

## 혼합분포 문항반응모형 적용을 위한 잠재집단의 이질성 수준 탐색\*

박 산 하\*\*

이화여자대학교

김 성 은

한국청소년정책연구원

본 연구는 혼합분포 문항반응모형 적용 시 집단의 이질성에 따른 잠재집단 추정의 정확성을 확인하여, 모형 활용을 위해 요구되는 이질성 수준을 제안하는 데 목적이 있다. 특히 기존 혼합분포 문항반응모형 연구에서 다루지 않은 하위집단 간 거리로 집단의 이질성을 측정하였다. 2-모수 혼합분포 문항반응모형을 사용하였으며, 모의실험을 통해 다양한 검사 조건(잠재집단의 수, 피험자 수, 잠재집단의 비율, 모수분포 평균의 차이)에서 잠재집단 간 마할라노비스 거리에 따른 잠재집단 수 결정의 정확성과 피험자 분류 정확성을 확인하였다. 연구결과, 잠재집단 간 마할라노비스 거리가 3 이상인 조건에서 잠재집단 수 결정 정확도와 피험자 분류 정확도가 눈에 띄게 높아졌다. 본 연구는 추후 혼합분포 문항반응모형을 활용하고자 하는 연구자들에게 잠재집단 추정의 정확성을 예측하기 위한 정보와 이를 바탕으로 모형 활용을 위해 요구되는 집단의 이질성 수준에 대한 정보를 제공하는 데 의의가 있다.

주제어 : 혼합분포 문항반응모형(mixture item response model), 잠재집단, 마할라노비스 거리

---

\* 이 논문은 제1저자의 박사학위논문의 일부를 발췌, 수정·보완하여 재구성한 것임.

\*\* 교신저자 : 박산하, 이화여자대학교 교육학과, fvmru@hanmail.net

## I. 서 론

사회과학에서 관심 있는 연구대상은 한 집단 안에 여러 가지 하위집단이 존재하기도 한다(Collins & Lanza, 2010). 한 가지 현상에 대하여 하위집단은 각기 다른 특성을 가질 수 있지만 하나의 동질집단으로 가정한다면, 하위집단별 특성은 드러나지 못한 채 사라지게 된다. 예를 들어 청소년의 정서·행동 문제는 성별에 따라 다른 양상으로 나타날 수 있다. 하지만 성별이라는 하위집단을 인지하지 못한 채 동질한 하나의 모집단을 가정하고 분석한다면, 자료가 가지고 있는 참 의미를 파악하지 못하게 된다. 이때, 하위집단은 직접 관찰할 수 있는 성별, 학교, 국가와 같은 명시적 집단일 수도 있고 관찰된 응답에 있어 유사한 특징을 가지는 비가시적 잠재집단(latent class)일 수도 있다. 잠재집단은 Cohen과 Bolt(2005), Maij-de Meij, Kelderman과 van der Flier(2010)에서 모집단의 이질성을 설명하는 데 유용하게 활용되었고, 자료를 설명하고 현상을 해석하는데 명시적 집단보다 더 효과적인 것으로 나타나기도 하였다.

잠재집단의 존재를 탐지하고 이를 포함하여 문항 및 피험자 특성을 분석하고자 한다면 혼합분포 문항반응모형(mixture item response model)을 사용할 수 있다. 혼합분포 문항반응모형은 혼합모형과 문항반응모형을 결합한 모형으로 각각의 모형보다 잠재된 하위집단이 존재하는 자료에 더 적합한 모형이다(Muthén & Asparouhov, 2006). 또한, 두 모형을 하나의 모형으로 결합함으로써 두 단계의 분석에서 나타날 수 있는 오차의 중복 문제를 줄일 수 있다. 혼합분포 문항반응모형은 혼합모형을 통해 피험자가 가지고 있는 특성에 따라 잠재집단을 분류함으로써 인간 지향적 접근(person-oriented approach)을 시도하고, 문항반응모형을 통해 문항과 잠재능력의 관계를 규명함으로써 변수 지향적 접근(variable-oriented approach)을 시도한다(Bergman & Magnusson, 1997). 이로써 혼합분포 문항반응모형은 다양한 관점으로 자료가 가진 정보를 더욱 풍부하게 이해할 수 있도록 한다.

만약 잠재집단이 존재하지 않는 동질집단의 자료이거나, 존재하더라도 이질성이 크지 않아 거의 동질한 집단의 자료로 볼 수 있다면 혼합모형을 사용하기보다는 하나의 모집단을 가정한 모형을 사용하는 것이 자료를 설명하는 데 더 적합할 것이다. 그러나 이질적인 잠재집단이 존재하는 자료에 문항반응모형을 활용하고자 한다면 혼합분포 문항반응모형을 사용하는 것이 적절하다. 그렇기 때문에 연구자가 관심 있는 집단이 이질적인지, 이질적이라면 그 수준이 어느 정도인지 확인하는 것은 모형 선택을 위해 반드시 필요한 단계이다. 또한 잠재집단의 이질성은 피험자 분류의 성공을 결정하는 중요한 요인이다(Lubke & Tueller, 2010). 이러한 이유로 기존의 혼합모형 연구들은 잠재집단의 이질성을 주요 관심 변수로 연구에 포함하고 있다(예: Blackwell, 2011; Choi, Paek, & Cho, 2017; Granado, 2015). 하지만 혼합

분포 문항반응모형에서 잠재집단의 이질성은 주로 문항 모수에 의한 차이로만 다루어지고 있으며(예: Choi, Paek, & Cho, 2017; Finch & French, 2012), 특히 모형 활용을 위해 요구되는 잠재집단 간 최소 이질성 수준을 탐색하는 연구는 부족한 실정이다.

본 연구는 혼합분포 문항반응모형 적용 시 잠재집단의 이질성에 따른 피험자 분류의 정확성을 비교하여, 모형 활용을 위해 요구되는 이질성 수준을 제안하고자 한다. 이때 기존 혼합분포 문항반응모형 연구에서 다루지 않은 하위집단 간 마할라노비스 거리를 통해 집단의 이질성을 측정함으로써 하나의 모수에 의한 이질성이 아닌 모형에 영향을 미치는 여러 요인을 모두 고려한 응답자료 자체의 이질성이 피험자 분류에 미치는 영향을 확인하고자 하였다.

## II. 이론적 배경

### 1. 혼합분포 문항반응모형

혼합분포 문항반응모형은 관찰된 자료가 동질한 모집단에 기인한 것이 아니라 둘 이상의 잠재집단의 자료가 혼합되어 있다고 가정한다(Rost, 1990). 따라서 모집단의 이질성을 다룰 수 있어 덜 제한적이며, 다양한 실재를 반영할 수 있다. 하위 잠재집단이 존재하는 모집단의 분석은 두 단계로 이루어진다. 첫 번째는 자료를 동질한 하위 잠재집단으로 분류하는 작업이며, 두 번째는 각 잠재집단에 대해 모형의 모수를 추정하는 작업이다. 이 두 단계는 혼합분포 문항반응모형을 통해 동시에 이루어진다. 일반적인 혼합분포 문항반응모형에 따른 주변 응답 확률(marginal response probability)은 식(1)과 같이 표현할 수 있다.

$$P(x_{ij} = 1) = \sum_{g=1}^G \pi_g \left( c_{ig} + (1 - c_{ig}) \frac{1}{1 + e^{-a_{ig}(\theta_{jg} - b_{ig})}} \right) \quad (1) \quad (1)$$

$$(\text{단, 모든 잠재집단에서 } 0 \leq \pi_g \leq 1, \sum_{g=1}^G \pi_g = 1)$$

여기서  $x_{ij}$ 은 피험자  $j$ 의 문항  $i$ 의 반응이고,  $G$ 는 잠재집단의 수,  $\pi_g$ 는 잠재집단  $g$ 에 속할 확률을 나타낸다.  $\theta_{jg}$ 는 피험자가 잠재집단  $g$ 에 속한다고 가정했을 때, 피험자  $j$ 의 능력

을 나타내는 잠재집단별 개인의 능력 모수이다. 이와 유사하게  $a_{ig}$ 는 잠재집단  $g$ 의  $i$  문항의 변별도,  $b_{ig}$ 는 잠재집단  $g$ 의 문항  $i$ 의 난이도,  $c_{ig}$ 는 잠재집단  $g$ 의 문항  $i$ 의 추측도이며 잠재집단별 문항 모수를 나타낸다. 기본적인 문항반응모형에서와 동일하게 문항 추측도인  $c$ 를 0으로 고정하면 2-모수 혼합분포 문항반응모형이 되고, 추측도인  $c$ 를 0으로, 변별도인  $a$ 가 동일하다고 고정하면 1-모수 혼합분포 문항반응모형이 된다. 각 문항모수의 의미 역시 기본적인 문항반응모형과 동일하다.

한편, 피험자는 응답 벡터를 기반으로 계산된 각 잠재집단에 속할 확률을 가지게 되며, 식(2)로 나타낼 수 있다.

$$p(g|X) = \frac{\pi_g p(X|g)}{\sum_{g=1}^G \pi_g p(X|g)} \quad (2)$$

$p(g|X)$ 는 응답 벡터  $X$ 가 나타났을 때 잠재집단  $g$ 에 속할 확률을,  $p(X|g)$ 는 잠재집단  $g$ 에서 응답 벡터  $X$ 가 나타날 확률을 의미한다. 모형 추정을 통해 피험자가 각 잠재집단에 속할 확률을 산출한 뒤 가장 확률이 높은 잠재집단에 할당된다(Rost, 1997). 모형은 각 피험자에 대하여 두 가지 진단적 정보를 제공한다. 하나는 개인이 각 잠재집단에 속할 비율이며, 두 번째는 능력 추정치이다.

혼합분포 문항반응모형은 하위집단의 이질성을 모형에 반영하여 잠재집단을 구분하고, 잠재집단의 특성을 확인하는 데 사용할 뿐 아니라(예: Finch & Pierson, 2011; Muthén & Asparouhov, 2006), 차별기능문항 등을 활용하여 검사의 공정성을 확인하고(예: Cho, Suh, & Lee, 2016; Kang & Cohen, 2007), 문제해결전략의 차이를 찾거나(예: Cho et al., 2010; Mislevy & Verhelst, 1987), 속도성 검사를 분석하는(예: Suh, Cho, & Wollack, 2012) 등 다양한 활용이 가능하여 최근 들어 많은 연구가 진행되고 있다. 이 외에도 혼합분포 문항반응모형 선택을 위한 적합도 지수의 정확성 비교(예: Choi, Paek, & Cho, 2017; Li, Kim, & Cho, 2009), 잠재집단의 분포가 모형 선택에 미치는 효과(예: Sen, Cohen, & Kim, 2016; Subedi, 2009), 잠재집단의 비교를 위해 공통 척도로 맞추는 방법(예: Choi, 2014; Paek & Cho, 2015), 결측값이 미치는 영향(예: Lee, 2012) 등 혼합분포 문항반응모형을 이해하기 위한 다양한 연구들도 이루어지고 있다. 혼합분포 문항반응모형 추정 방법에 대한 보다 자세한 내용은 Cho, Cohen과 Kim(2013), Finch와 French(2012) 등을 참고할 수 있으며, De Ayala와 Santiago(2017)를 통해 분석에 필요한 명령문(Syntax)과 실제자료 분석결과(Output)의 해석방법을 확인할 수 있다.

## 2. 마할라노비스거리

일반적으로 마할라노비스 거리(Mahalanobis distance; MD)는 유클리드 거리와 함께 공간상의 길이를 측정하기 위해 활용된다. 특히 마할라노비스의 거리는 유클리드 거리(Euclidean distance)와 달리 각 변수의 표준편차를 고려한 상대적인 거리로, 자료와 중심(centroid) 사이의 통계적인 거리를 나타낸다. 그러므로 변수의 수가 많은 다차원 상의 공간에서 거리를 측정하는데 유용하게 사용된다.

마할라노비스 거리는 잠재집단 중심(centroid) 사이의 통계적 거리를 측정하기 위해 사용함으로써 잠재집단 간의 거리 즉, 잠재집단의 이질성을 측정하는 지표로 사용하기도 한다(Lubke & Muthén, 2007; Lubke & Neale, 2006; Peugh & Fan, 2013). 이때, 두 집단 사이의 마할라노비스 거리를 구하는 방법은 식(3)과 같다(Lubke & Tueller, 2010).

$$MD = \sqrt{(\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2)} \quad (3)$$

여기서  $\mu_1$ 은 첫 번째 잠재집단의  $i$ 차원 문항 평균 벡터를,  $\mu_2$ 은 두 번째 잠재집단의  $i$ 차원 문항 평균 벡터를 나타내며,  $\Sigma$ 는  $i \times i$ 차원의 공분산 행렬을 나타낸다. 여러 개의 모수로 이루어진 문항반응모형의 경우, 마할라노비스 거리를 사용하여 잠재집단의 이질성을 측정하면, 문항의 모수의 차이만이(예: Choi, Paek, & Cho, 2017; Finch & French, 2012) 아닌, 응답 자료에 영향을 준 모든 요인을 반영한 응답 자료 자체의 이질성을 고려할 수 있다.

## III. 연구방법

본 연구는 혼합분포 문항반응모형 적용 시 집단의 이질성에 따른 잠재집단 추정의 정확성을 확인하고자 다양한 검사조건을 가정한 모의실험을 수행하였다. 이 모의실험에서 잠재집단 간의 이질성 정도는 마할라노비스 거리로 판단하였다. 모의실험을 통하여 마할라노비스 거리에 따라 첫째, 잠재집단 수 결정의 정확도는 어떻게 달라지는지, 둘째, 피험자 분류 정확도(엔트로피, 잠재집단 일치율, 피험자 모수 추정치 평균의 RMSE)는 어떠한지를 확인하고자 하였다. 또한 추가 분석으로서 잠재집단 간 문항 모수 패턴은 거리에 따라 얼마나 같거나 다른지도 그래프를 통해 확인하였다.

## 1. 모의자료 생성조건

모의자료는 피험자 수, 잠재집단의 비율, 모수분포 평균의 차이, 잠재집단의 마할라노비스 거리, 잠재집단의 수를 고려하여 생성하였다. 그 밖에 검사 길이, 자료 생성 모형 등 고정된 조건은 분석에 불안정성을 초래하지 않도록 설정하였다. 검사의 길이는 성취도 검사에서 적당한 문항 수인 25문항으로 고정하였다(Cho, Cohen, & Kim, 2014). 자료 생성 모형은 모든 문항이 동일한 변별도를 갖지 않는 이상 모수 추정에 어려움이 있을 수 있는 라쉬 모형(성태제, 2016)과 추측도 모수 추정의 불안정성(DeMars, 2001; van der Linden & Hambleton, 1997), 특정 조건에서 예측할 수 없는 높은 문항 부적합(Chon, Lee, & Dunbar, 2010; Sinharay, & Lu, 2008) 등의 문제를 지닌 3-모수 모형을 제외하고 혼합분포 문항반응모형 중 잠재집단 회복률이 가장 높은 2-모수 혼합분포 문항반응모형(Li et al., 2009)을 사용하여 자료를 생성하였다.

피험자 수는 이전 혼합분포 문항반응모형 연구에서 보편적으로 2가지 또는 3가지 조건으로 조절하고 있다. 피험자 수가 적은 조건은 약 500명, 일반적인 조건은 약 1,000명, 충분한 조건은 약 2,500명으로 연구를 수행하였다(Cho, Cohen, & Kim, 2013; Choi, Peak, & Cho, 2017; Finch & French, 2012; Sen, Cohen, & Kim, 2016). 본 연구에서는 피험자 수를 일반적인 조건과 충분한 조건으로 하고자 1,000명과 2,500명으로 설정하였다.

잠재집단은 두 개 또는 세 개가 존재하도록 하였다. 비율은 균등한 조건과 균등하지 않은 조건으로 생성하였다. 균등하지 않은 조건은 두 집단인 경우 Finch와 French(2012)를 참고하여 0.25, 0.75로, 세 집단인 경우 0.25, 0.25, 0.5의 비율로 집단을 구성하였다.

모수 분포는 잠재집단 간 평균 차이가 없는 경우와 있는 경우로 분리하였다. 집단의 능력 수준과 난이도 수준이 일치할 때 추정의 오차가 최소이기 때문에(De Ayala, 2009) 추정의 정확성을 위해 피험자 모수와 난이도 모수의 분포를 동일하게 설정하였다. 잠재집단이 두 개인 경우를 분석한 결과, 잠재집단 수 결정에서나 잠재집단 분류의 정확성에서 모수분포 평균의 차이가 결과에 영향을 주지 못하였다. 따라서 잠재집단이 세 개인 경우에서 모수분포 평균의 차이는 조건으로 고려하지 않았고, 세 집단의 평균 차이가 없도록 고정하였다. 모수 분포 평균의 차이를 조절하기 위해 자료 생성에 사용한 분포는 <표 1>과 같다.

잠재집단의 이질성은 마할라노비스 거리를 활용하여 잠재집단 중심점(centroid) 간 거리를 조절하였다. Lubke와 Neale(2006)은 잠재변수모형(latent variable models)에서 집단의 분리를 조절하기 위하여 마할라노비스 거리를 사용하였는데 두 집단인 경우 1.5, 3으로 조절하였다. 균등한 비율의 잠재집단에 비교해 균등하지 않은 비율의 잠재집단은 집단의 분리를 위해 더 먼 거리가 필요하기 때문에(Lubke & Tueller, 2010) 본 연구에서는 잠재집단의 거리를 마할

〈표 1〉 모수 추출을 위해 가정한 분포

	class 1	class 2			
		모수분포 평균의 차이			
		0	0.5	1	1.5
a	U(1,2)	U(1,2)	U(1,2)	U(1,2)	U(1,2)
b	N(0,1)	N(0,1)	N(0.5,1)	N(1,1)	N(1.5,1)
$\theta$	N(0,1)	N(0,1)	N(0.5,1)	N(1,1)	N(1.5,1)

라노비스 거리 2, 3, 4, 5로 설정하고 자료를 생성하였다. 세 집단에서 마할라노비스 거리가 2란 첫 번째 집단과 두 번째 집단의 거리, 두 번째 집단과 세 번째 집단의 거리, 세 번째 집단과 첫 번째 집단의 거리가 각각 2인 조건을 의미한다.

위에서 언급한 것처럼 두 개의 잠재집단의 경우 문항 모수의 추출을 위해 분포를 가정하고 이를 바탕으로 문항 모수를 추출하였다. 하지만 잠재집단이 세 개인 경우 특정 거리를 갖는 자료 세트를 추출하기 용이하지 않았다. 이러한 이유로 잠재집단이 세 개인 조건의 자료 생성은 잠재집단이 두 개인 모의자료에서 잠재집단의 마할라노비스 거리가 2, 3, 4, 5일 때 생성된 자료의 잠재집단 간 문항 모수의 차이를 활용하였다. 즉, 잠재집단이 두 개인 경우는 문항 모수 분포를 가정하여 자료를 생성하였으나 세 개인 경우는 잠재집단이 두 개인 경우의 문항 모수를 참고하여 문항 모수를 결정하였고 이를 바탕으로 잠재집단이 세 개인 응답 자료를 생성하였다. 잠재집단이 세 개인 조건에서 사용한 문항 모수는 [부록 1]에서 확인할 수 있다.

결론적으로 잠재집단이 두 개인 경우, 자료 생성 조건은 피험자 수(1,000, 2,500), 잠재집단의 비율(균등, 불균등), 모수분포 평균의 차이(0, 0.5, 1, 1.5), 잠재집단 간 마할라노비스 거리(2, 3, 4, 5)를 고려하여 총 64가지( $=2*2*4*4$ ) 조건의 자료를 생성하였다. 잠재집단이 세 개인 경우는 자료 생성 조건은 피험자 수, 잠재집단의 비율, 잠재집단의 마할라노비스 거리를 고려하여 총 16가지( $=2*2*4$ ) 조건의 자료를 생성하였다. 자료 생성을 위해 R-software3.0.3을 사용하였으며, 구체적인 코드는 [부록 2]에서 확인할 수 있다. 자료 생성을 위한 대략적 과정



〔그림 1〕 자료 생성 과정

은 [그림 1]과 같다.

예를 들어 두 잠재집단이며, 피험자 1,000명, 잠재집단 비율이 균등, 모수분포 동일, 마할라노비스 거리가 2인 조건의 자료를 생성한다고 가정해보자. 이를 위해 두 잠재집단의 모수분포를 동일하게  $a1 \sim U(1, 2)$ ,  $b1 \sim N(0, 1)$ ,  $\theta1 \sim N(0, 1)$ ,  $a2 \sim U(1, 2)$ ,  $b2 \sim N(0, 1)$ ,  $\theta2 \sim N(0, 1)$ 으로 입력하고, 피험자 수와 잠재집단 비율도 입력한다. 다음으로 집단별로 해당 분포에서 변별도, 난이도, 피험자 모수를 추출한다. 추출된 모수를 바탕으로 각 집단의 응답 자료를 생성하고, 두 집단의 마할라노비스 거리를 산출한다. 산출된 거리가 2가 아니라면 다시 집단별로 모수를 추출하여 이후 과정을 반복한다. 만약 산출된 거리가 2라면 해당 응답 자료와 모수를 저장하고 1개의 자료 세트를 얻게 된다. 자료 세트와 모수 세트가 50개가 저장될 때까지 동일한 과정을 반복한다. 이러한 50개의 모수 세트를 통해 마할라노비스 거리가 2인 다양한 검사 조건을 가정할 수 있다. 이후 저장된 자료 세트를 분석에 활용하였다.

## 2. 분석방법

생성한 자료의 분석은 문항반응모형과 혼합분포 문항반응모형을 통해 이루어졌다. 두 집단을 가정하고 생성한 자료는 2PL, MIX2PL\_2class, MIX2PL\_3class 모형을 활용하여 분석하였으며(64조건\*3모형), 세 집단을 가정하고 생성한 자료는 2PL, MIX2PL\_2class, MIX2PL\_3class, MIX2PL\_4class 모형으로 분석하였다(16조건\*4모형). 자료 분석을 위해 Mplus 7와 R-software3.0.3의 MplusAutomation Package version 0.6-3을 사용하였다.

위 모형을 통해 모수를 추정하는 방법으로는 강건 최대우도 추정(robust maximum likelihood; MLR)방법을 선택하였다. 최대우도 추정은 정규성(normality)을 만족하지 않는 이분(dichotomous) 변수를 분석하는 경우  $\chi^2$  값의 과대추정(overestimation)과 표준오차의 과소추정(underestimation) 문제를 보일 수 있으므로 혼합분포 문항반응모형의 추정은 강건한 최대우도 방법을 활용하였다. 이때, 국소 최대치(local maxima) 문제가 나타날 수 있으므로 본 연구에서는 자료 생성 시 정해진 잠재집단의 수보다 더 많은 집단을 가정하는 모형을 수행하는 경우에는 800개의 무선 초깃값을 사용하였고, 나머지 경우에는 Mplus 프로그램에서 고정된 20개의 무선 초깃값을 사용하였다. 최대 우도 추정방법에서 나타날 수 있는 국소 최대치 문제 외에도, 잠재집단분석에서는 라벨전환(label switching) 현상이 있을 수 있다. 본 연구에서는 이를 확인하기 위해 생성된 참(true) 잠재집단과 추정된 잠재집단의 교차표를 작성하였고, 이러한 현상이 나타난 자료 세트는 라벨을 다시 붙이는 절차(relabelling)를 거쳐 수정하였다.

자료를 생성한 후에는 잠재집단 수 결정 정확성과 피험자 분류 정확성이라는 두 개의 평가준거를 가지고 모형 적용이 적절한 마할라노비스 거리를 찾기 위하여 모형적합도 지수,



엔트로피, 잠재집단 일치율, 피험자 모수 추정치 평균의 RMSE 등을 산출하였다. 또한 추가적 분석으로써 잠재집단 간의 마할라노비스 거리에 따라 추출된 모수의 패턴은 어떠한지 살펴보기 위해 [그림 1]의 자료생성 과정을 통해 추출된 모수를 그래프를 그려 확인하였다. 본 연구에서 문항모수의 패턴이란 25개 문항의 모수(예: 문항난이도)를 사용하여 그리는 그래프로 이해할 수 있는데, 이 그래프의 모양이 잠재집단 간 얼마나 유사하거나 다른지를 확인하였다.

### 3. 평가준거

#### 1) 잠재집단 수 결정 정확성

최적의 잠재집단 수를 결정하는 다양한 연구가 있었음에도 이를 결정하기 위한 추론 과정에 대한 일반화된 합의는 이루어지지 않은 상태이기 때문에(Tein, Cox, & Cham, 2013) 대부분의 연구에서 최적의 잠재집단의 수를 결정하기 위해 다양한 통계적 지표를 조합하여 잠재집단의 수를 추론하고 있다. 따라서 본 연구에서는 최적의 잠재집단 수를 결정하기 위해 정보기반지수(information criteria; IC)인 AIC(Akaike, 1987), BIC(Schwarz, 1978), ABIC(Schloer, 1987), AICC(Sugiura 1978)를 사용하여 해당 자료에 적합한 모형을 결정하였다.

#### 2) 피험자 분류 정확성

다양한 잠재집단 간 마할라노비스 거리 따른 피험자 분류 정확도를 확인하기 위해 엔트로피와 잠재집단 일치율, 피험자 분포 평균의 RMSE 등을 확인한다. 엔트로피(entropy)는 추정된 혼합모형을 통해 얻은 잠재집단의 분류 정확도를 표준화한 지수이다. 따라서 엔트로피를 활용하여 각 잠재집단이 분명히 분류되어 있는지를 판단할 수 있으며, 모형을 평가하기 위한 지수로 활용할 수 있다(Crosby et al., 2009). 엔트로피는 0~1 사이의 값을 가지며, 값이 클수록 잠재집단의 분류가 정확함을 의미한다(Celeux & Soromenho, 1996; Henson, Reise, & Kim, 2007). 엔트로피는 다음 식(4)를 통해 산출할 수 있다(Asparouhov, & Muthén, 2014).

$$entropy = 1 + \frac{1}{n \log(G)} \left( \sum_{i=1}^n \sum_{g=1}^G \hat{p}_{ig} \log(\hat{p}_{ig}) \right) \quad (4)$$

위 식에서  $n$ 은 피험자 수,  $G$ 는 잠재집단의 수,  $\hat{p}_{ig}$ 는 집단  $g$ 에 속하는 학생  $i$ 의 추정된 조건부 확률을 의미한다.

잠재집단 일치율은 처음에 생성된 참(true) 잠재집단과 추정된 잠재집단이 일치하는 비율

로 산출하였다. 잠재집단의 일치율이 높을수록 피험자 분류가 잘 된 것으로 판단할 수 있다.

피험자 분류 정확성을 확인하기 위하여 피험자 모수 추출 시 가정한 분포의 평균과 추정된 피험자 모수 평균의 RMSE(root mean square error)를 산출하였다. RMSE가 작을수록 추출한 분포와 추정된 분포가 유사함을 의미하며 이는 잠재집단의 분류가 잘된 것을 보여준다. 피험자 모수 추정치 평균의 RMSE는 식(5)를 통해 산출한다.

$$\text{피험자 모수 추정치 평균 RMSE} = \sqrt{\frac{\sum_{r=1}^R \sum_{c=1}^C (\widehat{mean}_{cr} - mean_{cr})^2}{RC}} \quad (5)$$

식(5)에서 R은 자료의 총 세트 수, 즉 50이며, C는 잠재집단의 수,  $mean_{cr}$ 는 자료 세트 r, 잠재집단 c의 피험자 모수를 추출한 분포의 평균,  $\widehat{mean}_{cr}$ 는 자료 세트 r, 잠재집단 c의 추정된 피험자 모수 평균을 의미한다.

## IV. 연구결과

### 1. 잠재집단 수 결정

잠재집단 수 결정이 정확하게 이루어지는지를 확인하기 위해 모형 적합도 지수인 AIC, BIC, AICC, ABIC를 활용하였고, ‘생성한 잠재집단 수+1’개의 모형 적합도를 비교하여, 적합도가 가장 작은 모형을 선택하였다. 일부 혼합분포 문항반응모형 분석에 국소 최대치 가능성 등과 관련된 오류와 경고가 나타났다. 조건에 따른 오류와 경고 수, 이를 제외하고 결과에 제시한 자료 세트 수는 박산하(2018)에 제시되어 있다. 조건별로 총 자료 세트 중 잠재집단 수를 맞힌 지수 개수의 비율이 <표 2>, <표 3>에 제시되어 있다. 예를 들어, 잠재집단 수가 두 개이고, 균등하며, 잠재집단 간 거리는 2이고 모수분포 평균의 차이가 0, 1,000명의 피험자인 조건에서 맞힌 지수 개수가 2개인 경우가 33.3, 3개인 경우가 66.7이라 함은 4개의 지수 중 맞힌 지수가 2개인 세트가 총 세트 중 33.3%, 4개 중 맞힌 지수가 3개인 세트는 66.7%라는 뜻이다.

<표 2>와 같이 잠재집단이 두 개인 경우, 1,000명의 피험자 조건부터 살펴보면, 잠재집단 간 마할라노비스 거리가 멀수록 적합도 지수가 비교적 일관된 결과를 보임을 알 수 있다. 잠재집단의 거리가 3 이상인 대부분 조건에서 맞힌 지수가 3개 이상인 경우가 95% 이상 나

〈표 2〉 잠재집단 수를 맞힌 지수 개수의 비율(두 개의 잠재집단)

비율	MD	평균 차이	맞힌 적합도 지수 개수(1,000명)				맞힌 적합도 지수 개수(2,500명)			
			1	2	3	4	1	2	3	4
균등	2	0	0.0	33.3	66.7	0.0	0.0	96.7	3.3	0.0
		0.5	0.0	26.1	73.9	0.0	0.0	100.0	0.0	0.0
		1	8.6	25.7	65.7	0.0	0.0	100.0	0.0	0.0
		1.5	3.2	32.3	64.5	0.0	0.0	100.0	0.0	0.0
	3	0	0.0	0.0	100.0	0.0	0.0	96.3	3.7	0.0
		0.5	0.0	2.8	97.2	0.0	0.0	92.3	7.7	0.0
		1	0.0	6.9	93.1	0.0	0.0	96.4	3.6	0.0
		1.5	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0
	4	0	0.0	3.1	93.8	3.1	0.0	91.3	8.7	0.0
		0.5	0.0	3.8	96.2	0.0	0.0	79.2	20.8	0.0
		1	0.0	0.0	96.6	3.4	0.0	92.9	7.1	0.0
		1.5	0.0	3.7	92.6	3.7	0.0	86.7	13.3	0.0
	5	0	0.0	4.0	96.0	0.0	0.0	72.0	28.0	0.0
		0.5	0.0	0.0	93.1	6.9	0.0	74.1	25.9	0.0
		1	0.0	0.0	96.9	3.1	0.0	82.8	17.2	0.0
		1.5	0.0	0.0	100.0	0.0	0.0	71.4	28.6	0.0
불균등	2	0	4.2	91.7	0.0	4.2	0.0	93.9	6.1	0.0
		0.5	10.7	89.3	0.0	0.0	0.0	93.5	6.5	0.0
		1	7.1	89.3	3.6	0.0	0.0	90.0	10.0	0.0
		1.5	3.7	88.9	7.4	0.0	0.0	90.6	9.4	0.0
	3	0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0
		0.5	0.0	0.0	92.6	7.4	0.0	85.2	14.8	0.0
		1	0.0	6.5	93.5	0.0	0.0	96.4	3.6	0.0
		1.5	0.0	0.0	100.0	0.0	0.0	93.9	6.1	0.0
	4	0	0.0	6.7	93.3	0.0	0.0	75.0	25.0	0.0
		0.5	0.0	3.4	86.2	10.3	0.0	69.2	26.9	0.0
		1	0.0	0.0	100.0	0.0	0.0	70.4	29.6	0.0
		1.5	0.0	4.0	92.0	4.0	0.0	82.6	17.4	0.0
	5	0	0.0	0.0	93.8	6.3	0.0	64.0	36.0	0.0
		0.5	0.0	4.2	95.8	0.0	0.0	64.0	32.0	4.0
		1	0.0	3.3	93.3	3.3	0.0	76.9	23.1	0.0
		1.5	0.0	0.0	96.4	3.6	0.0	77.8	22.2	0.0

〈표 3〉 잠재집단 수를 맞힌 지수 개수의 비율(세 개의 잠재집단)

비율	MD	맞힌 적합도 지수 개수(1,000명)				맞힌 적합도 지수 개수(2,500명)			
		1	2	3	4	1	2	3	4
균등	2	18.2	81.8	0.0	0.0	3.7	70.4	25.9	0.0
	3	0.0	0.0	92.3	7.7	0.0	39.3	69.7	0.0
	4	0.0	0.0	100.0	0.0	0.0	23.3	76.7	0.0
	5	0.0	0.0	88.2	11.8	0.0	21.7	60.9	17.4
불균등	2	39.1	60.9	0.0	0.0	40.0	60.0	0.0	0.0
	3	0.0	0.0	100.0	0.0	0.0	45.2	54.8	0.0
	4	0.0	0.0	96.4	3.6	0.0	33.3	66.7	0.0
	5	0.0	0.0	74.1	25.9	0.0	19.4	71.0	9.7

타났다. 하지만 마할라노비스 거리가 2인 조건에서는 맞힌 지수의 개수가 대부분 적게 나타났다. 잠재집단의 비율에 따른 차이도 존재하였다. 거리가 2이고 잠재집단의 비율이 0.5인 조건에서는 3개 이상 맞힌 경우가 64.5%~73.9%로 나타났다. 거리가 2이고 잠재집단의 비율이 0.25인 조건에서는 맞힌 지수가 2개 이하인 경우가 대부분으로 나타났다. 2,500명의 피험자 조건에서도 잠재집단 간 마할라노비스 거리가 멀수록 맞힌 지수 개수가 비교적 일관된 결과를 보임을 알 수 있다. 하지만 1,000명의 조건과 달리 2,500명 조건에서는 맞힌 지수가 2개인 비율이 가장 높았다. 잠재집단의 거리가 2와 3 이하인 대부분 조건에서 맞힌 지수가 2개인 경우가 90% 이상 나타났다. 마할라노비스 거리가 4, 5인 조건은 맞힌 지수가 3개 이상인 비율이 증가하였다.

<표 3>과 같이 잠재집단이 세 개인 경우에도 맞힌 지수 개수가 증가하는 경향은 마할라노비스 거리에 따라 비교적 분명하게 나타났다. 잠재집단의 거리가 2인 경우 맞힌 지수 개수 비교적 적고, 잠재집단의 거리가 3 이상인 경우 맞힌 지수 개수 비교적 많았다. 거리에 따른 경향은 피험자 수가 적을 때 더 도드라졌다. 피험자 수가 1,000명이고 거리가 3 이상인 조건에서는 맞힌 지수가 3개 이상인 비율은 100%였다. 피험자 수가 2,500명이고 거리가 3 이상인 조건에서는 거리가 늘어날수록 3개 이상 맞힌 지수 개수가 증가하였다. 마할라노비스 거리가 2인 조건에서는 맞힌 지수 개수가 1개뿐인 경우도 상당한 비율(3.7%~40.0%)로 나타났다.

## 2. 피험자 분류

피험자의 분류가 정확하게 이루어지는지 확인하기 위하여 엔트로피와 잠재집단이 일치하는 비율, 피험자 분포 평균의 RMSE를 확인하였다. 일부 혼합분포 문항반응모형 분석에 국소 최대치 가능성 등 오류와 경고가 나타났다. 해당 자료는 이후 분석에서 제외되었고, 50세트가 충족되지 못한 조건은 자료를 추가로 생성하여 분석하였다.

### 1) 엔트로피

추정된 잠재집단의 분류 정확도를 표준화한 수치인 엔트로피를 활용하여 피험자의 분류 정확성을 확인하였다. 50개 세트의 평균 엔트로피는 <표 4>와 같다.

모든 조건에서 엔트로피는 0.52~0.96 사이에 존재하였다. 엔트로피는 잠재집단 간 마할라

〈표 4〉 엔트로피 평균

N	비율	MD	2class				평균	3class
			모수분포 평균 차이					
			0	0.5	1	1.5		
1,000	균등	2	0.56	0.56	0.57	0.56	0.56	0.57
		3	0.78	0.78	0.78	0.79	0.78	0.77
		4	0.91	0.91	0.91	0.91	0.91	0.89
		5	0.96	0.96	0.96	0.96	0.96	0.96
	불균등	2	0.64	0.64	0.64	0.64	0.64	0.60
		3	0.82	0.82	0.82	0.82	0.82	0.79
		4	0.92	0.92	0.92	0.92	0.92	0.90
		5	0.96	0.96	0.96	0.96	0.96	0.96
2,500	균등	2	0.54	0.53	0.53	0.53	0.53	0.52
		3	0.78	0.78	0.78	0.78	0.78	0.76
		4	0.91	0.91	0.91	0.91	0.91	0.89
		5	0.96	0.96	0.96	0.96	0.96	0.96
	불균등	2	0.62	0.61	0.62	0.61	0.62	0.54
		3	0.81	0.81	0.81	0.81	0.81	0.77
		4	0.92	0.92	0.92	0.93	0.92	0.89
		5	0.96	0.96	0.96	0.96	0.96	0.96

노비스 거리에 따라 명확한 차이를 보였고 이러한 경향은 다른 검사 조건과 관계없이 나타났다. 잠재집단 수, 피험자 수, 모수분포 평균의 차이는 엔트로피에 거의 영향을 미치지 못함을 알 수 있다. 잠재집단의 거리가 2인 조건에서 엔트로피 평균은 0.57, 거리가 3인 조건에서는 0.79, 거리가 4인 조건에서는 0.90, 거리가 5인 조건에서는 0.96으로 분석되었다.

## 2) 잠재집단 일치율

잠재집단이 일치한다는 것은 분석을 통해 처음 생성 시 할당된 잠재집단에 동일하게 할당되었음을 의미한다. 세트별로 잠재집단의 일치율을 분석하였으며, 50세트의 평균을 산출하였다. 구체적인 잠재집단 일치율은 <표 5>와 같다.

모든 조건에서 잠재집단 일치율은 82.3~99.0 사이에 존재하였다. 잠재집단 일치율은 잠재집단 간 마할라노비스 거리에 따라 명확한 차이를 보였고 이러한 경향은 다른 검사 조건과 관계없이 나타났다. 모수분포 평균의 차이는 잠재집단 일치율에 거의 영향을 미치지 못함을 알 수 있다. 잠재집단의 거리가 2인 조건에서 잠재집단 일치율 평균은 78.6, 거리가 3인 조건에서는 90.7, 거리가 4인 조건에서는 96.0, 거리가 5인 조건에서는 98.4로 분석되었다.

## 3) 피험자 모수 추정치의 RMSE

피험자 모수를 추출한 분포의 평균과 추정된 피험자 분포의 평균 차이를 통해 잠재집단 분류의 정확성을 확인하였다. 평균의 RMSE가 작을수록 가정한 피험자 분포의 평균과 추정된 피험자 분포의 평균이 유사함을 의미한다. 피험자 모수 추정치 평균의 RMSE는 <표 6>과 같다.

모든 조건에서 RMSE는 0.011~0.055 사이에 존재하였다. RMSE는 잠재집단 간 마할라노비스 거리에 따라 명확한 차이를 보였고 이러한 경향은 다른 검사 조건과 관계없이 나타났다. 모수분포 평균의 차이에 따른 일관된 RMSE의 변화는 없었다. 잠재집단의 거리가 2인 조건에서 RMSE 평균은 0.046, 거리가 3인 조건에서는 0.027, 거리가 4인 조건에서는 0.020, 거리가 5인 조건에서는 0.014로 분석되었다.

## 3. 거리별 모수 패턴 차이

잠재집단 간 마할라노비스 거리에 따라 추출된 모수의 경향을 확인하기 위하여 피험자 수 1,000명, 균등한 잠재집단 비율, 잠재집단 수는 2개인 조건에서의 경향을 제시한다. 문항 변별도는 거리에 따라 특별한 경향이 없었고, 피험자 모수도 거리와 관계없이 가정한 분포에 따라 피험자 모수가 추출되었다. 문항 난이도 모수 패턴은 마할라노비스 거리에 따라 차

〈표 5〉 잠재집단 일치율

N	비율	MD	집단크기	2class				평균	3class
				모수분포    평균    차이					
				0	0.5	1	1.5		
1,000	균등	2		82.6	82.3	82.4	82.3	82.4	69.9
		3		92.9	92.7	92.8	92.9	92.8	86.8
		4		97.2	97.2	97.1	97.2	97.2	94.1
		5		98.7	98.6	98.7	98.6	98.7	97.8
	불균등	2	small	67.2	68.6	69.1	69.6	68.6	62.7
			large	90.6	90.7	90.8	90.3	90.6	79.3
			total	84.7	85.2	85.3	85.1	85.1	71.0
		3	small	85.7	85.9	86.4	86.5	86.1	83.1
			large	96.5	96.4	96.1	96.0	96.3	91.2
			total	93.8	93.8	93.7	93.6	93.7	87.1
		4	small	94.6	94.1	94.7	94.4	94.5	92.4
			large	98.4	98.3	98.5	98.5	98.4	96.4
			total	97.5	97.3	97.6	97.5	97.5	94.4
		5	small	97.2	97.0	97.4	97.4	97.3	97.1
			large	99.3	99.2	99.2	99.3	99.3	98.8
			total	98.8	98.7	98.8	98.8	98.8	97.9
2,500	균등	2		84.2	83.9	84.1	84.1	84.1	74.1
		3		93.3	93.3	93.4	93.3	93.3	88.4
		4		97.4	97.4	97.4	97.4	97.4	94.7
		5		98.7	98.8	98.7	98.8	98.8	98.2
	불균등	2	small	68.2	69.6	67.5	70.4	68.9	67.7
			large	93.4	93.0	93.7	92.6	93.2	82.8
			total	87.1	87.1	87.2	87.0	87.1	75.2
		3	small	87.0	86.5	86.5	87.2	86.8	85.3
			large	96.9	97.0	96.9	96.8	96.9	92.6
			total	94.4	94.4	94.3	94.4	94.4	88.9
		4	small	94.5	94.7	94.8	94.7	94.7	93.5
			large	98.8	98.7	98.7	98.9	98.8	96.8
			total	97.7	97.7	97.7	97.8	97.7	95.1
		5	small	97.4	97.2	97.4	97.2	97.3	97.5
			large	99.3	99.5	99.5	99.3	99.4	99
			total	98.9	98.9	99.0	98.8	98.9	98.2

〈표 6〉 피험자 모수 추정치의 RMSE

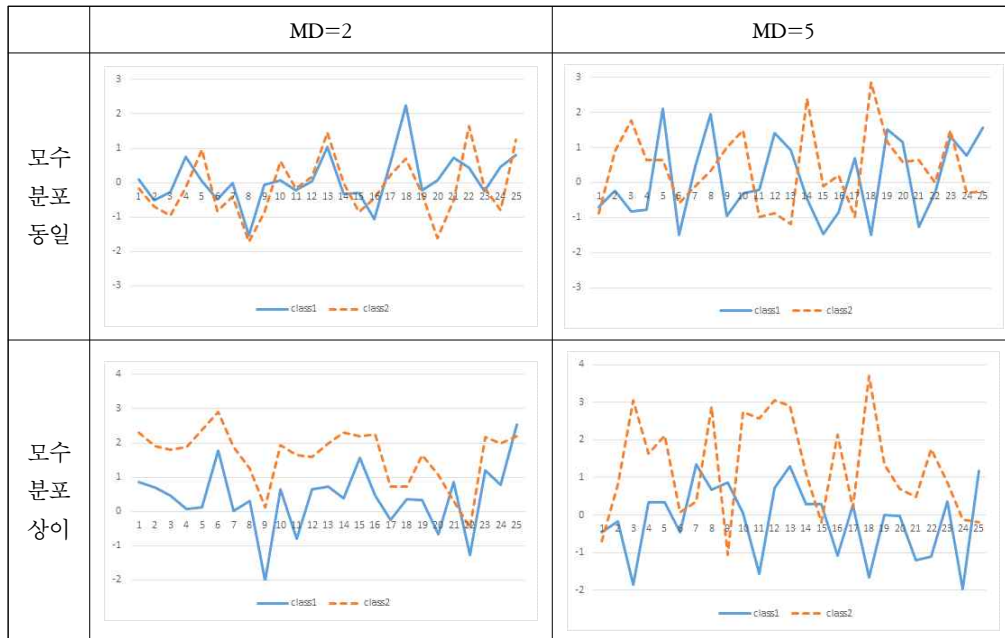
N	비율	MD	2class				평균	3class
			모수분포 평균 차이					
			0	0.5	1	1.5		
1,000	균등	2	0.037	0.041	0.046	0.041	0.041	0.044
		3	0.029	0.021	0.030	0.029	0.027	0.027
		4	0.019	0.020	0.018	0.020	0.019	0.020
		5	0.015	0.014	0.012	0.015	0.014	0.015
	불균등	2	0.049	0.052	0.047	0.055	0.051	0.048
		3	0.031	0.036	0.035	0.032	0.034	0.024
		4	0.020	0.024	0.021	0.022	0.022	0.023
		5	0.015	0.019	0.014	0.014	0.016	0.015
2,500	균등	2	0.041	0.045	0.039	0.043	0.042	0.048
		3	0.026	0.027	0.029	0.030	0.028	0.022
		4	0.017	0.018	0.016	0.017	0.017	0.017
		5	0.015	0.012	0.015	0.011	0.013	0.014
	불균등	2	0.043	0.050	0.047	0.041	0.045	0.049
		3	0.023	0.025	0.032	0.034	0.029	0.021
		4	0.023	0.022	0.021	0.015	0.020	0.018
		5	0.017	0.015	0.012	0.015	0.015	0.011

이가 나타났는데, 예시는 [그림 2]와 같다.

모수분포가 동일한 조건에서 문항 난이도는  $N(0, 1)$ 을 가정하였기 때문에 거리가 2인 조건에서나 거리가 5인 조건에서 모두  $X=0$  축 근처에 나타났다. 그러나 그래프의 유형은 다르게 분석되었다. 거리가 2인 조건(MD=2)의 문항 난이도는 첫 번째 잠재집단(class1)과 두 번째 잠재집단(class2)의 문항 난이도가 비교적 유사하게 평행을 이루며 나타났다. 반면 거리가 5인 조건(MD=5)의 문항 난이도는 첫 번째 잠재집단과 두 번째 잠재집단의 문항 난이도 그래프가 계속 교차하며 잠재집단에 따라 패턴이 매우 이질적으로 나타난다.

모수분포가 다른 조건에서도 이러한 현상을 유사하게 나타났다. 문항 난이도는 첫 번째 집단에서는  $N(0, 1)$ 을, 두 번째 집단에서는  $N(1.5, 1)$ 을 가정하였기 때문에 첫 번째 집단은  $X=0$  축 근처에, 두 번째 집단은 그래프가  $X=1.5$  축 근처에 높게 나타났다. 이는 거리가 2





(그림 2) 문항 난이도 패턴 예시

인 조건과 거리가 5인 조건에서 동일하였다. 그러나 난이도 그래프의 유형은 거리가 2인 조건에서는 두 집단이 거의 유사한, 평행에 가까운 패턴으로 나타났으나, 거리가 5인 조건에서는 매우 다른 패턴으로 나타남을 알 수 있다.

## V. 논의 및 결론

본 연구는 혼합분포 문항반응모형 적용 시 집단의 이질성에 따른 잠재집단 추정의 정확성을 확인하여, 모형 활용을 위해 요구되는 이질성 수준을 제안하고자 다양한 검사조건을 가정한 모의실험을 수행하였다. 그 결과, 잠재집단의 거리가 2인 경우, 잠재집단 수 결정 정확도와 피험자 분류 정확도에서 수용하기 어려운 수준을 산출하였으나, 잠재집단의 거리가 3 이상인 경우, 비교적 안정된 수치를 보임을 확인하였다.

이 결과에 터하여 혼합분포 문항반응모형 적용이 고민되는 상황에서 잠재집단 간 마할라노비스 거리를 산출해 모형을 결정할 수 있다. 본 연구에서 혼합분포 문항반응모형 활용을 위해 요구되는 집단 간 거리 수준을 확인하여 잠재집단 간 마할라노비스 거리가 3 이상인 자료에서 비교적 정확한 잠재집단 추정이 가능함을 보였다. 연구결과를 바탕으로 거리가 3

이상인 자료의 경우, 혼합분포 문항반응모형을 적용한다면 두 개 이상의 적합도 지수가 일치하는 결과를 보이며, 엔트로피가 0.8, 잠재집단 일치율이 90%, 피험자 모수 추정치 평균에 대한 RMSE는 0.03 정도의 잠재집단 분류 정확성을 나타낼 것을 예측할 수 있다.

이는 잠재집단 분류 정확성을 판단하는 지표로 기존에 사용하는 모형 적합도와 엔트로피 외에 잠재집단 간 마할라노비스 거리를 사용할 수 있음을 의미하기도 한다. 혼합분포 문항 반응모형을 통해 실제 자료를 분석하면, 참(true) 잠재집단을 알 수 없으므로 간접적인 방법인 모형 적합도와 엔트로피를 가지고 잠재집단 분류의 정확성을 예측해왔다. 하지만 모형 적합도 지수를 사용하는 경우 모형 자체에 대한 적합도라기보다는 다른 모형과의 비교를 위한 상대적 지표이므로 해당 모형의 분류 정확성을 확인하기는 어렵다. 또한, 엔트로피는 각 잠재집단으로 분류된 확률을 기반으로 산출된 지수으로써 직접적인 분류 정확도를 의미하는 수치는 아니다. 일부 연구에서 특정 준거 값(예:  $>0.8$ )을 좋은 잠재집단 분류를 의미하는 값으로 사용하고 있지만, 이 역시 높은 엔트로피가 높은 잠재집단 일치율과 관련이 있다는 연구(Lubke & Muthén, 2007)에 근거하며, 널리 인정되는 준거 값에 관한 연구는 아직 부족한 실정이다. 적절한 엔트로피 값에 대한 일반적인 합의가 부족하므로 Pastor 외(2007)는 잠재집단 분류의 정확성을 확인하는데 하나의 엔트로피 값에 대해 해석하기보다는 다른 모형과의 비교를 통해 사용할 것을 제안하기도 하였다. 또한, Wu(2009)는 특정 조건에서 엔트로피의 경향과 분류 정확도가 반드시 일치하는 것은 아니라고 하였다.

본 연구는 잠재집단 분류 정확성을 판단하는 지표로 잠재집단 간 마할라노비스 거리도 사용할 수 있음을 제안한다. 마할라노비스 거리는 엔트로피나 모형 적합도와 달리 직관적으로 이해하기 용이하며, 이론적 배경에 의해 어느 정도 예측하는 것이 가능하다. 엔트로피와 적합도 지수만으로 잠재집단 분류 정확성을 예측하기 모호한 경우 잠재집단 간의 거리를 산출해 봄으로써 자료에 대한 이해를 높이고, 분류 정확성을 예측하는 또 다른 지표로 활용할 수 있을 것이다.

또한 잠재집단 간 거리별 자료를 확인한 결과, 가까운 거리는 문항 난이도 패턴(그래프의 모양)이 유사했고, 먼 거리는 문항별 난이도 패턴이 서로 매우 다르며, 때로는 그래프가 교차하는 모양을 보였다. 이 결과를 통해 잠재집단 간 거리는 개별 문항 모수의 차이라기보다는 문항 모수 패턴의 차이에 의한 것으로 추론해 볼 수 있다. 결과적으로, 잠재집단 모수분포 평균의 차이는 잠재집단 수 결정 정확성과 잠재집단 분류 정확성에 영향을 주지 못했고, 문항 모수 패턴의 차이로 유발된 잠재집단의 거리는 잠재집단 수 결정 정확성과 잠재집단 분류 정확성에 큰 영향을 준 것으로 보인다. 이러한 결과는 난이도 차이의 패턴이 가장 이질적인 완전 교차(fully crossing) 유형에서 적합도 지수의 수행이 가장 좋았고, 문항 난이도 차이의 패턴이 가장 유사한 완전 평행(fully parallel) 유형에서의 수행이 가장 좋지 못하다는

Choi, Paek과 Cho(2017)의 결과를 지지한다.

본 연구는 대규모 표본의 검사를 고려하여 자료를 생성하였기 때문에 검사 결과 역시 대규모 자료에만 국한되어 활용 가능하다. 또한, 문항 수를 고정하여 모의실험을 진행하였기 때문에 다양한 검사 길이를 고려하지 못하였고, 세 잠재집단 분석을 가정한 모의실험에서 검사 조건으로 동일한 잠재집단 간 마할라노비스 거리만 포함하였다. 그러므로 연구결과를 더 일반적으로 적용할 수 있도록 다양한 검사 조건을 반영한 연구가 수행될 필요가 있다.

## 참고문헌

- 성태제(2016). **문항반응이론의 이해와 적용**. 제2판. 서울: 교육과학사.
- 박산하(2018). 혼합분포 문항반응모형 적용을 위한 잠재집단의 이질성 수준 탐색. 이화여자 대학교 대학원 박사학위논문.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52(3), 317-332.
- Asparouhov, T., & Muthén, B. (2014). *Variable-specific entropy contribution*. Technical appendix. Los Angeles: Muthén & Muthén. Retrived from <http://www.statmodel.com/download/UnivariateEntropy.pdf>
- Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology*, 9(2), 291-319.
- Blackwell, K. C. (2011). *The Sensitivity of Confirmatory Factor Analytic Fit Indices to Violations of Factorial Invariance across Latent Classes: A Simulation Study* (Doctoral dissertation, Arizona State University, AZ). Retrived from <https://repository.asu.edu>
- Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13(2), 195-212.
- Cohen, A. S., & Bolt, D. M. (2005). A mixture model analysis of differential item functioning. *Journal of Educational Measurement*, 42(2), 133-148.
- Cho, S. J., Cohen, A. S., & Kim, S. H. (2013). Markov chain Monte Carlo estimation of a mixture item response theory model. *Journal of Statistical Computation and Simulation*, 83(2), 278-306.
- Cho, S. J., Cohen, A. S., & Kim, S. H. (2014). A mixture group bifactor model for binary responses. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3), 375-395.
- Cho, S. J., Cohen, A. S., Kim, S. H., & Bottge, B. (2010). Latent transition analysis with a mixture item response theory measurement model. *Applied Psychological Measurement*, 34(7), 483-504.
- Cho, S. J., Suh, Y., & Lee, W. Y. (2016). After differential item functioning Is detected: IRT item

- calibration and scoring in the presence of DIF. *Applied Psychological Measurement*, 40(8), 573-591.
- Choi, I. H., Paek, I., & Cho, S. J. (2017). The impact of various class-distinction features on model selection in the mixture Rasch model. *The Journal of Experimental Education*, 85(3), 411-424.
- Choi, Y. J. (2014). *Metric identification in mixture IRT models* (Doctoral dissertation, University of Georgia, GA). Retrived from <https://www.uga.edu/>
- Chon, K. H., Lee, W. C., & Dunbar, S. B. (2010). A comparison of item fit statistics for mixed IRT models. *Journal of Educational Measurement*, 47(3), 318-338.
- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis*. Hoboken, NJ: Wiley.
- Crosby, R. D., Wonderlich, S. A., Engel, S. G., Simonich, H., Smyth, J., & Mitchell, J. E. (2009). Daily mood patterns and bulimic behaviors in the natural environment. *Behaviour Research and Therapy*, 47(3), 181-188.
- De Ayala, R. J. (2009). *Theory and practice of item response theory*. New York: Guilford.
- De Ayala, R. J., & Santiago, S. Y. (2017). An introduction to mixture item response theory models. *Journal of School Psychology*, 60, 25-40.
- DeMars, C. (2001). Group differences based on IRT scores: Does the model matter?. *Educational and Psychological Measurement*, 61(1), 60-70.
- Finch, W. H., & French, B. F. (2012). Parameter estimation with mixture item response theory models: A Monte Carlo comparison of maximum likelihood and Bayesian methods. *Journal of Modern Applied Statistical Methods*, 11(1), 167-178.
- Finch, W. H., & Pierson, E. E. (2011). A mixture IRT analysis of risky youth behavior. *Frontiers in psychology*, 2, 1-10. <https://doi.org/10.3389/fpsyg.2011.00098>
- Granado, E. A. (2015). *Comparing Three Effect Sizes for Latent Class Analysis* (Doctoral dissertation, University of North Texas, TX). Available from ProQuest Dissertations and Theses database. (UMI No. 10075932)
- Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(2), 202-226.
- Kang, T., & Cohen, A. S. (2007). IRT model selection methods for dichotomous items. *Applied Psychological Measurement*, 31(4), 331-358.
- Lee, S. (2012). *The impact of missing data on the dichotomous mixture IRT models* (Doctoral dissertation, University of Georgia, GA). Retrived from <https://www.uga.edu/>
- Li, F., Cohen, A. S., Kim, S. H., & Cho, S. J. (2009). Model selection methods for mixture

- dichotomous IRT models. *Applied Psychological Measurement*, 33(5), 353-373.
- Lubke, G., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling*, 14(1), 26-47.
- Lubke, G. H., & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood. *Multivariate Behavioral Research*, 41(4), 499-532.
- Lubke, G., & Tueller, S. (2010). Latent class detection and class assignment: A comparison of the MAXEIG taxometric procedure and factor mixture modeling approaches. *Structural Equation Modeling*, 17(4), 605-628.
- Maij-de Meij, A. M., Kelderman, H., & van der Flier, H. (2010). Improvement in detection of differential item functioning using a mixture item response theory model. *Multivariate Behavioral Research*, 45(6), 975-999.
- Mislevy, R. J., & Verhelst, N. (1987). *Modeling item responses when different subjects employ different solution strategies* (RR-87-47-ONR). ETS Research Report Series.
- Muthén, B., & Asparouhov, T. (2006). Item response mixture modeling: Application to tobacco dependence criteria. *Addictive Behaviors*, 31(6), 1050-1066.
- Paek, I., & Cho, S. J. (2015). A note on parameter estimate comparability: across latent classes in mixture IRT modeling. *Applied Psychological Measurement*, 39(2), 135-143.
- Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology*, 32(1), 8-47.
- Peugh, J., & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(4), 616-639.
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14(3), 271-282.
- Rost, J. (1997). Logistic mixture models. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 449-463). New York: Springer-Verlag.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52(3), 333-343.
- Sen, S., Cohen, A. S., & Kim, S. H. (2016). The impact of non-normality on extraction of spurious latent classes in mixture IRT models. *Applied Psychological Measurement*, 40(2), 98-113.
- Sinharay, S., & Lu, Y. (2008). A further look at the correlation between item parameters and item fit statistics. *Journal of Educational Measurement*, 45(1), 1-15.

- Subedi, D. R. (2009). *Investigating unobserved heterogeneity using item response theory mixture models*. (Doctoral dissertation, Michigan State University, MI). Available from ProQuest Dissertations and Theses database. (UMI No. 3381368)
- Suh, Y., Cho, S. J., & Wollack, J. A. (2012). A comparison of item calibration procedures in the presence of test speededness. *Journal of Educational Measurement*, 49(3), 285-311.
- Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics-Theory and Methods* 7(1), 13-26.
- Tein, J. Y., Cox, S., & Cham, H. (2013). Statistical Power to detect the correct number of classes in latent class analysis. *Structural Equation Modeling*, 20(4), 640-657.
- Van Der Linden, W. J., & Hambleton, R. K. (1997). Item response theory: Brief history, common models, and extensions. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 1-28). New York: Springer-Verlag.
- Wu, Q. (2009). *Class extraction and classification accuracy in latent class models* (Doctoral dissertation, The Pennsylvania State University, PA). Retrieved from <https://etda.libraries.psu.edu/catalog>

© 논문접수: 2018. 8. 6 / 수정본 접수: 2018. 9. 9 / 게재승인: 2018. 9. 16

— 저 자 소 개 —

- 박산하 : 이화여자대학교에서 교육측정 및 평가 박사학위를 취득하였으며, 현재 이화여자대학교에서 시간강사로 재직 중임. 연구관심 분야는 문항반응이론, 검사 동등화, 표준화검사 제작 등임. fjvmru@hanmail.net
- 김성은 : 이화여자대학교에서 교육측정 및 평가 박사학위를 취득하였으며, 현재 한국청소년정책연구원에서 부연구위원으로 재직 중임. 연구관심 분야는 인지진단모형, 문항반응모형, 아동 및 청소년패널자료 분석 등임. miilli@nypi.re.kr

〈ABSTRACT〉

## Exploration of the Degree of Heterogeneity Between Latent Classes When Applying a Mixture IRT

San-Ha PARK

Ewha Womans University

Sung-Eun Kim

National Youth Policy Institute

The purpose of this study is to suggest a degree of heterogeneity between classes in order to utilize a model, after confirming the accuracy of latent class estimation, based on distance, when utilizing mixture IRT. The heterogeneity between the classes was measured by the distance between latent classes, which is not a factor being dealt within the existing mixture IRT research. In order to confirm this, a simulation study was used that allowed for the assumptions of the following varying conditions; numbers of classes (2class, 3class), sample sizes (1,000, 2,500), mixing proportions (equal, unequal) the difference of mean parameter distributions (0, 0.5, 1, 1.5), and the distance between the latent classes (2, 3, 4, 5). The accuracy of correct model selections based on the heterogeneity of latent classes and the accuracy of examinee classification, were confirmed by the simulation study. In conclusion The results show that there is a consistent tendency that if the Mahalanobis distance between latent classes is 3 or more, then the higher the accuracy of examinee classifications and correct model selections. This study provided information to researchers who intend to utilize mixture IRT; the information to predict the accuracy of latent class estimation and based on this, information about the degree of heterogeneity between classes required to utilize a model.

*Keywords : Mixture item response model, Latent class, Mahalanobis distance*

[부록 1] 잠재집단이 세 개인 경우 사용한 문항 모수

〈표〉 잠재집단이 세 개인 경우 자료 생성을 위해 사용한 문항 모수

문항	변별도			난이도											
	c1	c2	c3	MD2			MD3			MD4			MD5		
				c1	c2	c3	c1	c2	c3	c1	c2	c3	c1	c2	c3
1	1	1	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	1	1	1	0.0	0.0	0.0	2.0	1.0	1.0	2.0	1.0	1.0	2.0	0.0	2.0
3	1	1	1	0.0	0.0	0.0	1.5	0.5	0.5	1.5	0.5	0.5	1.5	0.5	-0.5
4	1	1	1	-1.0	-1.0	-1.0	1.0	0.0	0.0	1.0	0.0	-1.0	1.0	0.0	-1.0
5	1	1	1	-1.5	-0.5	-0.5	0.5	-1.5	-1.5	1.5	-1.5	-1.5	1.5	-1.5	-1.5
6	1	1	1	-1.0	0.0	0.0	-0.5	1.5	1.5	-0.5	1.5	1.5	-1.5	1.5	1.5
7	1	1	1	-0.5	0.0	0.0	-1.0	0.0	0.0	-1.0	0.0	1.0	-1.0	0.0	1.0
8	1	1	1	0.0	1.0	1.0	-1.5	-0.5	-0.5	-1.5	-0.5	-0.5	-1.5	-0.5	0.5
9	2	2	2	0.5	0.0	0.0	-2.0	-1.0	-1.0	-2.0	-1.0	-1.0	-2.0	0.0	-2.0
10	2	2	2	1.0	0.0	0.0	1.0	2.0	1.0	1.0	2.0	1.0	2.0	2.0	0.0
11	2	2	2	1.5	0.5	0.5	0.5	1.5	0.5	0.5	1.5	0.5	-0.5	1.5	0.5
12	2	2	2	-0.5	-1.5	-0.5	0.0	1.0	0.0	-1.0	1.0	0.0	-1.0	1.0	0.0
13	2	2	2	0.0	-1.0	0.0	-1.5	0.5	-1.5	-1.5	1.5	-1.5	-1.5	1.5	-1.5
14	2	2	2	0.0	-0.5	0.0	1.5	-0.5	1.5	1.5	-0.5	1.5	1.5	-1.5	1.5
15	2	2	2	1.0	0.0	1.0	0.0	-1.0	0.0	1.0	-1.0	0.0	1.0	-1.0	0.0
16	2	2	2	0.0	0.5	0.0	-0.5	-1.5	-0.5	-0.5	-1.5	-0.5	0.5	-1.5	-0.5
17	1	2	2	0.0	1.0	0.0	-1.0	-2.0	-1.0	-1.0	-2.0	-1.0	-2.0	-2.0	0.0
18	1	2	2	0.5	1.5	0.5	1.0	1.0	2.0	1.0	1.0	2.0	0.0	2.0	2.0
19	1	2	2	-0.5	-0.5	-1.5	0.5	0.5	1.5	0.5	0.5	1.5	0.5	-0.5	1.5
20	2	1	2	0.0	0.0	-1.0	0.0	0.0	1.0	0.0	-1.0	1.0	0.0	-1.0	1.0
21	2	1	2	0.0	0.0	-0.5	-1.5	-1.5	0.5	-1.5	-1.5	1.5	-1.5	-1.5	1.5
22	2	1	2	1.0	1.0	0.0	1.5	1.5	-0.5	1.5	1.5	-0.5	1.5	1.5	-1.5
23	2	2	1	0.0	0.0	0.5	0.0	0.0	-1.0	0.0	1.0	-1.0	0.0	1.0	-1.0
24	2	2	1	0.0	0.0	1.0	-0.5	-0.5	-1.5	-0.5	-0.5	-1.5	-0.5	0.5	-1.5
25	2	2	1	0.5	0.5	1.5	-1.0	-1.0	-2.0	-1.0	-1.0	-2.0	0.0	-2.0	-2.0
평균	1.6	1.6	1.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0



## [부록 2] 자료 생성을 위한 R코드 예시

<pre> class2&lt;-function(nSub,ratio,mean_dif,Ma.Distance){   MD&lt;-0   repeat{     b1&lt;-rnorm(25,0,1)     b2&lt;-rnorm(25,mean_dif,1)     a1&lt;-runif(25,1,2)     a2&lt;-runif(25,1,2)     para&lt;-c(a1,b1,a2,b2)     theta1&lt;-mat.or.vec(nSub*ratio,1)     theta2&lt;-mat.or.vec(nSub*(1-ratio),1)     theta1&lt;-rnorm(nSub*ratio,0,1)     theta2&lt;-rnorm(nSub*(1-ratio),mean_dif,1)     thetatem&lt;-c(theta1,theta2)     response&lt;-mat.or.vec(nSub,25+2)     for (row in 1:nSub) {       if (row&lt;=nSub*ratio) {         class&lt;-1         b&lt;-b1         a&lt;-a1         theta&lt;-thetatem[row]       }       else if (row&gt;nSub*ratio) {         class&lt;-2         b&lt;-b2         a&lt;-a2         theta&lt;-thetatem[row]       }       for (column in 1:25) {         p=exp(a[column]*(theta-b[column]))/(1+exp(a[column]           *(theta-b[column])))         r&lt;-runif(1,0,1)         if (r&lt;=p) {           response[row,column]&lt;-1 </pre>	<pre>         }         else {           response[row,column]&lt;-0         }       }       response[row,25+1]&lt;-class       response[row,25+2]&lt;-theta     }     ### Mahalanobis distance ###     class&lt;-response[,25+1]     theta&lt;-response[,25+2]     mat1&lt;-response[which(class==1),-c(25+1,25+2)]     mat2&lt;-response[which(class==2),-c(25+1,25+2)]     mat1.1&lt;-scale(mat1,center=T,scale=F)     mat2.1&lt;-scale(mat2,center=T,scale=F)     mat1.2&lt;-cov(mat1.1,method="pearson")     mat2.2&lt;-cov(mat2.1,method="pearson")     ##the pooled covariance##     n1&lt;-nrow(mat1)     n2&lt;-nrow(mat2)     n3&lt;-n1+n2     mat3&lt;-((n1/n3)*mat1.2) + ((n2/n3)*mat2.2)     mat4&lt;-solve(mat3)     mat5&lt;-as.matrix((colMeans(mat1)-colMeans(mat2)))     mat6&lt;-t(mat5)%*%mat4     MD&lt;-sqrt(mat6 %*% mat5)     print(MD)     if(round(MD,1)==Ma.Distance) break   }   return(list(response=response,para=para)) } </pre>
---	--