

## 내용 및 인지 영역을 함께 고려한 평가 데이터 분석을 위한 Q행렬 기반 다차원 문항반응모형의 활용 방안 연구: TIMSS 2019 수학 평가 분석

김 래 영 (서울대학교, 강사)  
황 수 빈 (서울대학교, 대학원생)  
이 슬 기 (서울대학교, 대학원생)  
유 연 주 (서울대학교, 교수)<sup>†</sup>

본 연구의 목적은 수학 평가 데이터를 내용 및 인지 영역을 함께 고려하여 분석하기 위해, da Silva(2019)가 제시한 Q행렬 기반의 다차원 문항반응모형(M2PL-Q)의 활용 방안을 제시하고, 이를 TIMSS 2019 8학년 수학 평가 데이터에 적용하여 분석한 결과를 제시하는 것이다. 연구 결과 M2PL-Q 모형을 통해 학생의 능력 수준을 내용 영역과 인지 영역에 걸쳐서 추정할 수 있음을 확인하였으며, 각 영역에 대한 능력 수준이 서로 연관되어 있는 양상을 확인하였다. 또한, 문항의 특성을 영역별로 구분하여 추출할 수 있으며, 문항에 따라 내용 영역과 인지 영역이 문제 해결에 미치는 영향이 서로 다를 수 있음을 확인하였다. 본 연구는 기존의 방법에서는 별도로 분석되던 내용 영역과 인지 영역을 하나의 분석 모형에 포함하여 평가 데이터를 종합적으로 분석하는 방안을 제시했다는 점에서 의의를 지닌다. 각 영역에 대해 추정된 능력 수준을 개별 학생에 대한 진단에 활용하면, 학생이 세부적인 내용 및 인지 영역에서 보이는 강점과 약점을 파악하여 학습을 지원할 수 있을 것으로 기대한다. 또한, 각 평가 문항의 세부적인 특성을 고려하여 평가의 상황과 목적에 따라 적절하게 활용함으로써, 평가의 타당성과 효율성을 높이고 학생의 능력 수준을 보다 효과적으로 진단할 수 있을 것으로 기대한다.

### I. 서론

학생들의 수학 교과에 대한 능력과 성취는 개념적 지식에 해당하는 내용 영역과 개념적 지식을 활용하는 방법에 해당하는 인지 영역에 복합적인 영향을 받아서 향상된다(박지현, 김수진, 2015; 한채린, 박만구, 2015; Pogoy et al., 2015). 이를 고려하여, 학교 현장에서 인지 영역을 내용 영역과 함께 다루고 가르칠 필요성이 강조되어 왔으며(Harks et al., 2014), 최근에는 교육과정 문서에서도 알아야 하는 교과 지식을 통해 내용 영역을 명시할 뿐 아니라 교과 지식으로 무엇을 어떻게 해야 하는지와 같은 진술을 통해 인지 영역을 함께 강조하고 있다(송미영, 김선희, 2007). 우리나라의 2009 개정 수학과 교육과정에서는 수학적 문제해결, 수학적 추론, 수학적 의사소통을 포함하는 수학적 과정을 제시하여 인지 영역에 해당하는 과정적 지식을 강조하였으며, 이는 2015 개정 교육과정의 수학 교과 역량으로도 연결된다(박지현, 김수진, 2015). 2022 개정 교육과정에서는 내용 체계를 지식·이해, 과정·기능, 가치·태도로 나누어 제시하고 있으며, 이중 과정·기능이 지식을 습득하는 데 활용되는 사고 및 탐구 과정에 해당한다는 점에서 인지 영역과 관련된다(교육부, 2022). 또한, 우리나라의 국가수준 학업성취도 평가나 수학·과학 성취도 추이변화 국제비교 연구(Trends in International Mathematics and Science Study;

\* 접수일(2024년 8월 26일), 심사(수정)일(2024년 9월 11일), 게재확정일(2024년 9월 26일)

\* MSC2020분류 : 97D60

\* 주제어 : 내용 영역, 인지 영역, 문항반응이론, Q행렬, TIMSS 2019

† 교신저자 : yyoo@snu.ac.kr

TIMSS와 국제학업성취도평가(Programme for International Student Assessment; PISA)와 같은 국제 수준의 학업성취도평가에서도 내용 영역과 인지 영역을 함께 고려하여 평가들을 구성하고 있다(임해미 외, 2018; Mullis & Martin, 2017; OECD, 2023).

평가 결과 데이터를 바탕으로 학생들의 능력을 통계적인 모형을 이용하여 분석하는 연구들에서도 내용 영역 뿐 아니라 인지 영역이 강조되고 있다(Harks et al., 2014). 평가 데이터를 분석할 때 인지 영역의 구분을 고려하는 것이 모형의 적합성을 높일 수 있음이 보고되어 왔으며(Wu & Adams, 2006; Young et al., 2008), 인지 영역을 복수의 하위영역으로 구분하여 각 영역에 대한 학생들의 능력을 분석하고 영역 사이의 관계를 확인하는 연구들도 이루어져 왔다(Arkan, 2015; Wu & Adams, 2006; 박지현 & 김수진, 2015). 한편, 대부분의 연구들은 평가 데이터를 분석할 때 내용 영역과 인지 영역 중 하나의 영역에 대해서만 분석을 진행하고 있다. 학생들이 수학 문항을 해결하기 위해서는 내용 영역에 해당하는 개념적 지식과 인지 영역에 해당하는 과정적 역량이 모두 필요하다는 점에서(Anderson et al., 2001), 평가 데이터는 두 영역을 함께 고려하여 분석하는 것이 필요하다. 또한, 학생들의 내용 영역에 대한 능력과 인지 영역에 대한 능력은 서로 밀접하게 연관되어 있으므로(George & Robitzsch, 2018), 두 영역을 모두 포함하여 각 영역의 능력 사이의 관계를 반영한 분석이 이루어질 필요가 있다.

내용 영역과 인지 영역의 구분은 문항의 특성을 분석할 때도 중요한 역할을 할 수 있다. 수학 능력을 측정하는 많은 평가들에서는 평가들을 구성할 때 문항의 특성을 내용 영역과 인지 영역의 두 차원으로 나누어 구성하고 있다(Balfaqeeh et al. 2022; Gierl et al., 2003). 예를 들어, TIMSS 2019의 8학년 수학 평가들에서는 내용 영역을 수(Number), 대수(Algebra), 기하(Geometry), 자료와 확률(Data and Probability)의 네 가지 영역으로, 인지 영역을 알기(Knowing), 적용하기(Applying), 추론하기(Reasoning)의 세 가지 영역으로 구성한 뒤, 각 평가 문항을 내용 영역 중 하나와 인지 영역 중 하나에 대응시켜 평가들을 개발하고 있다. 한편, 동일한 내용 및 인지 영역에 해당하는 문항이라고 하더라도, 어떤 문항은 내용 영역과 관련하여 개념적 지식을 잘 알고 있는 것이 문제 해결을 위해 더 중요할 수 있고, 어떤 문항은 인지 영역과 관련하여 개념을 적용하고 활용하는 것이 문제 해결을 위해 더 중요할 수 있다. 한편, 이러한 문항의 세부적인 특성을 전문가가 직접 판단하는 것은 시간과 비용이 소모되는 일이며, 전문가가 판단한 문항의 특성이 실제 문항의 특성과 다를 수도 있다. 평가 데이터를 내용 및 인지 영역을 모두 포함한 모형을 이용하여 분석하면, 이와 같은 문항의 내용 및 인지 영역에 대해 구분된 특성을 데이터에 기반하여 추정할 수 있다. 또한, 세부적으로 추정된 문항의 특성을 바탕으로 학생의 내용 및 인지 영역에 대한 능력을 보다 높은 정확도로 추정할 수 있으며, 추정된 문항의 특성은 문항을 개발하고 활용하는 상황에서도 유용한 정보를 제공할 수 있다.

내용 영역과 인지 영역을 구분하여 학생의 능력과 문항의 특성을 분석하기 위해서는 다음과 같은 사항을 반영한 분석이 필요하다. 첫째, 학생의 능력이 내용 영역과 인지 영역을 모두 포함하여 구성될 수 있도록 다차원적인 가정을 동반하여 분석해야 한다. 둘째, 문항들의 특성이 해당하는 내용 영역 및 인지 영역으로 조합되어 구성되어 있는 상황을 고려하여 분석해야 한다. 기존의 평가 데이터 분석에 많이 활용되는 문항반응이론(Item Response Theory; IRT)을 확장한 다차원 문항반응이론(Multidimensional IRT; MIRT)의 경우, 학생의 능력과 문항의 특성이 다차원적으로 구성되어 있다고 가정하긴 하나, 각 문항이 하나의 영역에만 해당한다고 가정하므로 문항이 복수의 영역의 조합에 해당하는 상황을 분석하는 데는 한계가 존재한다(da Silva et al., 2019; Harks et al., 2014; Jang, 2022; Zhang, 2004). 또 다른 분석 모형인 인지진단모형(Cognitive Diagnostic Model; CDM)을 활용하면 위의 두 가지 사항을 만족하는 분석이 가능하지만, 인지진단모형에서는 학생의 능력이 각 검사 영역에 대하여 숙달과 미숙달의 이분적인 구분으로 구성되어 있다고 가정하게 된다. 인지진단모형은 이와 같은 이분적인 가정을 바탕으로 모형이 포함하는 모수들을 안정적으로 추정할 수 있다는 장점이 있지만, 이분적인 구분으로는 각 영역에 대한 학생의 능력과 관련하여 세부적인 정보를 제공하기는 힘들다는 한계가 존재한다(Shu et al., 2023).

이에 본 논문에서는 da Silva 외(2019)가 다차원 문항반응이론에 Q행렬을 적용하여 제시한 M2PL-Q 모형(multidimensional two-parameter logistic model with a Q-matrix)을 TIMSS 2019 8학년 수학 평가 데이터에 적용하고, 이를 통해 학생들의 능력과 문항의 특성을 내용 영역과 인지 영역을 함께 고려해서 분석을 진행하고자 한다. M2PL-Q 모형을 이용할 경우, 다차원 문항반응이론을 기반으로 학생의 능력과 문항의 특성을 다차원적인 가정과 함께 분석할 수도 있고, 문항과 검사 영역 사이의 관계를 나타내는 행렬인 Q행렬을 기반으로 각 문항이 여러 검사 영역에 걸쳐있는 상황을 반영하여 분석할 수도 있다. 또한, 학생의 각 영역에 대한 능력이 연속적인 값을 갖는다고 가정하므로, 각 내용 영역과 인지 영역에 대한 학생의 능력을 세밀하게 분석할 수도 있다. 따라서 본 연구에서는 M2PL-Q 모형으로 TIMSS 2019 8학년 수학 평가 데이터를 분석하여 학생들의 내용 및 인지 영역에 대한 능력을 추정하고, 추정 결과를 바탕으로 우리나라 학생들의 각 영역에 대한 능력을 확인하고 내용 및 인지 영역 사이의 관계를 분석해보고자 한다. 또한, TIMSS 2019 8학년 수학 평가 문항의 특성을 내용 영역과 인지 영역을 구분하여 분석하고, 각 영역에 대한 문항 특성들 사이의 관계를 살펴보고자 한다.

## II. 연구의 배경

### 1. 문항반응이론

문항반응이론(Item Response Theory; IRT)은 문항에 대한 응답이 학생의 능력 수준 및 문항의 특성의 두 가지에 의하여 결정된다고 가정하는 모형이다(Embretson & Reise, 2000). 문항반응이론에서는 각 학생의 잠재적인 능력 수준이 단일한 차원의 값으로 구성된다고 가정한다. 문항의 특성은 모형에 따라 다르게 가정되며, 보통 난이도(difficulty), 변별도(discrimination), 추측도(guessing) 등으로 이루어진다. 문항의 특성이 난이도만으로 구성되는 모형을 1-모수(1PL) 모형, 난이도와 변별도로 구성되는 모형을 2-모수(2PL) 모형, 난이도와 변별도 및 추측도의 세 가지로 구성되는 모형을 3-모수(3PL) 모형이라고 부른다. 예를 들어, 2-모수 모형에서는 학생이 문항  $j$ 의 정답을 맞힐 확률을 학생의 능력  $\theta$ , 문항  $j$ 의 난이도  $b_j$ 와 변별도  $a_j$ 에 대해 아래와 같이 표현한다.

$$P(X_j = 1 | \theta, a_j, b_j) = \frac{1}{1 + \exp[-a_j(\theta - b_j)]}$$

위의 식에서 문항의 응답  $X_j$ 는 문항  $j$ 를 맞혔을 경우 1이고 틀렸을 경우 0으로 표현된다. 문항의 특성은 위의 확률식을  $\theta$ 에 대한 함수로 바라보고 해석할 수 있다. 문항의 난이도  $b_j$ 는 문항  $j$ 를 맞힐 확률이 0.5가 되도록 하는  $\theta$ 의 값에 해당하며, 난이도가 높을수록 문항을 맞히기 어려움을 의미한다. 문항의 변별도  $a_j$ 는 변곡점( $\theta = b_j$ 인 지점)에서의 함수의 변화율에 해당하며, 변별도가 높을수록 문항  $j$ 를 맞히기 위하여  $\theta$ 가 많은 영향을 미친다는 것을 의미한다. 또한 변별도가 높을수록 능력이 높은 학생과 낮은 학생을 잘 구분할 수 있는 문항임을 의미하기도 한다.

문항반응이론에서는 기본적으로 검사가 단일 영역을 측정하는 문항들로 구성되어 있다고 가정하고, 학생의 능력도 단일 차원의 값으로 구성된다고 가정한다. 한편, 실제 검사는 단일한 영역을 측정하는 문항들로만 구성되기보다 여러 세부 영역을 측정하도록 구성되는 경우가 많으므로, 학생의 능력 수준 또한 검사와 관련된 세부 영역들에 대하여 다차원적인 프로파일로 존재한다고 가정하는 것이 보다 실제 상황을 잘 반영할 수 있다(Ackerman, 1994; Buck, 1994; Wu & Adams, 2006). 이와 같은 다차원적인 가정을 동반하여 분석하기 위하여 다차원 문항반응이론(Multi-dimensional IRT; MIRT)이나 인지진단모형(Cognitive Diagnostic Model; CDM) 등이 제시되고 연구되어 왔다. 한편, 다차원 문항반응이론은 학생의 능력이 복수의 검사 영역에 대하여 다차원적으로 존재한다고 가정하긴 하나, 각 문항은 검사 영역 중 하나에만 관련된다고 가정하는 등 문항이 내용 및 인지

영역과 같이 여러 영역에 걸쳐있는 상황을 분석하는 데는 한계가 존재한다(da Silva et al., 2019; Harks et al., 2014; Jang, 2022; Zhang, 2004). 이로 인해 다차원 문항반응이론을 활용하여 평가 데이터를 분석한 연구들에서는 내용 혹은 인지 영역 중 하나의 영역을 선택하여 모형을 구성하고, 각 평가 문항이 선택된 영역에 포함된 세부 영역 중 하나에만 관련되어 있는 상황으로 분석을 진행하고 있다(Rijmen et al., 2014; von Davier, 2020). 또한, 인지진단모형의 경우 문항과 영역 사이의 관계를 나타내는 Q행렬을 도입하여 문항이 여러 영역에 걸쳐있는 상황을 반영할 수 있지만, 각 영역에 해당하는 학생의 능력이 연속적인 값이 아니라 숙달과 미숙달의 이분적인 값으로 구성된다고 가정한다. 하지만 이와 같은 숙달 혹은 미숙달의 이분적인 구분은 각 영역에 대한 학생의 능력을 세밀하게 진단할 수는 없다는 한계가 존재하고, 이로 인해 많은 측정 상황에서는 학생의 능력 수준이 이분적이기 보다 연속적인 값을 갖는다고 가정하게 된다(da Silva et al., 2019; Shu et al., 2023).

## 2. Q행렬을 도입한 다차원 문항반응이론

da Silva 외(2019)는 인지진단모형에서 사용되는 Q행렬을 다차원 문항반응이론에 도입하여, 검사 영역 및 학생의 능력이 다차원적으로 구성된다고 가정하면서, Q행렬을 구성하는 영역을 바탕으로 학생의 능력을 해석할 수 있는 M2PL-Q 모형(Multidimensional 2PL model with Q-matrix)을 제시하였다. 검사 영역이  $K$ 개로 구성되어 있다고 가정할 때, M2PL-Q 모형에서는 학생의 능력  $\theta$ 가 아래와 같이  $K$ 차원의 프로파일로 구성되며 각 차원의 능력 수준이 연속적인 값을 갖는다고 가정한다.

$$\theta = (\theta_1, \theta_2, \dots, \theta_K)$$

다음으로, 학생이 문항에 응답할 확률을 모델링하기 위하여 Q행렬(Tatsuoka, 1983)을 구성한다. Q행렬은 문항과 검사 영역 사이의 관계를 나타내는 행렬로, 문항의 개수  $J$ 와 검사 영역의 개수  $K$ 에 대하여  $J \times K$  행렬로 표현된다. Q행렬의  $(j, k)$  성분  $q_{jk}$ 는 아래와 같이 문항  $j$ 가 영역  $k$ 와 관련이 있으면 1로 표시되고 관련이 없으면 0으로 표시된다.

$$Q = (q_{jk}), q_{jk} = \begin{cases} 1, & \text{문항 } j \text{가 영역 } k \text{와 관련 있는 경우} \\ 0, & \text{그렇지 않은 경우} \end{cases}$$

기존의 다차원 문항반응이론의 경우 각 평가 문항이 여러 영역 중 하나에만 해당한다고 가정하므로, Q행렬을 구성하면 각 행이 하나의 성분만 1이고 나머지 성분은 모두 0이게 된다. M2PL-Q 모형은 하나의 평가 문항이 여러 영역과 관련될 수 있다고 가정하므로 Q행렬의 각 행이 두 개 이상의 성분에서 1의 값을 가질 수 있다는 점에서 차이가 존재하며, M2PL-Q 모형이 기존의 다차원 문항반응이론을 보다 다양한 상황에 적용할 수 있도록 확장한 것으로 바라볼 수 있다.

M2PL-Q 모형에서는 문항반응이론의 2-모수 모형과 마찬가지로 문항의 특성으로 난이도와 변별도를 도입한다. 문항의 난이도는 각 문항에 대하여 단일한 값으로 존재한다고 가정되며, 문항의 변별도는  $K$ 개의 영역 중 문항이 해당하는 영역, 즉 Q행렬에서 1로 표시된 영역에 대해서만 존재한다고 가정된다. 따라서 문항  $j$ 의 특성은 난이도  $b_j$  및  $k$  번째 영역에 대한 변별도  $a_{jk}$ 에 대하여 아래와 같이 표현되며

$$\xi_j = (a_{j1}, a_{j2}, \dots, a_{jK}, b_j)$$

문항이  $k$  번째 영역에 해당하지 않는 경우  $a_{jk} = 0$ 의 값을 갖게 된다.

위의 가정을 바탕으로 M2PL-Q 모형에서 능력이  $\theta$ 인 학생이 문항  $j$ 를 맞힐 확률은 아래와 같이 로지스틱 함수(logistic function)를 이용하여 표현된다.

$$P(X_{ij} = 1 | \theta, \xi_j) = \frac{1}{1 + \exp(-\eta_{ij})}$$

$$\eta_{ij} = \sum_{k=1}^K q_{jk} a_{jk} \theta_{ik} - b_j$$

위의 식에서, 난이도  $b_j$ 는 문항  $j$ 가 해당하는 모든 영역에 대해 능력이 0인 학생이 문항  $j$ 를 맞힐 확률을 결정한다.  $k$  번째 영역에 대한 변별도  $a_{jk}$ 는 해당 영역의 능력  $\theta_{ik}$ 에 가중치로 적용되므로,  $a_{jk}$ 가 높을수록 문항을 맞히기 위해  $\theta_{ik}$ 가 많은 영향을 미친다는 것을 의미하며,  $\theta_{ik}$ 가 높은 학생과 낮은 학생을 잘 구분할 수 있는 문항임을 의미한다. M2PL-Q 모형에서는 학생의 문항 응답 데이터와 Q행렬을 통해서 각 학생의  $K$ 개 영역에 대한 능력 수준 및 각 문항의 특성을 추정하게 된다.

이상에서 살펴본 바와 같이 M2PL-Q 모형은 Q행렬을 도입하여 평가 문항이 여러 영역에 걸쳐서 해당하는 상황을 분석할 수 있으며, 문항의 변별도가 문항이 해당하는 영역들에 대하여 개별적으로 존재하므로 문항의 특성을 영역별로 구분하여 파악할 수 있다. 따라서 본 연구에서는 TIMSS 2019 8학년 수학 평가 데이터에 M2PL-Q 모형을 적용하여 내용 영역과 인지 영역을 모두 포함하도록 모형을 구성하고, 각 평가 문항이 내용 및 인지 영역에 걸쳐 복수의 영역에 관련되어 있는 상황을 반영하여 분석을 진행하였다. 이를 통해, 학생의 내용 및 인지 영역에 대한 능력 수준을 함께 진단하고, 문항의 특성을 내용 영역과 인지 영역을 구분하여 추정하고자 하였다.

### 3. 연구방법

#### 가. 연구 자료

본 연구에서는 TIMSS 2019 8학년(중학교 2학년) 수학 평가 결과를 바탕으로 우리나라 학생들의 잠재 능력 및 문항 특성을 분석하였다. TIMSS는 4학년과 8학년 학생들을 대상으로 수학 및 과학 성취도를 조사하는 국제 비교 연구로, TIMSS 2019에서는 수학 평가틀을 평가해야 할 내용을 구체화한 내용 영역과 평가해야 할 사고과정에 해당하는 인지 영역으로 구분한다(Mullis & Martin, 2017). TIMSS 2019 8학년 수학에서는 <표 II-1>과 같이 내용 영역이 수(Number), 대수(Algebra), 기하(Geometry), 자료와 확률(Data and Probability)의 네 영역으로 구성되어 있으며, 인지 영역이 알기(Knowing), 적용하기(Applying), 추론하기(Reasoning)의 세 영역으로 구성되어 있다(Mullis & Martin, 2017). 내용 영역 중 수 영역은 정수, 분수와 소수, 비·비례·백분율의 하위영역으로 이루어져 있으며, 대수 영역은 식·연산·방정식, 관계와 함수의 하위영역으로 이루어져 있다. 기하 영역은 기하 도형과 측정이라는 하위요소로 구성되어 있으며, 자료와 확률은 자료, 확률의 두 하위요소로 구성되어 있다. 인지 영역 중 알기 영역은 학생들이 알아야 할 사실, 개념, 절차를 포함하며, 회상하기, 인식하기, 분류·정렬하기, 계산하기, 인출하기, 측정하기의 6가지의 하위요소로 구성되어 있다. 적용하기 영역은 문제해결 과정에서 지식이나 개념의 이해를 적용하는 능력에 초점을 맞춘 것으로, 결정하기, 표현·모델링하기, 실행하기의 5가지의 하위영역으로 구성되어 있다. 추론하기 영역은 다양한 상황, 맥락에서의 다단계 문제해결 과정을 포함하는 것으로, 분석하기, 통합·종합하기, 평가하기, 결론짓기, 일반화하기, 정당화하기의 5가지의 하위요소로 구성되어 있다.

&lt;표 II-1&gt; TIMSS 2019 8학년 수학 평가들의 내용 및 인지 영역과 하위요소

영역		하위요소
내용	수 (Number)	정수, 분수와 소수, 비·비례·백분율
	대수 (Algebra)	식·연산·방정식, 관계와 함수
	기하 (Geometry)	기하 도형과 측정
	자료와 확률 (Data and Probability)	자료, 확률
인지	알기 (Knowing)	회상하기, 인식하기, 분류·정렬하기, 계산하기, 인출하기, 측정하기
	적용하기 (Applying)	결정하기, 표현·모델링하기, 실행하기
	추론하기 (Reasoning)	분석하기, 통합·종합하기, 평가하기, 결론짓기, 일반화하기, 정당화하기

TIMSS 2019 8학년 수학 평가의 각 문항은 4개의 내용 영역 중 하나와 3개의 인지 영역 중 하나에 해당하게 된다. 분석에는 8학년 수학 평가 문항 중 응답이 존재하지 않는 문항을 제외하고, 소문항들을 합쳐서 하나의 응답으로 기록된 경우는 하나의 문항으로 취급하여 총 206개의 문항을 분석에 활용하였다(Fishbein et al., 2020). <표 II-2>는 분석에 활용된 문항에 대하여 각 영역에 해당하는 문항의 수와 비율을 나타낸 것이다(Foy et al., 2020).

&lt;표 II-2&gt; TIMSS 2019에서 사용된 문항 중 각 내용 및 인지 영역에 해당하는 문항의 수

영역		인지 영역			합계
		알기	적용하기	추론하기	
내용 영역	수	26	28	9	63 (30.6%)
	대수	24	24	13	61 (29.6%)
	기하	4	21	18	43 (20.9%)
	자료와 확률	10	23	6	39 (18.9%)
합계		64 (31.1%)	96 (46.6%)	46 (22.3%)	206

#### 나. 연구 대상

본 연구에서는 TIMSS 2019 8학년 수학 평가 결과를 바탕으로 우리나라 중학교 2학년 학생들의 각 영역별 능력 수준을 다른 국가 학생들과 비교하고자, 한국, 싱가포르, 대만, 일본의 4개 국가 학생들의 응답 데이터를 분석하였다. 전반적인 영역에 대한 수학 성취도가 비슷한 국가들에 대해 세부적인 내용 및 인지 영역의 능력 수준 차이를 비교하기 위하여, 전체 영역에 대한 평균 수학 성취도를 기준으로 하여 4개 국가를 연구 대상으로 선정하였다. 우리나라를 포함하여 선정된 4개 국가는 Mullis et al. (2020)의 Exhibit 3.1에 나타난 평균 수학 성취도가 가장 높은 4개 국가에 해당하며, 각 국가의 8학년 수학 평가에 참여한 학생 수는 각각 3,861명, 4,853명, 4,915명, 4,446명이다. TIMSS에서는 두 개의 문항군(block)으로 이루어진 여러 소책자(booklet)가 존재하고, 각 학생들은 소책자 중 하나를 배정받아 소책자에 해당하는 문항들을 해결하게 된다. 8학년 수학 평가의 경우 총 14개의 문항군이 있으며, 두 개의 문항군으로 구성된 14개의 소책자가 존재한다. 각 소책자를 구성하는 문항군 및 분석

문항 수는 <표 II-3>과 같다. 각 학생의 문항 응답은 학생이 배정받은 소책자에 포함된 문항에 대하여 맞으면 1, 틀렸으면 0으로 기록하였다. 서술형 문항의 경우 부분적으로 맞힌 응답(partially correct response)은 오답으로 취급하고 완전히 맞힌 응답만을 정답으로 취급하였으며, 해당하는 소책자에 포함된 문항이 결측치로 기록된 경우에는 오답으로 처리하였다. 각 학생이 해당하는 소책자에 포함된 문항이 아닌 경우에는 모두 결측치로 기록하였고, 결과적으로 블록희소행렬(block sparse matrix) 형태의 문항 응답 데이터를 분석하였다.

<표 II-3> 각 소책자를 구성하는 문항군 및 문항 수

소책자	소책자 1	소책자 2	소책자 3	소책자 4	소책자 5	소책자 6	소책자 7
문항군	ME01	ME02	ME03	ME04	ME05	ME06	ME07
	ME02	ME03	ME04	ME05	ME06	ME07	ME08
문항 수	31	30	30	30	28	28	31
소책자	소책자 8	소책자 9	소책자 10	소책자 11	소책자 12	소책자 13	소책자 14
문항군	ME08	ME09	ME10	ME11	ME12	ME13	ME14
	ME09	ME10	ME11	ME12	ME13	ME14	ME01
문항 수	29	27	28	29	28	30	33

#### 다. 분석 방법 및 절차

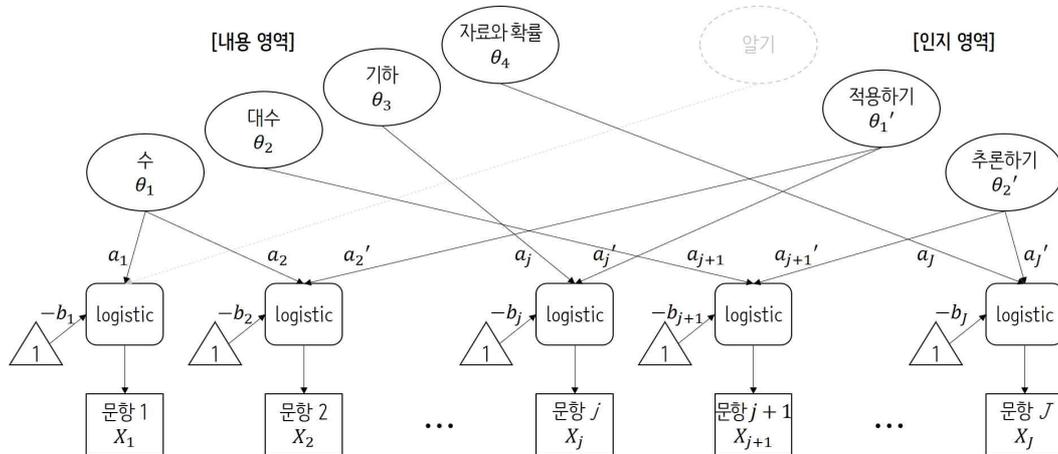
TIMSS 2019 8학년 수학 평가 데이터를 내용 영역과 인지 영역을 함께 고려하여 분석을 진행하기 위하여, Q행렬 및 학생의 능력 수준을 내용 영역과 인지 영역을 모두 포함하도록 구성하였다. 한편, 본 연구에서는 내용 영역에 대한 능력 수준을 각 영역의 개념적 지식을 얼마나 잘 알고 있는지를 의미하는 것으로 바라보고자 하였다. 이 경우 알기 영역은 학생들이 알아야 할 사실, 개념, 절차에 해당한다는 점에서 내용 영역과 관련되므로 (Novikasari, 2016), 알기 능력은 각 내용 영역의 능력에 포함된다고 바라보고 알기 영역에 해당하는 문항은 내용 영역에 대한 학생의 능력 수준과 문항의 변별도로 응답 확률이 결정된다고 가정하였다. 이에, Q행렬 및 학생의 능력 수준을 알기를 제외하고 수, 대수, 기하, 자료와 확률의 4개 내용 영역과 적용하기, 추론하기의 2개 인지 영역을 포함하여 총 6개 영역으로 구성하였다.

결과적으로 구성된 Q행렬은 <표 II-4>와 같으며, 연구에서 활용한 M2PL-Q 모형의 설계를 그림으로 나타내면 [그림 II-1]과 같다. Q행렬은 알기를 제외한 6개의 영역 중 문항이 해당하는 영역은 1로 표기하고 해당하지 않는 영역은 0으로 표기하여 구성하였다. 학생  $i$ 의 능력 수준은  $\theta_i = (\theta_{i1}, \theta_{i2}, \theta_{i3}, \theta_{i4}, \theta_{i1}', \theta_{i2}')$ 와 같이 4개의 내용 영역과 2개의 인지 영역에 대한 총 6개의 능력 수준을 포함하는 프로파일로 구성하였다. 이때,  $\theta_i$ 의 각 성분은 학생  $i$ 의 수, 대수, 기하, 자료와 확률, 적용하기, 추론하기에 해당하는 잠재 능력 수준을 의미하게 된다. M2PL-Q 모형에서는 문항의 특성이 난이도와 문항이 해당하는 영역에 대한 변별도로 구성된다. 인지 영역이 알기에 해당하는 문항의 경우 4가지 내용 영역 중 하나에만 해당하므로, 난이도와 해당하는 내용 영역 변별도의 두 가지 특성으로 구성된다. 또한, 인지 영역이 적용하기 또는 추론하기에 해당하는 문항의 경우 4가지 내용 영역 중 하나와 2가지 인지 영역 중 하나에 해당하므로, 난이도와 해당하는 내용 영역 변별도 및 인지 영역 변별도의 세 가지 특성으로 구성된다. [그림 II-1]은 각 문항에 대한 응답 확률이 문항의 난이도, 문항이 해당하는 내용 영역의 능력 수준과 변별도, 문항이 해당하는 인지 영역의 능력 수준과 변별도를 바탕으로 로지스틱 함수를 적용하여 결정되는 상황을 도식으로 나타낸 것이다. 알기 영역에 대해서는 능력 수준과 변별도가 모두 존재하지 않는다고 가정하였으므로 회색 점선으로 표시하였다. 예를 들어, 문항 1의 경우 내용 영역은 수에 해당하고

인지 영역은 알기에 해당하므로, 문항의 난이도  $b_1$ , 학생의 수 영역에 대한 능력 수준  $\theta_1$  및 문항의 내용 영역 변별도  $a_1$ 로 문항의 응답 확률이 결정된다. 또한, 문항 2의 경우 내용 영역은 수, 인지 영역은 적용하기에 해당하므로, 문항의 난이도  $b_2$ , 학생의 수 영역에 대한 능력 수준  $\theta_1$  및 문항의 내용 영역 변별도  $a_2$ , 학생의 적용하기 영역에 대한 능력 수준  $\theta_1'$  및 문항의 인지 영역 변별도  $a_2'$ 로 문항의 응답 확률이 결정된다.

<표 II-4> 분석에 사용된 Q행렬 중 일부

문항	내용 영역				인지 영역	
	수	대수	기하	자료와 확률	적용하기	추론하기
1	1	0	0	0	0	0
2	1	0	0	0	1	0
3	1	0	0	0	1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
205	0	0	0	1	1	0
206	0	0	0	1	0	1



[그림 II-1] 분석에 사용된 M2PL-Q 모형의 설계

다음으로, 모수를 추정하기 위해 우리나라를 포함한 4개 국가의 학생들이 각 문항에 보인 응답으로부터 문항 응답 데이터를 구성하였고, 문항 응답 행렬에서 결측치가 아닌 응답만을 이용하여 추정을 진행하였다. 학생의 각 영역에 대한 능력 수준의 사전분포는 표준정규분포(standard normal distribution)로 설정하였고, 문항의 난이도에 대한 사전분포는 평균이 0이고 표준편차가 1인 정규분포(normal distribution)로 설정하였으며, 문항의 변별도에 대한 사전분포는 표준정규분포에서 양수 값만 가지도록 제한한 절단정규분포(truncated normal distribution)로 설정하였다.

$$\theta_{ik}, \theta_{ik}' \sim normal(0, 1)$$

$$b_j \sim normal(0, 1)$$

$$a_{jk}, a_{jk}' \sim normal(0, 1) \quad T(0, \infty)$$

위의 식에서  $\theta_{ik}$ 와  $\theta_{ik}'$ 은 각각 학생  $i$ 의 내용 영역  $k$  및 인지 영역  $k'$ 에 대한 능력 수준을 의미한다.  $b_j$ 는 문항  $j$ 의 난이도를,  $a_{jk}$ 와  $a_{jk}'$ 는 문항  $j$ 의 내용 영역  $k$  및 인지 영역  $k'$ 에 대한 변별도를 의미한다.

모형의 추정을 위해 MCMC(Markov Chain Monte Carlo) 알고리즘을 총 20,000회의 샘플링을 통해 진행하였으며, 앞의 10,000회는 burn-in 구간으로 제외하고 뒤의 10,000회 샘플의 평균으로 모수를 추정하였다. 총 20,000회의 샘플링을 통한 수렴 여부는 Gelman-Rubin 수렴진단 통계량을 통해 판단하였다. 한편, 문항반응모형에서는 학생의 능력 수준의 척도에 따라 문항의 변별도 및 난이도의 척도가 달라지게 된다(Natesan et al., 2016). 따라서, 각 영역에 대한 학생의 능력 수준을 평균이 0이고 표준편차가 1이 되도록 표준화하고, 그에 맞게 문항의 변별도와 난이도의 척도를 문항 응답 확률이 유지되도록 조정하여 최종 문항 모수 추정값을 계산하였다.

M2PL-Q 모형의 설계 및 구현에는 R 프로그램(버전 4.3.1)을 사용하였으며, MCMC 알고리즘을 이용한 모수 추정 과정은 JAGS 프로그램(Plummer, 2017)과 R 프로그램의 rjags 패키지(Su & Yajima, 2020)를 이용하여 진행하였다. 분석에 사용한 M2PL-Q 모형의 코드는 <부록 1>에서 확인할 수 있다.

### III. 연구 결과 및 논의

#### 1. 내용 및 인지 영역에 대한 능력 수준 분석

##### 가. 4개 국가에 대한 평균 능력 수준

M2PL-Q 모형으로 추정한 학생들의 각 영역 능력 수준에 대하여 분석을 진행하였고, 분석 대상인 한국, 싱가포르, 대만, 일본 4개 국가 학생들의 각 영역에 대한 능력 수준의 추정 결과는 <표 III-1>과 같다. 우리나라 학생들의 능력 수준 평균은 내용 영역 중 수 영역(0.04)과 기하 영역(0.02)에서는 양수로 나와서 4개 국가 평균보다 높은 것을 확인할 수 있다. 반면, 내용 영역 중 대수 영역(-0.05)과 자료와 확률 영역(-0.04), 그리고 두 인지 영역인 적용하기 영역(-0.05)과 추론하기 영역(-0.08)은 음수로 나와서 4개 국가 평균보다 다소 낮은 것을 확인할 수 있다. 특히, 두 인지 영역의 경우 4개 국가 중 가장 낮은 평균 능력 수준을 보이는 것을 확인할 수 있다.

<표 III-1> 4개 국가 학생들의 각 영역에 대한 능력 수준 평균(표준편차)

국가	내용 영역				인지 영역	
	수	대수	기하	자료와 확률	적용하기	추론하기
한국	0.04 (0.98)	-0.05 (1.06)	0.02 (0.98)	-0.04 (0.99)	-0.05 (1.01)	-0.08 (0.99)
싱가포르	0.04 (0.97)	0.01 (0.98)	-0.13 (1.02)	0.09 (1.03)	0.02 (0.98)	0.05 (1.03)
대만	0.10 (1.04)	0.05 (1.00)	0.01 (1.04)	-0.13 (1.01)	-0.03 (1.03)	0.02 (0.98)
일본	-0.19 (0.98)	-0.02 (0.96)	0.10 (0.93)	0.09 (0.94)	0.05 (0.98)	-0.01 (1.00)

##### 나. 각 영역의 능력 수준에 대한 상관관계

다음으로, 내용 및 인지 영역들 사이의 관계를 분석하기 위하여, 학생들의 각 영역 능력 수준의 상관관계를 피어슨 상관계수(Pearson correlation coefficient)를 통해 확인해보았고 그 결과는 <표 III-2>와 같다. 우선 기하 및 자료와 확률 사이를 제외한 모든 경우에서 상관계수가 0.4 이상으로 나와 중간 이상(Dancey & Reidy, 2017, p. 182)의 상관관계가 존재하는 것을 확인할 수 있으며, 이는 내용 영역 사이와 인지 영역 사이뿐 아니라 내용

영역과 인지 영역 사이에도 상관이 있음을 의미한다. 구체적으로, 내용 영역들 사이의 상관관계를 확인해 보면, 수와 대수 사이의 상관계수가 0.628로 가장 높은 것을 확인할 수 있고, 이는 수 영역과 대수 영역이 가장 밀접하게 연관되어 있음을 의미한다. 반면, 자료와 확률의 경우 다른 내용 영역들과의 상관계수가 비교적 낮음을 확인할 수 있었고, 이는 내용 영역 중 자료와 확률이 상대적으로 독립적인 영역에 해당함을 의미한다. 인지 영역의 경우 적용하기와 추론하기 사이의 상관계수가 0.577로 나와서 내용 영역들 사이의 상관관계 중 수와 대수 사이의 상관계수를 제외한 모든 조합보다 높은 상관계수를 보임을 확인할 수 있었다. 이는 인지 영역 사이의 연관성이 내용 영역 사이의 연관성보다 높음을 의미한다. 마지막으로 내용 영역과 인지 영역 사이의 상관관계를 확인한 결과, 적용하기가 추론하기보다 내용 영역들과의 상관계수가 전체적으로 높게 나온 것을 확인할 수 있었다.

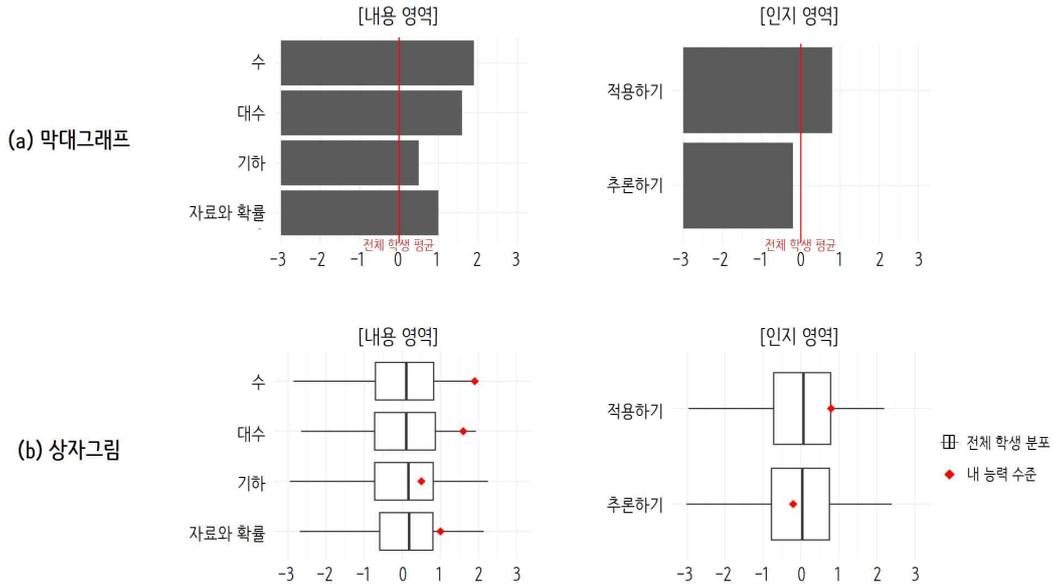
<표 III-2> 각 영역의 능력 수준에 대한 상관관계

		내용 영역				인지 영역	
		수	대수	기하	자료와 확률	적용하기	추론하기
내용 영역	수	-	0.628	0.503	0.463	0.615	0.548
	대수	0.628	-	0.511	0.453	0.629	0.596
	기하	0.503	0.511	-	0.369	0.549	0.503
	자료와 확률	0.463	0.453	0.369	-	0.486	0.436
인지 영역	적용하기	0.615	0.629	0.549	0.486	-	0.577
	추론하기	0.548	0.596	0.503	0.436	0.577	-

#### 다. 능력 수준 추정 결과의 개인별 진단 활용 방법

설계된 M2PL-Q 모형은 내용 및 인지 영역에 대한 능력 수준을 동시에 추정하므로, 학생들에게 내용 및 인지 영역에 대한 능력 수준을 함께 진단할 수 있다. 또한, 복수의 영역에 대한 학생의 능력 수준을 연속적인 값으로 추정해주기 때문에, 개별 학생에게 각 영역의 능력 수준에 대한 진단을 구체적인 수치와 함께 제공할 수 있다. [그림 III-1]은 개별 학생에 대한 능력 수준을 시각화하여 진단한 예시이다. 우선 (a)는 막대그래프를 이용하여 학생의 영역별 능력 수준을 나타내고 전체 학생의 평균 능력 수준을 함께 표시한 예시로, 학생은 각 영역에 대한 자신의 능력 수준을 전체 학생에 대한 평균과 비교하여 확인할 수 있다. 다음으로 (b)는 전체 학생의 능력 수준을 영역별로 상자 그림으로 나타내고 해당 학생의 능력 수준을 상자 그림 위에 표시한 예시로, 학생은 전체 학생들의 능력 분포를 바탕으로 자신의 위치를 확인할 수 있다.

본 연구에서는 학생들의 영역별 능력 수준을 평균이 0이고 표준편차가 1이도록 척도를 설정하였기에, [그림 III-1]에서도 설정된 척도에 맞게 수치가 표시되어 있다. 한편, 학생들에게 보다 직관적인 수치로 영역별 능력 수준에 대한 정보를 제공하기 위한 등의 목적으로, 능력 수준의 척도를 조정하여 진단 결과를 제공할 수도 있다. 예를 들어 TIMSS에서는 문항반응이론을 이용하여 추정한 학생들의 능력 수준을 평균이 500점이고 표준편차가 100점이도록 척도를 조정하여 제공한다(Foy et al., 2020). 또한, 모든 문항을 틀렸을 때의 영역별 능력 수준을 0점으로 설정하고, 모든 문항을 맞혔을 때의 영역별 능력 수준을 100점으로 설정하면, 학생들에게 보다 익숙할 수 있는 100점 만점의 점수로 능력 수준을 진단할 수도 있다. 이와 같이 진단이나 평가의 상황과 목적에 맞게 척도를 설정하면, 능력 수준에 대한 진단 결과를 학생들이 직관적으로 이해할 수 있도록 제공할 수도 있게 된다.



[그림 III-1] 개별 학생의 능력 수준에 대한 진단 예시

## 2. 문항 모수 추정 결과 분석

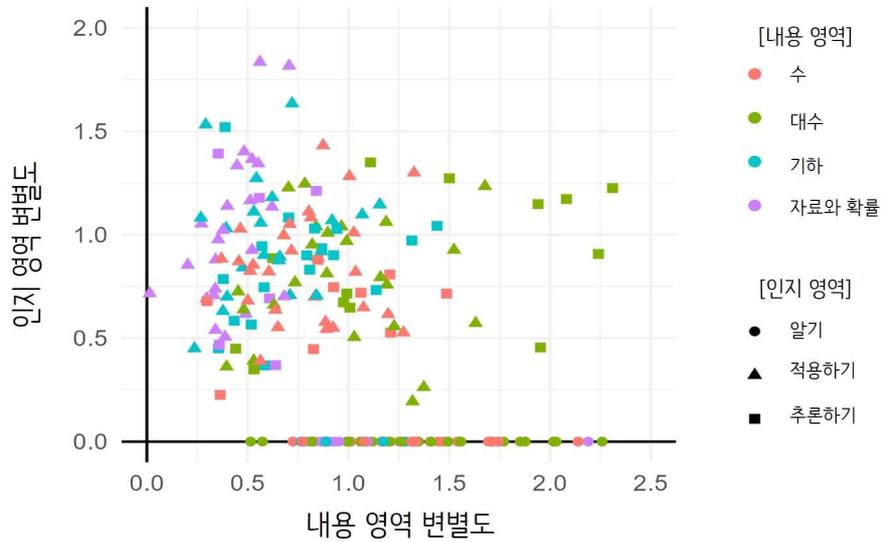
### 가. 난이도 추정 결과

모형 추정 결과로부터 얻은 전체 문항의 난이도 평균은  $-1.06$ 이고, 표준편차는  $1.02$ 로 나타났다. 난이도의 평균이 음의 값으로 나타난 것은 학생들이 평가 문항을 전반적으로 잘 맞혔기 때문이며, 이는 연구 대상으로 성취가 가장 높은 4개 국가를 선정했기 때문에 나타난 결과로 볼 수 있다. 4개의 내용 영역에 대하여 각 영역에 해당하는 문항의 난이도 평균을 비교한 결과, 수, 대수, 기하, 자료와 확률 영역에서 각각  $-1.18$ ,  $-1.04$ ,  $-0.87$ ,  $-1.12$ 로 나타난 것을 확인할 수 있었다. 이는 학생들이 기하, 대수, 자료와 확률, 수 영역의 순서로 각 영역에 해당하는 문항들을 어려워했음을 의미한다. 또한, 3개의 인지 영역에 대하여 각 영역에 해당하는 문항의 난이도 평균을 비교한 결과, 알기, 적용하기, 추론하기 영역에서 각각  $-1.42$ ,  $-1.08$ ,  $-0.50$ 로 나타난 것을 확인할 수 있었다. 이는 학생들이 추론하기, 적용하기, 알기 순서로 각 영역에 해당하는 문항들을 어려워했음을 의미한다.

### 나. 변별도 추정 결과

다음으로, 추정 결과로부터 얻은 각 문항의 내용 영역 변별도와 인지 영역 변별도를 산점도로 나타내면 [그림 III-2]와 같다. 그림에서 가로축이 내용 영역 변별도, 세로축이 인지 영역 변별도를 의미하며, 인지 영역이 알기에 해당하는 문항의 경우 내용 영역 변별도만이 존재하므로 가로축 위에 그 값을 표시하였다. 그림을 통해 내용 및 인지 영역의 변별도가 문항마다 다르게 나타나는 것을 확인할 수 있으며, 이는 문항마다 정답 확률에 각 영역의 능력 수준이 미치는 영향력이 서로 다를 수 있음을 의미한다. 특히, 내용 영역 변별도가 인지 영역 변별도보다 높은 경우도 존재하고, 반대로 인지 영역 변별도가 내용 영역 변별도보다 높은 경우도 존재하는 것을 통해, 문항의 정답 확률에 내용 영역과 인지 영역 중 특정 영역이 더 많은 영향을 미칠 수 있음을 확인할 수 있다.

4개의 내용 영역에 대하여 각 영역에 해당하는 문항의 내용 영역 변별도 평균을 확인한 결과, 수, 대수, 기하, 자료와 확률 영역에서 각각 1.03, 1.20, 0.71, 0.65로 나타난 것을 확인할 수 있었다. 이는 분석에 사용된 문항 중 수 및 대수 영역에 해당하는 문항들이 기하 및 자료와 확률 영역에 해당하는 문항들보다 각 내용 영역의 능력 수준을 변별하기 용이한 문항들이었음을 의미한다. 또한, 알기를 제외한 2개의 인지 영역에 대하여 각 영역에 해당하는 문항의 인지 영역 변별도 평균을 확인한 결과, 적용하기 영역이 0.90으로 추론하기 영역의 0.82보다 다소 높았으나 그 차이가 내용 영역 변별도에서만큼 크지는 않은 것으로 나타났다.



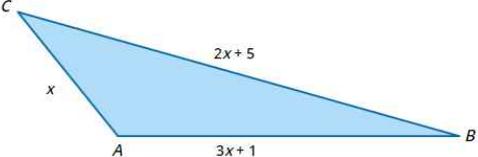
[그림 III-2] 문항의 내용 영역 변별도 및 인지 영역 변별도에 대한 산점도

**다. 세부 문항에 대한 특성 분석**

마지막으로, 각 문항에 대한 난이도 및 변별도를 추정 결과를 구체적인 문항 내용과 함께 분석해보았다. TIMSS 2019에서는 전체 문항 중 ME02 및 ME06 문항군에 해당하는 문항만을 공개하고 있으므로, 두 문항군에 해당하는 총 29개 문항에 대하여 분석을 진행하였다. 본 절에서는 내용 및 인지 영역이 같으나 문항 모수가 다르게 추정된 두 쌍의 문항에 대한 분석 결과를 제시하며, 전체 공개 문항에 대한 난이도 및 변별도 추정 결과는 <부록 2>에서 확인할 수 있다.

<표 III-3>은 내용 영역이 대수, 인지 영역이 적용하기로 같은 ME72098 문항과 ME62241 문항의 세부 내용과 추정된 문항 모수를 나타낸 것이다. ME72098 문항은 주어진 삼각형의 둘레를 통해 일차방정식을 세우고 이를 해결하여 한 변의 길이를 구하는 문제이며, ME62241 문항은 주어진 설명을 따라 핸드폰의 가격에 대한 등식을 세우는 문제이다. 두 문항의 난이도 추정값은 각각 -0.760와 -1.264로 ME72098 문항이 ME62241 문항보다 어려운 것을 알 수 있다. 두 문항의 인지 영역 변별도의 경우 각각 1.234와 1.246으로 크게 다르지 않으나, 내용 영역 변별도의 경우 각각 1.678과 0.784로 ME72098 문항이 더 높은 것으로 나타났다. 이와 같은 결과는 ME72098 문항이 ME62241 문항보다 어려운 문항이고 대수 영역에 대한 내용 지식을 더 많이 요구하는 문항임을 의미하며, ME62241 문항은 등식을 세우기만 하면 되지만 ME72098 문항은 방정식을 세우고 연산을 거쳐 푸는 과정이 모두 포함되어 있기 때문에 해석될 수 있다.

<표 III-3> ME72098 문항과 ME62241 문항의 세부 내용 및 추정된 문항 모수

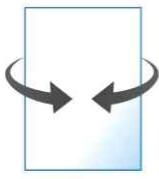
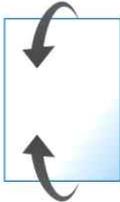
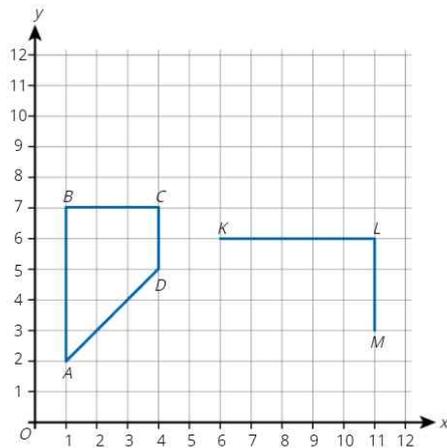
ME72098		ME62241	
<p>The perimeter of triangle ABC is 21 cm.</p>  <p>What is the value of <math>x</math>?</p> <p><math>x =</math> <input type="text"/> cm</p>		<p>Roy buys cell phones for <math>x</math> zeds each and sells them to make a profit. He determines his selling price for each phone, <math>y</math> zeds, by doubling the price he paid and subtracting 3 zeds.</p> <p>Write an equation that shows <math>y</math> in terms of <math>x</math>.</p> <p>Equation: <input type="text"/></p>	
난이도	-0.760	난이도	-1.264
내용 영역 및 변별도	Algebra 1.678	내용 영역 및 변별도	Algebra 0.784
인지 영역 및 변별도	Applying 1.234	인지 영역 및 변별도	Applying 1.246

SOURCE: TIMSS 2019 Assessment. Copyright © 2021 International Association for the Evaluation of Educational Achievement (IEA).

Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

다음으로 <표 III-4>는 동일한 직사각형을 다른 방향으로 회전하여 만들어지는 열린 원기둥을 비교하는 ME72180 문항과 두 다각형이 합동이 되도록 하는 점의 위치를 찾는 ME72198 문항에 대하여 세부 내용 및 문항 모수 추정 결과를 나타낸 것으로, 두 문항은 내용 영역이 기하이고 인지 영역이 추론하기로 같은 문항에 해당한다. 두 문항의 난이도는 각각 -0.823와 -1.794로 ME72180 문항이 ME72198 문항보다 어려운 것을 알 수 있으며, ME72180 문항은 학생들이 까다로워하는 입체도형을 다루고 있고 ME72198 문항은 평면도형을 다루고 있다는 점이 난이도 차이의 원인 중 하나일 것으로 추측된다. 두 문항의 변별도를 비교해보면, 내용 영역 변별도는 각각 0.586과 1.440이고 인지 영역 변별도는 각각 0.369와 1.043으로, 두 영역 모두에 대하여 ME72198 문항이 ME72180 문항보다 변별도가 높은 것을 확인할 수 있다. 이는 ME72180 문항의 경우 회전체를 상상하는 직관적인 풀이 방법이 가능한 반면, ME72198 문항의 경우 다각형의 합동에 대한 내용 지식을 바탕으로 나머지 한 점의 위치를 추론하는 과정에서 보다 체계적인 풀이 방법을 요구하기 때문일 수 있다. 또한, ME72180 문항은 “크다”, “같다”, “작다” 중 하나를 선택하는 형태의 문항인 반면, ME72198 문항의 경우 점의 좌표를 입력하는 구성형 문항이어서 학생들의 능력을 변별하기 보다 용이했다는 점도 영향을 미쳤을 것으로 예상된다.

<표 III-4> ME72180 문항과 ME72198 문항의 세부 내용 및 추정된 문항 모수

ME72180		ME72198	
<p>Soh and Ben have identical rectangular pieces of paper. They use different ways to roll their papers into cylinders so that the opposite sides of the paper touch as shown below.</p> <div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>Soh's Method</p>  </div> <div style="text-align: center;"> <p>Ben's Method</p>  </div> </div> <p>Compare the properties of the two cylinders. Use the drop-down menus.</p> <p><b>Height</b> Soh's cylinder <input type="text" value="Choose one"/> Ben's cylinder</p> <p><b>Diameter</b> Soh's cylinder <input type="text" value="Choose one"/> Ben's cylinder</p> <p><b>Surface Area (with open ends)</b> Soh's cylinder <input type="text" value="Choose one"/> Ben's cylinder</p>		<p>Woo drew trapezoid <math>ABCD</math>. He then started drawing a <b>congruent</b> trapezoid <math>KLMN</math>.</p> <div style="display: flex; align-items: center;">    </div>  <p>What will be the coordinates of point <math>N</math> when Woo completes the figure?</p> <p>Answer: ( <input type="text" value=""/> , <input type="text" value=""/> )</p>	
난이도	-0.823	난이도	-1.794
내용 영역 및 변별도	Geometry 0.586	내용 영역 및 변별도	Geometry 1.440
인지 영역 및 변별도	Reasoning 0.369	인지 영역 및 변별도	Reasoning 1.043

SOURCE: TIMSS 2019 Assessment. Copyright © 2021 International Association for the Evaluation of Educational Achievement (IEA).

Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

#### IV. 결론 및 제언

본 연구에서는 TIMSS 2019 8학년 수학 데이터를 M2PL-Q 모형을 통해 내용 영역과 인지 영역을 함께 고려하여 분석을 진행하였다. 한국, 싱가포르, 대만, 일본 4개 국가의 응답 데이터를 M2PL-Q 모형으로 분석하여, 학생들의 능력 수준과 문항의 특성을 각 내용 및 인지 영역에 대하여 확인하였다. 추정된 학생들의 능력 수준을 바탕으로 우리나라 학생들의 각 영역에 대한 수준을 다른 국가들과 비교하여 분석하고 영역들 사이의 관계를 파악하였으며, 문항의 난이도, 내용 영역 변별도, 인지 영역 변별도를 추정하여 그 결과를 분석하였다. 본 연구에

서 도출된 결론은 다음과 같다.

첫째, da Silva 외(2019)가 제시한 M2PL-Q 모형을 활용하여 평가 데이터를 내용 영역과 인지 영역을 구분하여 함께 분석할 수 있음을 확인하였다. 내용 영역과 인지 영역을 함께 고려하여 평가 데이터를 분석할 필요성이 제시되어 왔으나, 기존에 평가 데이터의 분석에 주로 활용되던 다차원 문항반응이론이나 인지진단모형이 지닌 한계로 인하여 제한된 분석만이 이루어져 왔다. 본 연구는 내용 및 인지 영역과 같이 평가 영역이 복수의 영역에 대하여 구성된 상황에서 응답 데이터를 분석할 때, M2PL-Q 모형을 활용해서 복수의 영역에 걸친 문항의 특성을 확인하고 학생의 능력을 추정할 수 있음을 확인했다는 점에서 중요한 의의가 있다. 또한, M2PL-Q 모형은 문항과 영역 사이의 관계에 대한 Q행렬이 규명되어 있으면 학생들의 응답 데이터로부터 각 영역에 대한 학생의 능력 수준과 문항의 특성을 추정할 수 있으므로, 본 연구에서 분석한 TIMSS 2019 8학년 수학 평가 외의 다른 평가에 대해서도 평가틀을 바탕으로 내용 및 인지 영역에 대한 Q행렬을 구성하고 결과를 분석할 수 있다. 뿐만 아니라, 본 연구에서 분석한 내용 및 인지 영역이 아니라 다른 영역의 조합으로 문항이 구분된 상황에서도 영역들을 함께 고려하여 분석이 가능할 것으로 기대한다.

둘째, M2PL-Q 모형으로 추정된 결과를 바탕으로 학생들의 내용 및 인지 영역에 대한 능력 수준을 분석한 결과, 우리나라 학생들의 평균적인 능력 수준을 다른 국가들과 비교하여 확인할 수 있었으며, 내용 및 인지 영역의 능력 사이의 관계를 확인할 수 있었다. 우선, 4개 국가의 평균 능력 수준과 비교하였을 때, 우리나라 학생들은 내용 영역 중 수 영역 및 기하 영역에 대하여 상대적으로 높은 능력을 보였으며, 내용 영역 중 대수 영역 및 자료와 확률 영역에 대해서는 상대적으로 낮은 능력을 보였다. 한편, 본 연구에서 분석한 TIMSS 2019 8학년 검사에 참여한 학생들은 본검사가 시행된 2018학년도 기준 중학교 2학년에 해당하며, 검사 시기까지 2009 개정 교육과정에 따라 학습한 학생들이다(상경아 외, 2020). 자료와 확률 영역과 관련하여 2015 개정 수학과 교육과정과 2022 개정 수학과 교육과정에서 통계적 소양이 강조되고 있으므로(권점례, 2024; 이경화 외, 2021; 탁병주, 2018), 해당 교육과정으로 학습한 학생들을 대상으로 한 추후 TIMSS 데이터를 분석하는 후속 연구를 통해, 자료와 확률 영역에 대한 우리나라 학생들의 수준 변화를 지속적으로 관찰할 필요가 있다. 또한, 인지 영역인 적용하기 및 추론하기 영역에 대해서는 우리나라 학생들이 4개 국가 중 가장 낮은 평균을 보이는 것으로 나타났다. 우리나라 학생들이 수학 교과에서 알기 영역에 비하여 적용하기와 추론하기 영역에 상대적으로 낮은 성취를 보인다는 결과는 기존의 TIMSS 2019 결과 분석에서도 4학년과 8학년에 걸쳐 동일하게 나타난 것으로(상경아 외, 2020), 이에 대한 원인과 해결 방안을 마련하기 위한 후속 연구가 필요할 것으로 보인다.

다음으로, 추정된 학생들의 능력 수준에 대한 상관계수를 계산하여 각 영역 사이의 관계를 분석한 결과, 거의 모든 영역들 사이에 중간 이상의 상관관계가 존재함을 확인할 수 있었다. 이와 같은 결과는 내용 영역과 인지 영역이 각각 내부적으로 연관되어 있을 뿐 아니라 내용 영역과 인지 영역 사이에도 연관성이 존재한다는 선행 연구의 주장과도 일치한다(George & Robitzsch, 2018; Harks et al., 2014; Niss, 2003). 또한, 내용 영역 중 수와 대수 영역에 대한 상관관계가 가장 높게 나타나 두 영역이 밀접한 관계에 있음을 확인할 수 있었고, 이와 같은 결과는 수 영역이 대수적 추론과 문제해결의 기반이 된다는 점에서 유의미한 연관성이 존재하기 때문일 수 있을 것으로 생각된다(Ameron, 2002). 자료와 확률의 경우 다른 내용 영역과의 관계가 상대적으로 약함을 확인할 수 있었고, 이는 통계의 아이디어를 수학에서는 사용하지 않는 등 통계가 수학의 다른 영역과는 다른 특성을 지닌다는 점이 그 원인일 수 있을 것으로 예상된다(Moore, 1992). 인지 영역인 적용하기와 추론하기 사이의 상관관계는 내용 영역들 사이의 상관관계와 비교하여 높게 나온 것을 통해, 내용 영역이 인지 영역에 비해 세부 영역 사이의 관계가 상대적으로 독립적인 것을 확인할 수 있었다. 마지막으로 내용 영역과 인지 영역 사이의 상관관계를 확인한 결과, 적용하기가 추론하기보다 내용 영역과의 상관관계가 전체적으로 높게 나타났으며, 이는 추론하기가 적용하기보다 내용 지식을 고차원적인 수준에서 다루고 있기 때문인 결과로 예상된다(Delil, 2019).

또한, 영역별로 추정된 학생의 능력 수준을 개별 학생에 대한 진단 결과로 제공할 수 있음을 확인하였다. 기

존의 문항반응이론의 경우 능력 수준을 전체 검사 결과에 대해 하나의 값으로 제시하므로, 세부 영역에 대한 학생의 수준을 확인하기 어렵다는 한계가 존재한다. 다차원 문항반응이론을 활용하더라도 각 문항이 하나의 영역에만 해당한다는 가정으로 인해, 내용 및 인지 영역과 같이 조합된 영역에 대해 능력 수준을 추정하기에는 어려움이 존재한다. 본 연구에서 활용한 M2PL-Q 모형을 이용하면 내용 영역과 인지 영역을 함께 고려하여 능력 수준을 추정할 수 있으므로, 학생의 능력에 대해 영역별로 보다 구체적이고 정확한 정보를 제공할 수 있다. 개별 학생에 대한 진단 결과를 통해 학생이 어느 내용 및 인지 영역에서 강점이나 약점을 보이는지 알 수 있고, 이와 같은 정보를 온라인 학습 시스템 등에서 활용하면 보완이 필요한 영역에 대해서 학습 콘텐츠를 맞춤형으로 제공할 수 있다. 학생의 능력 수준은 IV장에서 제시한 것과 같이 막대그래프나 상자그림 등을 활용하여 시각화할 수 있고, 이를 대시보드 등을 통해 학생에게 제공할 수도 있다. 또한, 교사는 전체 학생에 대한 진단 결과를 바탕으로 학생들의 영역별 능력 수준을 확인하고 교수학습 계획을 세우는 데 활용할 수도 있을 것으로 기대한다.

셋째, 추정된 문항의 난이도 및 변별도를 바탕으로 각 문항의 특성을 분석한 결과, 문항에 따라 난이도, 내용 영역 변별도, 인지 영역 변별도가 다르게 나타나는 양상을 확인할 수 있었다. 각 내용 및 인지 영역별로 추정된 난이도의 평균을 비교한 결과, 내용 영역의 경우 기하, 대수, 자료와 확률, 수 영역의 순서로 난이도가 높은 것을 확인할 수 있었다. 인지 영역의 경우 추론하기, 적용하기, 알기 영역의 순서로 난이도가 높은 것을 확인할 수 있었으며, 이와 같은 결과의 원인으로서는 일반적으로 인지 영역 중 알기가 가장 낮은 수준의 능력에 해당하며 추론하기가 가장 높은 수준의 능력에 해당하기 때문임을 예상해볼 수 있다(Delil, 2019). 또한, 추정된 내용 영역 변별도 및 인지 영역 변별도의 비교를 통해서, 문항에 따라 내용 영역에 대한 능력과 인지 영역에 대한 능력이 문제 해결에 서로 다른 영향을 미칠 수 있음을 확인할 수 있었다. 내용 영역 변별도가 더 높아서 내용 영역에 대한 능력이 해결에 많은 영향을 미치는 문항도 존재하고, 인지 영역 변별도가 더 높아서 인지 영역에 대한 능력이 해결에 많은 영향을 미치는 문항도 존재하는 것으로 나타났으며, 공개문항에 대해서는 내용 및 인지 영역 변별도 추정 결과를 문항의 세부 내용과 함께 확인할 수도 있었다. 이와 같은 결과는 수학 문항이 해당하는 내용 및 인지 영역이나 세부 내용에 따라 난이도 및 변별도와 같은 문항의 특성이 달라질 수 있으며, 본 논문에서 사용한 M2PL-Q 모형을 통해 이와 같은 문항의 특성을 추정하고 분석할 수 있음을 의미한다. 수학 평가를 개발하고 분석하는 개발자 및 연구자들은 이러한 분석 방법을 통해 문항의 특성을 보다 세부적으로 파악하여 문항에 대한 다양한 정보를 얻을 수 있게 된다. 또한, 각 문항의 특성을 고려하여 평가의 상황 및 목적에 따라 적절하게 문항을 활용하면 평가의 타당성과 효율성을 높일 수 있으며, 평가 결과를 바탕으로 학생의 내용 및 인지 영역에 대한 능력 수준을 보다 효과적으로 추정하여 정확한 진단 결과를 제공할 수 있을 것이다.

본 연구는 내용 영역과 인지 영역을 함께 고려하여 학생의 능력과 문항의 특성을 추정하고, 이와 같은 분석을 위한 M2PL-Q 모형의 활용 가능성을 확인하였다는 점에서 중요한 의의가 있다. 한편, M2PL-Q 모형의 추정을 위하여 MCMC 알고리즘을 이용하였고, MCMC 알고리즘은 국제 수준의 학업성취도와 같은 큰 데이터를 분석하기 위하여 많은 시간이 소요된다는 점에서 한계가 존재한다. 보다 짧은 시간을 통해 모형의 모수를 추정할 수 있는 방법에 대한 연구가 이루어진다면, M2PL-Q 모형의 활용 가능성을 더욱 높일 수 있을 것으로 기대한다.

## 참 고 문 헌

- 교육부. (2022). **수학과 교육과정**. 교육부 고시 제 2022-33호 [별책 8].
- Ministry of Education (2022). *Mathematics curriculum*. Notification of Ministry of Education No. 2022-33 [Vol 8].
- 권점례(2024). TIMSS 수학 평가에서 학교급 전환에 따른 학생들의 학업성취 변화 추이 분석. **수학교육논문집**, **38**(2), 121-144. <https://doi.org/10.7468/jksmee.2024.38.2.121>
- Kwon, J. R. (2024). Analysis of the trend of mathematical achievement of students according to school grade change in TIMSS. *Communications of Mathematical Education*, *38*(2), 121-144. <https://doi.org/10.7468/jksmee.2024.38.2.121>
- 박지현, 김수진. (2015). 수학 인지적 속성에 따른 TIMSS 2011 8학년 성취도 상위국 특성 비교. **수학교육학연구**, **25**(3), 303-321.
- Park, J. H., & Kim, S. (2015). The analysis of characteristic achievement of TIMSS 2011 G8 high-performing countries according to the mathematics cognitive attributes. *Journal of Educational Research in Mathematics*, *25*(3), 303-321.
- 상경아, 김경희, 박상욱, 전성균, 박미미, 이재원, 민여준. (2020). **수학·과학 성취도 추이변화 국제 비교 연구: TIMSS 2019 결과 분석** (RRE 2020-10). 한국교육과정평가원.
- Sang, K. A., Kim K. H., Park S. W., Jeon S. K., Park, M. M., & Lee, J. W. (2020). *An international comparative study on the trend of mathematical and scientific achievement: TIMSS 2019* (RRE 2020-10). Korea Institute of Curriculum and Evaluation.
- 송미영, 김선희. (2007). 중학교 수학 교육과정의 내용 및 인지행동의 위계성 조사. **학교수학**, **9**(2), 223-240.
- Song, M. Y., & Kim, S. H. (2007). Investigating the hierarchical nature of content and cognitive domains in the mathematics curriculum for Korean middle school students via assessment items. *School Mathematics*, *9*(2), 223-240.
- 이경화, 유연주, 탁병주. (2021). 데이터 기반 통계교육을 위한 수학과 교육과정 재구조화 방향 탐색. **학교수학**, **23**(3), 361-386. <http://doi.org/10.29275/sm.2021.09.23.3.361>
- Lee, K. H., Yoo, Y. J., & Tak, B. (2021). Towards data-driven statistics education: An exploration of restructuring the mathematics curriculum. *School Mathematics*, *23*(3), 361-386. <http://doi.org/10.29275/sm.2021.09.23.3.361>
- 임혜미, 김성경, 박지현. (2018). 2015 개정 수학과 교육과정의 수학 교과 역량을 반영한 국가수준 학업성취도 평가의 평가틀 및 예시 문항 개발. **학교수학**, **20**(1), 65-82. <http://doi.org/10.29275/sm.2018.03.20.1.65>
- Rim, H., Kim, S. K., & Park, J. H. (2018). Development of assessment framework and items of NAEA considering the math competencies of the 2015 revised mathematics curriculum. *School Mathematics*, *20*(1), 65-82. <http://doi.org/10.29275/sm.2018.03.20.1.65>
- 탁병주. (2018). 통계적 소양으로서 자료의 분류 및 표현 활동의 의의 분석: 초등학교 1~2학년군 수학과 교육과정을 중심으로. **한국초등수학교육학회지**, **22**(3), 221-240.
- Tak, B. (2018). An analysis on classifying and representing data as statistical literacy: Focusing on elementary mathematics curriculum for 1st and 2nd grades. *Journal of Elementary Mathematics Education in Korea*, *22*(3), 221-240.
- 한채린, 박만구. (2015). 수학 학습 평가틀 비교 분석 : NAEP 2015, TIMSS 2015, PISA 2015를 중심으로. **수학교육**, **54**(3), 261-282. <https://doi.org/10.7468/mathedu.2015.54.3.261>
- Han, C., & Park, M. (2015). A comparison study on mathematics assessment frameworks: Focusing on NAEP 2015, TIMSS 2015 and PISA 2015. *The Mathematics Education*, *54*(3), 261-282. <https://doi.org/10.7468/mathedu.2015.54.3.261>
- Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, *7*(4), 255-278. [https://doi.org/10.1207/s1532690xam0704\\_03](https://doi.org/10.1207/s1532690xam0704_03)

- i.org/10.1207/s15324818ame0704\_1
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., and Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Ankan, S. (2015). Construct validity of TIMSS 2011 mathematics cognitive domains for Turkish students. *International Online Journal of Educational Sciences*, 7(1), 29-44.
- Balfaqeeh, A., Mansour, N., & Forawi, S. (2022). Factors influencing students' achievements in the content and cognitive domains in TIMSS 4th grade science and mathematics in the United Arab Emirates. *Education Sciences*, 12(9), 618. <http://dx.doi.org/10.3390/educsci12090618>
- Buck, G. (1994). The appropriacy of psychometric measurement models for testing second language listening comprehension. *Language Testing*, 11(2), 145-170. <https://doi.org/10.1177/026553229401100204>
- da Silva, M. A., Liu, R., Huggins-Manley, A. C., & Bazan, J. L. (2019). Incorporating the q-matrix into multidimensional item response theory models. *Educational and Psychological Measurement*, 79(4), 665-687. <https://doi.org/10.1177%2F0013164418814898>
- Dancey, C. P., & Reidy, J. (2017). *Statistics without maths for psychology*. Pearson.
- Delil, A. (2019). How fifth graders are assessed through central exams in Turkey: A comparison with TIMSS 2019 Assessment Framework. *International Online Journal of Educational Sciences*, 3(11), 222-234.
- Embretson, S. E., & Reise, S. (2000). Item response theory as model-based measurement. In Embretson, S. E., & Reise, S. (Eds.), *Item response theory for psychologists* (pp. 158-186). Lawrence Erlbaum Associates.
- Fishbein, B., Foy, P., & Tyack, L. (2020). Reviewing the TIMSS 2019 achievement item statistics. In Martin, M. O., von Davier, M., & Mullis, I. V. S. (Eds.), *Methods and procedures: TIMSS 2019 Technical report* (pp. 10.1-10.70). TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/timss2019/methods/chapter-10.html>
- Foy, P., Fishbein, B., von Davier, M., & Yin, L. (2020). Implementing the TIMSS 2019 scaling methodology. In Martin, M. O., von Davier, M., & Mullis, I. V. S. (Eds.), *Methods and procedures: TIMSS 2019 technical report* (pp. 12.1-12.146). TIMSS & PIRLS International Study Center, Boston College. <https://timssandpirls.bc.edu/timss2019/methods/chapter-12.html>
- George, A. C., & Robitzsch, A. (2018). Focusing on interactions between content and cognition: a new perspective on gender differences in mathematical sub-competencies. *Applied Measurement in Education*, 31(1), 79-97. <https://doi.org/10.1080/08957347.2017.1391260>
- Gierl, M. J., Bisanz, J., Bisanz, G. L., & Boughton, K. A. (2003). Identifying content and cognitive skills that produce gender differences in mathematics: A demonstration of the multidimensionality-based DIF analysis paradigm. *Journal of Educational Measurement*, 40(4), 281-306.
- Harks, B., Klieme, E., Hartig, J., & Leiss, D. (2014). Separating cognitive and content domains in mathematical competence. *Educational Assessment*, 19(4), 243-266. <https://doi.org/10.1080/10627197.2014.964114>
- Jang, Y. J. (2022). *Reliability and validity evidence of diagnostic methods: Comparison of diagnostic classification models and item response theory-based methods* [Unpublished doctoral dissertation, University of Minnesota].

- Moore, D. (1992). Teaching statistics as a respectable subject. In F. Gordon & S. Gordon (Eds.), *Statistics for the twenty-first century* (p. 14-25). The Mathematical Association of America.
- Mullis, I. V. S., & Martin, M. O. (2017). *TIMSS 2019 assessment frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study.
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Retrieved from Boston College, TIMSS & PIRLS International Study.
- Natesan, P., Nandakumar, R., Minka, T., & Rubright, J. D. (2016). Bayesian prior choice in IRT estimation using MCMC and variational Bayes. *Frontiers in psychology, 7*, 1422. <https://doi.org/10.3389/fpsyg.2016.01422>
- Niss, M. (2003). Mathematical competencies and the learning of mathematics: The Danish KOM project. In A. Gagatsis & S. Papastavridis (Eds.), *Mediterranean Conference on Mathematical Education* (pp. 115-124). Athens, Greece: Hellenic Mathematical Society and Cyprus Mathematical Society.
- Novikasari, I. (2016). The improvement of mathematics content knowledge on elementary school teacher candidates in problem based learning-models. *International Journal of Education and Research, 4*(17), 153-162.
- OECD. (2023). *PISA 2022 assessment and analytical framework*. Retrieved from <https://doi.org/10.1787/19963777>
- Plummer, M. (2017). *JAGS Version 4.3.0 user manual*. Retrieved from <https://sourceforge.net/projects/mcmc-jags/files/Manuals/4.x/>
- Shu, T., Luo, G., Luo, Z., Yu, X., Guo, X., & Li, Y. (2023). An explicit form with continuous attribute profile of the partial mastery DINA model. *Journal of Educational and Behavioral Statistics, 48*(5), 573-602.
- Su, Y. S., & Yajima, M. (2020). *R2jags: Using R to run 'JAGS'*. R package version 0.6-1. Retrieved from <https://CRAN.R-project.org/package=R2jags>
- Tatsuoka, K. K. (1983). Rule-space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement, 20*, 345-354.
- von Davier, M. (2020). TIMSS 2019 scaling methodology: Item response theory, population models, and linking across modes. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report* (pp. 11.1-11.25). TIMSS & PIRLS International Study Center, Boston College. <https://timssandpirls.bc.edu/timss2019/methods/chapter-11.html>
- Wu, M., & Adams, R. (2006). Modelling mathematics problem solving item responses using a multidimensional IRT model. *Mathematics Education Research Journal, 18*(2), 93-113.
- Young, J. W., Cho, Y., Ling, G., Cline, F., Steinberg, J., & Stone, E. (2008). Validity and fairness of state standards-based assessments for English language learners. *Educational Assessment, 13*, 170-192.
- Zhang, J. (2004). *Comparison of unidimensional and multidimensional approaches to IRT parameter estimation* (ETS Research Report 04-44). Educational Testing Service.

## A study on the application of M2PL-Q model for analyzing assessment data considering both content and cognitive domains: An analysis of TIMSS 2019 mathematics data

**Kim, Rae Yeong**

Seoul National University  
E-mail : pianist9430@snu.ac.kr

**Hwang, Su Bhin**

Graduate School of Seoul National University  
E-mail : sbhwangt@snu.ac.kr

**Lee, Seul Gi**

Graduate School of Seoul National University  
E-mail : wisdom1129@snu.ac.kr

**Yoo, Yun Joo<sup>†</sup>**

Seoul National University  
E-mail : yyoo@snu.ac.kr

This study aims to propose a method for analyzing mathematics assessment data that integrates both content and cognitive domains, utilizing the multidimensional two-parameter logistic model with a Q-matrix (M2PL-Q; da Silva, 2019). The method was applied to the TIMSS 2019 8th-grade mathematics assessment data. The results demonstrate that the M2PL-Q model effectively estimates students' ability levels across both domains, highlighting the interrelationships between abilities in each domain. Additionally, the M2PL-Q model was found to be effective in estimating item characteristics by differentiating between content and cognitive domain, revealing that their influence on problem-solving can vary across items. This study is significant in that it offers a comprehensive analytical approach that incorporates both content and cognitive domains, which were traditionally analyzed separately. By using the estimated ability levels for individual student diagnostics, students' strengths and weaknesses in specific content and cognitive areas can be identified, supporting more targeted learning interventions. Furthermore, by considering the detailed characteristics of each assessment item and applying them appropriately based on the context and purpose of the assessment, the validity and efficiency of assessments can be enhanced, leading to more accurate diagnoses of students' ability levels.

---

\* 2020 Mathematics Subject Classification : 97D60

\* Key words : content domain, cognitive domain, item response theory, Q-matrix, TIMSS 2019

<sup>†</sup> corresponding author

<부록 1> 분석에 활용한 M2PL-Q 모형의 rjags 코드

---

```

# N : 전체 문항 응답의 수
# I : 전체 학생의 수
# K : 영역의 개수
# J : 전체 문항의 수

# Y : 문항 응답 관측값(정답 유무에 따라 0 또는 1의 값을 갖는 N 크기의 벡터)
# Y_prob : 문항 응답 확률(0에서 1 사이의 값을 갖는 N 크기의 벡터)
# disc : 문항 변별도(J×K 크기의 행렬)
# diff : 문항 난이도(J 크기의 벡터)
# theta : 능력 수준(I×K 크기의 행렬)
# ind_j : 각 관측값이 해당하는 문항의 번호(N 크기의 벡터)
# ind_i : 각 관측값이 해당하는 학생(N 크기의 벡터)
# Q : Q행렬(J×K 크기의 행렬)

for (n in 1:N) {
  logit(Y_prob[n]) <- inprod(disc[ind_j[n],1:K], theta[ind_i[n],1:K]) - diff[ind_j[n]]
  Y[n] ~ dbern(Y_prob[n])
}
for (i in 1:I) {
  for (k in 1:K) {
    theta[i,k] ~ dnorm(0, 1)
  }
}
for (j in 1:J) {
  for (k in 1:K) {
    disc[j,k] <- xdisc[j,k] * Q[j,k]
    xdisc[j,k] ~ dlnorm(0, 4)
  }
  diff[j] ~ dnorm(0, 1)
}

```

---

## &lt;부록 2&gt; TIMSS 2019 8학년 수학 공개문항에 대한 모수 추정 결과

문항 ID	난이도	내용 영역 및 변별도		인지 영역 및 변별도	
ME72007	0.323	Number	1.405	Knowing	-
ME72025	-1.304	Number	0.833	Applying	0.699
ME72017	0.061	Number	1.210	Reasoning	0.527
ME72190	-2.387	Number	1.446	Knowing	-
ME72068	-2.828	Algebra	1.542	Knowing	-
ME72076	-1.304	Algebra	1.471	Knowing	-
ME72056	-0.637	Algebra	1.187	Applying	1.061
ME72098	-0.760	Algebra	1.678	Applying	1.234
ME72103	-0.917	Algebra	0.735	Applying	0.769
ME72121	-2.638	Geometry	0.656	Applying	0.875
ME72180	-0.823	Geometry	0.586	Reasoning	0.369
ME72198	-1.794	Geometry	1.440	Reasoning	1.043
ME72227	-1.347	Data and Probability	1.197	Knowing	-
ME72170	-1.314	Data and Probability	0.490	Applying	0.616
ME72209	0.724	Data and Probability	0.609	Reasoning	0.692
ME62150	-2.409	Number	1.744	Knowing	-
ME62335	-3.308	Number	1.346	Knowing	-
ME62219	-1.165	Number	0.804	Applying	1.112
ME62002	-0.231	Number	0.364	Reasoning	0.226
ME62149	-0.559	Algebra	0.480	Applying	0.638
ME62241	-1.264	Algebra	0.784	Applying	1.246
ME62105	-0.794	Algebra	1.941	Reasoning	1.148
ME62040	-0.471	Geometry	0.236	Applying	0.450
ME62288	0.806	Geometry	0.720	Applying	1.634
ME62173	-1.771	Geometry	1.315	Reasoning	0.972
ME62133	-0.830	Data and Probability	0.341	Applying	0.879
ME62123A	-2.016	Data and Probability	0.905	Knowing	-
ME62123B	-1.045	Data and Probability	0.377	Applying	1.022