



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](#)

강 태 훈 교수 지도
석사학위 청구논문

Q 행렬 타당화를 위한
LLTM 활용 가능성 탐색

2021

성신여자대학교 대학원
교육학과
김 현 미

Q 행렬 타당화를 위한
LLTM 활용 가능성 탐색

강태훈 교수 지도

이 논문을 석사학위논문으로 제출함

2020년 11월

성신여자대학교 대학원

교육학과

김 현 미

인 준 서

김현미의 석사학위 논문으로 인준함

2020년 11월

심사위원장_____ (인)

심 사 위 원_____ (인)

심 사 위 원_____ (인)

성신여자대학교 대학원

논문개요

우리나라의 기존 평가 방식인 고전검사이론과 문항반응이론은 학생의 현재 능력 수준을 파악하는 데에는 효과적이지만, 교수·학습의 개선을 위해 구체적인 정보를 제공하며 피드백을 하는 측면에서는 부족하다. 이에 따라 피험자의 강약점을 파악함으로써 구체적인 피드백을 제공할 수 있는 인지진단모형이 최근 주목받고 있다(이영주, 2014). 인지진단모형은 학생들의 지식 구조의 상태와 과정능력을 측정하기 위해 만들어진 모형으로 학생들의 인지적 강점과 약점에 대한 정보를 제공한다(Leighton & Gierl, 2007).

인지진단모형은 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). 인지진단모형은 각 문항들과 문항이 측정하고자 하는 인지요소 간의 관계를 토대로 학생들의 각 문항에 대한 응답을 분석하여 개별학생들의 인지상태 프로파일을 제공한다(김희경 외, 2012). 따라서 목적에 적합한 인지진단모형에 따른 검사를 진행하기 위해서는 각 인지요소를 측정하기 적절한 문항을 사용해야할 뿐 아니라 인지요소와 문항의 관계가 올바르게 가정되어야 한다.

정확하지 않은 Q 행렬은 잘못된 피험자의 능력측정, 잘못된 모델 모수 측정, 좋지 못한 모델 적합성 등의 원치 않는 결과를 초래할 수 있으므로 인지진단모형이 성공적으로 활용되기 위해서는 Q 행렬의 타당성을 검증할 수 있는 적절한 방법의 제작이 매우 중요하다(Rupp & Templin, 2008; Feng, 2013).

본 연구의 목적은 Q 행렬 타당화 기법으로 LLTM(Linear Logistic Test Model)을 활용가능성을 확인하고자 실제 검사 자료에 대중적으로 쓰이는 자카드 계수, 중다회귀와 비교하여 LLTM의 수행을 확인해보았다.

Q행렬 타당화 기법으로써 LLTM의 활용 가능성을 알아보기 위해 구체적으로 중학교 영어, 수학 실제 검사 자료에 적용하여 연구를 진행하였다. 중학교 3학년 영어검사와 중학교 2학년 수학 검사에 Q행렬 타당화를 진행하면서 자카드 계수, 중다회귀, LLTM을 적용하여 Q행렬 타당화 수행 결과를 비교해볼 수 있었다.

실제 자료를 분석해본 결과를 보면,

첫째, 영어 검사 자료는 총 5개의 인지요소로 구성되어 있고 Q 행렬 타당화를 진행한 결과, 자카드 계수, 중다회귀, LLTM 세 방법 모두 부적절한 인지요소가 없다는 일관된 결과를 보여 Q 행렬 타당화 기법으로써 LLTM을 활용할 수 있는 가능성을 알 수 있었다.

둘째, 총 8개의 인지요소로 구성되어 있는 수학 검사 자료를 Q 행렬 타당화 과정에서 자카드 계수, 중다회귀, LLTM을 사용한 결과, 자카드 계수에서만 인지요소 Q1이 부적절하다고 나왔으나 중다회귀와 LLTM은 일치한 결과를 보였다. 실제 검사 자료를 활용하여 Q 행렬 타당화를 하는 방법으로 LLTM의 가능성을 탐색하였다.

셋째, 실제 검사 자료에 Q 행렬 타당화 기법으로 LLTM을 사용해봄으로써 Q 행렬 타당화 기법으로써 LLTM의 활용가능성을 확인하였다. Q 행렬을 제작할 때, Q 행렬 타당화는 굉장히 중요한 과정이기 때문에 기존에 수행하던 Q 행렬 타당화 기법인 자카드 계수, 중다회귀 등과 같은 방법과 LLTM을 함께 사용하면 더 정교한 Q 행렬을 제작할 수 있을 것이다.

중학교 영어, 수학 검사 자료의 Q 행렬에 자카드 계수, 중다회귀, LLTM을 사용하여 Q 행렬 타당성을 확인한 결과 세 방법 모두 일관된 결과를 보

여 LLTM의 Q 행렬 타당화 기법으로써의 가능성을 제시하였다.

주요어: Q 행렬 타당화, 인지진단모형, DINA, DINO, 자카드 계수, 중다 회귀, LLTM

목 차

논문개요

I. 서론	1
II. 이론적 배경	3
1. 용어의 정리	3
가. 인지요소	3
나. Q 행렬	5
2. 인지진단이론	6
가. 고전검사이론과 문항반응이론의 차별점	7
나. 인지진단모형	8
3. Q 행렬 정의 및 타당화	13
가. 자카드 계수	15
나. 중다회귀	17
다. LLTM	17
III. 연구 방법	21
1. 실제 자료	21
가. 영어 검사 자료	21
나. 수학 검사 자료	24
2. 결과 비교 방법	28

IV. 연구 결과	29
1. 실제 자료를 활용한 Q 행렬 타당화	29
가. 영어 검사 자료 Q 행렬 타당화	29
나. 수학 검사 자료 Q 행렬 타당화	31
IV. 결론 및 논의	35

참고문헌

ABSTRACT

표 목 차

<표Ⅱ-1> Q 행렬 예시	6
<표Ⅱ-2> Q 행렬 예시	14
<표Ⅱ-3> 읽기이해 검사 8가지 인지요소 Q 행렬 자카드 계수	16
<표Ⅲ-1> 18개 문항 영어 검사 자료의 문항분석 결과 및 검사 구성	22
<표Ⅲ-2> 18개 문항 영어 검사의 Q 행렬	23
<표Ⅲ-3> 영어 검사의 세분화 된 인지요소	23
<표Ⅲ-4> 21개 문항 수학 검사 자료의 문항분석 결과 및 검사 구성	25
<표Ⅲ-5> 21개 문항 수학 검사의 Q 행렬	26
<표Ⅲ-6> 세분화 된 인지요소(수학)	27
<표Ⅳ-1> 영어 검사 자카드 계수 산출 결과	29
<표Ⅳ-2> 영어 검사 중다회귀 분석 결과	30
<표Ⅳ-3> 영어 검사 LLTM 분석 결과	30
<표Ⅳ-4> 수학 검사 자카드 계수 산출 결과	31
<표Ⅳ-5> 수학 검사 중다회귀 분석 결과	32
<표Ⅳ-6> Q1 삭제 후 자카드 계수 산출 결과	33
<표Ⅳ-7> Q1 삭제 후 중다회귀분석 결과	33
<표Ⅳ-8> 수학 검사 LLTM 분석 결과	34

그 립 목 차

<그림 II-1> 인지요소 위계 유형	4
<그림 II-2> 문항특성곡선 예시	19

I. 서 론

1. 연구의 필요성 및 목적

우리나라의 기존 평가 방식인 고전검사이론과 문항반응이론 등은 학생의 현재 능력 수준을 파악하는 데에는 효과적이지만, 교수·학습의 개선을 위해 구체적인 정보를 제공하며 피드백을 하는 측면에서는 부족하다. 이에 따라 피험자의 강약점을 파악함으로써 구체적인 피드백을 제공할 수 있는 인지진단모형이 최근 주목받고 있다(이영주, 2014). 인지진단모형은 학생들의 지식 구조의 상태와 과정능력을 측정하기 위해 만들어진 모형으로 학생들의 인지적 강점과 약점에 대한 정보를 제공한다(Leighton & Gierl, 2007).

인지진단모형은 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). 인지진단모형은 각 문항들과 문항이 측정하고자 하는 인지요소 간의 관계를 토대로 학생들의 각 문항에 대한 응답을 분석하여 개별학생들의 인지상태 프로파일을 제공한다(김희경 외, 2012). 따라서 목적에 적합한 인지진단모형에 따른 검사를 진행하기 위해서는 각 인지요소를 측정하기 적절한 문항을 사용해야할 뿐 아니라 인지요소와 문항의 관계가 올바르게 가정되어야 한다.

인지진단모형의 Q 행렬이란 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 인지요소와 문항간의 관계성이 관찰을 통하여 분명하

게 드러나지 않는 경우도 있기 때문에 전문가들의 결정으로 만들어진 Q 행렬을 전적으로 받아들이기 어려운 측면이 있다. 그럼에도 불구하고 교육현장에서 이루어지는 인지진단모형에 따른 평가 방식은 Q 행렬 타당화를 전문가의 검토에 주로 의존하는 경향이 있다.

이에 따라 기존에 제작된 Q 행렬을 타당화하는 방법들에 관한 연구들이 진행되었는데 행렬간의 교집합의 크기를 합집합의 크기로 나누어 유사도를 나타내는 자카드 계수(Jacard coefficient) 산출과 김성훈(2005)과 Buck & Tatsuoka(1998)가 Q-행렬 타당화를 위해 실시한 중다회귀 분석 등이 개발되었다(김명연, 2016).

정확하지 않은 Q 행렬은 잘못된 피험자의 능력측정, 잘못된 모델 모수 측정, 좋지 못한 모델 적합성 등의 원치 않는 결과를 초래할 수 있으므로 인지진단평가가 성공적으로 활용되기 위해서는 Q 행렬의 타당성을 검증할 수 있는 적절한 방법의 제작이 매우 중요하다(Rupp & Templin, 2008; Feng, 2013). 본 연구에서는 Q 행렬의 타당성을 검사하는 방법으로 LLTM(Linear Logistic Test Model)의 사용 가능성을 제시하고 그 수행을 기존 방법들과 비교해보고자 한다.

본 연구의 연구 문제는 다음과 같다. 첫째, LLTM이 Q 행렬 타당화 기법으로 작성할 수 있는 이론적 근거에 대해 탐색한다. 둘째, 두 가지 실제자료에 기존의 Q 행렬 타당화 기법과 유사한 수행을 보이는지 확인하고자 한다.

Ⅱ. 이론적 배경

1. 용어의 정리

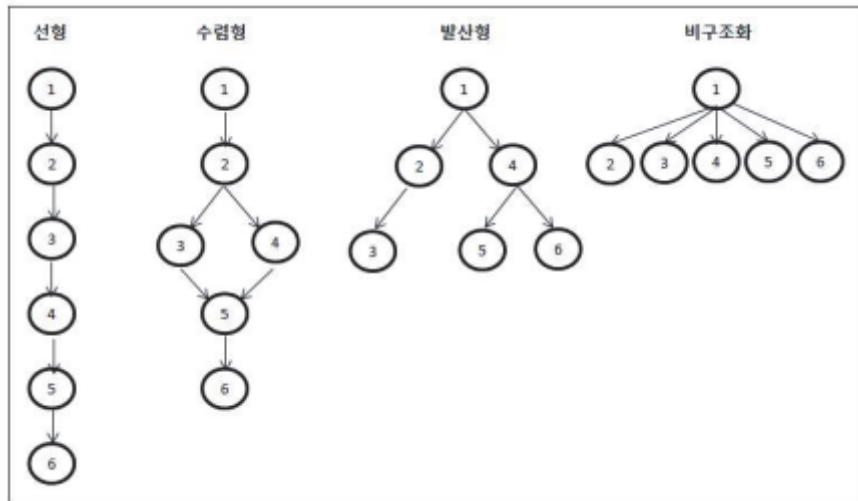
가. 인지요소

인지진단은 여러 개의 인지요소의 숙달 여부를 파악하는 과정이라 할 수 있다. 인지요소(attribute)는 학생의 진단적 상태를 표현하는 범주적인 잠재 변수(categorical latent variable)이다(김성은 외, 2012). 여기서 ‘범주적’이란 분절적 수준에서 하나의 유한한 수를 의미하며, ‘잠재’는 직접적으로 관찰할 수 없음을, ‘변수’는 학생 간 변화하는 상태를 의미한다(Rupp, Templin, & Henson, 2010).

인지요소를 숙달했다는 의미는 해당 지식 또는 기술을 이해하고 습득했음을 의미한다. 검사에서 특정 문항에 정답하기 위해서는 특정 인지요소를 숙달해야 하고, 인지요소를 숙달한 학생은 해당하는 문항에 정답인 반응을 할 것이다.

이러한 인지요소들 간에는 함축하는 의미에 따라 위계가 존재하기도 하는데, 인지요소의 위계는 피험자의 집단 내에서 인지요소의 의존에 대해 설명한다(인보란, 2014). 예를 들면 만약 인지요소 위계가 첫 번째 인지요소 숙달이 두 번째 인지요소를 필수적으로 숙달해야 한다면 첫 번째 인지요소는 일반적으로 두 번째 인지요소를 숙달하기 전에 숙달되어야 한다는 의미이고 첫 번째 인지요소가 미숙달 상태라면 당연히 두 번째 인지요소 역시 미숙달 상태라고 볼 수 있다.

Gierl, Leighton와 Hunka(2007)는 인지요소에 위계가 있다고 말하며, 그 유형을 4가지로 나누어 설명하였다.



출처: Gierl, Leighton, & Hunka(2007)

<그림 II-1> 인지요소의 위계 유형

Gierl, Leighton과 Hunka(2007)은 인지요소의 위계 유형에는 선형, 수렴형, 발산형, 비구조화 네 가지 유형을 제시하였는데, 인지요소 위계 중 선형은 모든 인지요소들이 나란히 연결되어 있다. 따라서 나열된 인지요소 중 마지막의 인지요소를 피험자가 숙달했다면 그 앞의 모든 인지요소를 숙달한 것이다. 예를 들어 위의 그림에서 선형 인지요소 위계에서는 피험자가 여섯 번째 인지요소를 숙달했다면 여섯 번째 인지요소를 제외한 그 앞의 인지요소 전부를 모두 숙달했다고 볼 수 있다. 수렴형 인지요소 위계는 여러 가지 인지요소 중 하나의 인지 요소가 전제조건인 다른 인지요소를 가지는 것이다. 경로 하위의 인지요소를 숙달했다고 한다면 그 위의 상위 인지요소들

중 하나는 숙달한 것이다. 예를 들어 위의 그림에서 수렴형 인지요소 위계에서는 피험자가 인지요소 5를 숙달했다고 하면, 인지요소 위계 순서에 따라 인지요소 1, 2, 3, 4를 전부 숙달했다고 볼 수 있다. 발산형 인지요소 위계는 하나의 인지요소에서 여러 갈래로 구분되어 인지요소의 갈래가 나뉜다. 위에 그림에서 살펴보면 발산형 인지요소 위계는 인지요소 1에서 시작되어 인지요소 2와 4 두 가지로 갈라진다. 발산형 인지요소 위계를 나눠서 살펴보면 인지요소 1, 2, 3은 그 안에서 선형 위계를 가지고 있고, 인지요소 4, 5, 6은 전체 발산형 위계 안에서 또 하나의 발산형 위계를 가진다. 따라서 인지요소 4, 5, 또는 6을 숙달했다면 인지요소 1은 당연히 숙달된 상태일 것이다. 비구조화 인지요소 위계는 하나의 인지요소가 여러 가지 인지요소 전제조건을 가지는 유형을 말한다. 하위 인지요소를 하나라도 숙달했다면 상위 인지요소를 숙달했다는 것을 의미하고, 그러나 상위요소를 숙달했다고 반드시 하위요소를 숙달했다는 것은 아니다. 위의 그림에서 비구조화 인지요소 위계에서는 다른 인지요소를 숙달했다고 말하려면 인지요소 1을 필수적으로 숙달해야만 하는 구조이다.

나. Q 행렬

인지진단모형의 Q 행렬은 문항과 인지요소 간의 관계를 구조화한 행렬이며, 열은 인지요소, 행은 문항을 의미한다(Tatsuoka,1983). 해당 문제를 해결하는데 해당 인지요소가 필요하다면 1, 해당 인지요소가 필요하지 않다면 0으로 이분변수로 표현한다. 인지진단모형을 적용하기에 앞서 검사에서 측정하려는 인지요소가 먼저 결정되어야 하며, 좀 더 구체적으로 문항 수준에서 측정하려는 인지요소가 규명되어야 한다(이영주, 2014).

<표 II-1> Q 행렬 예시

	덧셈	뺄셈	곱셈	나눗셈
1번: $7 + 4 - 1 = ?$	1	1	0	0
2번: $9 - 2 \times 4 = ?$	0	1	1	0
3번: $6 \div 4 + 2 = ?$	1	0	0	1
4번: $2 \times 9 \div 4 = ?$	0	0	1	1

<표 II-1>는 사칙 연산 검사 4문항의 Q 행렬을 표현한 예시이다. 이 검사는 4개의 덧셈, 뺄셈, 곱셈, 나눗셈을 인지요소를 측정하는 문항으로 이루어져있다. 각 문항은 2개 이상의 인지요소를 측정하고 있다. 예시를 보면 Q 행렬을 통해 4번 문항의 풀이에는 인지요소 곱셈과 나눗셈이 요구됨을 알 수 있다.

2. 인지진단이론

인지과학은 측정·평가분야에서 기본적인 지식과 인지기능(cognitive skills)의 이해와 더불어 측정 내용에 대한 인지적 정보를 줄 수 있는 이론적 토대를 만들어 왔으며, 피험자에 대한 학습 정보 보완을 위하여 인지진단이론(cognitive Diagnosis Theory)을 발전시켰다(Chimpman et al, 1995, 재인용). 인지진단이론은 검사에 의해 측정되는 핵심속성(attribute)을 학습자가 어느 정도 숙달했는지 알려줄 수 있는 이론이다(DiBello et al, 1995, 재인용). 여기서 핵심속성이란 학생들이 문항을 정확히 푸는데 필요한 능력(ability), 지식(knowledge), 기능(skills)이나 인지과정(cognitive process) 등이 있는데 이는 매우 다양하다(Tatsuoka, 1983, 재인용). 인지진단이론은 개

별 학생의 지식 및 기능에 대한 숙달 상태를 진단하여 검사에 의해 측정되는 유용한 정보를 학생이나 교사, 학부모 등에게 제공한다는 장점이 있다(김수진 외, 2008). Dibello(1995)는 인지진단이론의 목적을 다음과 같이 말한다. 개개인의 학생이 어떤 인지요소를 숙달하였는지 알려주어 교사, 학부모, 학생에게 학습의 진단적인 정보를 제공하는 것이다. 따라서 교육 현장에서 인지진단이론에 기초하여 시험 문항을 제작하고, 평가 결과를 분석한다면 학생들이 문제를 해결할 때 필요한 인지요소 대한 정보를 학생과 학부모, 교사에게 제공하여 학습에 발전적인 도움을 줄 수 있을 것이다.

가. 고전검사이론과 문항반응이론의 차별점

고전검사이론은 문항반응이론과 비교하여 훨씬 간단한 절차로 문항을 분석할 수 있어 검사에 적용하는 것과 결과 해석이 용이하다는 장점이 있다. 그러나 고전검사이론은 문항난이도, 문항변별도와 같은 문항의 고유한 특성이 피험자 집단의 특성에 의하여 변화되며, 피험자의 능력이 검사도구의 특성에 따라 달리 추정되는 문제점을 내포하고 있다(이은정, 2018). 또한, 피험자들의 능력을 비교하는 과정에서 총점에 의존하므로 문항 특성을 반영하지 않아 정확성이 결여되는 문제점이 있다. 고전검사이론이 이러한 한계점을 가지고 있어 문제점을 보완하기 위하여 문항반응이론이 대두되었다.

문항반응이론은 총점에 근거하여 피험자의 능력을 추정하고 문항을 분석하는 고전검사이론과 달리 각각의 문항에 근거하여 분석한다. 그러므로 문항반응이론에서는 각 문항마다 고유한 문항특성곡선이 존재하고, 문항특성곡선에 따라 문항을 분석한다. 문항반응이론을 전개하기 위해서는 지역독립성 가정(local independence assumption)과 일차원성 가정(unidimensionality assumption), 이 두 가지 조건이 충족되어야 한다. 지역독립성 가정이란 피

험자가 하나의 문항에 정답할 확률은 그 검사의 다른 문항에 정답할 확률에 영향을 미치지 않는다는 가정이다(정희영 외, 2004). 일차원성 가정은 한 문항은 하나의 특성만 측정해야 한다는 가정이다. 지역독립성 가정은 1번 문항이 2번 문항에 정답할 수 있는 단서를 포함한다거나 정답을 돕는 힌트가 제공되면 안된다는 내용이다. 일차원성 가정이 충족되면 지역독립성 또한 당연히 이루어지나 지역독립성이 이루어졌다고 해서 반드시 일차원성 가정이 충족되는 것은 아니다(정희영 외, 2004).

나. 인지진단모형

인지진단모형은 여러 학자들에 의하여 규칙장 모형(Tatsuoka, 1983), LCM(Haertel, 1989), RUM(DiBello et al., 1995), NIDA(Maris, 1999), DINA(de la Torre & Douglas, 2004; Haertel, 1989), NIDO(Junker & Sijtsma, 2001), C-RUM(Hartz, 2002), Fusion Model(DiBello et al., 1995; Hartz, 2002), GDM(von Davier, 2005), DINO(Templin & Henson, 2006), LCDM(Henson, Templin, & Willse, 2009), G-DINA(de la Torre, 2011)등 62개 이상의 인지진단모형들이 개발되어 왔다(Fu & Li, 2007).

Rupp 외(2010)는 여러 인지진단모형을 다음과 같은 특성에 따라 분류하였다. 첫째, 관찰된 응답변수가 이분 또는 다분 응답변수인가, 둘째, 잠재변수가 이분 또는 다분 인가, 셋째, 모형의 유형이 보상적 모형(compensatory model) 또는 비보상적 모형(noncompensatory model)인가에 대하여 분류되어진다(김지효, 2013).

응답변수와 잠재변수가 모두 이분 변수일 때, 모든 인지진단모형의 적용이 가능하다. 인지진단모형은 정답과 오답에 따라 1 또는 0인 자료에 적용할 수 있으며, 각 인지요소의 숙달여부에 따라 '숙달' 또는 '미숙달'로 정보

제공이 가능하다. 응답변수와 잠재변수가 다분변수인 경우에는 BIN, LCDM, GDM, G-DINA모형 등에 적용이 가능하다.

잠재변수가 다분 변수인 경우에 적용이 가능한 모형은 잠재변수가 이분인 경우도 가능하다는 점에서 인지진단모형의 적용범위가 넓고 유연성이 있으나, 모형이 매우 복잡한 단점이 있다(Rupp et al., 2010).

보상적 모형은 문항에 정답하기 위해 필요한 인지요소들을 모두 숙달하지 않아도 숙달한 인지요소가 숙달하지 못한 인지요소의 부재를 채워줄 수 있다고 가정한다. 이러한 가정을 하는 모형은 비결합적 규칙(disjunctive rule)을 따른다고 한다. 비결합적 규칙이 적용되는 모형은 문항을 푸는데 필요한 인지요소 중에서 모든 인지요소를 숙달하지 않고 몇 개의 인지요소만 숙달하여도 정답을 할 확률이 높아지는 특성이 있다(Rupp et al., 2010). 대표적으로 비결합적 규칙을 따르는 보상적 모형은 DINO, NIDO, C-RUM이 있다.

비보상적 모형은 문항에 정답하기 위해 필요한 인지요소를 전부 숙달하여야 하고, 숙달하지 못한 인지요소가 하나라도 있다면 숙달된 다른 요소가 그 인지요소의 결핍을 보완하지 못한다는 가정을 가지고 있다. 예를 들어, 덧셈, 뺄셈에 관련한 문항이 있다면 ‘덧셈’과 ‘뺄셈’ 두 가지 인지요소를 모두 숙달했을 때만 문항에 정답했다고 본다. 즉, 덧셈과 뺄셈 두 가지 중 ‘뺄셈’을 숙달하지 못했다면 ‘덧셈’이 ‘뺄셈’을 보완할 수 없어 이 문항의 정답을 못했다고 보는 것이다. 이것을 결합적 규칙(conjunctive rule)을 따르는 모형이라고 한다. 결합적 규칙이란 문항에 정답하기 위해 필요한 모든 인지요소를 숙달하였을 때 문항에 정답을 할 수 있다고 여기는 것이다(Rupp et al., 2010). 대표적인 결합적 규칙 하의 비보상적 모형으로는 DINA, NIDA, NC-RUM이 있다.

1) DINA

DINA(deterministic input, noisy“and”gate)모형(de la Torre & Douglas, 2004; Haertel, 1989; Junker & Sijtsma, 2001)은 결합적 규칙을 따르는 비보상적 모형이다. 즉, 문항에서 요구하는 인지요소를 전부 숙달한 학생은 문항에 정답할 것이라고 생각하지만, 요구되는 인지요소 중 하나라도 숙달하지 못한다면 오답할 것이라고 예상한다. DINA모형에서 피험자가 해당 문항에 정답할 확률을 구하려면 3가지 구성요소가 필요하다. 첫 번째 요소는 잠재변수인 η_{ij} 이며 다음의 식과 같이 정의된다.

$$\eta_{ij} = \prod_{k=1}^K (a_{ik})^{q_{jk}}$$

η_{ij} 는 응시생 i 가 문항 j 에 정답하기 위해 필요한 모든 인지요소들을 숙달이 되었는지 아닌지에 대한 이분법적 값을 의미한다. 모든 인지요소를 숙달했다면 $\eta_{ij}=1$, 요구되는 인지요소 중에서 하나라도 숙달하지 못하면 $\eta_{ij}=0$ 이 된다. 그러므로 $\eta_{ij}=0$ 인 집단의 응시생들이 서로 다른 인지요소패턴을 갖고 있다 하더라도 분별할 수 없다. q_{jk} 는 문항 j 에 정답하기 위해서 인지요소 k 가 요구되면 1, 아니면 0이 된다. a_{ik} 는 응시생 i 의 인지요소 k 에 대한 숙달여부로, 인지요소 k 를 숙달하였으면 $a_{ik}=1$ 이 되며, 인지요소 k 를 숙달하지 못하였으면 $a_{ik}=0$ 이 된다. 즉, 위의 식은 응시생 i 가 문항 j 의 정답을 하기 위해서는 그 문항에 정답하기 위해 필요한 모든 인지요소를 갖춰야 하며, 인지요소 중 하나라도 숙달하지 못한다면 오답을 한다는 것을 말한다.

그러나 학생의 실제적인 문항반응은 잡음요인(noisy)으로 인해 이론적으로 기대한 반응과 다른 결과가 나타날 수 있다. 즉, 해당 인지요소를 숙달하

지 않았더라도 추측하여 문항에 정답하거나 인지요소를 숙달했더라도 실수로 문항에 정답하지 않은 경우가 있을 수 있다.

DINA모형은 추측모수(g_j)와 부주의 오류모수(s_j)을 고려하여 정답을 할 확률을 구한다. 추측모수는 문항이 요구하는 모든 인지요소들을 숙달하지 않았는데 정답을 맞힐 확률과 Q 행렬에 포함되지 않은 또 다른 인지요소를 사용하였을 가능성까지 포함한다(de la Torre & Douglas, 2004). 반면, 부주의 오류모수는 문항에 정답하기 위해 필요한 모든 인지요소들을 숙달했지만 오답을 하는 경우를 말한다. 다음 식은 DINA모형에서 추측모수인 g_j 와 부주의 오류모수인 s_j 을 수식으로 나타낸 것이다.

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

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

다음 식은 응시생 i 의 잠재변수 η_{ij} 와 문항 j 의 추측모수, 부주의 오류모수를 동시에 고려할 때 문항 j 에 정답을 할 확률을 나타낸다. X_{ij} 는 문항 $J(j = 1, \dots, J)$ 에 대하여 i 번째 응시생의 응답 반응이며, 정답과 오답 반응은 1 또는 0으로 표시된다.

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

만약에 모든 인지요소를 숙달하였다면($\eta_{ij} = 1$), $(1 - s_j)$ 가 정답할 확률이 되며, 필요한 인지 요소 중 하나라도 결핍이 있다면($\eta_{ij} = 0$), g_j 가 정답을 맞힐 확률이 된다.

2) DINO

DINO(deterministic input, noisy“or”gate)모형(Templin & Henson, 2006)은 DINA모형과 마찬가지로 문항 수준에서 추측 모수와 부주의 오류 모수를 고려한다. DINO모형은 DINA모형과 달리 각 인지요소 사이의 관계가 ‘AND’가 아닌 ‘OR’의 관계라는 점이다. 따라서, DINO모형은 비보상적인 DINA모형과는 다르게 보상적 모형으로 결핍된 인지요소가 있다하더라도 숙달된 인지요소가 보상해줄 수 있다. DINO모형을 통해 응시생이 정답할 확률을 구하기 위해서는 3가지 구성요소가 필요하다. 첫 번째 구성요소는 잠재변수인 ζ_{ij} 이다

$$\zeta_{ij} = 1 - \prod_{k=1}^K (1 - a_{ik})^{q_{jk}}$$

ζ_{ij} 는 응시생 i 가 문항 j 에 필요한 인지요소들 중 한개 이상의 인지요소를 숙달했는지 아닌지에 대한 이분법적 값을 의미한다. 즉, 문항 j 에 필요한 인지요소들 중 한개 이상을 숙달했으면 $\zeta_{ij}=1$ 이고 숙달하지 않았다면 $\zeta_{ij}=0$ 이 된다. 그러므로 필요한 인지요소들이 전부 충족되지 않더라도 부족한 인지요소를 숙달된 인지요소가 보상할 수 있다. 따라서 DINO모형은 인지요소를 적게 가진 피험자도 문항에 대해 정답할 확률이 높아지는 것을 기대할 수 있다.

DINO모형에서는 ζ_{ij} 를 사용하여 추측모수와 부주의 오류모수를 계산한다. 다음 식은 DINO모형에서의 추측모수와 부주의 오류모수를 수식으로 나타낸 것이다.

$$g_j = P(X_{ij} = 1 | \zeta_{ij} = 0)$$

$$g_j = P(X_{ij} = 1 | \zeta_{ij} = 0)$$

DINO모형에서 잠재변수(ζ_{ij})와 추측모수, 부주의 오류모수를 한번에 고려하였을 때, 피험자가 정답할 확률은 아래의 식과 같다. X_{ij} 는 문항 $J(j=1, \dots, J)$ 에 대하여 응시생 i 의 관찰된 응답반응이며, $(1-s_j)$ 는 문항 j 에 대하여 부주의 오류모수가 아닌 확률이며, g_j 는 문항 j 에 대하여 추측한 확률이다.

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

만약, 문항에 정답하기 위해 요구되는 인지요소들 중 숙달된 인지요소가 한 개 이상일 때($\zeta_{ij}=1$), $(1-s_j)$ 는 정답할 확률이 되지만, 모든 인지요소가 미숙달인 경우 경우($\zeta_{ij}=0$), $(1-s_j)$ 부분이 1이 되면서 추측모수(g_j)가 정답할 확률이 된다.

3. Q 행렬 정의 및 타당화

인지진단모형의 중요한 요소 중 다른 하나는 Q 행렬이다(Tatsuoka, 1983). Q 행렬은 문항들에 의해 인지요소가 측정될 수 있도록 각 문항에 대응되는 인지요소가 0과 1의 행렬 형태로 구성된 형식을 가진다(권승아, 2017). 만약 문항 i 를 해결하는 데 인지요소 a 가 필요하다면 Q 행렬 중 ia 셀은 1로 표기하고 문항 i 의 나머지 셀은 0으로 표기한다(Madison & Bradshaw, 2015). 일반적으로 문항은 행에 인지요소는 열에 기입하는 구조

적 형태를 보이며, 통계적 측면에서의 Q 행렬은 부하 행렬(loading matrix) 또는 패턴 행렬(pattern matrix)로서 Q 행렬을 통해 각 문항이 어떤 잠재변수와 관련이 있는가를 알 수 있다(Rupp, Templin & Henson, 2010). Q 행렬의 예시는 <표 II-2>와 같다.

<표 II-2> Q 행렬 예시

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

위의 <표 II-2>와 같이 6개의 문항과 4개의 인지요소로 이루어진 검사가 있다면 문항 1은 인지요소 3을 측정할 수 있고, 문항 2는 인지요소 2와 3의 측정이 가능하다. 또한 문항 3은 인지요소 2를 측정할 수 있으며, 문항 4는 인지요소 1, 2, 3을 모두 측정할 수 있다.

Q 행렬의 작성은 인지진단모형의 추론과정에서 매우 중요한 역할을 담당하기 때문에 Q 행렬 작성자는 일반적으로 교사 및 교과전문가가 제공하는 학습 내용과 이론에 대한 통찰력이 필요하다(권승아, 2017). 동시에 문항에 작용하는 인지 및 정신능력에 대한 질적 탐구 과정도 동반되어야 Q 행렬 작성에 필요한 내용적 검증 과정을 통해 Q 행렬의 내용 타당도(Boorsboom & Mellenberg, 2007)를 확보할 수 있다. 만약 Q 행렬이 부적절하다면 결국 인지진단모형을 통한 추정과 잠재집단 분류 결과의 타당성은 훼손된다(DeCarlo, 2011). 위와 같은 이유로 Q 행렬 타당성을 증가시켜 더욱 정교한 Q 행렬을 작성할 수 있도록 전문가 논의 뿐만 아니라 다양한 통계적 방법이 제안되고 있다.

가. 자카드 계수

자카드 계수는 1901년 Jaccard에 의해 개발되었고, 두 가지 자료의 유사성이 어느 정도인지 파악할 수 있다. 김수진 외(2013), 인보란(2014), 김명연(2016) 등의 연구에서 두 가지 인지요소 간의 유사성을 판별하여 인지요소의 삭제 여부를 확인하기 위해 사용되었으며, 아래의 식으로 계산된다.

$$J(A,B) = |A \cap B| / |A \cup B|$$

위의 식에 따르면 인지요소 a, b에 대한 자카드 계수를 계산하기 위해서는 인지요소 a와 b를 동시에 측정하는 문항의 수를 인지요소 a, b를 각각 측정하는 문항과 동시에 측정하는 문항의 수로 나누어 구할 수 있다. 만일 자카드 계수가 .20이라면 전체 문항 중 20%의 문항이 인지요소 a, b 두 가지를 동시에 측정한다고 볼 수 있으며, 계수가 .90이면 전체 문항 중 90%의 문항이 인지요소 두 가지를 동시에 측정한다고 볼 수 있다. 자카드 계수를 산출한 결과, 동시에 측정되는 인지요소의 내용을 고려하여 하나의 인지요소를 삭제하거나 하나로 통합하는 것이 바람직 할 수 있다.

Henning(2004)의 연구에서 군집분석을 진행하면서 자카드 계수가 .50이상 일 때, 유사한 대상이 모인 집단으로 보았고, Q 행렬 타당화에 자카드 계수를 사용한 최근 연구들에서도 이 기준을 활용하고 있다.

한편, Q 행렬 타당화를 실시하면서 자카드 계수를 사용하는 경우 주의할 점이 있는데, 자카드 계수가 높게 나와 많은 문항에서 중복 측정되고 있는 인지요소가 있다고 추정되더라도 수치에만 의존하여 이들은 통합하는 것이 아니라 전문가의 판단에 따라 문항의 내용과 인지요소 사이의 관계를 고려해야한다. 문항의 정답을 맞추기 위해 항상 두 개의 인지요소가 동시에 필

요하다고 해서 동일한 요소라고 판단할 수는 없다. 또한, 만약 인지요소가 보상적인 관계를 가진다면 인지요소 두 가지 중 하나만 숙달해도 정답률은 올라갈 것이다. 그렇기 때문에 하나의 문항에서 두 가지의 인지요소를 측정하는 것으로 Q 행렬을 제작하였더라도 실제로 두 가지 인지요소를 모두 숙달한 피험자가 드문 경우에 있을 수도 있다.

아래의 <표 II-3>은 읽기이해 검사를 Q 행렬 타당화하면서 자카드 계수를 사용한 예시이다. 8개의 인지요소 간의 유사성을 산출한 결과 인지요소 2와 인지요소 3의 유사성이 0.785로 나타났는데 이것은 인지요소 2와 3이 관련된 문항 중 78.5%에 해당하는 문항에서 두 가지 인지요소가 동시에 나타난 것을 의미한다(이재호, 2016). Henning(2004)의 기준에 따라서 인지요소 2와 3은 유사성이 높아 수정이 필요하다고 판단할 수 있다.

<표 II-3> 읽기이해 검사 8가지 인지요소 Q 행렬 자카드 계수(이재호, 2016)

인지 요소	1	2	3	4	5	6	7	8
2	0.000							
3	0.000	0.785						
4	0.000	0.357	0.272					
5	0.000	0.142	0.5	0.000				
6	0.3	0.000	0.000	0.000	0.000			
7	0.3	0.000	0.000	0.000	0.000	0.000		
8	0.000	0.214	0.000	0.000	0.000	0.000	0.000	

나. 중다회귀

Buck& Tatsuoka(1998), 김성훈(2005), 김희경 외(2012)는 Q 행렬 타당화를 위해 중다회귀분석을 실시하였다. 중다회귀분석은 종속변수는 문항의 정답률로, 독립변수는 인지요소별 숙달확률로 하여 각 인지요소가 검사의 정답률에 얼마나 기여하는지를 알아보기 위해 적용되었다. 중다회귀분석을 한결 인지요소 중 정답률과 부적 관계를 보인다면 삭제할 수 있다.

중다회귀분석은 선형관계를 가정하긴 하나 인지요소의 정답률과 숙달확률은 비선형적 관계를 가질 수 있다. 이에 따라 정답률을 로짓변환하여 분석을 실시하기도 하였다(김희경 외, 2012). 그러나 이런 경우 인지요소 간 상관성이 높으면 다중공선성에 의한 불안정성 또한 고려해야 한다. 인지진단모형에서 인지요소 사이의 다양한 관계가 가정될 수 있는데 하나의 인지요소가 다른 인지요소의 전제조건이 되거나 숙달이 전후 순서를 가질 수 있다.

다. LLTM(Linear Logistic Test Model)

Rasch 모형은 문항반응이론의 한 모형으로 1960년 Rasch에 의해 개발되었다(지은림, 채선희, 2000). Rasch 모형은 문항 변별도, 문항 추측도는 배제하고, 문항 난이도만 고려하는 1-모수 모형이다. Rasch 모형에서는 피험자 능력이 문항 난이도 보다 높으면 피험자가 문항에 정답할 확률이 높아진다고 예상하고, 피험자 능력이 문항 난이도 보다 낮으면 피험자가 정답할 확률은 낮아지고, 오답할 확률이 높아진다고 예상한다.

피험자의 능력을 추정하기 위해 Rasch 모형에서는 로지스틱 함수를 사용한다. Rasch 모형에서는 피험자 능력 추정과 문항 난이도 측정을 위해 처리가 필요한데 첫 번째로, 피험자 능력과 문항 난이도의 차이를 비교한 다음

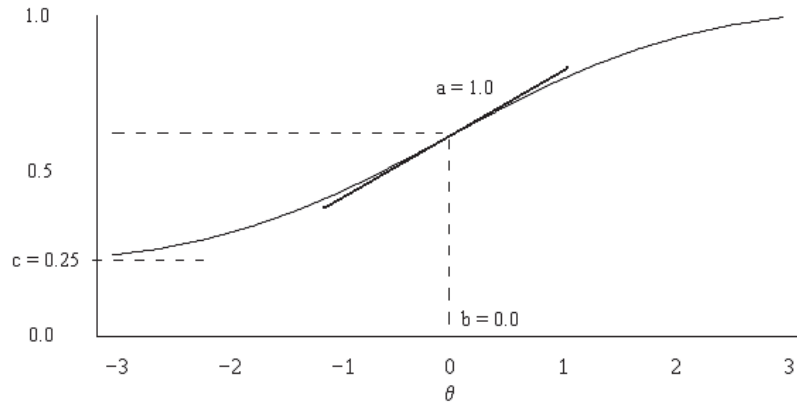
으로 이러한 차이에 따라서 피험자가 정답할 확률을 구하고, 마지막으로 피험자가 정답할 확률을 로짓 점수로 변환한다.

Rasch 모형에서는 피험자 v 의 능력을 β 하고 하며 문항 i 의 난이도를 δ 라로 표시한다. 또한, 문항 i 에 대한 피험자 v 의 반응을 X_{vi} 로 나타낸다. 피험자의 반응(X_{vi})은 피험자 능력(β_v)과 문항의 난이도 중 어느 것이 얼마나 더 높은가에 의해 결정되는데, 이 차이에 따라 정답을 할 확률이 어떻게 변할 것인가를 아래 식으로 나타낸 것이 Rasch 모형이다(지은림 외, 2000).

$$p(X_{vi}) = \frac{\exp(\beta_v - \delta_i)}{1 + \exp(\beta_v - \delta_i)}$$

여기서, $P(X_{vi})$ 은 문항에서 피험자가 정답하여 1점을 받을 확률을 뜻한다. Rasch 모형은 피험자 능력과 문항 난이도가 동일한 수준에 위치할 때 피험자가 정답할 확률이 0.5가 된다. 만약, 피험자 능력이 문항 난이도보다 낮다면 정답을 할 확률이 0.5보다 낮아지고, 피험자 능력이 문항 난이도보다 높을수록 정답할 확률이 0.5보다 높아진다. 이런 방법에 따르면 피험자 능력이 문항 난이도 보다 높으면 정답을 할 확률이 1이 될 것이고, 반대의 경우에는 정답을 할 확률은 0이 될 것이라는 이전의 생각과 차이가 있다(안정민, 2012).

Rasch 모형으로 피험자 v 가 문항 i 에서 정답을 할 확률을 계산할 수 있는데 정답할 확률은 비선형적으로 문항특성 곡선과 같은 모양을 띤다. 문항 특성 곡선의 예시는 다음과 같다.



<그림 II-2> 문항특성곡선 예시

따라서 정답할 확률이 비선형적이기 때문에 피험자가 정답할 확률을 로지스틱 변환(logistic transformation)하여 피험자 능력 추정치를 계산하게 된다. 한편 문항 난이도를 위해서는 오답할 확률을 로지스틱 변환하여 추정치를 구하게 된다(김미예, 2002).

LLTM(Linear Logistic Test Model)은 Rasch 모형을 기반으로 문항 난이도 모수를 문항 특성 혹은 관련 공변인들의 선형 결합으로 표현하는 초기 심리측정학 모형의 일종이다. Rasch 모형에서 문항 i 의 난이도를 β_i 라고 적는다면, LLTM에서는 이를 $\beta'_i = \sum_{k=1}^K q_{ik}\eta_k$ 로 규정한다. 여기서 q_{ik} 는 문항 i 와 관련될 수 있는 k 개의 문항 특성들($k=1, \dots, K$; 해당 특성을 가지면 1, 가지고 있지 않으면 0)을 의미하며 η_k 는 이러한 선형결합에서의 계수로서 해당 문항 특성 k 가 문항 i 의 난이도에 기여하는 정도를 뜻한다.

흔히 Rasch 모형의 확장이 LLTM이라고 말하기도 하지만 전체 문항의 수보다 K 가 작다면 오히려 LLTM 하에서 추정해야 할 모형 모수가 같을 수도 있기 때문에, 모형 복잡성 측면에서 볼 때 많은 경우 LLTM이 Rasch

모형보다 더 단순하게 된다. LLTM은 문항 난이도의 변산이 모형에서 고려된 문항 특성 혹은 공변인들에 의해 완벽하게 설명된다는 매우 강한 가정을 갖기 때문에 현실적이지 못하다는 비판을 흔히 받는다(Hartig, Frey, Nold, & Klieme, 2012; Rijmen, de Boeck & Leuven, 2002).

Baghaei, P., & Hohensinn, C(2017)은 Q 행렬을 타당화 하는 과정에서 LLTM과 RM(Rasch Model)의 모형 적합성을 비교하는 시뮬레이션 연구를 진행하였는데 본 연구에서는 일반적으로 사용되고 있는 Q 행렬 방법인 자카드 계수, 중다회귀 방법과 비교하는 연구를 진행하고자 한다.

Ⅲ. 연구 방법

1. 실제 자료

가. 영어 검사 자료

본 연구에서는 Q 행렬 타당화에서 LLTM의 활용가능성을 탐색하기 위하여 중학교 3학년 영어 검사 자료를 사용하였다. 영어검사는 총 18문항으로 구성된 검사로 2014년 한 시도교육청 수준에서 실시된 표준화 학업성취도 검사로서 피험자 수는 약 22,000명이었다. 이 중에서 미응답자 등을 제외하고 무선적으로 표집한 3,000명의 응답자료를 본 연구의 분석 대상으로 사용하였다. 검사 신뢰도를 구한 결과 크론박 알파가 0.85로 높게 산출되었으며, 18점 만점으로 볼 때 평균과 표준편차는 각각 8.20 및 4.59로 나타났다.

그밖에 고전검사이론에 따른 문항분석 결과는 <표 III-1>에서 제시한 바와 같다. 정답률로 구한 문항 난이도를 보면 18개 문항 중에서 가장 어려운 문항은 9번이었으며 가장 쉬운 문항은 3번이었다. 문항 및 검사 점수간 양류 상관계수로 구한 문항 변별도의 경우 가장 높은 문항은 8번이었으며 가장 낮은 문항은 14번이었지만, 모든 문항의 변별도가 0.3보다 큰 것으로 나타나서 치명적인 문제는 없다고 볼 수 있었다. 문항 난이도의 평균은 0.46이었으며 문항 변별도 평균은 0.53이었다. 항목 제거시 알파를 볼 때 해당 문항이 제거된다면 검사 신뢰도가 가장 작아지는 것은 8번 문항으로 나타나서 해당 문항이 전체 검사 신뢰도에 가장 큰 기여를 하는 것으로 나타났다. 검사 영역 측면에서 볼 때 18문항 중 읽기에 해당하는 문항이 13개 그리고 쓰기에 해당하는 문항이 5개였다.

<표 III-1> 18개 문항 영어 검사 자료의 문항분석 결과 및 검사 구성

문항	난이도	변별도	내용	영역
1	.68	0.60	적절한 의견 말하기	읽기
2	.50	0.46	적절한 의견 말하기	읽기
3	.74	0.55	표에 대한 설명하기	읽기
4	.52	0.48	분위기 찾기	읽기
5	.48	0.52	적절한 어휘 찾기	읽기
6	.49	0.60	변화 찾기	읽기
7	.49	0.50	문장 배열하기	쓰기
8	.47	0.64	목적 파악	읽기
9	.29	0.49	문장 넣기	쓰기
10	.38	0.59	빈칸 넣기	읽기
11	.35	0.60	글의 요지	읽기
12	.52	0.58	제목 추론	읽기
13	.42	0.51	지시대명사의 대상	읽기
14	.30	0.38	문맥 흐름 파악	쓰기
15	.40	0.49	제목 추론	읽기
16	.37	0.47	세부내용 일치	읽기
17	.38	0.45	문장 배열하기	쓰기
18	.43	0.57	본문내용 요약하기	쓰기

본 연구에서 사용하고 있는 영어 학업성취도 검사에 대하여, 김명연(2016)이 중학교 영어 교사 2명 및 해당 과목 전공 대학원생 2명의 도움을 받아 작성한 Q-행렬은 <표 III-2>와 같다.

<표 III-2> 18개 문항 영어 검사의 Q 행렬 (김명연, 2016, p.51)

문항	인식하기	이해하기	추론하기	평가하기	구성하기
1	0	1	1	0	1
2	0	1	1	0	1
3	0	1	0	1	0
4	0	1	1	0	0
5	1	0	0	0	0
6	0	1	1	0	0
7	0	1	0	0	1
8	0	1	1	0	0
9	0	1	0	0	1
10	0	1	1	0	0
11	0	1	1	0	0
12	0	1	1	0	0
13	0	1	0	0	0
14	0	1	1	1	0
15	0	1	1	0	0
16	1	1	0	1	0
17	0	1	1	0	1
18	0	1	1	0	1

<표 III-3> 영어 검사의 세분화 된 인지요소

인식하기	이해하기			추론하기		평가하기	해석하기		
	일반적 내용의 이해	논리적 구조의 이해	세부 정보의 이해	연역적으로 추론하기	귀납적으로 추론하기		요약하기	응답하기	글의 순서 배열하기

총 5개의 인지요소(인식하기, 이해하기, 추론하기, 평가하기, 해석하기)가 도출되었다. 이 중 ‘이해하기’는 다시 ‘일반적 내용의 이해’, ‘논리적 구조의 이해’, 그리

고 ‘세부 정보의 이해’로 세분화될 수 있으며, ‘추론하기’의 경우 ‘연역적으로 추론하기’와 ‘귀납적으로 추론하기’로 나누어 질 수 있고, 마지막으로 ‘해석하기’의 경우 ‘요약하기’, ‘응답하기’, ‘글의 순서배열 하기’ 등으로 세분화될 수 있다. 본 연구에서는 이렇게 모두 세분화된 수준의 인지요소들을 모두 사용하기보다는 분석 편의를 위하여 원래 추출된 5개 인지요소 체계를 사용하여 Q 행렬 타당화를 자카드 계수, 중다회귀, LLTM을 사용하여 진행하려고 한다.

나. 수학 검사 자료

본 연구에서는 Q 행렬 타당화에서 LLTM의 활용가능성을 탐색하기 위하여 중학교 2학년 수학 검사 자료를 사용하였다. 수학검사는 총 21문항으로 구성된 검사로 2014년 한 시도교육청 수준에서 실시된 표준화 학업성취도 검사로서 피험자 수는 약 22,000명이었다. 이 중에서 미응답자 등을 제외하고 무선적으로 표집한 2,000명의 응답자료를 본 연구의 분석 대상으로 사용하였다. 검사 신뢰도를 구한 결과 크론박 알파가 0.75로 높게 산출되었으며, 21점 만점으로 볼 때 평균과 표준편차는 각각 12.38 및 3.78로 나타났다.

그밖에 고전검사이론에 따른 문항분석 결과는 <표 III-4>에서 제시한 바와 같다. 정답률로 구한 문항 난이도를 보면 21개 문항 중에서 가장 어려운 문항은 19번이었으며 가장 쉬운 문항은 1번이었다. 문항 및 검사 점수간 양류 상관계수로 구한 문항 변별도의 경우 가장 높은 문항은 14번이었으며 가장 낮은 문항은 1번이었지만, 대체로 문항 변별도는 0.3이상으로 나타나 큰 문제는 없는 것으로 보여진다. 문항 난이도의 평균은 0.59이었으며 문항 변별도 평균은 0.40이었다. 검사 영역 측면에서 볼 때 21문항 중 개념이해에 해당하는 문항이 10개, 문제해결에 해당하는 문항이 6개, 절차적 지식에 해당하는 문항이 4개 그리고 측정에 해당하는 문항이 1개였다.

<표 III-4> 21개 문항 수학 검사 자료의 문항분석 결과 및 검사 구성

문항	난이도	변별도	내용	영역
1	.92	.16	수와 조작	절차적 지식
2	.89	.26	측정	개념이해
3	.84	.34	기하학	개념이해
4	.83	.47	수와 조작	개념이해
5	.77	.51	자료 분석	문제해결
6	.82	.50	수와 조작	문제해결
7	.68	.43	기하학	문제해결
8	.66	.42	측정	문제해결
9	.68	.39	수와 조작	절차적 지식
10	.73	.50	대수학과 함수	개념이해
11	.81	.49	수와 조작	개념이해
12	.49	.52	수와 조작	절차적 지식
13	.38	.38	수와 조작	문제해결
14	.59	.52	대수학과 함수	개념이해
15	.50	.41	대수학과 함수	개념이해
16	.31	.30	확장, 개방형 문제해결	측정
17	.45	.46	기하학	개념이해
18	.25	.30	측정	개념이해
19	.24	.30	측정	문제해결
20	.35	.41	대수학과 함수	개념이해
21	.32	.38	문제해결	절차적 지식

본 연구에서 사용하고 있는 수학 학업성취도 검사에 대하여, 김명연(2016)이 중학교 수학 교사 2명 및 해당 과목 전공 대학원생 2명의 도움을 받아 작성한 Q-행렬은 <표 III-5>와 같다.

<표 III-5> 21개 문항 수학 검사의 Q 행렬 (김명연, 2016)

문항	계산하기	인식하기	재발견하기	선택하기	이해하기	추론하기	표현하기	해석하기
1	1	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	1
3	0	0	0	0	1	0	1	0
4	0	0	0	0	1	0	0	0
5	1	0	0	0	0	0	1	1
6	1	0	0	0	1	0	1	0
7	0	0	0	0	0	1	1	0
8	1	0	0	0	1	0	0	1
9	1	0	0	0	1	0	1	0
10	1	1	0	0	0	0	1	0
11	1	0	0	0	1	0	1	0
12	1	0	0	0	1	0	1	0
13	1	0	1	0	0	0	1	1
14	1	0	0	0	1	0	0	0
15	0	0	1	0	1	0	0	1
16	0	0	0	1	1	0	0	0
17	0	0	0	0	1	1	0	0
18	1	0	0	0	1	0	0	0
19	1	0	0	0	1	1	1	1
20	1	0	0	0	1	1	1	1
21	1	0	0	0	1	0	1	0

<표Ⅲ- 6> 세분화 된 인지요소(수학)

계산하기		인식하기	재발견하기	선택하기	이해하기		추론하기			표현하기	해석하기		
사 칙 계 산	정형 적인 대수 적 절차				개 념 과 성 질 의 이 해	원 리 와 법 칙 의 이 해	귀 납 적 으 로 추 론 하 기	직 관 적 으 로 추 론 하 기	연 역 적 으 로 정 당 화 하 기		합 수 와 도 형 의 방 정 식 영 역 의 자 료 해 석 하 기	도 형 다 이 어 그 램 의 정 보 해 석 하 기	통 계 적 자 료 해 석 하 기

총 8개의 인지요소(계산하기, 인식하기, 재발견하기, 선택하기, 이해하기, 추론하기, 표현하기, 해석하기)가 도출되었다. 이 중 ‘계산하기’는 다시 ‘사칙계산’, ‘정형적인 대수적 절차’로 세분화될 수 있으며, ‘이해하기’의 경우 ‘개념과 성질의 이해’와 ‘원리와 법칙의 이해’로 나누어 질 수 있고, ‘추론하기’의 경우 ‘귀납적으로 추론하기’, ‘직관적으로 추론하기’, ‘연역적으로 정당화하기’로 나누어질 수 있고, 마지막으로 ‘해석하기’는 ‘함수와 도형의 방정식 영역의 자료 해석하기’, ‘도형 다이어그램의 정보 해석하기’, ‘통계적 자료 해석 하기’ 등으로 세분화될 수 있다. 본 연구에서는 이렇게 모두 세분화된 수준의 인지요소들을 모두 사용하기보다는 분석 편의를 위하여 원래 추출된 8개 인지요소 체계를 사용하여 Q 행렬 타당화를 자카드 계수, 중다회귀, LLTM을 사용하여 진행하려고 한다.

2. 결과 비교 방법

실제 자료인 중학교 3학년 영어 검사 자료와 중학교 2학년 수학 검사 자료를 활용하여 Q 행렬 타당화 방법인 자카드 계수, 중다회귀, LLTM을 각각 적용해보고 각 인지요소가 각각의 Q 행렬 타당화 방법에 따라 일치하는지 살펴보고, 한 가지 방법에서 오류가 발생했을 시 다른 방법에서도 오류를 발견하는지 확인해보고자 한다.

IV. 연구 결과

1. 실제 자료를 활용한 Q 행렬 타당화

가. 영어 검사 자료 Q 행렬 타당화

Q 행렬은 인지진단모형에서 중요한 핵심으로써 타당화 검증이 필수적으로 수반된다. 개발된 Q 행렬 초안을 확정하기 전에 Q 행렬 타당화를 위한 통계 분석을 진행하였다. 타당화를 위한 통계 분석 방법은 행렬간의 교집합의 크기를 합집합의 크기로 나누어 유사도를 나타내는 자카드 계수(Jacard coefficient) 산출, 김성훈(2005)과 Buck & Tatsuoka(1998)가 Q 행렬 타당화를 위해 실시한 중다회귀분석, LLTM(linear logistic test model, LLTM; fisher, 1973; 2005; De Boeck & Wilson, 2004)을 통해 타당화 분석을 실시하였다.

자카드 계수를 통해 Q 행렬에서의 각 인지요소 사이의 유사성을 확인하여 인지요소와 문항 간의 관계 설정이 타당한가를 분석하고, 서로 겹치거나 유사한 인지요소가 있어 인지요소의 삭제 여부를 살펴보기 위해 산출된다. 영어 검사 자료의 자카드 계수 산출 결과는 다음의 <표 IV-1>과 같다.

<표 IV-1> 영어 검사 자카드 계수 산출 결과

	Q1	Q2	Q3	Q4
Q2	0.1			
Q3	0.0	0.4		
Q4	0.2	0.2	0.1	
Q5	0.0	0.4	0.3	0.0

Henning의 기준으로 0.50을 넘는지 자카드 계수를 확인해보았을 때, 0.50을 넘는 경우가 없기 때문에 인지요소 간 유사성은 없다고 판단하였다. 따라서, 영어 검사의 Q 행렬이 타당함을 알 수 있었다. 중다회귀분석 결과는 아래 <표IV-2>와 같다.

<표 IV-2> 영어 검사 중다회귀 분석 결과

모형	비표준화계수		표준화계수	t	유의확률
	B	표준오차	오류		
(상수)	.60	.91		.67	.52
Q1	-.68	.71	-.44	-.95	.36
Q2	-.75	.92	-.36	-.81	.43
Q3	-.09	.34	-.09	-.25	.80
Q4	.30	.46	.23	.64	.54
Q5	.05	.31	.05	.17	.87

Q 행렬 타당화를 하기 위해서 중다회귀분석을 진행한 결과는 위의 표와 같다. Q 행렬 내에 인지요소가 정답률에 부적 영향을 미치는지 확인해본 결과, 유의하게 부적 영향을 미치는 경우는 발견되지 않았기 때문에 필요하지 않은 인지 요소는 없는 것으로 판단되어 추가적인 삭제는 필요하지 않음을 확인하였다.

다음으로 LLTM으로 문항모수를 측정하기 위하여 WinBUGS 프로그램을 통하여 추정하였다. 결과는 <표 IV-3>과 같다.

<표 IV-3> 영어 검사 LLTM 분석 결과

node	mean	standard error	2.5%	median	97.5%
eta[1]	0.3811	0.03382	0.3157	0.381	0.4477
eta[2]	0.2576	0.03077	0.1964	0.2575	0.3172
eta[3]	0.052	0.0244	0.0033	0.05185	0.0519
eta[4]	-0.2351	0.0244	-0.2981	-0.2351	-0.1718
eta[5]	-0.07902	0.02223	-0.122	-0.2351	-0.03466

LLTM 분석 결과 상에서 eta[1]~[5]값을 확인해보았을 때, 95% 신뢰구간 사이에 0이 포함되지 않기 때문에 eta[1]~[5]값이 유의미한 차이가 있는 것으로 보아 Q 행렬을 수정하지 않아도 되는 것을 확인할 수 있다.

나. 수학 검사 자료 Q 행렬 타당화

개발된 Q 행렬 초안을 확정하기 전에 Q 행렬 타당화를 위한 통계 분석을 진행하였다. 타당화를 위한 통계 분석 방법은 행렬간의 교집합의 크기를 합집합의 크기로 나누어 유사도를 나타내는 자카드 계수(Jacard coefficient) 산출, 김성훈(2005)과 Buck & Tatsuoka(1998)가 Q 행렬 타당화를 위해 실시한 중다회귀분석, LLTM(linear logistic test model, LLTM; fisher, 1973; 2005; De Boeck & Wilson, 2004)을 통해 타당화 분석을 실시하였다.

자카드 계수와 중다회귀를 통해 Q 행렬에서 인지요소 간 유사성을 확인하여 문항과 인지요소 간 관계 설정이 타당한지를 분석하였고, 이는 서로 겹치거나 유사한 인지요소가 있는지를 살펴보기 위해 산출된다. 그 결과는 <표 IV-4>, <표 IV-5>와 같다.

<표 IV-4> 수학 검사 자카드 계수 산출 결과

	Q1	Q2	Q3	Q4	Q5	Q6	Q7
Q2	0.1						
Q3	0.1	0.0					
Q4	0.0	0.0	0.0				
Q5	0.5	0.0	0.1	0.1			
Q6	0.1	0.0	0.0	0.0	0.2		
Q7	0.6	0.1	0.1	0.0	0.4	0.2	
Q8	0.3	0.0	0.4	0.0	0.2	0.2	0.3

<표 IV-5> 수학 검사 중다회귀 분석 결과

모형	비표준화계수		표준화계수	t	유의확률
	B	표준오차 오류	베타		
(상수)	.30	.47		.64	.53
Q1	-.09	.40	-.08	-.21	.84
Q2	-.29	.72	-.12	-.41	.69
Q3	.50	.64	.34	.78	.45
Q4	-.20	.71	-.08	-.29	.78
Q5	-.10	.36	-.09	-.28	.79
Q6	-.19	.44	-.145	-.44	.67
Q7	.08	.32	.073	.24	.82
Q8	-.11	.43	-.10	-.25	.81

Q 행렬 타당화를 하기 위해서 중다회귀분석을 진행한 결과는 위의 표와 같다. Q 행렬 내에 인지요소가 정답률에 부적 영향을 미치는지 확인해본 결과, 유의하게 부적 영향을 미치는 경우는 발견되지 않았기 때문에 필요하지 않은 인지 요소는 없는 것으로 판단되어 추가적인 삭제는 필요하지 않음을 확인하였다.

그러나 Henning의 기준으로 선행 연구에서 또한 자카드 계수가 0.50보다 크다면 유사성이 있다고 판단하여 인지요소 Q1과 Q5, Q1과 Q7이 유사하다고 판단할 수 있다. 따라서 Q1 계산하기 영역을 삭제한 후 다시 모형에 적용하고, 실제 자료를 추정하였다. Q1을 삭제한 후에 산출한 자카드 계수는 <표IV-6>과 같다.

<표 IV-6> Q1 삭제 후 자카드 계수 산출 결과

	Q2	Q3	Q4	Q5	Q6	Q7
Q3	0.0					
Q4	0.0	0.0				
Q5	0.0	0.1	0.1			
Q6	0.0	0.0	0.0	0.2		
Q7	0.1	0.1	0.0	0.4	0.2	
Q8	0.0	0.4	0.0	0.2	0.2	0.3

Q1을 삭제한 후 자카드 계수가 0.50 이상인 값이 없어 인지요소간 유사성이 없는 것을 알 수 있다. 또한 Q1을 삭제한 후 추가적으로 중다회귀분석을 실시하였다.

<표 IV-7> Q1 삭제 후 중다회귀분석 결과

모형	비표준화계수		표준화계수	t	유의확률
	B	표준오차 오류	베타		
(상수)	2.17	.90		2.42	.03
Q2	-.71	1.29	-.14	-.55	.59
Q3	-.69	.99	-.23	-.70	.50
Q4	-1.86	1.15	-.37	-1.62	.13
Q5	-1.11	.74	-.45	-1.50	.16
Q6	-1.08	.66	-.40	-1.65	.12
Q7	-.46	.54	-.22	-.85	.41
Q8	-.43	.66	-.19	-.65	.53

인지요소 Q1을 삭제한 후, 중다회귀분석을 진행한 결과는 위의 표와 같다. Q 행렬 내에 인지요소가 정답률에 부적 영향을 미치는지 확인해본 결과, 유의하게 부적 영향을 미치는 경우는 발견되지 않았기 때문에 필요하지 않은 인지 요소는 없는 것으로 판단되어 추가적인 삭제는 필요하지 않음을 확인하였다.

다음으로 LLTM으로 문항모수를 측정하기 위하여 WinBUGS 프로그램을 통하여 추정하였다. 결과는 <표 IV-8>과 같다.

<표 IV-8> 수학 검사 LLTM 분석 결과

node	mean	standard error	2.5%	median	97.5%
eta[1]	0.5218	0.0290	0.4642	0.5221	0.579
eta[2]	-0.4881	0.0589	-0.601	-0.4885	-0.3738
eta[3]	0.4326	0.0457	0.3405	0.4327	0.5229
eta[4]	1.591	0.0553	1.487	1.592	1.701
eta[5]	0.0927	0.0232	0.0473	0.0925	0.1379
eta[6]	1.166	0.0340	1.099	1.166	1.232
eta[7]	-0.4232	0.0256	-0.4725	-0.4234	-0.3725
eta[8]	-0.1448	0.0345	-0.2133	-0.1445	-0.0792

LLTM 분석 결과 상에서 eta[1]~[8]값의 95%의 신뢰구간 사이에 0이 포함되지 않기 때문에 유의미한 차이가 있는 것으로 보아 Q 행렬을 수정할 필요가 없는 것을 알 수 있다.

V. 결론 및 논의

본 연구에서의 목적은 실제 검사 자료의 Q 행렬 타당화를 진행함에 있어 LLTM(Linear Logistic Test Model)의 활용 가능성을 탐색해보는 것에 있다. 검사 자료는 중학교 3학년 영어 검사와 중학교 2학년 수학 검사에서 각각 3,000명, 2,000명을 임의 표집하였고, 자카드 계수, 중다회귀, LLTM을 사용하여 Q 행렬 타당화를 진행하였다.

본 연구의 결과를 바탕으로 논의하면 다음과 같다.

첫째, 영어 검사 자료는 총 5개의 인지요소로 구성되어 있고 Q 행렬 타당화를 진행한 결과, 자카드 계수, 중다회귀, LLTM 세 방법 모두 부적절한 인지요소가 없다는 일관된 결과를 보여 Q 행렬 타당화 기법으로써 LLTM을 활용할 수 있는 가능성을 알 수 있었다.

둘째, 총 8개의 인지요소로 구성되어 있는 수학 검사 자료를 Q 행렬 타당화 과정에서 자카드 계수, 중다회귀, LLTM을 사용한 결과, 자카드 계수에서만 인지요소 Q1이 부적절하다고 나왔으나 중다회귀와 LLTM은 일치한 결과를 보였다. 실제 검사 자료를 활용하여 Q 행렬 타당화를 하는 방법으로 LLTM의 가능성을 탐색하였다.

셋째, 실제 검사 자료에 Q 행렬 타당화 기법으로 LLTM을 사용해봄으로써 Q 행렬 타당화 기법으로써 LLTM의 활용가능성을 확인하였다. Q 행렬을 제작할 때, Q 행렬 타당화는 굉장히 중요한 과정이기 때문에 기존에 수행하던 Q 행렬 타당화 기법인 자카드 계수, 중다회귀 등과 같은 방법과 LLTM을 함께 사용하면 더 정교한 Q 행렬을 제작할 수 있을 것이다.

본 연구 LLTM(Linear Logistic Test Model)을 실제 영어, 수학 인지진단 검사 자료의 Q 행렬 타당화 기법으로 적용함으로써 LLTM의 Q 행렬 타당

화 활용 가능성을 확인하였다는 것에 그 의의가 있으나 다음과 같은 한계가 있다.

첫째, 본 연구에서는 중학교 2학년 수학검사 자료와 중학교 3학년 영어 검사 자료에 한정적으로 Q 행렬 타당화를 실시하였으나 추후에 수학, 영어 뿐만 아닌 다른 과목 검사 자료를 분석하고, 더 많은 인지요소를 보유한 검사 자료를 활용하여 보다 더 많은 실제 검사 자료에 LLTM을 적용해볼 필요가 있다.

둘째, 실제 검사 자료에 대해 Q 행렬 타당화를 진행한 결과 자카드 계수, 중다회귀와 비교하였을 때, LLTM을 사용하여 Q 행렬 타당화가 원활히 이루어지고 있음을 보았다. 다만 실제 자료에 한정하여 진행하였기 때문에 Q 행렬 타당화에서 LLTM의 활용성을 검증하기 위해서는 추가적인 시뮬레이션 연구를 진행할 필요가 있다. 예를 들어 Q 행렬에서 오지정한 요소를 포함하는 비율, 오지정 종류(과잉, 과소, 혼합) 등을 설정하여 시뮬레이션 연구를 진행할 수 있을 것이다.

셋째, LLTM은 문항 난이도의 변산이 모형에서 고려된 문항 특성 혹은 공변인들에 의해 완벽하게 설명된다는 매우 강한 가정을 갖기 때문에 현실적이지 못하다는 비판을 흔히 받는다. 따라서 약점을 보완한 RE-LLTM에 대한 수행이 필요할 것으로 보인다. RE-LLTM 방법이 LLTM에 대해 갖는 차별점은 문항 난이도에 대한 설명에서 문항 난이도의 선형결합에 더하여 잔차변산을 고려한다는 것이다. 식으로 표현하면 $\beta_i = \sum_{k=1}^K q_{ik}\eta_k + \epsilon_i = \beta'_i + \epsilon_i$ 과 같으며, 결과적으로 문항 모수와 피험자 모수 모두 무선적 효과를 갖는 것이기 때문에 교차적 무선 효과모형(Janssen et al., 2004)이 된다.

넷째, 본 연구에서는 Q 행렬 타당화 기법으로 자카드 계수, 중다회귀, LLTM을 소개하고 있는데 국내·외에서 Q 행렬 타당화에 관련한 연구가 활발히 진행되면서 RSS, 인공신경망 등 다양한 Q 행렬 타당화 기법들이 사용

되고 있다. LLTM의 활용 가능성을 살펴보기 위해선 추가적으로 다른 Q 행렬 타당화 방법과의 비교가 필요할 것으로 보인다.

다섯째, 보다 복잡한 위계적인 인지요소 구조를 가진 Q 행렬에 적용해보고, 다양한 인지진단 모형을 고려하여 실질적으로 문항 모수, 피험자 모수에 어떤 영향을 미치는지 살펴볼 필요가 있다.

참 고 문 헌

- 김성은(2015). 다집단 인지진단모형의 적용과 Q 행렬 타당화 방법 간의 회복률 비교: 정서·행동문제 검사에의 적용. 이화여자대학교 대학원.
- 김성훈(2005). 인지구조모형에 근거한 학생의 지식상태 진단. *교육학연구*, 43(1), 81-107.
- 김명연(2016). 보상성·비보상성 가정이 심리측정 모형을 통한 다차원적 능력 추정에 미치는 영향. 성신여자대학교 대학원.
- 김희경, 한정아, 최숙기, 김부미(2012). 인지진단모형을 적용한 학업성취 프로파일 분석 및 결과 보고 방안. *한국교육과정평가원 연구보고 RRE 2012-7*.
- 안희상(2017). 인지진단 검사의 Q 행렬 타당성 분석 방법 연구. 서울대학교 대학원.
- 양희원, 이규민, 강상진(2019). Q 행렬 오지정 및 인지요소 위계구조가 인지진단모형 추정 양호도에 미치는 영향. *교육평가연구* 제32권 제1호, 53-79.
- 이영주(2014). 인공신경망에 근거한 인지진단모형 Q 행렬의 타당성 평가. 이화여자대학교 대학원.
- 지은림, 채선희(2000). Rasch 모형의 이론과 실제. 서울: 교육과학사.
- Boorsboom, D., & Mellenberg, G. D. (2007). Test validity in cognitive assessment. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications*, 85-115. U.K.: Cambridge University Press

- Buck, G., & Tatsuoka, K.(1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test, *Language Testing*, 15(2), 119-257.
- DeCarlo, T. (2011). On the Analysis of Fraction Subtraction Data: The DINA Model, Classification, Latent Class Sizes, and the Q-Matrix. *Applied Psychological Measurement*, 35(1), 8-26.
- DiBello, L., Stout, W., & Roussos, L. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361-389). Hillsdale, NJ: Erlbaum
- Madison, M. J., & Bradshaw, L. P. (2015). The effects of Q-matrix design on classification accuracy in the log-linear cognitive diagnosis model. *Educational and Psychological Measurement*, 75(3), 491-511.
- Rupp, A. A., & Templin, J. L. (2008). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, 68(1), 78-96.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. N.Y.: Guilford Press.
- 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.

ABSTRACT

An exploratory study on the usability of the LLTM as a Q matrix validation method

Hyunmi Kim
Department of Education
Graduation School of
Sungshin University

Korea's existing evaluation methods such as classical examination theory and question-and-answer theory are effective in determining the current level of ability of subjects, but lack in terms of providing specific information and giving feedback to improve teaching and learning. As a result, cognitive diagnostic models that can provide specific feedback by identifying the subjects' strengths and weaknesses have recently attracted attention (Lee Young-joo, 2014). The cognitive assessment is an evaluation method designed to measure the status of students' knowledge structure and their ability to process and provides information on their cognitive strengths and weaknesses (Leighton & Gierl, 2007).

The cognitive assessment considers the prior knowledge and functions

required for learning as a combination of several cognitive factors, allowing students to receive feedback on which factors they have mastered and which ones they have not mastered (Jang, 2008). The cognitive assessment provides a profile of individual students' cognitive status by analyzing their responses to each question based on the relationship between each question and the cognitive factors they wish to measure (Kim Hee-kyung et al., 2012). Therefore, in order to proceed with the cognitive examination suitable for the purpose, the appropriate questions to measure each recognition element must be used, as well as the relationship between the recognition element and the question must be correctly assumed.

The Q matrix of cognitive diagnostic tests is a matrix that indicates the relationship between the test questions and elements, and is produced by experts who know the test content based on theories and other practical considerations (Dibello et al., 2007). However, because the relationship between the cognitive component and the question is not clearly apparent, it is difficult to guarantee that the Q matrix made by the expert's decision is appropriate and may differ from the true Q matrix of the actual test. Nevertheless, cognitive assessment conducted at educational sites relies mainly on expert review levels for the rationalization of the Q matrix.

Accordingly, studies were conducted on how to statistically justify the existing Q matrix, such as the calculation of the Jacard coefficient, which shows the similarity by dividing the size of the intersection by the size of the assembly, and the analysis of the multiple return conducted by

Kim Sung-hoon (2005) and Buck & Tatsuoka (1998) to justify the Q-matrix.

Inaccurate Q matrices can have unwanted consequences, such as capability measurements of the wrong subjects, incorrect model parameters measurements, poor model fit, and so on, so it is very important to create an appropriate way to verify the validity of the Q matrix for successful use (Rup & Templin, 2008; Feng, 2013). In this study, we would like to present the availability of the Linear Logistic Test Model (LLTM) as a way to examine the validity of the Q matrix and compare its performance with the existing methods.

Key words: Q-matrix validation, cognitive diagnosis model, DINA, DINO, Jacard coefficient, multiple regression, LLTM