개의 인지요소와 관련되었으며 인지요소3(일차방정식의 해의 의미)과 인지요소4(일차방정식 풀기)를 동시에 담고 있는 유일한 문항이기도 하다(<표 Ⅳ-11> 참조). 다른 문항들은 주어진 일차방정식의 해를 구하거나, 주어진 해를 식에 대입하여 문제를 푸는 두 가지 문제풀이 중 한 가지를 요구하는 반면 문항8은 이 두 가지를 모두 필요로 한다. 문항8에 필요한 모든 인지요소를 습득한 학생이라 할지라도 하나의 일차방정식을 풀어서 얻은 해를 다른 식에 대입해야 해야 한다는 전략을 세우지 못하는 학생은 이 문제를 풀 수 없다. 이러한 문항8의 특징은 기대반응과 실제응답이 다른 경우들을 발생시켰을 것이라 예상된다.

문항21은 인지요소1(사칙계산 및 식의계산), 인지요소2(등식의 이해), 인지요소6(부등식의 이해), 인지요소8(일차부등식 풀기), 인지요소9(일차

방정식 활용), 인지요소10(대수적 모델링), 총 6개의 인지요소를 필요로 하는 복잡한 문항으로 같은 유형의 문항으로는 문항22와 문항23이 있다.

21. 무게가 같은 금속판 세 개가 있다. 금속판 한 개의 무게를 추 8g과 비교해보니 다음과 같았다.  금속판 세 개 전체의 무게를 추 20g과 비교해보니 다음과 같았다.  다음 중 금속판 한 개의 무게로 가능한 것은 어느 것인가? ① 5g ② 6g ③ 7g ④ 8g ⑤ 9g	22. 사진을 인화하는데 기본 10장은 5000원이고, 한 장 더 추가 시 250원을 받는다고 한다. 최소 몇 장 이상 뽑아야 1장의 가격이 300원 이하가 되는가? ① 40장 ② 50장 ③ 60장 ④ 70장 ⑤ 80장
	23. 현지네 반 학생들이 미술관에 갔더니 1인당 입장료가 7000원이고 20명 이상인 경우 10% 할인, 30명 이상인 경우 20% 할인 된 가격의 단체권을 구입하여 입장할 수 있다고 한다. 20명 이상 30명 미만의 단체는 최소 몇 명 이상일 때 30명의 단체권을 구입하여 입장하는 것이 유리한지 구하면? ① 25명 이상 ② 26명 이상 ③ 27명 이상 ④ 28명 이상 ⑤ 29명 이상

위에서 분석한 문항20의 경우 같은 구성의 문항18, 문항19에 비해 정답률이 낮아 $AVERSS_j$값이 커졌다면, 문항21은 같은 구성의 문항22, 문항23에 비해 정답률이 높아서 $AVERSS_j$이 커지는 효과를 가져왔다. 문

항21의 정답률이 같은 유형의 다른 문항들에 비해 높은 이유는 문장으로 서술된 상황을 부등식으로 세워야 하는 문항22, 문항23과 달리 문항21의 상황은 그림으로 제시되어 직관적으로 다가오기 때문이다. 더불어 보기에서 주어진 5가지 숫자를 상황에 대입해봄으로써 정답을 얻을 수 있기 때문에 부등식을 형식적으로 세우지 못한 학생들도 문제를 풀 수 있다.

그리고 문항21은 문항22, 문항23과 또 다른 차이점이 존재한다. 문항21을 정확히 풀어내기 위해서는 금속판 한 개의 무게를 미지수로 두고, 2가지 상황으로부터 두 개의 일차부등식을 세워 풀어내야 한다. 따라서 연립부등식의 개념이 필요하다. 즉, 문항21에는 Q행렬에서 나타내지 않고 있는 요소가 추가적으로 존재한다. 이 또한 문항21의 $AVERSS_j$값을 증가시키는 요인으로 작용했을 것으로 예상된다.

문항 14는 인지요소1(사칙계산 및 식의계산), 인지요소2(등식의 이해), 인지요소6(부등식의 이해)로 구성되어있는 문항이다. 이 문항을 풀기 위해서는 실제로 인지요소6(부등식의 이해)만이 필요해 보인다. 그러나 인지요소6(부등식의 이해)이 있으려면 인지요소1(사칙계산 및 식의계산), 인지요소2(등식의 이해)이 선행되어야한다는 위계관계가 설정되어 있기 때문에 추가적으로 이 두 개의 인지요소를 포함하게 되었다(윤지영, 2015). 문항14가 인지요소1(사칙계산 및 식의계산), 인지요소2(등식의 이해)과 직접적 관련이 없다는 것은 Q행렬이 문항14의 특성을 잘 나타내지 않고 있다고 볼 수 있다.

13. 다음 중 두 수의 대소 관계를 바르게 나타낸 것은? ① $5<3$ ② $1<-2$ ③ $-3>-1$ ④ $1+3<4$ ⑤ $2-3\le 0$	14. 다음 중 'a는 3보다 크지 않다'를 부등호를 사용하여 바르게 나타낸 것은? ① $a<3$ ② $a>3$ ③ $a\le 3$ ④ $a\ge 3$ ⑤ $a\ne 3$

문항14와 같은 유형으로 분류되는 문항에는 문항13이 있다. 문항13은 숫자만으로 이루어진 5개의 부등식의 참•거짓을 판별하는 문제이므로, 부등호의 개념을 이해하기만 하면 풀 수 있는 가장 기본적인 문제이다. 보기를 판별하는 과정에 계산이 포함되므로 인지요소1(사칙계산 및 식의 계산)을 필요로 하지만, 인지요소2(등식의 이해)를 직접적으로 포함하고 있지는 않다. 문항13와 문항14는 같은 유형으로 분류되어있지만, 숫자만으로 이루어진 문항13과 달리 문항14에는 a라는 변수가 등장하므로 피험자들이 체감하는 난이도에는 차이가 있을 수 있다.

또한 문항14의 또 다른 특징으로는 명제의 부정을 물어보는 문제로 볼 수 있다는 것이다. '3보다 크지 않다'를 수학적으로 표현하는 과정에서 언어적으로 '크다'의 부정을 물어보는 수학적 의사소통과 관계가 있다. 이러한 수학적 의사소통 능력 또한 명시되어있는 인지요소가 아닌 잔여능력이므로 오차를 발생시켰을 것이다.

이렇듯 $AVERSS_j$값이 0.3보다 큰 문항들을 분석해보니 Q행렬이 해당 문항들을 제대로 표현하지 못하고 있다는 공통점을 찾을 수 있었다. 이는 $AVERSS_j$값이 'Q행렬이 문항을 얼마나 잘 표현하고 있는지', 즉, 'Q행렬이 타당한지'를 드러내는 지표로 사용가능함을 보여준다.

질적 분석을 통해 알 수 있었던 가장 중요한 결과는 $AVERSS_j$값이 큰 문항들은 문항-인지요소관계가 아닌 다른 부분에서 문제를 가지고 있다는 것이다. 3.1의 방정식과 부등식 평가에서 측정하고자 하는 인지요소는 내용중심 인지요소로 그 나타내는 바가 명확하기 때문에 각 문항들이 어떤 인지요소를 담고 있는지 논란의 여지가 거의 없었다. 그럼에도 불구하고 Q행렬이 타당도가 낮았던 이유는 기존 10개의 인지요소 만으로 문항을 표현할 수 없는 경우들이 발생하였기 때문이다. 이러한 경우들을 기존의 Q행렬 정련방법들로 개선될 수 없는 부분이다. 기존의 Q행렬 정련방법들은 전문가가 제시한 인지요소의 구성은 전적으로 신뢰하고 검사의 모든 문항이 제시된 인지요소들의 조합으로 표현 가능함을 가정한다. 따라서 Q행렬에 명시되지 않은 새로운 요소가 필요한 경우에 이에 대한 개선은 이루어지지 않은 채 잘못된 정련을 시행할 수 있다. 반면,

$AVERSS$를 활용한 검사의 개선 방법에서는 Q벡터의 타당성이 낮은 문항들을 질적 분석함으로써 Q행렬의 문제점을 보다 정확히 진단할 수 있다.

3.2. 방정식과 부등식 검사의 개선

본 연구에서는 $AVERSS_j$값이 어떤 컷오프 값보다 큰 문항들을 분석한 후에 취할 수 있는 검사의 개선 방안으로 5가지를 제시하였다. 4.3.1절에서 $AVERSS_j$값이 큰 문항을 분석한 결과 이 문항들은 제시된 10개의 인지요소만으로 표현되기 어려운 다른 세부능력을 필요로 함을 알 수 있었다. 이러한 상황에서 검사를 개선하는 방향으로는 *개선방안3, 개선방안4, 개선방안5* 세 가지 방법이 있다.

먼저 문항들의 문제가 될 만한 부분을 제거하여 문항이 해당 인지요소만의 조합으로 설명 가능하도록 바꾸는 *개선방안3*을 적용하면, 문자 상수에 대한 이해가 필요했던 문항20을 다음과 같이 변형할 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 $x<-\frac{1}{a}$	20. 일차부등식 $-3x+2>1$를 풀면? ① $x<-\frac{1}{3}$ ② $x>-\frac{1}{3}$ ③ $x<-\frac{5}{3}$ ④ $x>\frac{1}{3}$ ⑤ $x<\frac{1}{3}$

이렇게 바꾸어준 문항20의 경우 더 이상 논란의 여지를 가지지 않으며, 같은 인지요소 조합을 가졌던 문항18, 문항19와의 난이도 차이도 존재하지 않는다. 다른 $AVERSS_j$값이 큰 문항들도 아래와 유사한 방법으로 추가적으로 필요한 요인을 제거시킴으로써 Q행렬에 명시된 인지요소의 조합만을 활용하여 문항을 풀 수 있도록 문항을 변형할 수 있다.

이렇듯 *개선방안3*에서는 $AVERSS_j$값이 큰 문항들을 특이점을 제거된 형태로 변형한다. 이럴 경우에는 인지요소의 개수나 Q행렬이 바뀌지 않으므로 추정 과정의 위험성은 존재하지 않는다. 그러나 이는 검사를 제작하는 단계에서 가능한 개선 방향이기 때문에 이미 진행된 검사를 분석해야 하는 현재 상황에서는 시행하기 어렵다.

기존의 검사를 그대로 활용하면서 검사를 개선하는 방안에는 *개선방안4*와 *개선방안5*가 있다. *개선방안4*는 $AVERSS_j$값이 큰 문항들에게서 새로 발견된 세부능력들을 인지요소로 추가해주는 것이다. Ⅳ-3.2절에 따르면 $AVERSS_j$값이 0.3보다 큰 문항 4개 중 2개(문항20, 문항8)가 '문자 상수의 이해'라는 공통된 추가 능력을 필요로 한다. 또한 이는 5번째로 $AVERSS_j$값이 큰 문항17에서도 필요한 능력이기도 하다. 따라서 *개선방안4*를 적용하여 검사를 개선하는 방법으로 인지요소 목록에 '문자 상수의 이해'를 추가하는 것을 고려해 볼 수 있다. 이렇게 새로운 인지요소를 추가해주는 방법은 $AVERSS_j$값이 큰 3개의 문항들의 문제점을 한 번에 해결해 준다. 그러나 '문자 상수의 이해'라는 새로운 인지요소는 단 3개의 문항에만 필요한데다가 기존 인지요소들과 위계구조로 연결되지않기 때문에 가능한 인지요소벡터의 개수를 2배 증가시킨다. 따라서 '문자 상수의 이해' 인지요소를 추가하여 Q행렬을 변형할 경우 피험자의 인지요소벡터를 분류하는 추정과정에 문제가 생길 수 있다. 이 문제점을 해결하기 위해서는 '문자 상수의 이해'와 관련된 문항들을 몇 개 더 추가하여 각각의 인지요소벡터에 해당되는 문항이 필요한 만큼 존재하도록 검사를 변형시켜야 한다. 그러므로 이미 진행된 검사를 그대로 활용하여 분석해야 하는 현재 상황에서는 새로운 인지요소를 추가하는 것은 오히려 측정을 방해할 수 있다.

위의 두 가지 개선이 불가능할 때, 택할 수 있는 개선 방안으로는 문제점을 가진 문항들을 제거하는 *개선방안5*가 있다. 본 연구에서는 인지요소의 개수를 늘리는 대신 $AVERSS_j$값이 큰 문항들을 제거하는 개선을 시행하였다. 다음 <표 Ⅳ-12>의 6가지 검사는 Ⅳ-3.1절에서 다루었던 4개의 문항(문항8, 문항14, 문항20, 문항21)을 공통점을 가진 것들끼리 조합해서 이들을 기존검사에서 제거하여 만든 검사들이다.

	제거 문항	공통점
검사1	20	$AVERSS_j$값이 가장 큰 문항
검사2	17,20	문자 상수 포함 (조작 필요)
검사3	8,17,20	문자 상수 포함
검사4	8,21	연립 방정식 개념
검사5	14	명제의 부정
검사6	8,14,17,20,21	$AVERSS_j$값이 큰 5개의 문항 제거

<표 Ⅳ-12> 문항 제거 후 검사의 재구성

제거문항에 문항17이 포함된 이유는 문항17이 문항20과 마찬가지로 문자 상수를 포함하고 있을뿐더러 문항17의 $AVERSS_j$값 역시 약 0.266으로 4.3.1절에서 분석한 4개의 문항 다음으로 크기 때문이다.

검사1은 문항별 $AVERSS_j$값이 가장 컸던 문항20만을 기존검사에서 제거한 검사이다. 검사2는 문자 상수가 등장하고 이를 조작하는 과정이 필요한 문항17과 문항20을 기존검사에서 제거한 검사이고, 검사3은 문자 상수가 등장하는 문항8, 문항17, 문항20을 제거하고 남은 문항들로 구성된 검사이다. 검사4는 연립방정식의 개념이 필요한, 즉 문항에 식이 2개 등장하는 경우인 문항8과 문항21을 기존검사에서 제거한 검사이다. 검사5는 명제의 부정을 이해할 수 있는지를 담고 있었던 문항14만을 제거한 검사이다. 마지막으로 검사6은 위에 등장한 5개의 문항을 모두 제거한 검사이다. 위 여섯 가지 검사에 대하여 피험자의 인지요소벡터를 다시 분류한 후, 이를 이용하여 검사의 $AVERSS$를 구한 결과를 <표 Ⅳ-13>에서 확인할 수 있다.

검사 종류	*AVERSS*
기존 검사	0.1885
검사1	0.1753
검사2	0.1696
검사3	0.1628
검사4	0.1701
검사5	0.1816
검사6	0.1392

<표 Ⅳ-13> 검사 1~6의 *AVERSS*

검사의 문항에 비해 인지요소의 개수가 많을수록 검사의 타당도는 유리하다. 요인분석의 원리에 따르면 결과를 설명하는 요인이 많을수록 상황을 더 잘 설명할 수 있는 모델을 세울 수 있으므로 모델 핏(model fit)이 더 좋을 수밖에 없기 때문이다. 그러나 요인이 많다는 것은 정해진 데이터 안에서 추정해야할 변수가 많다는 의미이므로 추정이 정확하게 이루어지지 않을 수도 있다는 불확실성이 뒤따른다. 즉, 요인이 많은 수록 결과를 더 잘 설명할 수 있는 모델을 세울 수 있지만, 요인의 개수가 너무 많아지면 모수를 추정하는 과정에서 오류가 발생할 확률 역시 크다. 그러므로 모델을 세울 때에는 비교적 상황을 잘 설명할 수 있으면서 추정 또한 어느 정도 신뢰할 수 있는 접합 점을 찾는 것이 매우 중요하다.

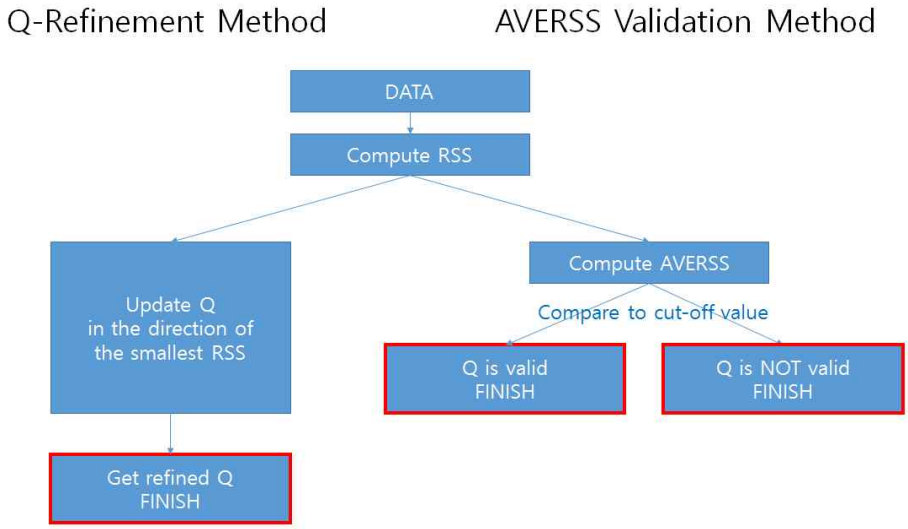
검사6의 경우 *AVERSS*값이 가장 작으므로 Q행렬의 타당도가 가장 높다고 할 수 있다. 특이점을 가지고 있는 문항을 모두 제거하였으므로 기존검사에 비해서 *AVERSS*값이 크게 감소하였다. 그러나 인지요소의 개수는 고정된 상태에서 문항의 개수만 20개로 줄어들어 문항 수에 대한 인지요소 수의 비가 증가하였으므로 타당도가 증가하는 것은 당연한 결과이기도 하다. 또한, 문항 20개로 10개의 인지요소의 숙달 여부를 판정해야 하므로 피험자의 능력 추정이 잘못 이루어졌을 위험성이 크다.

한 문항만 제거하였을 경우, 문항14를 제거한 검사5보다는 문항20을

제외한 검사1의 타당도가 더 높았다. 두 문항을 제거하였을 경우, 문자 상수의 조작을 필요로 하는 문항들(문항17, 문항20)을 제외한 검사2의 타당도가 연립방정식의 개념을 필요로 하는 문항들(문항8, 문항21)을 제외한 검사4의 타당도 보다 높았다. 문자 상수가 등장하는 세 개의 문항을 제거한 검사3의 경우 검사2보다$AVERSS$값이 0.7가량 감소했다. 본 연구에서는 추정의 안정성과 Q행렬의 타당성 두 가지를 고려하였을 때, 검사2 혹은 검사3을 활용하는 것을 제안한다.

4. 두 타당화 방법 비교

기존 Chia-Yi Chiu의 Q행렬 정련방법과 $AVERSS$를 이용한 타당성 분석 방법은 RSS라는 타당성지수를 활용하는 양적 분석을 포함한 다는 점에서 공통점을 가지지만, 그 접근 방향은 전혀 다르다. 기존 Q행렬 정련방법은 데이터를 통해서 Q행렬을 개선해나가는 탐색적 접근의 방법인 반면 $AVERSS$ 타당성 분석 방법의 경우 전문가가 제시한 Q행렬의 타당성을 입증하고자 하는 확인적 접근의 방법이다. 다음의 [그림 Ⅳ-5]는 두 타당화 방법의 프로세스를 보여준다.



[그림 Ⅳ-5] Chia-Yi Chiu의 Q행렬 정련방법과 $AVERSS$ 타당성 분석 방법의 프로세스

만약 정확한 Q행렬에 Chia-Yi Chiu의 Q행렬 정련방법을 적용시킬 경우, 모든 Q벡터의 문항별 RSS_j값이 최소가 아닌 이상 정련이 일어난다. 반면 $AVERSS$ 타당성 분석에서는 모의실험 데이터 분석 결과 정확한 Q행렬에 대하여 95%의 확률로 타당하다고 판정한다. 반대로 어떤 검사의 Q행렬이 잘못된 경우, Chia-Yi Chiu의 Q행렬 정련방법을 적용하였을 때

데이터의 오류에 의해 잘못 정련이 일어날 가능성이 있는 반면(Ⅳ-2절 참조), *AVERSS* 타당성 분석에서는 *AVERSS*값이 컷오프 값보다 크다면 타당하지 않다고 판정하는데 그치고 정련과정은 일어나지 않는다. 모의 실험 데이터 분석 결과 *AVERSS* 타당성 분석 방법은 Q행렬의 원소가 3%이상 잘못된 경우 93% 이상의 확률로 타당하지 않다고 판정한다.

Chia-Yi Chiu의 Q행렬 정련 방법을 포함한 기존의 정련 방법들은 검사데이터를 가장 잘 설명할 수 있는 Q행렬을 타당성지수를 사용하여 찾아나갔다면, 이 새로운 방법에서는 검사 데이터의 오차를 어느 정도 고려하여 타당성지수가 컷오프 값보다 작기만 하다면 전문가가 제시한 Q행렬이 타당하다고 간주한다. 즉, *AVERSS* 타당성 분석 방법은 데이터의 허용 가능한 오차범위를 자체적으로 가지고 있다. 실제 인지진단 검사에서는 피험자의 추측 및 실수, 제한 시간 등의 다양한 변인이 존재하고 이는 기대반응과 실제반응이 다른 경우, 즉, 데이터의 오류를 발생시킨다. 이러한 측면에서 *AVERSS* 타당성 분석 방법은 검사데이터의 오류에 민감하게 영향을 받는 기존의 정련 방법들보다 실질적으로 인지진단 검사에 적용 가능한 방법이라 할 수 있다.

그리고 Ⅳ-3절에서는 *AVERSS*를 활용하여 검사의 개선하는 프로세스를 제안하였다. 이는 단순히 Q행렬을 개선하는 것이 아니라 검사 자체를 재구성하는 방법이다. 기존 Q행렬 정련방법들은 검사가 고정된 상태에서 데이터를 올바르게 분석하기 위해 사용되므로 검사의 문항, 추출된 인지요소, 피험자의 응답 데이터 자체는 변하지 않는 상태에서 문항과 인지요소의 관계만을 개선한다. 반면 *AVERSS*를 활용한 검사의 개선방법은 검사를 제작하는 단계에서 적용 가능한 방법으로 검사 문항 혹은 인지요소의 구성도 가변적인 것으로 취급한다. 즉, 이 방법에서 말하는 개선은 검사 문항을 변형시키거나 인지요소를 추가하는 것 등을 포함한다. 그런데 어떤 인지진단 검사의 문항-인지요소의 타당성이 떨어지는 경우, 문항과 인지요소와 관계가 잘못 설정되었을 수도 있지만, 문항이 기존 인지요소들의 조합으로 설명되지 않을 수도 있으며, 실제로 Ⅳ-3.1절에서 문항-인지요소 관계의 타당성이 낮은 문항들을 분석해본 결과 해당

문항들은 모두 새로운 인지요소를 필요로 하였다. 그러므로 $AVERSS$를 활용한 검사의 개선방법은 문항과 인지요소의 구성 자체는 고정된 상태에서 이미 제시된 인지요소들과 문항 사이의 관계에만 주목하는 기존의 Q행렬 정련방법들보다 더욱 본질적인 개선이 가능하다.

5. 이탈모수가 $AVERSS$에 미치는 영향

RSS_{ij}값은 피험자i의 문항 j에 대한 기대반응(η_{ij})과 실제응답(Y_{ij})의 차이를 나타내는 수치이다. 따라서 RSS_{ij}값이 1이 되는 상황은 이탈상황과 추측상황으로 요약될 수 있다.

	$Y_{ij}=1$	$Y_{ij}=0$
$\eta_{ij}=1$	$RSS_{ij}=0$	$RSS_{ij}=1$ (이탈)
$\eta_{ij}=0$	$RSS_{ij}=1$ (추측)	$RSS_{ij}=0$

<표 Ⅳ-14> 이탈-추측 상황과 RSS의 관계

그러므로 검사를 구성하는 문항들의 이탈모수나 추측모수가 크다는 것은 $RSS_{ij}=1$일 확률 높다는 의미이고, RSS_{ij}의 평균인 $AVERSS$ 역시 큰 값을 가질 확률이 높다. <표 Ⅳ-15>와 [그림 Ⅳ-6]에서 검사 문항의 이탈모수가 클수록 $AVERSS$도 큰 값을 가짐을 확인할 수 있다.

이탈 모수	$AVERSS$의 평균	$AVERSS$의 컷오프 값
0.05	0.11648964	0.121760
0.1	0.15105620	0.156282
0.2	0.22018128	0.224802

<표 Ⅳ-15> [그림Ⅳ-6] 검사 문항의 이탈모수에 따른 $AVERSS$의 평균 및 컷오프 값

그러나 문항 모수가 큰 것이 반드시 검사의 제작과정에 오류가 있음을 의미하는 것은 아니다. 따라서 문항 모수의 크기가 다른 검사들에 같은 기준을 적용하여 $AVERSS$를 분석하는 것은 옳지 않다. 본 연구에서

는 검사의 문항 모수가 다른 3가지 경우에 대하여 각각 $AVERSS$값의 컷오프 값을 제시함으로써 $AVERSS$를 활용한 타당성 분석 방법의 틀을 구축하였다. 실제 검사에서는 이탈모수가 0.05, 0.1, 0.2뿐 아니라 여러 가지 값을 가질 수 있으므로 본 연구에서 제시한 3개의 컷오프 값들은 대략적인 기준을 제시하는 정도라 할 수 있다.

$AVERSS$ 타당성 분석 방법이 다양한 검사상황에 맞게 활용되기 위해서는 문항 모수에 따라 컷오프 값이 어떻게 변하는지에 대한 연구가 진행되어야 한다. 본 연구에서는 추측모수는 일괄적으로 0.2로 통일하여 이탈모수에 따른 변화에만 초점을 두었으나 $AVERSS$의 분포에 추측모수가 미치는 영향 역시 이탈모수 만큼이나 클 것이라 예상된다. 따라서 다양한 이탈-추측모수 조합의 상황에서 $AVERSS$의 분포를 분석할 필요가 있다. 또한 본 연구에서 진행한 모의실험에서는 검사 문항별로 이탈, 추측모수의 차이를 두지 않았으나 실제 검사에서는 각 문항마다 문항모수가 다르므로 다양한 이탈-추측모수를 가지는 문항들로 이루어진 검사의 경우 어떠한 컷오프 값을 택해야 하는지에 대한 논의 역시 필요하다.

$AVERSS$ 타당성 분석 방법을 실제 검사에 적용하는 것에 있어서 가장 어려운 점은 $AVERSS$의 분포가 이탈-추측모수에 민감하게 반응한다는 것이다. $AVERSS$의 분포를 살펴보면(Ⅳ-2절 참조) 잘못된 Q행렬의 $AVERSS$값을 구했을 때 $AVERSS$값이 증가하는 정도에 비하여 검사 문항들의 이탈모수에 의해 받는 영향이 훨씬 더 크다는 것을 확인할 수 있다. $AVERSS$ 타당성 분석 방법에 사용되는 컷오프 값은 검사 문항들의 이탈-추측모수에 따라 크게 달라지기 때문에 $AVERSS$타당성 분석 방법을 적용하기 위해서는 검사 문항들의 이탈-추측모수를 정확하게 측정하는 과정이 선행되어야 한다.

결과적으로 $AVERSS$ 타당성 분석 방법은 이탈-추측모수의 영향을 크게 받는다는 것에서 두 가지 불편함이 존재한다. 첫 번째는 다양한 이탈-추측모수 조합을 가지는 검사 상황에 따라 $AVERSS$의 분포에 대한 연구가 이루어져야 한다는 것이고, 두 번째는 검사 문항의 이탈-추측모수를 정확하게 측정하는 과정이 선행되어야 한다는 것이다.

Ⅴ. 논의 및 결론

1. 연구 결과 요약 및 논의

본 연구의 목적은 수학 인지진단 검사를 제작하는데 필요한 Q행렬의 타당성을 판정하는 양적인 방법을 개발하는 것에 있다. 이에 대해 기존의 Q행렬 정련 방법의 문제점에 대한 분석이 이루어졌으며, 이를 보완한 *AVERSS* 타당성 분석 방법을 고안하였다. 또한, 모의실험 데이터를 이용하여 이 새로운 방법에 대한 효율성을 살펴보았으며, *AVERSS*를 활용하여 인지진단 검사를 개선하는 프로세스를 제시하였다. 다음은 연구 결과로부터 얻어낸 본 연구의 결론이다.

첫째, 실제 인지진단 검사에 사용된 Q행렬에 Chia-Yi Chiu의 Q행렬 정련방법을 적용해본 결과, 정련이 제대로 일어나지 않음을 알 수 있었다. 이를 위해 936명을 대상으로 하는 일차방정식과 일차부등식 검사의 자료가 활용되었다. 이 검사의 Q행렬에 Q행렬 정련방법 적용한 결과 25개의 문항 중 16개의 문항에 대한 Q벡터가 변경되었으며, 문항과 인지요소를 대조해본 결과 문항과 전혀 상관없는 인지요소가 추가되거나 필수적인 인지요소가 제거되는 등 이론적으로 설명할 수 없는 정련이 일어났음을 알 수 있었다. 본 연구에서는 이러한 문제점의 원인으로 타당성지수가 가장 작을 때가 정확한 Q행렬이라는 Q행렬 정련방법의 가정을 지목하고, 유사한 가정을 가지는 Hamming모형의 실효성을 살펴보았다. Hamming모형으로 일차방정식과 일차부등식 검사를 분석한 결과 피험자의 인지요소벡터 분류가 제대로 일어나지 않음을 확인할 수 있었다.

이는 실제 인지진단 검사에 탐색적 접근의 Q행렬 정련방법들을 적용하는 것이 가지는 위험성을 제기한다. 탐색적 접근의 Q행렬 정련방법은 검사 데이터를 측정할 때 생긴 여러 가지 오류로 인해 잘못된 Q행렬을 개선된 Q행렬로 제시할 수 있다. 따라서 다양한 변인에 따른 다양한 오류가 발생 가능한 실제 인지진단 검사에 탐색적 접근의 Q행렬 정련방법

을 적용한 결과는 이론과 전혀 맞지 않을 가능성이 있다.

둘째, 본 연구에서 개발한 확인적 접근의 $AVERSS$ 타당성 분석 방법의 효율성과 적용가능성을 확인하였다. 본 연구에서는 Chia-Yi Chiu의 Q행렬 정련방법에서 사용한 문항별 RSS_j값 대신 Q행렬 전체의 타당성을 대표하는 타당성지수 $AVERSS$를 정의하였다. $AVERSS$는 전체 피험자와 문항에 대하여 이상반응과 실제응답이 다를 경우의 비율을 나타내는 값으로 검사의 규모에 의존적이지 않는 타당성지수이다. 타당성 판정에 필요한 $AVERSS$의 컷오프 값을 얻기 위해서 모의실험 데이터로부터 $AVERSS$의 분포를 분석한 결과 검사 문항의 문항모수에 따라 $AVERSS$의 분포가 달라짐을 확인할 수 있었다. 본 연구에서는 추측모수가 0.2로 고정된 상태에서 3가지 이탈모수 상황에 따른 $AVERSS$ 컷오프 값을 $AVERSS$ 분포의 95 백분위수의 값으로 각각 다르게 제시하였다. 이렇게 제시한 컷오프 값에 대하여 $AVERSS$ 타당성 분석방법의 효율성을 살펴본 결과, 3%이상 잘못된 Q행렬에 대하여 93% 이상의 확률로 타당하지 않다고 판정함을 알 수 있었다. 또한 $AVERSS$ 타당성 분석방법을 방정식과 부등식 검사에 적용해봄으로써 검사 문항 모수에 대한 정보가 주어질 경우 실제 인지진단 검사에 적용 가능한 방법임을 확인하였다.

이 연구는 확인적 접근으로 Q행렬의 타당성을 보는 양적인 방법을 개발하였다는 점에서 그 의의가 있다. 이전까지는 Q행렬의 타당성에 관하여 검사 데이터를 통한 통계적 분석 없이 이론만을 근거로 하는 분석이나, 데이터의 오류에 대한 방어능력이 약한 탐색적 접근의 방법을 통해 분석이 이루어졌다. $AVERSS$ 타당성 분석 방법은 전문가가 이론을 근거로 제시한 Q행렬을 검사 데이터를 사용하여 확인하는 확인적 접근의 양적인 방법으로 실제 인지진단 검사에 보다 적합한 형태의 방법이다.

마지막으로 $AVERSS$를 활용하여 검사를 개선하는 프로세스를 제시하였고, 이 개선방법의 적용가능성을 확인하였다. 본 연구에서는 문항별 $AVERSS_j$ 분석을 통하여 타당성이 낮은 Q벡터에 해당하는 문항들을 질적 분석하고, 문항분석 결과에 따른 5가지의 개선방안을 상황에 맞게 적용하여 검사를 개선하는 검사의 개선방법을 구축하였다. 그리고 실제 방

정식과 부등식 검사를 이 프로세스를 활용하여 개선해보았다. 이 검사에 사용 문항들 중에서 $AVERSS_j$값이 0.3보다 큰 5개의 문항을 분석한 결과 모두 기존의 인지요소 이외의 새로운 인지요소를 필요로 함을 알 수 있었다. 이럴 경우 적용 가능하다 3가지 개선방안 중에서 문항을 제거하는 *개선방안5*를 적용하였고, 유사한 특이점을 가지는 문항끼리 그룹지어 기존검사에서 제거한 6개의 검사에 대하여 Q행렬의 타당성을 분석한 결과 기존 검사보다 타당도가 높아졌음을 확인할 수 있었다. 이때 문항제거 후 Q행렬의 타당성을 분석하는 과정에서는 $AVERSS$ 타당성 분석 방법이 활용되었다.

$AVERSS$를 활용한 검사의 개선 방법이 가지는 특이점은 Q행렬의 타당성지수를 통해 '문항-요소의 관계' 이상의 것들을 보고자했다는 것이다. 이는 타당성지수 $AVERSS$를 Q행렬에 타당도를 보는 것 외에 수학인지진단 검사에 활용할 수 있는 또 다른 측면을 제시하기도 한다.

또한, $AVERSS$를 활용한 검사의 개선 방법은 검사를 제작하는 과정에서 사전검사의 연구결과를 사용하여 검사를 개선하는 가이드라인을 제시했다는 점에서 의미가 있다. 기존의 Q행렬 정련방법들은 고정된 검사에 대하여 문항-인지요소의 관계만을 개선하였다면, $AVERSS$를 통한 검사의 개선방법은 문항 혹은 인지요소의 구성도 가변적인 것으로 취급하기 때문에 보다 본질적인 개선이 가능하다. 실제 인지진단 검사에서는 어떤 문항에 대한 Q벡터의 타당성이 떨어질 경우, 문항-인지요소의 관계가 잘못 설정되었을 수도 있지만 새로운 인지요소를 필요로 하는 경우가 종종 있다. 이러한 검사에 대하여 기존의 Q행렬 정련방법을 적용할 경우 정확한 문제점을 찾아낼 수 없을뿐더러, 해당 문항으로 인해 발생한 오류로 인해 잘못된 정련이 일어날 가능성이 크다. 이러한 측면에서 $AVERSS$를 통한 검사의 개선방법은 문항의 질적 분석을 통한 실질적인 개선이 가능하다는 장점이 있다.

2. 연구의 제한점 및 제언

본 연구는 실제 수학 인지진단 검사에 적용 가능한 Q행렬의 타당성을 분석하는 확인적 접근의 양적인 방법을 개발하였다는 것에 그 의의가 있으나 다음과 같은 한계를 가지고 있다.

첫째, 기존 Q행렬 정련방법과 *AVERSS* 타당성 분석 방법의 실효성을 논의함에 있어서 충분한 적용연구를 거치지 못하였다. 어떤 Q행렬 타당화 방법이 실제 인지진단 검사에 적용 가능한지를 살펴보기 위해서는 다양한 인지진단 검사에 대하여 적용하여 그 실효성을 분석하는 절차가 필요하다. 본 연구에서 분석에 활용된 실제 인지진단 검사는 방정식과 부등식 평가 단 한가지로 이 평가의 Q행렬에 대하여 *AVERSS* 타당성 분석 방법의 적용가능성을 알아보았다. 물론 모의실험 데이터를 활용하여 *AVERSS* 타당성 분석 방법의 효율성을 논하였으나, 실제 인지진단 검사에서는 보다 예측 불가능한 요인들이 작용 가능하므로 실제 검사의 Q행렬에 대한 이 새로운 타당성 분석 방법의 유효성은 달라질 수 있다. 따라서 *AVERSS* 타당성 분석 방법의 실효성 입증을 위해 다양한 실제 인지진단 검사에 대한 적용 연구가 필요하다.

둘째, 타당성지수 *AVERSS*의 분포에 대한 연구가 더욱 필요하다. 본 연구에서는 가상데이터를 통하여 이탈모수에 따른 *AVERSS*의 분포를 분석하였는데, 그 결과 이탈모수가 클수록 *AVERSS*값도 커짐을 확인하였다. 그러나 이탈모수와 유사한 영향을 미칠 수 있는 추측모수로부터 받는 영향에 대한 분석이나, 다양한 이탈-추측모수 조합에 대하여 *AVERSS*의 분포를 알아보는 과정은 이루어지지 않았다. 본 연구에서는 검사 문항들이 동일한 문항모수를 갖는 상황에서 3가지 이탈-추측모수 조합에 대한 *AVERSS* 컷오프 값만을 제시하였고, 이는 *AVERSS*의 컷오프 값에 대한 가이드라인을 제공하는 정도라 할 수 있다.

AVERSS 타당성 분석 방법은 문항모수에 따라 다른 컷오프 값을 사용하기 때문에 이 방법의 구축을 위해 다양한 문항모수 조합에 따른 *AVERSS*값의 컷오프 값이 제시되어야 한다. 이를 위해 다양한 문항모수

조합의 검사에 대하여 $AVERSS$의 분포를 알아보는 연구가 필요하며, 검사문항들의 이탈-추측모수가 주어졌을 때 $AVERSS$의 컷오프 값을 구할 수 있도록 도표나 관계식이 제시되어야 한다.

셋째, $AVERSS$ 타당성 분석 방법은 검사 문항들의 문항모수로부터 매우 큰 영향을 받는 다는 단점을 가진다. 본 연구에서는 $AVERSS$ 분포가 검사 문항의 이탈모수로부터 매우 큰 영향을 받음을 확인하였다. 물론 $AVERSS$ 타당성 분석 방법은 잘못된 Q행렬을 안정적으로 타당하지 않다고 판정할 수 있는 방법임이 확인되었으나, 이는 고정된 문항모수 조합의 검사 데이터에 대하여 $AVERSS$가 매우 작은 표준편차를 가지고 있기 때문이다. 실제로 잘못된 Q행렬의 $AVERSS$값은 정확한 Q행렬로부터 얻은 $AVERSS$ 분포의 평균값과 큰 차이가 나지 않았고, 이 차이는 검사의 이탈모수에 의해 $AVERSS$의 분포가 변화하는 정도에 비해 매우 미미했다. 즉 $AVERSS$ 타당성지수는 Q행렬의 오류에 비해 검사의 문항모수로부터 받는 영향이 더 크고, 이는 정확한 검사의 문항모수 정보로부터 $AVERSS$의 컷오프 값을 얻어야 함을 의미한다. 따라서 $AVERSS$ 타당성 분석 방법은 검사 문항들의 문항모수에 대한 정확한 측정이 선행되어야만 사용가능하다.

이러한 문제점은 피험자의 전반적인 문제풀이능력을 나타내는 연속적 모수의 필요성을 제시한다. Q행렬의 타당성을 분석하는 과정은 이상적인 모델 적용 결과와 실제 결과의 차이, 즉, 기대반응과 실제응답이 다른 정도에 대한 분석에서 출발한다. 그런데 DINA 모델의 경우 기대반응과 실제응답이 다른 상황을 각 검사 문항의 특성인 문항의 이탈모수와 추측모수로만 설명하기 때문에 검사 문항들의 이탈-추측모수가 타당성 분석 과정에 매우 큰 영향을 미칠 수밖에 없는 구조이다. 타당성 분석 과정에 있어서 검사 문항 모수의 역할을 줄임으로써 안정성을 확보하기 위해서는 피험자의 전반적 문제풀이능력에 관련된 연속적 모수가 필요하다. 이러한 측면에서 피험자의 일반적인 능력을 나타내는 연속적 모수를 사용하는 Fusion모델을 중심으로 Q행렬 타당성 분석 방법에 대한 연구의 필요성을 남겨둔다.

<참고문헌>

김동화(2003). 현행 입시제도에 따른 수학학습불안과 학습성취도에 미치는 영향. 교육이론과 실천, 13(2). 41-56

김희경, 한정아, 최숙기, 김부미(2012). 인지진단모형을 적용한 학업성취 프로파일 분석 및 결과 보고 방안, 한국교육과정평가원.

김희경, 김부미(2013). 인지진단모형을 활용한 수학 학업성취 결과 분석 - 2011년 국가수준 학업성취도 평가 자료를 중심으로. 학교수학, 15(2). 289-314

윤지영(2015). 요소 간 위계 방식과 인공신경망을 적용한 수학 인지진단 평가 연구, 서울대학교 석사학위논문.

이영주(2014). 인공신경망에 근거한 인지진단모형 Q행렬의 타당성 평가. 이화여자대학교 대학원 박사학위논문.

장상호(2004). 학교교육의 정상화를 위한 교육평가의 재구성. 교육원리연구, 9(1). 83-176

최승현, 박상욱, 황혜정(2014). PISA와 TIMSS 결과에 나타난 우리나라 학생의 정의적 성취 실태 분석-수학 교과를 중심으로-. 한국학교수학회논문집 17(1). 23-43

최승현, 황혜정(2014). 수학 교과에서의 학생의 정의적 특성 요인의 성취 실태-국내 중등 수업 사례를 중심으로-. 한국수학교육학회지 시리즈 E-<수학교육논문집>, 28(2). 235-253

한혜숙, 최계현(2011) 중등 수학 교사들의 정의적 특성에 대한 인식과 수업 실태 분석. 한국학교수학회논문집 14(4) 491-518

Baker, F. B. (2001). The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation.

Baker, F. B. & Kim, S.-H. (2004). Item Response Theory. Marcel Dekker, Inc.

Barnes, T. (2010). Novel derivation and application of skill matrices: The q-matrix method. In C. Ramero, S. Vemtora, M.

Pechemizkiy, & R.S.J.de Baker (Eds.), *Handbook of educational data mining (pp.159-172)*. Boca Raton, FL: Champman & Hall

Chiu, C.-Y., & Douglas, J. A. (2013). A nonparametric approach to cognitive diagnosis by proximity to ideal item response patterns. *Journal of Classification, 30.* 225-250.

Chiu, C.-Y. (2013). Statistical Refinement of the Q-Matrix in Cognitive Diagnosis. *Applied Psychological Measurement, 37(8).* 598-618

Close, N.C.(2012). An Exploratory Technique for Finding the Q-matrix in Cognitive Diagnostic Assessment: Combining Theory with Data. *Unpublisched doctoral dissertation,* University of Minnesota, Minneapolis, MN.

Croker, L. & Algina, J. (1986). Introduction to Classical and Modern Test Theory, CENGAGE Learning

DiBello, L.V., Roussos, L. A., & Stout(2007). Review of Cognitively Diagnostic Assessment and a Summary of Psychometric Models. *Handbook of Statistics, 26*. Elsevier

DeCarlo, L.T.(2012). Recognizing Uncertainty in the Q-Matrix via a Bayesian Extension of the DINA Model, *Applied Psychological Measurement, 36(6)*. 447-468.

de la Torre (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 & Chiu, C.-Y. (2015). A General Method of Empirical Q-matrix Validation, *The Psychometric Society.*

Desmarais, M. C., Beheshti, B., & Xu, P. (2013). The refinement of a Q-matrix: Assessing methods to validate tasks to skills mapping.

Devellis, R. F. (2006). Classical Test Theory. *Medical Care, 44(11)*. Measurement in a Multi-Ethnic Society. S50-S59

Dogan, E. & Tatsuoka, K. (2007). An international comparison using a

diagnostic testing model: Turkish students' profile of mathematical skills on TIMSS-R, *Educ Stud Math, 68*. 263-272

Dylan Wiliam(2011). What is assessment for learning?. *Studies in Educational Evaluation, 37*. 3-14.

Eggen, P. D., & Kauchak, D.P.(2010). Educational Psychology: Windows on Classroom (8th ed.). *NJ: Pearson Education, Inc*.

FENG, Y.(2013). Estimation and Q-matrix validation for diagnostic classification models. (Master's thesis). *University of South Carolina Scholar Commons*.

Haertel, E.H. (1984). An application of latent class models to assessment data. *Applied Psychological Measurement, 8.* 333-346

Hartz, S. M. (2002). A Bayesian framework for the Unified Model for assessing cognitive abilities: Blending theory with practicality. Unpublished doctoral dissertation, University of Illinois, Champaign, IL.

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*. 191-210

Jang, E. E. (2008). A framework for cognitive diagnostic assessment. In C. A, Chapelle, Y.-R. Chung, & J. Xu (Eds.). *Towards adaptive CALL: Natural language processing for diagnostic language assessment (pp.117-131).* Ames, IA: Iowa State University

Junker, B., Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*, 258-272.

Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The Attribute Hierarchy Method for Cognitive Assessment: A Variation on Tatsuoka's Rule- Space Approach. *Journal of Educational Measurement, 41(3),* 205-237.

Leighton, J., & Gierl, M. (Eds.). (2007). Cognitive diagnostic assessment for education: Theory and applications. *Cambridge University Press*.

Liu, J., Xu, G., & Ying, Z. (2012). Data-driven learning of Q-matrix. *Applied Psychological Measurement, 36.* 609-618

Nyquist, J.B., (2003). The benefits of reconstruing feedback as a larger system of formative assessment: A meta-analysis. Unpublished Master of Science thesis, Vanderbilt University.

Romero, S. J., Ordoñez, X. G., Ponsoda, V., & Revuelta, J. (2014). Detection of Q-matrix misspecification using two criteria for validation of cognitive structures under the Least Squares Distance Model. *Psicológica, 35*. 149-169

Rupp, A. A., & Templin, J. (2008). The effects of Q-Matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement, 68.* 78-96.

Tatsuoka, K. K. (1983). An Approach for Dealing with Misconceptions Based on Item Response Theory. *Journal of Educational Measurement, 20(4)*. 345-354

Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational Statistics, 10(1)*. 55-73

Tatsuoka, K. K. & Tatsuoka, M. M. (1987). Bug distribution and pattern classification. *Psychometrika 52*, 193- 206.

Tatsuoka, K. K. (1991). Boolean algebra applied to determination of the universal set of knowledge states (Technical Report ONR-91-1). Princeton, NJ: Educational Testing Service.

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

Methods. In press.

Thompson, B. (2004). Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications. *American Psychological Association.*

Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of psychology, 18.* 161-169

Wainer, H. & Braun, H. I. (1988). Test Validity. Routledge.

[부 록] 잘못된 Q행렬 Q1~Q15에 관한 정보

a. Q1 (약 1% 변형)

변형 원소 개수 : 3개 (문항-요소관계 추가 : 3개 / 삭제 : 0개)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	1	0	0	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0	0	0
문항3	1	1	0	0	0	0	0	0	0	0
문항4	1	1	0	0	0	0	0	0	0	0
문항5	1	1	1	0	0	0	0	0	0	0
문항6	1	1	1	0	0	0	0	0	0	0
문항7	1	1	0	1	0	0	0	0	0	0
문항8	1	1	1	1	0	0	0	0	0	0
문항9	1	1	1	0	0	1	0	0	0	1
문항10	1	0	0	0	0	0	0	0	0	1
문항11	1	1	0	1	1	0	0	0	0	1
문항12	1	1	0	1	1	0	0	0	0	1
문항13	1	1	0	0	0	1	0	0	0	0
문항14	1	1	0	0	0	1	0	0	0	0
문항15	1	1	0	0	0	1	1	1	0	0
문항16	1	1	0	0	0	1	1	0	0	0
문항17	1	1	0	0	0	1	1	1	0	0
문항18	1	1	0	0	0	1	0	1	0	0
문항19	1	1	0	0	0	1	0	1	0	0
문항20	1	1	0	0	0	1	0	1	0	0
문항21	1	1	1	0	0	1	0	1	1	1
문항22	1	1	0	0	0	1	0	1	1	1
문항23	1	1	0	0	0	1	0	1	1	1
문항24	1	1	0	1	1	0	0	0	0	1
문항25	1	1	0	0	0	1	0	1	1	1

b. Q2 (약 1% 변형)

변형 원소 개수 : 2개 (문항-요소관계 추가 : 1개 / 삭제 : 1개)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	1	0	0	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0	0	0
문항3	1	1	0	0	0	0	0	0	0	0
문항4	1	1	0	0	0	0	0	0	0	0
문항5	1	1	1	0	0	0	0	0	0	0
문항6	1	1	0	0	0	0	0	0	0	0
문항7	1	1	0	1	0	0	0	0	0	0
문항8	1	1	1	1	0	0	0	0	0	0
문항9	1	1	1	0	0	0	0	0	0	1
문항10	1	0	0	0	0	0	0	0	0	1
문항11	1	1	0	1	1	0	0	0	0	1
문항12	1	1	0	1	1	0	0	0	0	1
문항13	1	1	0	0	0	1	0	0	0	0
문항14	1	1	0	0	0	1	0	0	0	0
문항15	1	1	0	0	0	1	0	1	0	0
문항16	1	1	0	0	0	1	1	0	0	0
문항17	1	1	0	0	0	1	1	1	0	0
문항18	1	1	0	0	0	1	0	1	0	0
문항19	1	1	0	0	0	1	0	1	0	0
문항20	1	1	0	1	0	1	0	1	0	0
문항21	1	1	0	0	0	1	0	1	1	1
문항22	1	1	0	0	0	1	0	1	1	1
문항23	1	1	0	0	0	1	0	1	1	1
문항24	1	1	0	1	1	0	0	0	0	1
문항25	1	1	0	0	0	1	0	1	1	1

c. Q3 (약 1% 변형)

변형 원소 개수 : 3개 (문항-요소관계 추가 : 2개 / 삭제 : 1개)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	1	0	0	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0	0	0
문항3	1	1	0	0	0	0	0	0	0	0
문항4	1	1	0	0	0	0	0	0	0	0
문항5	1	1	1	0	0	0	0	0	0	0
문항6	1	1	1	0	0	0	0	0	0	0
문항7	1	1	0	1	0	0	0	0	0	0
문항8	1	1	1	1	0	0	0	0	0	0
문항9	1	1	1	0	0	0	0	0	0	1
문항10	1	1	0	0	0	0	0	0	0	1
문항11	1	1	0	1	1	0	0	0	0	1
문항12	1	1	0	1	1	1	0	0	0	1
문항13	1	1	0	0	0	1	0	0	0	0
문항14	1	1	0	0	0	1	0	0	0	0
문항15	1	1	0	0	0	1	0	1	0	0
문항16	1	1	0	0	0	1	1	0	0	0
문항17	1	1	0	0	0	1	1	1	0	0
문항18	1	1	0	0	0	1	0	1	0	0
문항19	1	1	0	0	0	1	0	1	0	0
문항20	1	1	0	0	0	1	0	1	0	0
문항21	1	1	0	0	0	1	0	1	0	1
문항22	1	1	0	0	0	1	0	1	1	1
문항23	1	1	0	0	0	1	0	1	1	1
문항24	1	1	0	1	1	0	0	0	0	1
문항25	1	1	0	0	0	1	0	1	1	1

d. Q4 (약 1% 변형)

변형 원소 개수 : 2개 (문항-요소관계 추가 : 2개 / 삭제 : 0개)

	A1	A2	A3	A4	A5	A6	A7	A8
문항1	1	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0
문항3	1	1	0	0	0	0	0	0
문항4	1	1	0	0	0	0	0	0
문항5	1	1	1	0	0	0	0	0
문항6	1	1	1	0	0	0	0	0
문항7	1	1	0	1	0	0	0	0
문항8	1	1	1	1	0	0	0	0
문항9	1	1	1	0	0	0	0	0
문항10	1	0	0	0	0	0	0	0
문항11	1	1	0	1	1	0	0	0
문항12	1	1	0	1	1	0	0	0
문항13	1	1	0	0	0	1	0	0
문항14	1	1	0	0	0	1	0	0
문항15	1	1	0	0	0	1	0	1
문항16	1	1	0	0	0	1	1	0
문항17	1	1	0	0	0	1	1	1
문항18	1	1	0	0	0	1	0	1
문항19	1	1	0	0	0	1	0	1
문항20	1	1	0	0	0	1	0	1
문항21	1	1	0	0	0	1	0	1
문항22	1	1	0	0	0	1	0	1
문항23	1	1	0	0	0	1	0	1
문항24	1	1	1	1	1	0	0	0
문항25	1	1	0	0	0	1	0	1

e. Q5 (약 1% 변형)

변형 원소 개수 : 3개 (문항-요소관계 추가 : 1개 / 삭제 : 2개)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	1	0	0	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0	0	0
문항3	1	0	0	0	0	0	0	0	0	0
문항4	1	1	0	0	0	0	0	0	0	0
문항5	1	1	1	0	0	0	0	0	0	0
문항6	1	1	1	0	0	0	0	0	0	0
문항7	1	1	0	1	0	0	0	0	0	0
문항8	1	1	1	1	0	0	0	0	0	0
문항9	1	1	1	0	0	0	0	0	0	1
문항10	1	0	0	0	0	0	0	0	0	1
문항11	1	1	0	1	1	0	0	0	0	1
문항12	1	1	0	1	1	0	0	0	0	1
문항13	1	1	0	0	0	1	0	0	0	0
문항14	1	1	0	0	0	1	0	0	0	0
문항15	1	1	0	0	0	1	0	1	0	0
문항16	1	1	0	0	0	1	1	0	0	0
문항17	1	1	0	0	0	1	1	1	0	0
문항18	1	1	0	0	0	1	0	1	0	0
문항19	1	1	0	0	0	1	0	1	0	0
문항20	1	1	0	1	0	1	0	1	0	0
문항21	1	1	0	0	0	1	0	1	1	1
문항22	1	1	0	0	0	1	0	1	1	1
문항23	1	1	0	0	0	1	0	1	1	1
문항24	1	1	0	1	1	0	0	0	0	1
문항25	1	1	0	0	0	1	0	1	0	1

f. Q6 (약 3% 변형)

변형 원소 개수 : 8개 (문항-요소관계 추가 : 5개 / 삭제 : 3개)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	1	0	0	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0	0	0
문항3	1	1	0	0	0	0	0	0	0	0
문항4	1	1	0	1	0	0	0	0	0	0
문항5	1	1	1	0	0	0	0	0	0	0
문항6	1	1	0	0	0	0	0	0	0	0
문항7	1	1	0	1	0	0	0	0	0	0
문항8	1	1	1	1	0	0	0	0	0	0
문항9	1	1	1	0	0	0	0	0	0	1
문항10	1	0	0	0	0	0	0	0	0	1
문항11	1	1	0	1	1	0	0	0	0	1
문항12	1	1	0	1	1	1	0	0	0	0
문항13	1	1	0	0	0	1	0	0	0	0
문항14	1	1	0	0	0	1	0	0	0	1
문항15	1	1	0	0	0	1	0	1	0	0
문항16	1	1	0	0	0	1	1	0	0	0
문항17	1	1	0	0	0	1	1	1	0	0
문항18	1	1	1	0	0	1	0	1	0	0
문항19	1	1	0	0	0	1	0	1	0	0
문항20	1	1	0	0	0	1	0	1	0	0
문항21	1	1	0	0	0	1	0	1	1	1
문항22	1	1	0	0	0	1	0	1	1	1
문항23	1	1	0	0	0	1	0	1	0	1
문항24	1	1	0	1	1	1	0	0	0	1
문항25	1	1	0	0	0	1	0	1	1	1

g. Q7 (약 3% 변형)

변형 원소 개수 : 7개 (문항-요소관계 추가 : 4개 / 삭제 : 3개)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	1	0	0	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0	0	0
문항3	1	1	0	0	0	0	0	0	0	0
문항4	1	1	0	0	0	0	0	0	0	0
문항5	1	1	1	0	0	0	0	0	0	0
문항6	1	1	1	0	0	0	0	0	0	0
문항7	1	1	0	0	0	0	0	0	0	0
문항8	1	1	0	1	0	0	0	0	0	0
문항9	1	1	1	0	0	0	0	0	0	1
문항10	1	1	0	0	0	0	0	0	0	1
문항11	1	1	0	1	1	0	0	0	0	1
문항12	1	1	0	1	1	0	0	0	0	1
문항13	1	1	0	1	0	1	0	0	0	0
문항14	1	1	0	0	0	1	0	0	0	0
문항15	1	1	0	0	0	1	0	1	0	0
문항16	1	1	0	0	0	1	0	0	0	0
문항17	1	1	0	0	0	1	1	1	0	0
문항18	1	1	0	0	0	1	0	1	0	0
문항19	1	1	0	0	0	1	0	1	0	0
문항20	1	1	1	0	0	1	0	1	0	0
문항21	1	1	0	0	0	1	0	1	1	1
문항22	1	1	0	0	0	1	0	1	1	1
문항23	1	1	0	0	0	1	1	1	1	1
문항24	1	1	0	1	1	0	0	0	0	1
문항25	1	1	0	0	0	1	0	1	1	1

h. Q8 (약 3% 변형)

변형 원소 개수 : 7개 (문항-요소관계 추가 : 5개 / 삭제 : 2개)

	A1	A2	A3	A4	A5	A6	A7	A8
문항1	1	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0
문항3	1	1	0	0	0	0	0	0
문항4	1	1	0	0	0	0	0	0
문항5	1	1	1	0	0	1	0	0
문항6	1	1	1	0	0	0	0	0
문항7	1	1	0	1	0	0	0	0
문항8	1	1	1	1	0	0	0	0
문항9	1	1	1	0	0	0	0	0
문항10	1	0	0	0	0	0	0	0
문항11	1	1	0	1	1	1	0	0
문항12	1	1	0	1	0	0	0	0
문항13	1	1	0	0	0	1	0	0
문항14	1	1	0	0	0	1	0	0
문항15	1	1	0	0	0	1	1	1
문항16	1	1	0	0	0	1	1	0
문항17	1	1	0	1	0	1	1	1
문항18	1	1	0	0	0	1	0	1
문항19	1	1	0	0	0	1	0	1
문항20	1	1	0	0	0	1	0	1
문항21	1	1	0	0	0	1	0	1
문항22	1	1	0	0	0	1	0	1
문항23	1	1	0	0	0	1	0	1
문항24	1	1	0	1	1	0	0	0
문항25	1	1	0	0	0	1	0	1

i. Q9 (약 3% 변형)

변형 원소 개수 : 8개 (문항-요소관계 추가 : 4개 / 삭제 : 4개)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	1	0	0	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0	0	0
문항3	1	1	0	0	0	0	0	0	0	0
문항4	1	1	0	0	0	0	0	0	0	0
문항5	1	1	1	0	0	0	0	0	0	0
문항6	1	1	1	0	0	0	0	0	0	0
문항7	1	1	0	0	0	0	0	0	0	1
문항8	1	1	1	1	1	0	0	0	0	0
문항9	1	1	0	0	0	0	0	0	0	1
문항10	1	0	0	0	0	0	0	0	0	1
문항11	1	1	0	1	1	0	0	0	0	1
문항12	1	1	0	1	1	0	0	0	0	1
문항13	1	1	0	0	0	1	0	0	0	0
문항14	1	1	0	0	0	0	0	0	0	0
문항15	1	1	0	0	0	1	0	1	0	0
문항16	1	1	0	0	0	1	1	0	0	0
문항17	1	1	0	0	0	1	1	1	0	0
문항18	1	1	0	0	0	1	0	1	0	0
문항19	1	1	0	0	0	1	0	1	0	0
문항20	1	1	0	0	0	1	0	0	0	0
문항21	1	1	0	0	0	1	0	1	1	1
문항22	1	1	0	0	0	1	0	1	1	1
문항23	1	1	0	0	0	1	1	1	1	1
문항24	1	1	0	1	1	0	0	0	0	1
문항25	1	1	1	0	0	1	0	1	1	1

j. Q10 (약 3% 변형)

변형 원소 개수 : 8개 (문항-요소관계 추가 : 5개 / 삭제 : 3개)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	1	0	0	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0	0	0
문항3	1	1	0	0	0	1	0	0	0	0
문항4	1	1	0	0	0	0	0	0	0	0
문항5	1	1	1	0	0	0	0	0	0	0
문항6	1	1	0	0	0	0	0	0	0	0
문항7	1	1	0	1	0	0	0	0	0	0
문항8	1	1	1	1	0	0	0	0	0	0
문항9	1	1	1	0	0	0	0	0	0	1
문항10	1	1	0	0	0	0	0	0	0	1
문항11	1	1	0	1	1	0	0	0	0	1
문항12	1	1	0	1	1	0	0	0	0	1
문항13	1	1	0	0	0	1	0	0	0	0
문항14	1	1	0	0	0	1	0	0	0	0
문항15	1	1	0	0	0	1	0	1	0	1
문항16	1	1	0	0	0	1	1	0	0	0
문항17	1	1	0	0	0	1	1	1	0	0
문항18	1	1	0	0	0	1	0	1	0	0
문항19	1	1	0	0	0	1	0	0	0	0
문항20	1	1	1	0	0	1	0	1	0	0
문항21	1	1	0	0	0	1	0	1	0	1
문항22	1	1	0	1	0	1	0	1	1	1
문항23	1	1	0	0	0	1	0	1	1	1
문항24	1	1	0	1	1	0	0	0	0	1
문항25	1	1	0	0	0	1	0	1	1	1

k. Q11 (약 5% 변형)

변형 원소 개수 : 12개 (문항-요소관계 추가 : 8개/ 삭제 : 4개)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	1	0	0	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0	0	0
문항3	1	1	0	0	0	0	0	0	0	0
문항4	1	1	0	0	0	0	0	0	0	0
문항5	1	1	1	0	0	0	0	0	0	1
문항6	1	1	1	0	0	0	0	0	0	0
문항7	1	1	0	1	0	0	0	0	0	0
문항8	1	1	1	1	0	0	0	0	0	0
문항9	1	1	1	0	0	0	0	0	0	1
문항10	1	1	0	0	0	1	0	0	0	1
문항11	1	1	0	1	0	0	0	0	0	1
문항12	1	1	0	1	1	0	0	0	0	1
문항13	1	1	0	0	0	1	0	0	0	0
문항14	1	1	0	1	1	1	0	0	0	0
문항15	1	1	0	0	0	1	0	1	0	0
문항16	1	1	0	0	0	1	1	0	0	0
문항17	1	1	0	0	0	1	1	1	0	0
문항18	1	1	0	0	0	0	0	0	0	0
문항19	1	1	0	0	0	1	1	1	0	0
문항20	1	1	0	0	0	1	0	1	0	0
문항21	1	1	0	0	0	1	0	1	1	1
문항22	1	1	1	0	0	1	0	1	1	1
문항23	1	1	0	0	0	1	0	1	1	1
문항24	1	1	0	1	1	1	0	0	0	1
문항25	1	1	0	0	0	1	0	1	0	1

l. Q12 (약 5% 변형)

변형 원소 개수 : 13개 (문항-요소관계 추가 : 8개/ 삭제 : 5개)

	A1	A2	A3	A4	A5	A6	A7	A8
문항1	1	1	1	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0
문항3	1	1	0	0	0	0	0	0
문항4	1	1	0	0	0	0	0	0
문항5	1	1	0	0	0	0	0	0
문항6	1	1	1	0	0	0	0	0
문항7	1	1	0	0	0	0	0	0
문항8	1	1	1	1	1	0	0	0
문항9	1	1	1	0	0	0	0	0
문항10	1	0	0	0	0	0	0	0
문항11	1	1	0	1	1	0	0	0
문항12	1	1	0	1	1	0	0	0
문항13	1	1	0	0	0	1	0	0
문항14	1	1	0	0	0	1	0	1
문항15	1	1	0	0	0	1	0	1
문항16	1	1	0	0	0	1	1	0
문항17	1	1	0	0	0	1	1	1
문항18	1	1	0	0	0	1	0	1
문항19	1	1	0	0	0	1	0	1
문항20	1	1	0	0	0	1	1	0
문항21	1	1	0	0	0	1	0	1
문항22	1	1	1	0	0	1	0	1
문항23	1	1	0	0	0	1	0	1
문항24	1	1	0	1	1	0	0	0
문항25	1	1	0	1	0	1	0	1

m. Q13 (약 5% 변형)

변형 원소 개수 : 12개 (문항-요소관계 추가 : 9개/ 삭제 : 3개)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	1	0	0	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0	0	0
문항3	1	0	0	0	0	0	0	0	0	0
문항4	1	1	0	0	0	0	0	0	0	0
문항5	1	1	1	0	0	0	0	0	0	0
문항6	1	1	1	1	0	0	0	0	0	0
문항7	1	1	0	1	1	0	0	0	0	0
문항8	1	1	1	1	0	0	0	0	0	0
문항9	1	1	1	0	0	0	0	0	0	1
문항10	1	0	0	0	0	0	0	0	0	1
문항11	1	1	0	1	1	0	0	0	0	1
문항12	1	1	0	1	1	1	0	0	0	1
문항13	1	1	0	1	0	1	0	0	0	0
문항14	1	1	0	0	0	0	0	0	0	1
문항15	1	1	0	0	0	1	0	1	0	0
문항16	1	1	0	0	0	1	1	1	0	0
문항17	1	1	1	0	0	1	1	1	0	0
문항18	1	1	0	0	0	1	0	1	0	0
문항19	1	1	0	0	0	1	0	1	1	1
문항20	1	1	0	0	0	1	0	1	0	0
문항21	1	1	0	0	0	1	0	1	0	1
문항22	1	1	0	0	0	1	0	1	1	1
문항23	1	1	0	0	0	1	0	1	1	1
문항24	1	1	0	1	1	0	0	0	0	1
문항25	1	1	0	0	0	1	0	1	1	1

n. Q14 (약 5% 변형)

변형 원소 개수 : 13개 (문항-요소관계 추가 : 7개/ 삭제 : 6개)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	1	0	0	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0	0	0
문항3	1	1	0	1	0	0	0	0	0	0
문항4	1	1	0	0	0	0	0	0	0	0
문항5	1	1	0	0	0	0	0	0	0	0
문항6	1	1	1	0	0	0	0	0	0	1
문항7	1	1	0	1	0	1	0	0	0	0
문항8	1	1	1	1	1	0	0	0	0	0
문항9	1	1	1	0	0	0	0	0	0	1
문항10	1	0	0	0	0	0	0	0	0	1
문항11	1	1	0	1	1	0	0	0	0	0
문항12	1	1	0	1	1	0	0	0	0	1
문항13	1	1	0	0	0	1	1	0	0	0
문항14	1	1	0	1	0	1	0	0	0	0
문항15	1	1	0	0	0	1	0	1	0	0
문항16	1	1	0	0	0	1	1	0	0	0
문항17	1	1	0	0	0	1	0	1	0	0
문항18	1	1	0	0	0	1	0	1	0	0
문항19	1	1	0	0	0	1	0	0	0	0
문항20	1	1	0	0	0	1	0	1	0	0
문항21	1	1	0	0	0	1	1	1	1	1
문항22	1	1	0	0	0	1	0	1	1	1
문항23	1	1	0	0	0	1	0	1	1	1
문항24	1	1	0	1	1	0	0	0	0	1
문항25	1	1	0	0	0	1	0	1	0	0

o. Q15 (약 5% 변형)

변형 원소 개수 : 13개 (문항-요소관계 추가 : 6개/ 삭제 : 7개)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
문항1	1	0	0	0	0	0	0	0	0	0
문항2	1	0	0	0	0	0	0	0	0	0
문항3	1	1	0	0	0	0	0	0	0	0
문항4	1	0	0	0	0	0	0	0	0	0
문항5	1	1	1	0	0	0	0	0	0	0
문항6	1	1	1	0	0	0	0	0	0	0
문항7	1	1	0	1	0	0	0	0	0	0
문항8	1	1	0	1	0	0	0	0	0	0
문항9	1	1	1	1	1	0	0	0	0	1
문항10	1	0	0	0	0	0	0	0	0	0
문항11	1	1	0	1	0	0	0	0	0	1
문항12	1	1	0	1	1	1	0	0	0	1
문항13	1	1	0	0	0	1	0	0	0	0
문항14	1	1	0	0	0	1	0	0	0	0
문항15	1	1	1	0	0	1	0	1	0	0
문항16	1	1	0	0	0	1	1	0	0	0
문항17	1	1	0	0	0	1	0	1	0	0
문항18	1	1	0	1	0	1	0	1	0	0
문항19	1	1	0	0	0	1	0	1	0	0
문항20	1	1	0	0	0	1	1	0	0	0
문항21	1	1	0	0	0	1	0	1	1	1
문항22	1	1	0	0	0	1	0	1	1	1
문항23	1	1	0	0	0	1	0	1	1	1
문항24	1	1	0	1	1	0	0	0	0	0
문항25	1	1	0	0	0	1	0	1	1	1

ABSTRACT

A Study of Q-matrix Validation Method of Cognitive Diagnostix Assessment

Ann, Hee Sang
Department of Mathematics Education
The Graduate School
Seoul National University

For many years, evaluation was accepted as processes of confirming whether students learn what they were taught properly when the sequence of instructional activities was completed. However, recently, there has been an increasing tendency to conceive of the purpose of assessment as activities that are intended to guide the learning towards the intended goal, and that take place during the learning process, not just measuring the ability of students for quantification and ranking. In these sense, *Cognitive Diagnostic Assessment*(CDA) which is able to provide concrete customized feedback about students' cognitive strengths and weaknesses has become focus of attention.

Cognitive diagnostic assessment assumes that cognitive state is composed of multi-dimensional combination of many attributes(skills), and estimate students' skill mastery from their item responses. The matrix of specification which represents relationships between each item and each skills is often called Q-matrix. Because the result of cognitive diagnostic assessment is decided by students' item responses and Q-matrix, the

analysis of assessment result would be incorrect if a misspecified Q-matrix was used. Therefore, to prove the reliability of interpretation of assessment result, the process of verifying the validity of Q-matrix is needed.

This study focuses on Chia-Yi Chiu's Q-matrix refinement method among various Q-matrix validation methods, analyzes problems of this method, and propose $AVERSS$ validation method which is a new validation method to improve the problems of Q-matrix refinement method. $AVERSS$ validation method is a quantity approach method which identifies the validity of a Q-matrix, and uses $AVERSS$ as the validation index. The Q-matrix of a mathematical assessments in Yoon(2015) was used for application research of $AVERSS$ validation method, and simulation study was conducted for analyze the effectiveness and accuracy.

Chia-Yi Chiu's Q-matrix refinement method is a method to improve original Q-matrix using RSS as a validation index. The result of applying this method to real cognitive diagnostic assessment shows that the refinement process derived the Q-matrix which is inconsistent with the theory. This problem could originate from the assumption of Q-matrix refinement method, because it was confirmed that Hamming method that uses similar assumption also has the effectiveness problem. This exposes problems of exploratory approach for analyzing Q-matrices of real cognitive diagnostic assessment.

In this study, $AVERSS$ validation method which is s confirmatory approach method was developed, and $AVERSS$ was defined for the validation index which represents the validity of whole Q-matrix. This method confirm the produced Q-matrix by comparing RSS of Q-matrix, which could be computed from the assessment data, to the cut-off value. A 95-percentile from the distribution of $AVERSS$ based on simulation data was used for the cut-off value of this method. The

result of applying *AVERSS* validation method to misspecified Q-matrices shows that this method notices the misspecification of Q-matrices with a probability more than 93% if the misspecified Q-matrices differs from the true Q-matrix more than 3%. It means *AVERSS* validation method determines the validity of Q-matrix effectively.

This study developed a new confirmatory approach Q-matrix validation method which is applicable to real cognitive diagnostic assessment, and compliments the problems of existing Q-matrix validation method which has exploratory perspective. In addition, this study shows potential of *AVERSS* which is usable not only for Q-matrix validation but also for improvement of mathematical cognitive diagnostic assessment.

Keywords : Cognitive diagnostic assessment, Q-matrix validation, Confirmatory

Student Number : 2014-20937