

구조방정식모형을 이용한 차별문항기능의 탐지: MACS와 MIMIC의 비교

윤 수 철*

서울대학교 언어교육원

이 순 목

성균관대학교 심리학과

구조방정식모형을 이용하여 차별문항기능을 탐지하는 방법으로 평균 및 공분산구조(MACS) 모형과 다지표-다원인(MIMIC) 모형의 두 가지를 들 수 있다. 두 모형은 모두 구조방정식모형의 특수한 경우에 해당하지만, 모형 간에 통계적 가정 및 자료의 입력 방식이 다르기 때문에 연구 상황에 따라 그 수행이 상이할 수 있다. 특히 MIMIC 모형은 MACS 모형에 비해 추가적인 가정을 필요로 하기 때문에, 가정이 위배될 경우 균일적 차별문항기능 탐지에 더 불리할 수 있다. 또한 MACS 모형과 달리 MIMIC 모형은 집단변수를 포함한 단일 입력자료를 사용하기 때문에, 집단 간에 표본크기가 상이할 경우 MIMIC 모형이 MACS 모형보다 균일적 차별문항기능 탐지에 더 유리할 것으로 예상할 수 있다. 이러한 가능성에 대해 여러 문헌에서 지적되었음에도 불구하고, 다양한 상황에서 두 모형의 차이를 체계적으로 비교한 연구는 발견되지 않았다. 따라서 다양한 연구 상황을 반영한 몬테 카를로 모의실험(Monte Carlo simulation)을 통해 균일적 차별문항기능 탐지에 대한 두 모형의 수행을 비교하였다. 구체적으로는 집단효과, 측정변수 신뢰도의 차이, 전체 표본크기, 표본크기의 비율, 차별문항기능의 크기, 차별문항기능 탐지 전략 등을 체계적으로 조작하였으며, 이에 따른 두 모형의 수행을 비교하였다. 비교 결과, 추가적인 가정이 위배되는 상황에서도 MIMIC 모형의 균일적 차별문항기능 탐지율이 MACS 모형에 비해 크게 저하되지 않았으며, 표본 크기가 집단 간에 다른 경우 MIMIC 모형이 MACS 모형보다 우수한 탐지율을 보였다. 단, 두 모형의 효과적인 사용을 위해서는 적절한 탐지 전략이 필요하므로, 이에 대해 논의한 후 현실적으로 바람직한 방안을 제안하였다.

주요어 : 차별문항기능, Differential Item Functioning, 구조방정식모형, MACS, MIMIC

* 교신저자: 윤수철, 서울대학교 언어교육원, (151-742) 서울시 관악구 관악로 1
Tel: 02-880-1408, E-mail: cinematopia@naver.com

어떤 검사를 통해 얻은 점수를 일반화하기 위해서는 측정 동일성(Measurement Equivalence/ Invariance; Vandenberg & Lance, 2000)의 성립이 필수적이다. 일반적으로 구성개념(construct)의 조작적 정의가 집단 간에 동일할 때(Little, 1997), 또는 구성개념의 점수가 동일한 사람들에게서 관찰점수 수준에서도 동일한 점수가 기대될 때(Drasgow, 1984; Raju, Raffitte, & Byrne, 2002) 측정 동일성이 성립하는 것으로 간주한다.

측정 동일성이 성립하지 않을 경우 동일한 검사를 사용하더라도 집단이나 상황에 따라 일관된 해석을 도출하기 어렵다. 예를 들어 어떠한 능력 검사가 특정 집단에 더 유리한 문항을 포함하고 있을 경우, 해당 집단의 응시자들은 실제 능력보다 더 높은 평가를 받게 될 것이다. 이 경우 검사 점수에는 측정하고자 했던 구성개념의 점수 뿐 아니라 집단의 특성에 따른 점수도 포함되며, 이는 검사의 타당도를 저해하는 직접적인 원인이 될 수 있다.

측정 동일성 연구는 구조방정식모형(Structural Equation Modeling) 또는 문항반응이론(Item Response Theory)의 틀 안에서 이루어져 왔다. 문항반응이론 기반의 연구에서는 전통적으로 문항 수준에서의 측정 동일성, 즉 차별문항기능(Differential Item Functioning: DIF)에 초점을 둔 반면, 구조방정식모형 기반의 연구들은 1990년대까지 주로 요인구조의 전반적인 동일성에 초점을 두고 이루어졌다. 그러나 2000년대부터는 구조방정식모형 기반의 차별문항기능 연구들이 급증하였으며, 문항반응이론과의 비교 연구도 활발하게 이루어졌다(예: Meade & Lautenschlager, 2004; Stark, Chernyshenko, & Drasgow, 2006).

차별문항기능의 정의

어떠한 검사를 통해 측정하고자 하는 구성개념의 점수가 동일한 응답자들이 각자 소속된 집단의 특성으로 인해 같은 문항에 다르게 반응하는 것을 차별문항기능이라고 한다(성태제, 1993; Holland & Thayer, 1993). 또한 차별기능(differential functioning)이 나타난 문항을 차별기능문항이라고 한다.

차별문항기능은 검사가 측정하고자 하는 구성개념과는 별개의 개념으로, 문항에 차별기능이 존재한다는 것은 해당 문항의 내용에 대하여 집단별로 다르게 반응할 만한 요소가 포함되어 있음을 의미한다. 이러한 요소는 성별, 훈련 여부, 문화, 가치관 등 다양한 측면에서 나타날 수 있다. 따라서 차별문항기능을 검사의 사용 목적에 따라 특정 집단의 유/불리와 직결된 부정적 의미로 받아들일 수도 있지만, 검사와 각 문항 간의 논리적 관계를 정교화할 수 있는 단서로 받아들일 수도 있다. 이러한 맥락에서, 차별문항기능을 검사가 측정하고자 하는 구성개념이 각 문항에 잘 반영되어 있는지, 그렇지 않다면 왜 그런지에 대한 구성개념 타당화(construct validation)의 한 과정으로 보는 관점도 제시되었다(김성훈, 1995; Smith, 2002).

차별문항기능과 분명히 구분되어야 하는 개념으로 집단효과(impact)를 들 수 있다. 집단효과는 검사가 측정하고자 하는 구성개념, 즉 이론변수의 점수가 집단 간에 상이함을 의미한다(Ackerman, 1992; Millsap & Everson, 1993). 공통요인모형에서는 집단 간 요인평균의 차이를 집단효과로 간주한다. 이 때 기준이 되는 집단을 참조집단(referent group), 비교 연구되는 집단을 관심집단(focal group)으로 정의한다.

차별문항기능의 분류

차별문항기능은 크게 균일적 차별문항기능(uniform DIF)과 비균일적 차별문항기능(nonuniform DIF)으로 구분할 수 있다(Mellenbergh, 1982). 균일적 차별문항기능은 문항의 곤란도 모수(difficulty parameter)가 집단에 따라 다른 경우로 정의되며, 이론변수 값의 모든 범위에서 한 집단의 문항 평균이 다른 집단보다 일관되게 높은 것을 의미한다. 반면 비균일적 차별문항기능은 문항의 곤란도 모수와는 관계없이 변별도 모수(discrimination parameter)가 집단에 따라 다른 경우로 정의되며, 이론변수의 일부 구간에서는 한 집단의 문항 평균이 다른 집단보다 높지만 다른 구간에서는 그 관계가 역전되는 것을 의미한다. 즉, 문항 점수와 이론변수 간에 선형 관계를 가정한다면, 균일적 차별문항기능은 집단 간에 절편이 다른 경우이며, 비균일적 차별문항기능은 절편과 관계없이 집단 간에 기울기가 다른 경우이다. 이를 도식으로 표시하면 그림

1과 같다.

차별문항기능의 탐지 방법

집단효과를 통제하기 위해 사용하는 변수를 대응변수(matching variable)라고 한다. 이 대응변수를 사용하는 논리에 따라 차별문항기능을 탐지하는 방식을 크게 측정변수 방식과 잠재변수 방식으로 구분할 수 있다(Millsap & Everson, 1993).

측정변수 방식으로 대표적인 것은 카이제곱(χ^2) 방법(Camilli, 1979), 맨텔-헨젤(Mantel-Haenszel) 방법(Holland & Thayer, 1988), 로지스틱 판별함수 방법(Miller & Spray, 1993) 등이 있다. 이러한 방법들은 상대적으로 사용이 간편하고 효율적이기 때문에 현재까지도 널리 사용되고 있으나, 대응변수로 사용되는 검사 총점에 측정오차(measurement error)가 포함되어 있다는 점에서 한계가 있다(Meredith & Millsap, 1992; Millsap & Everson, 1993). 특히 검사의 길이가 짧은 경우 적은 문항 수로 인해 신뢰도

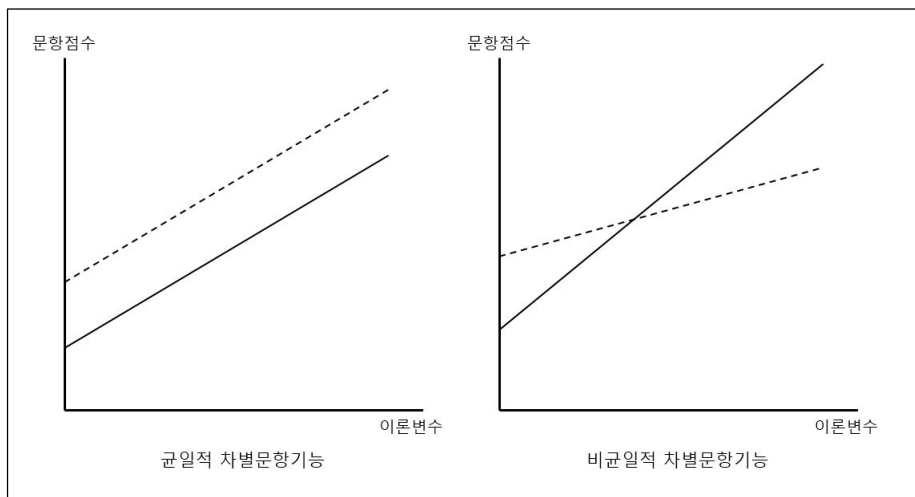


그림 1. 균일적 차별문항기능과 비균일적 차별문항기능

가 감소하고, 검사 총점의 측정오차가 증가하는 문제가 두드러진다. 이러한 경우 측정오차를 통제하는 잠재변수 방식을 사용하는 것이 더 적절하다(Gelin & Zumbo, 2007).

잠재변수 방식은 전통적으로 문항반응이론의 관점에서 발전해 왔으며, 그 중 Lord(1980)의 χ^2 방법, 우도비(Likelihood Ratio) 검증 방법(Thissen, Steinberg, & Wainer, 1988), SIBTEST 방법(Shealy & Stout, 1993) 등이 대표적이다. 문항반응이론에 기초한 방법과 구조방정식모형에 기초한 방법은 기본적으로 동일한 논리를 사용하지만, 문항반응이론에서는 잠재변수와 문항 간에 비선형 관계를, 구조방정식모형에서는 선형 관계를 가정한다는 점이 다르다. 또한 문항반응이론에서는 잠재변수의 분포를 추정해야 하기 때문에 정확한 추정을 위해 보다 긴 검사(30문항 이상)가 필요하다(Gelin & Zumbo, 2007). 심리학에서는 하나의 잠재변수에 대해 30문항 이하의 짧은 검사가 많이 사용되기 때문에, 이 경우 구조방정식모형을 사용하는 것이 더 적절하다(Gelin & Zumbo, 2007).

구조방정식모형에 기초한 차별문항기능 탐지 방법은 2000년대부터 본격적으로 연구되었다. 전통적으로 구조방정식모형의 관점에서는 ‘차별문항기능’이라는 용어보다는 부분측정동일성(partial measurement invariance)이라는 용어가 먼저 사용되었다. 부분측정동일성은 일부 문항(측정변수)으로 인해 전반적인 측정 동일성이 성립하지 않는 것을 의미하는 것으로, 차별문항기능과는 달리 요인구조의 동일성에 더 무게중심을 둔 개념이다. 측정변수가 문항일 경우 이것은 차별문항기능과 동일한 개념으로, 2000년대 초반부터 구조방정식모형 연구에서도 차별문항기능이라는 용어가 사용되었

다(예: Chan, 2000).

구조방정식모형의 관점에서 차별문항기능을 탐지하는 방법으로 평균 및 공분산구조(Mean And Covariance Structure: MACS; Sörbom, 1974; Vandenberg & Lance, 2000)모형과 다지표-다원인(Multiple Indicators Multiple Causes: MIMIC; Jöreskog & Goldberger, 1975; Muthén, 1979, 1983)모형의 두 가지를 들 수 있다. 두 모형 모두 구조방정식모형의 특수한 경우로, 다집단 상황에서의 논리를 서로 다른 방식으로 구현한 것이다.

평균 및 공분산구조 모형

평균 및 공분산구조 모형(이하 MACS 모형)은 이론모형(structural model)을 포함할 수도 있으나, 차별문항기능 탐지 시에는 측정모형(measurement model)에 대한 것이므로 복수의 집단에 대한 확인적 요인분석(Confirmatory Factor Analysis)을 의미한다. 다집단 상황에서의 측정모형은 식 (1)과 같이 표현되며, 양변에 공분산을 취해 식 (2)와 같은 공분산구조를, 평균을 취해 식 (3)과 같은 평균구조를 얻을 수 있다(이순목, 금은희, 이찬순, 2010).

$$y^{(g)} = \tau^{(g)} + \Lambda^{(g)}F^{(g)} + u^{(g)} \quad (1)$$

$$\Sigma^{(g)} = \Lambda^{(g)}\Phi^{(g)}\Lambda'^{(g)} + \Theta^{(g)} \quad (2)$$

$$\bar{y}^{(g)} = \tau^{(g)} + \Lambda^{(g)}\kappa^{(g)} \quad (3)$$

여기서 g = 집단

y = 측정변수 벡터

τ = 측정변수의 절편 벡터

Λ = 요인계수 행렬

F = 이론변수 벡터

- u = 측정오차 벡터
- Σ = 측정변수의 분산/공분산 행렬
- Φ = 이론변수의 분산/공분산 행렬
- Θ = 측정오차의 분산/공분산 행렬
- \bar{y} = 측정변수의 평균 벡터
- κ = 이론변수의 평균 벡터

MACS모형을 이용한 측정 동일성 검증은 여러 단계를 거치는데, 일부 절차를 제외하면 어느 정도 합의가 이루어져 있다. 일반적인 절차를 정리하면 표 1과 같다(Meredith, 1993; Schmitt & Kuljanin, 2006; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000).

표 1의 절차는 전반적인 측정 동일성에 대한 것으로, Byrne, Shavelson, & Muthén(1989)은 현실적으로 모든 측정변수에 대해 측정 동일성이 성립하거나 성립하지 않는 경우는 많지

않다고 주장하였다. 이를 시작으로 부분 측정 동일성(partial measurement invariance)의 개념이 등장하였다.

그런데 일부 측정변수에 대해 측정 동일성이 성립하지 않는 경우, 이는 부분 측정 동일성인 동시에 차별문화기능이 된다. 이러한 맥락에서 Chan(2000)은 부분 측정 동일성을 검증할 때 참조되는 수정지수(Modification Index: MI)를 이용해 차별문화기능을 탐지하는 방법을 소개하였는데, 이후의 연구에서는 부분 측정 동일성이라는 용어 대신 차별문화기능이라는 용어가 자주 사용되었다(예: Lee, 2009; Meade & Lautenschlager, 2004; Stark et al., 2006).

그림 1에서 볼 수 있듯, 균일적 차별문화기능은 절편에서의 차이로, 비균일적 차별문화기능은 기울기에서의 차이로 해석된다. 따라서 MACS모형의 경우 균일적 차별문화기능은

표 1. MACS를 이용한 측정 동일성 검증 절차

단계	검증 대상	설명
1	공분산행렬 동일성(the equality of covariance matrix)	표본공분산행렬이 집단 간에 동일함. 이 단계의 필요성에 대해서는 논란의 여지가 남아 있음.
2	측정틀 동일성(configural invariance)	요인에 대한 지표(indicator)들이 집단 간에 동일함.
3	측정단위 동일성(metric invariance)	각 요인계수(factor loading)가 집단 간에 동일함. 약한 요인 동일성(weak factorial invariance).
4	측정원점 동일성(scalar invariance)	각 측정원점(intercept 또는 scalar)이 집단 간에 동일함. 강한 요인 동일성(strong factorial invariance).
5	측정오차분산 동일성(uniqueness invariance)	각 측정오차분산(measurement error variance)이 집단 간에 동일함. 엄격한 요인 동일성(strict factorial invariance).
6	요인분산 동일성(factor variance invariance)	각 요인분산(factor variance)이 집단 간에 동일함.
7	요인공분산 동일성(factor covariance invariance)	각 요인공분산(factor covariance)이 집단 간에 동일함.
8	요인평균 동일성(factor mean invariance)	각 요인평균(factor mean)이 집단 간에 동일함.

부분 측정원점 동일성에, 비균일적 차별문항기능은 부분 측정단위 동일성에 해당한다(Chan, 2000). 또한 집단 간 요인평균의 차이는 집단효과(impact)에 해당한다.

에서는 여기에 이론변수의 예측변수 G 를 포함할 수 있다. 이러한 관계는 식 (4)를 통해 표현할 수 있다(Jöreskog & Goldberger, 1975; Muthén, 1983).

다지표-다원인 모형

$$y = \Lambda F + \beta' G + u \quad (4)$$

$$F = \alpha + \gamma' G + \zeta \quad (5)$$

다지표-다원인 모형(이하 MIMIC 모형)은 구조방정식모형의 특수한 경우로, 확인적 요인 분석 모형의 이론변수를 다른 측정변수를 통해 설명하는 형태를 가지고 있다. MIMIC 모형에서 측정변수 중 일부는 이론변수를 반영하는 지표(indicator)의 역할을 수행하지만 다른 일부는 이론변수의 원인 또는 선행변수(predictor)가 될 수 있다. 예를 들어 ‘가정 불화’, ‘재정적 문제’, ‘도박장과의 거리’라는 세 측정변수는 모두 ‘도박중독’이라는 이론변수와 함수관계에 있으나, ‘가정 불화’와 ‘재정적 문제’는 ‘도박중독’을 나타내는 지표로 볼 수 있는 반면 ‘도박장과의 거리’는 도박중독의 예측변수로 간주하는 것이 타당하다. 이러한 MIMIC 모형의 도식은 그림 2와 같다.

- 여기서 y = 측정변수 벡터
- Λ = 요인계수 행렬
- F = 이론변수 벡터
- β = 측정변수 y 에 대한 예측변수 G 의 계수 행렬
- G = F 에 대한 예측변수 벡터
- u = 측정오차 벡터
- α = 이론모형의 절편 벡터 (그림에서는 생략)
- γ = F 에 대한 G 의 계수 행렬
- ζ = 설명오차(disturbance) 벡터

그림 2에서 볼 수 있듯이, 이론변수 F 와 측정변수 $y_i (i = 1, 2, \dots, p)$ 간의 관계는 통상적인 확인적 요인분석 모형이다. MIMIC 모형

식 (4)에서 G 는 이산변수 또는 연속변수이며, 여러 개일 수도 있다. G 가 ‘0’과 ‘1’의 값을 가지며 두 개의 집단을 나타내는 가변수(dummy variable)일 경우, γ 는 두 집단 간의 이론변수 평균 차이(집단효과)가 된다. 만약 β 가

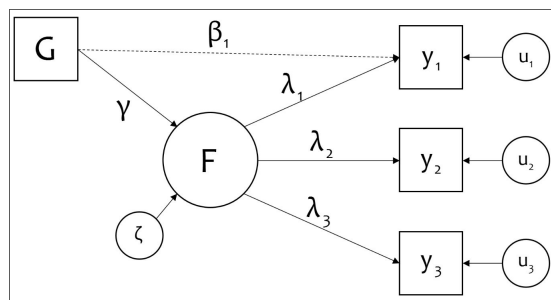


그림 2. MIMIC 모형의 도식적 표현

통계적으로 유의하다면(그림 2 참조), 대응변수의 역할을 하는 F 의 효과를 통제된 상태에서 측정변수 y 에 대해 집단 G 의 직접효과가 존재하는 것이므로 균일적 차별문항기능의 정의와 일치한다. G 가 가변수일 때 β 는 두 집단 간 절편의 차이를 의미하기 때문이다.

MIMIC 모형에서는 기본적으로 비균일적 차별문항기능을 탐지하기 어렵다. 예측변수 G 를 이론변수로 간주하고 이론변수 F 와의 상호작용 항을 통해 비균일적 차별문항기능을 탐지하는 절차(Restricted Factor Analysis; RFA)가 개발되었으나(Oort, 1998; Barendse, Oort, & Garst, 2010), 아직 충분히 연구되지는 않았다.

MACS 모형과 MIMIC 모형의 비교

균일적 차별문항기능 탐지 시 MACS 모형과 MIMIC 모형은 본질적으로 동일한 논리를 서로 다른 방식으로 전개한 것으로 볼 수 있다. MACS 모형에서 식 (3)을 이용하면 참조집단($g=0$)과 관심집단($g=1$) 각각의 측정변수 평균을 다음과 같이 표현할 수 있다.

$$\begin{aligned} \bar{y}_1^{(0)} &= \tau_1^{(0)} + \lambda_1^{(0)} \kappa^{(0)} && \text{(참조집단의 평균)} \\ \bar{y}_1^{(1)} &= \tau_1^{(1)} + \lambda_1^{(1)} \kappa^{(1)} && \text{(관심집단의 평균)} \end{aligned}$$

MACS 모형에서 균일적 차별문항기능은 부분 측정원점 동일성에 해당하기 때문에, 측정단위 동일성은 이미 성립한 것으로 본다. 따라서 요인계수는 두 집단에서 동일하고 ($\lambda_1^{(0)} = \lambda_1^{(1)} = \lambda_1$), 두 집단의 측정변수 평균 차이는 식 (6)과 같이 표현된다.

$$\bar{y}_1^{(1)} - \bar{y}_1^{(0)} = \tau_1^{(1)} - \tau_1^{(0)} + \lambda_1 (\kappa^{(1)} - \kappa^{(0)}) \quad (6)$$

즉 측정변수의 평균 차이를 측정원점의 차이에 의한 부분($\tau^{(1)} - \tau^{(0)}$)과 요인평균 차이에 의한 부분($\kappa^{(1)} - \kappa^{(0)}$)으로 분할할 수 있다. 이 때 모형 식별(identification)과정에서 참조집단의 요인평균 $\kappa^{(0)}$ 는 0으로 고정되기 때문에 실질적으로 $\kappa^{(1)}$ 은 요인평균 차이, 즉 집단효과에 해당한다. 측정원점의 차이와 집단효과를 각각 τ^* , κ^* 로 축약하면 식 (7)과 같이 표현할 수 있다.

$$\bar{y}_1^{(1)} - \bar{y}_1^{(0)} = \tau_1^* + \lambda_1 \kappa^* \quad (7)$$

한편 MIMIC 모형의 방정식에서 식 (5)의 F 를 식 (4)에 대입하여 전개하면 하나의 측정변수 y_1 에 대한 식을 다음과 같이 표현할 수 있다.

$$y_1 = \lambda_1 \alpha + \lambda_1 \gamma G + \lambda_1 \zeta + \beta_1 G + u_1 \quad (8)$$

G 가 두 집단을 나타내는 가변수이므로 참조집단의 경우 $G=0$, 관심집단의 경우 $G=1$ 을 대입하면 각 집단에서의 측정변수 평균을 다음과 같이 표현할 수 있다.

$$\begin{aligned} \bar{y}_1^{(0)} &= \lambda_1 \alpha && \text{(참조집단의 평균)} \\ \bar{y}_1^{(1)} &= \lambda_1 \alpha + \lambda_1 \gamma + \beta_1 && \text{(관심집단의 평균)} \end{aligned}$$

따라서 두 집단의 측정변수 평균 차이는 식 (9)와 같다.

$$\begin{aligned} \bar{y}_1^{(1)} - \bar{y}_1^{(0)} &= \lambda_1 \gamma + \beta_1 \\ &= \lambda_1 \kappa^* + \tau_1^* \end{aligned} \quad (9)$$

식 (9)에서 γ 는 요인평균의 차이로 식 (7)의 κ^* 에 해당한다. 또한 β_1 은 균일적 차별문항 기능을 의미하는 τ^* 에 대응된다. 따라서 식 (7)과 식 (9)는 동일해지며, 가정의 차이가 있을 뿐 MACS 모형과 MIMIC 모형에서 균일적 차별문항기능을 탐지하는 논리는 근본적으로 동일하다고 볼 수 있다.

연구문제

앞서 언급한 것처럼 MACS 모형과 MIMIC 모형은 균일적 차별문항기능 탐지에 있어 동일한 논리에 기반한 것으로 볼 수 있으나, 두 가지 면에서 중요한 차이를 보인다.

첫째, 표 1에 제시된 것처럼 MACS 모형에서는 요인계수, 측정오차분산, 요인분산 모두 동일성 검증의 대상이 된다. 즉 위의 동일성이 성립하지 않을 경우 이들의 동일화제약(equality constraint)을 해제하고 검증을 진행하는 것이 가능하다. 그러나 MIMIC 모형에서는 요인계수, 측정오차분산, 요인분산 모두 집단 간에 공통적으로 추정되기 때문에, 각각의 집단 간 동일성이 모형의 가정(assumption)이 된다. 요인계수 동일성(측정단위 동일성)의 경우 MACS 모형에서도 균일적 차별문항기능 탐지 이전에 성립해야 하는 조건이지만, 측정오차분산 동일성과 요인분산 동일성의 경우 MACS 모형에서는 성립하지 않아도 무방하다. 따라서 균일적 차별문항기능의 탐지 시 이 두 가정의 위배가 MIMIC 모형에 대해 상대적으로 더 큰 영향을 미칠 수 있다. 이에 대해서는 몇몇 문헌에서 지적한 바 있으나(예: Brown, 2006, Hancock, Lawrence, & Nevitt, 2000), 아직 체계적으로 검토한 연구는 발견되지 않았다.

이러한 부정적 영향은 집단효과와 존재 여

부에 따라 달라질 것으로 보인다. 식 (2)와 같이 요인분산과 측정오차분산은 요인계수의 추정에 영향을 미치고, 이 요인계수는 식 (7) 또는 식 (9)에서와 같이 집단효과와 결합되어 균일적 차별문항기능(τ_1^* 또는 β_1)의 탐지에 영향을 미친다. 그러나 집단효과가 0일 경우에는 식 (7)과 (9)에서 ‘측정변수 평균 차이=절편 차이’로 나타나기 때문에, 요인분산 동일성 또는 측정오차분산 동일성 위배로 인해 요인계수의 추정이 왜곡되더라도 균일적 차별문항기능 탐지에 큰 영향을 미치지 않을 것이다. 즉 집단 간 표본크기가 동일한 상황에서 균일적 차별문항기능을 탐지할 때, 집단효과가 존재할 경우에는 가정 위배로 인해 MIMIC 모형의 수행이 더 크게 저하되지만, 집단효과가 존재하지 않을 경우에는 가정이 위배되더라도 MIMIC 모형과 MACS 모형의 수행이 유사할 것으로 예상된다.

여기서 요인분산 또는 측정오차분산과 관련해 한 가지 언급할 필요가 있다. 식 (2)를 통해 도출할 수 있듯이, 1요인 측정모형에서 i 번째 측정변수의 분산은 $\sigma_{ii} = \lambda_i\phi + \theta_i$ 로 표현할 수 있다. 따라서 요인계수와 요인분산, 측정오차분산은 불가분의 관계이며, 요인계수와 요인분산이 동일한 경우 측정오차분산의 차이는 측정변수의 신뢰도¹⁾의 차이로 나타난다. 이러한 맥락에서 본 연구에서는 세 가지 가정 중 측정오차분산의 동일성이 위배되는 경우로 그 범위를 제한하고, 이후에는 측정오

1) 측정변수의 신뢰도는 탐색적 요인분석에서의 공통분(communality) 또는 확인적 요인분석에서의 측정변수 R^2 와 동일한 개념으로 볼 수 있으며,

1요인 모형에서는 $\rho = \frac{\lambda^2\phi}{\lambda^2\phi + \theta}$ 로 정의한다

(Jöreskog, 1971; Rock, Werts, & Flaugher, 1978).

차분산의 차이 대신 측정변수 신뢰도의 차이로 기술하겠다. 현실적으로 측정오차분산의 차이보다는 측정변수 신뢰도의 차이로 해석하는 것이 더 유용하기 때문이다.

둘째, MACS 모형에서는 집단별 공분산행렬을 사용하지만(식 (2) 참조), MIMIC 모형에서는 집단변수를 포함한 단일 공분산행렬을 사용한다. 측정 동일성 연구에서 집단 간에 표본 크기가 다를 경우 통계적 검정력이 감소하는 경향이 나타나는데(Hancock et al., 2000; Kaplan & George, 1995; Lubke & Dolan, 2003), 집단 간에 표본 크기의 차이가 클 경우 집단별 공분산행렬을 사용하는 MACS 모형의 균일적 차별문항기능 탐지율이 상대적으로 더 크게 감소할 것으로 기대된다. 반면 MIMIC 모형은 자료를 나누지 않고 전체 표본을 사용하기 때문에, 집단 간 표본크기의 차이가 크더라도 MACS 모형에 비해 균일적 차별문항기능 탐지율이 덜 감소할 것으로 기대할 수 있다.

실제 자료에 대해 MACS 모형이나 MIMIC 모형을 이용하여 균일적 차별문항기능을 탐지할 경우, 연구 상황에 따라 적절한 모형을 사용하려면 여러 조건에 따른 두 모형의 특징을 명확하게 파악해 둘 필요가 있다. 특히 가정의 위배는 현실 자료에서 빈번하게 발생하므로, 가정 위배 시 두 모형의 수행 저하를 체계적으로 비교할 필요가 있다. 또한 표본크기와 같이 현실적인 제약이 작용하는 요인에 대해 MACS 모형(집단별 자료 사용)과 MIMIC 모형(전체 자료 사용)의 한계와 두 모형의 차이를 규명할 필요가 있다. 이러한 맥락에서, 본 연구에서는 균일적 차별문항기능 탐지 시 집단효과의 존재 유무, 측정변수 신뢰도의 차이, 표본 크기의 차이에 따라 MACS 모형과 MIMIC 모형의 수행이 어떻게 달라지는지를

몬테-카를로 모의실험(Monte Carlo simulation) 연구를 통해 관찰하고자 하였다.

방 법

연구 설계

균일적 차별문항기능 탐지 시 다양한 조건에 따른 MACS 모형과 MIMIC 모형의 수행 차이를 관찰하기 위해 몬테-카를로 모의실험을 실시하였다. 모의실험은 $3 \times 3 \times 3 \times 4 \times 4 \times (2) \times (2)$ 설계로 구성되었으며, 집단간 요인으로 집단효과의 크기(3수준), 측정변수 신뢰도의 차이(3수준), 집단 간 표본크기의 비율(3수준), 전체 표본크기(4수준), 균일적 차별문항기능의 크기(4수준)가 설정되었으며, 집단내 요인으로 모형(2수준)과 탐지 전략(2수준)이 설정되었다. 여기서 전체 표본크기와 균일적 차별문항기능의 크기는 이미 그 효과가 알려져 있으므로(예: Barendse et al., 2010; Stark et al., 2006) 구획변수(block variable)로서 설정되었다. 종속변수는 정확 탐지율(True Detection Rate)과 오류 탐지율(False Detection Rate)이 설정되었다. 전체 설계에 대한 요약은 표 2에 제시되어 있다.

연구도구

자료 생성을 위해 6개의 측정변수로 이루어진 1요인 측정모형을 설정하였으며, 총 6개의 측정변수 중 마지막 1개의 측정변수에만 균일적 차별문항기능이 존재하는 상황을 가정하였다. 두 집단 비교의 기준인 참조집단의 모형 설정 시 요인계수는 평균이 0.8이 되도록, 측정원점은 평균이 2.0이 되도록 설정하였으며

표 2. 연구설계 요약

집단 효과	관심집단의 측정변수 신뢰도	전체 표본크기	집단 간 표본크기의 비율													
			1:1			3:2			4:1							
			차별문항기능의 크기	차별문항기능의 크기	차별문항기능의 크기	없음	작음	중간	큼	없음	작음	중간	큼			
없음 ($d=0$)	높음	100														
		300														
		500														
		1,000														
	중간	100														
		300														
		500														
		1,000														
	낮음	100														
		300														
		500														
		1,000														
작은 집단효과 ($d=-0.2$)	높음	100														
		300														
		500														
		1,000														
	중간	100														
		300														
		500														
		1,000														
	낮음	100														
		300														
		500														
		1,000														
큰 집단효과 ($d=-0.8$)	높음	100														
		300														
		500														
		1,000														
	중간	100														
		300														
		500														
		1,000														
	낮음	100														
		300														
		500														
		1,000														

집단내 요인(4수준):

모형(MACS/MIMIC) × 차별기능 탐지전략(자유-기저/제약-기저)

주. 연구요인은 음영으로 표시하였음. 표시하지 않은 요인은 구획변수임.

표 3. 자료생성에 사용할 참조집단의 요인모형

요인	측정변수	요인계수(λ)	측정원점(τ)	측정오차분산(θ)	측정변수 신뢰도
F ($\phi = 1.0$)	y_1	$\lambda_1 = 1.0$	$\tau_1 = 2.0$	$\theta_1 = 0.43$	0.7
	y_2	$\lambda_2 = 0.6$	$\tau_2 = 1.5$	$\theta_2 = 1.44$	0.2
	y_3	$\lambda_3 = 0.8$	$\tau_3 = 2.5$	$\theta_3 = 0.64$	0.5
	y_4	$\lambda_4 = 0.6$	$\tau_4 = 1.0$	$\theta_4 = 0.84$	0.3
	y_5	$\lambda_5 = 1.0$	$\tau_5 = 3.0$	$\theta_5 = 1.00$	0.5
	y_6	$\lambda_6 = 0.8$	$\tau_6 = 2.0$	$\theta_6 = 0.16$	0.8

요인분산은 1로 설정하였다. 측정오차분산은 6개 측정변수의 신뢰도 평균이 중간 크기(0.5)가 되도록 설정하였다. 관심집단은 각 실험조건에 따라 해당되는 모수를 다르게 설정하였다. 참조집단의 요인모형에 설정된 모수는 표 3에 제시되어 있다.

독립변수

집단효과 크기. 식 (7), 식 (9)와 같이 집단 간에 요인계수가 동일할 경우 측정변수의 평균 차이는 집단효과와 균일적 차별문항기능의 조합에 의해 나타난다. 따라서 각 모형이 집단효과와 균일적 차별문항기능을 얼마나 잘 분리해 내는지를 관찰하기 위해 집단효과 크기를 체계적으로 조작하였다. 이를 위해 Cohen(1988)의 $d = \frac{\mu_1 - \mu_2}{\sigma}$ 를 이용하여 집단 효과가 존재하지 않는 경우, 집단효과가 작은 경우($d = 0.2$), 집단효과가 큰 경우($d = 0.8$)를 설정하였다. 즉 참조집단의 요인평균을 0으로, 관심집단의 요인평균을 각각 $\kappa^{(1)} = 0$, $\kappa^{(1)} = -0.2$, $\kappa^{(1)} = -0.8$ 로 설정하였다.

측정변수 신뢰도의 차이. 앞에서 언급한

것처럼, 측정오차분산의 차이를 통해 측정변수의 신뢰도를 조작하였다. 측정변수의 분산 중 측정오차분산의 비율은 “1-신뢰도”의 값을 의미하므로, 요인계수와 요인분산을 고정된 상태에서 측정변수의 신뢰도 차이를 조작하는 것은 측정오차분산의 차이를 조작하는 것과 동일한 의미이다.

일반적으로 0.2~0.4는 작은 공통분, 0.6~0.8은 큰 공통분(Hogarty, Hines, Kromrey, Ferron, & Mumford, 2005; MacCallum, Widaman, Zhang, & Hong, 1999; Pennell, 1968)으로 간주되는데, 표 3에서 참조집단의 6번째 측정변수 신뢰도는 0.8로 큰 공통분에 해당한다. 이것을 기준으로 관심집단의 6번째 측정변수 신뢰도를 0.8(높은 신뢰도, 참조집단과 동일), 0.5(중간 신뢰도), 0.3(낮은 신뢰도)의 세 수준으로 조작하였다. 그 결과 관심집단에서 6번째 측정변수의 측정오차분산이 각각 0.16(높은 신뢰도), 0.64(중간 신뢰도), 1.493(낮은 신뢰도)으로 설정되었다. 이에 대한 계산 과정은 부록 I에 제시하였다.

집단 간 표본크기의 비율. 앞에서 언급한 것처럼 균일적 차별문항기능 탐지 시 집단 간

표본크기의 차이에 따라 두 모형의 수행에 차이가 나타날 가능성이 높기 때문에, 이를 체계적으로 조작하였다. 이 때 표본크기 비율의 효과와 전체 표본크기의 효과가 혼입되는 것을 방지하기 위해(Kaplan & George, 1995), 전체 표본크기 100, 300, 500, 1000 각각에 대해 참조집단과 관심집단의 표본크기 비율을 1:1, 3:2, 4:1로 교차시켰다. 여기서 3:2(1.5배)와 4:1(4배)의 비율은 표본크기의 차이가 눈에 띄게 크지 않은 경우와 상당히 큰 경우를 반영한 것이다. 결과적으로 참조집단과 관심집단의 표본크기는 각각 50/50, 60/40, 80/20, 150/150, 180/120, 240/60, 250/250, 300/200, 400/100, 500/500, 600/400, 800/200의 12개 조건으로 설정되었다. 이 중 가장 작은 표본크기인 20은 확인적 요인분석에서 합리적인 하한선으로 간주되는 25(Bentler & Bonett, 1980; Marsh, Balla, & McDonald, 1988)보다는 다소 작지만, 관심집단이 소수집단일 경우 현실적으로 관찰 가능한 크기이다.

전체 표본크기. 전통적으로 표본크기는 차별분항기능 연구에서 가장 중요한 변수 중 하나이다. 다만 표본크기가 클수록 차별분항기능 탐지에 더 유리한 것은 자명한 사실이기 때문에 관심의 대상이 아니라 구획변수로서 설계에 포함되었다. 일반적으로 구조방정식모형에서 요구되는 표본크기는 측정변수의 수가 12개보다 작으면 최소 200, 12개 이상일 경우 최소 $1.5p(p+1)^2$ 이상이 되어야 한다(Jöreskog & Sörbom, 1989)고 알려져 있다. 그러나 현실적으로 심리학 연구에서 표본크기가 작은 경우가 많으므로, 현실을 반영하기 위해

2) p 는 측정변수의 수

200 미만의 표본크기를 고려할 필요가 있다. 기존의 연구에서는 집단별 표본크기를 100 이상으로 설정하는 경우가 많았으나(예: Barendse et al., 2010; Stark et al., 2006), 심리학 분야에서 100 미만의 표본 크기는 특별히 중요하기 때문에(Paxton, Curran, Bollen, Kirby, & Chen, 2001) 각 집단의 표본크기가 100 미만인 경우를 고려할 필요가 있다. 따라서 두 집단 전체의 표본크기를 100, 300, 500, 1000의 네 수준으로 설정하였다.

균일적 차별분항기능의 크기. 균일적 차별기능 역시 그 크기가 클수록 차별분항기능 탐지에 더 유리한 것은 자명한 사실이므로 구획변수로서 포함되었다. 집단효과와 마찬가지로 균일적 차별분항기능의 크기 역시 Cohen (1988)의 d 를 이용해 설정하였다. d 는 측정원점의 차이를 측정변수의 표준편차인 $\sqrt{var(y)} = \sqrt{\lambda^2\phi + \theta}$ 로 나누어 구할 수 있는데, 참조집단에서 6번째 측정변수의 측정원점이 2.0이므로 관심집단의 측정원점을 각각 동일기능(no DIF, $d = 0$), 작은 차별기능(small DIF, $d = 0.2$), 중간 차별기능(media DIF, $d = 0.5$), 큰 차별기능(large DIF, $d = 0.8$)에 해당하도록 설정하였다. 그런데 측정변수 신뢰도와 표본크기가 조건에 따라 다르게 설정되었으므로, 각 측정오차분산 및 표본 크기의 조합에 따라 통합표준편차(pooled SD)를 이용한 d 를 계산해 측정원점 차이를 조작하였다. 그 결과 관심집단에서 6번째 측정변수의 측정원점은 각 조건의 측정오차분산과 표본 크기에 따라 0.87에서 2.0까지 다양한 값으로 설정하였다. 조건에 따른 계산 결과는 부록 II에 제시하였다.

모형. 균일적 차별문항기능 탐지를 위해 MACS 모형과 MIMIC 모형의 두 가지가 사용되었다.

차별기능 탐지 전략. 일반적으로 차별문항기능을 탐지 시 $\Delta\chi^2$ 검증을 사용하는데, 이때의 기저(baseline)모형을 어떻게 설정하느냐에 따라 크게 제약-기저(constrained-baseline) 전략과 자유-기저(free-baseline) 전략으로 구분할 수 있다. 기저모형은 $\Delta\chi^2$ 검증 시 기준이 되는 모형을 의미하며, 제약-기저 전략에서는 측정원점 동일성 모형이, 자유-기저 전략에서는 측정단위 동일성 모형이 기저모형으로 설정된다. 즉 제약-기저 전략에서는 모든 측정원점에 동일화 제약(equality constraint)이 부여된 모형(기저모형)의 χ^2 와 검증 대상이 되는 측정변수의 측정원점이 집단별로 추정되는 모형(연구모형)의 χ^2 를 이용해 $\Delta\chi^2$ 검증을 실시하는 반면, 자유-기저 전략에서는 모든 측정원점이 집단별로 추정되는 모형(기저모형)의 χ^2 와 검증 대상이 되는 측정변수의 측정원점에 동일화제약이 부여된 모형(연구모형)의 χ^2 를 이용해 $\Delta\chi^2$ 검증을 실시한다. 두 전략 모두 자유도 1의 $\Delta\chi^2$ 검증에 해당하며, $\Delta\chi^2$ 가 통계적으로 유의할 경우 해당 문항에는 균일적 차별문항기능이 존재하는 것으로 해석된다.

Stark 등(2006)은 두 전략 중 자유-기저 전략이 통계적으로 더 적절하다고 주장하였다. 그 이유로는 첫째, 제약-기저 전략을 사용할 경우 수정지수(MI)를 참고하는 경우가 있는데, 수정지수의 사용은 자료에 의해 유도되는(data-driven) 과정이므로 우연의 이론화(capitalization on chance)가 발생할 위험을 내포하고 있다. 둘째, 제약-기저 전략에서는 검증 대상이 되는

문항 외의 다른 문항들에 차별기능이 존재하지 않는다는 강한 가정을 내포하고 있는데, 만약 이 가정이 위배될 경우 모수 추정치가 편향될 가능성이 있다. 또한 $\Delta\chi^2$ 검증은 중심적 χ^2 분포(central χ^2 distribution)를 바탕으로 실시하게 되는데, 기저모형이 자료에 잘 맞지 않을 경우 $\Delta\chi^2$ 가 중심적 χ^2 분포를 따르지 않기 때문에 $\Delta\chi^2$ 검증이 적절하지 않을 수 있다(Maydeu-Olivares & Cai, 2006).

두 가지 탐지 전략에 따른 효과는 MACS 모형에 대해 Stark 등(2006)에서 연구된 바 있다. 그러나 MIMIC 모형에 대해서는 탐지 전략에 따른 효과가 연구된 바 없기 때문에, 이를 집단내 요인으로서 연구설계에 포함시켰다.

종속변수

정확 탐지율(TDR). 조건별로 각각 1,000개의 표본을 확률표집하여 연구모형과 기저모형 간에 $\Delta\chi^2$ 검증을 실시하였다. 그 결과 $\Delta\chi^2$ 값이 통계적으로 유의하면 1, 유의하지 않으면 0으로 기록하였다. 그 결과 차별기능이 존재하는 문항에서 1이 관찰된 비율을 정확 탐지율(True Detection Rate: TDR)로 정의하였다(Yoon & Millsap, 2007). 정확 탐지율은 존재하는 차별문항기능을 올바르게 탐지할 확률을 의미하는 것으로, 통계적 검정력(statistical power)에 해당하는 개념이다. 본 연구에서 정확 탐지율이 높은 모형은 균일적 차별문항기능 탐지에 더 뛰어난 모형으로 평가할 수 있다.

오류 탐지율(FDR). 정확 탐지율과 동일한 방식으로 기저모형과 연구모형 간에 $\Delta\chi^2$ 검증을 실시하였다. 그 결과 차별기능이 존재하지 않는 문항에서 $\Delta\chi^2$ 값이 통계적으로

유의하게 나타난 비율을 오류 탐지율(False Detection Rate: FDR)로 정의하였다(Yoon & Millsap, 2007). 즉 오류 탐지율은 1종 오류에 해당하는 개념으로, 이것이 높게 나타날 경우 현실적으로 사용하기 어려운 모형으로 평가할 수 있다. 각 조건별로 참조문항(y_1)을 제외한 나머지 5개 문항(y_2 - y_6)에 대해 각각 $\Delta\chi^2$ 검증을 실시하므로, 각 문항별로 나타난 $\Delta\chi^2$ 검증 기각 비율을 평균해서 조건별 오류 탐지율을 산출하였다. 이는 실제로 차별문항기능을 탐지할 때 어떤 문항에 차별문항기능이 존재하는지를 미리 알 수 없는 상황을 반영한 것이다. 단, 차별문항기능이 존재하는 조건에서는 y_6 를 제외한 나머지 4개의 문항에서의 기각 비율을 평균하여 오류 탐지율을 산출하였다.

부가 측정치

앞에서 언급한 것처럼 측정오차분산이 집단 간에 다를 경우, 측정오차분산 동일성을 가정하는 MIMIC 모형의 요인계수 추정치가 왜곡될 것으로 예상하였다. 이를 확인하기 위해 차별문항기능 탐지 대상이 되는 6번째 문항의 요인계수에 대한 오차제곱평균(Mean Squared Error: MSE)을 계산하였다. 오차제곱평균은 다음과 같이 정의된다.

$$MSE = E[(\hat{\theta} - \theta)^2] \quad (9)$$

여기서 $\hat{\theta}$ 는 모수 추정치를, θ 는 모수를 의미한다. 즉 오차제곱평균은 각 조건별 반복시행을 통해 얻은 모수 추정치와 모수 간의 차이를 제공해서 평균한 것으로, 추정의 정확성을 평가하는데 사용된다.

자료 생성 및 결과값 계산

위에서 언급된 조건에 따라 모집단을 정의한 후 Mplus 6(Muthén & Muthén, 2010) 프로그램을 이용하여 조건 당 1,000개의 표본자료를 생성하였다(부록 III 참고). 이때 문항들은 다변량정규분포를 따르도록 설정되었다. 5개의 집단간 요인의 조합으로 총 432개의 조건이 나타나므로, 결과적으로 총 432,000개의 표본자료를 생성하였다. 이어서 각각의 표본자료들을 Mplus 6과 SAS 9.3(SAS Institute, 2010) 프로그램을 이용하여 두 모형(MACS 및 MIMIC)의 각 탐지전략에 합치시킨 후(부록 III 참고), 각 반복시행별로 모수 추정치 및 χ^2 통계치를 산출하였다. 산출된 모수 추정치와 χ^2 통계치를 SAS 9.3 프로그램으로 읽어들이고, SAS MACRO 및 DATA STEP을 통해 $\Delta\chi^2$ 검증을 실시하여 정확 탐지율과 오류 탐지율을 계산하였다.

χ^2 검증과 마찬가지로 $\Delta\chi^2$ 검증 역시 표본크기에 민감하기 때문에, 표본크기가 증가함에 따라 현실적으로 수용할 만한 차이도 쉽게 기각하는 문제가 여러 차례 지적된 바 있다(예: Bentler & Bonnett, 1980; Marsh et al., 1988). 본 연구에는 표본크기가 큰 조건이 포함되어 있으므로, 이러한 문제를 해결하기 위해 $\Delta\chi^2$ 검증 시의 임계치(critical value)를 통상적인 임계치보다 높게 설정하였다. 즉, 유의수준 .05에서의 임계치인 3.84 대신 Marsh와 Hocevar(1985)가 제안한 5를 기준으로 $\Delta\chi^2$ 검증을 실시하였다.

결 과

정확 탐지율

차별문항기능이 존재하는 각 조건에 대해 정확 탐지율을 계산한 결과는 표 4와 같다. 전반적으로 독립변수의 수준에 따라 결과가 일정한 양상으로 나타났기 때문에, 지면 제약 상 일부 중간 조건(작은 집단효과 또는 표본 크기 비율 3:2)에 대한 결과는 생략하였다³⁾.

표 4에서 볼 수 있듯, 전반적으로 전체 표본크기가 클수록, 차별문항기능의 크기가 클수록 정확 탐지율이 높아지는 것을 볼 수 있다. 이는 일반적인 현상이며, 이 두 가지 변수가 통제 대상이므로 구획변수로서 포함된 것과 관련이 있다.

그림 3에 제시한 것처럼, 전반적으로 측정변수 신뢰도의 차이(측정오차분산의 차이)가 증가할수록 정확 탐지율이 감소하는 것을 볼 수 있다. 이러한 감소 추세는 집단효과 및 표본크기의 비율에 따라 다르게 나타났다. 표본크기의 비율이 1:1인 경우, 예상했던 바와 같이 집단효과가 존재하지 않을 때에는 측정변수 신뢰도의 차이와 관계없이 MACS 모형과 MIMIC 모형의 수행이 거의 동일하나⁴⁾, 집단효과가 존재할 경우에는 측정변수 신뢰도의 차이가 클수록 MIMIC 모형의 수행이 상대적으로 더 크게 저하된 것을 볼 수 있다. 또한 이때 6번째 요인계수의 오차제곱평균을 계산한 결과는 그림 4와 같다. 이에 대해 4개 모

형(2개 모형 × 2개 탐지 전략)을 종속변수로 한 혼합설계 분산분석(mixed-design ANOVA)을 실시한 결과, 모형×집단효과×측정변수 신뢰도 차이의 삼원상호작용효과가 통계적으로 유의하지 않았다, $F=.000, n.s.$. 즉 측정변수 신뢰도의 차이가 증가함에 따라 MIMIC 모형의 추정이 MACS 모형보다 부정확해지는 현상이 집단효과의 존재 여부와는 무관한 것으로 나타났다. 따라서 집단효과의 존재 여부와 관계없이 MIMIC 모형의 가정이 위배될 경우 요인계수의 추정이 왜곡되지만, 이러한 왜곡은 집단효과가 존재할 경우에만 균일적 차별문항기능 탐지에 영향을 미칠 것이라는 추론이 가능하다. 다만 그 차이가 눈에 띄게 크지는 않으므로, MIMIC 모형이 측정오차분산 동일성의 가정 위배에 대해 강내성(robustness)을 가지고 있는 것으로 해석할 수 있다. 이러한 양상은 탐지 전략(제약-기저 전략, 자유-기저 전략)과 관계없이 동일하게 나타났다.

한편 그림 3에서 표본크기가 4:1인 경우, 측정오차분산이 동일한 경우를 제외하고 모든 조건에서 MIMIC 모형의 정확 탐지율이 MACS 모형에 비해 높게 나타나는 것을 볼 수 있다. 특히 측정변수 신뢰도의 차이(측정오차분산의 차이)가 증가할수록 그 차이가 더 크게 나타나는데, 측정변수 신뢰도가 동일한 경우에는 제약-기저 전략을 사용한 MACS 모형이 자유-기저 전략을 사용한 MIMIC 모형에 비해 우수한 수행을 보여주는 반면, 측정변수 신뢰도의 차이가 작은 경우부터는 그 관계가 역전되는 것을 볼 수 있다. 이는 집단 간에 표본크기가 다른 경우 표본자료를 나누어서 사용하는 MACS 모형이 균일적 차별문항기능 탐지에 더 불리하며, 특히 측정변수 신뢰도가 다른 경우 더 심각해짐을 의미한다.

3) 전체 결과는 교신저자에게 요청하면 확인 가능함.

4) 그림 3 중 좌상단은 MACS 모형과 MIMIC 모형의 그래프가 겹쳐서 표시됨. 위는 제약-기저 전략, 아래는 자유-기저 전략에 해당함.

표 4. 정확 탐지율

정확 탐지율	표본크기 비율 1:1										표본크기 비율 4:1									
	좌측 사별가능					우측 사별가능					좌측 사별가능					우측 사별가능				
	MACS (95)	MACS (90)	MACS (85)	MACS (80)	MACS (75)	MACS (95)	MACS (90)	MACS (85)	MACS (80)	MACS (75)	MACS (95)	MACS (90)	MACS (85)	MACS (80)	MACS (75)	MACS (95)	MACS (90)	MACS (85)	MACS (80)	MACS (75)
전체	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300
성인	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300
청소년	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300
1,000	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98
...																				
총합	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300
...																				
남성	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300
1,000	89	91	89	91	89	91	89	91	89	91	89	91	89	91	89	91	89	91	89	91
...																				
총합	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300
...																				
여성	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300
1,000	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
...																				
총합	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300
...																				
남성	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300
1,000	87	90	88	86	86	86	86	86	86	86	86	86	86	86	86	86	86	86	86	86
...																				
총합	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300	300

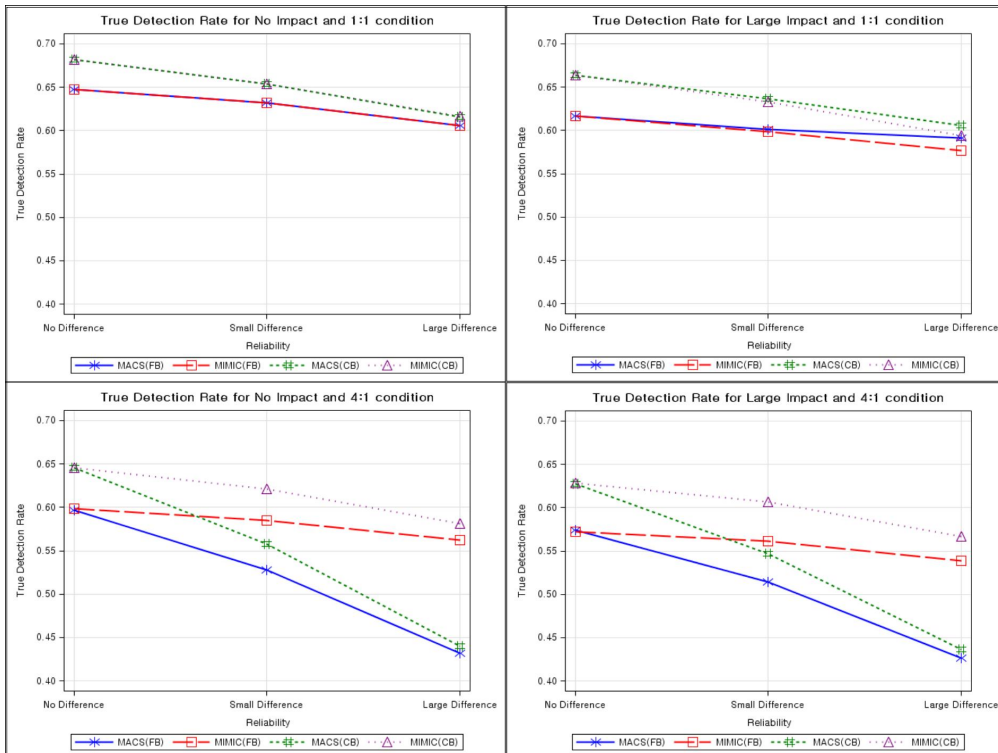


그림 3. 조건별 정확 탐지율

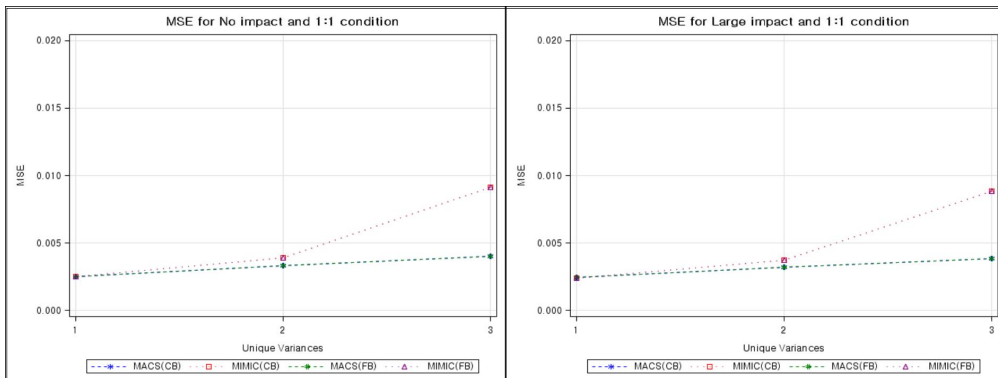


그림 4. 조건별 오차제곱평균(일부만 제시)

또한 표본크기가 다른 경우의 오차제곱평균 역시 그림 4와 동일한 양상으로 나타났음에도 MIMIC 모형의 수행이 우수한 것은, 가정 위배

에 따른 손실보다 표본크기 비율에 대한 강내성(robustness)의 이득이 더 큰 것으로 해석할 수 있다.

추가로 주목해야 할 부분은 탐지 전략에 따른 정확 탐지율의 차이이다. 그림 3에서 볼 수 있듯, 전반적으로 제약-기저 전략의 정확 탐지율이 자유-기저 전략의 정확 탐지율보다 높은 것으로 나타났다. Stark 등(2006)은 자유-기저 전략의 정확 탐지율이 제약-기저 전략보다 높은 것으로 보고하였는데, 이 차이는 균일적 차별기능을 가진 문항의 비율에 기인한 것으로 보인다. Stark 등은 15개의 문항 중 4개 문항(26.7%)에 차별기능이 존재하는 상황을 설정하였으나, 본 연구에서는 6개의 문항 중 1개 문항(16.7%)에만 차별기능이 존재하는 상황을 설정했다는 점이 다르다. 이와 관련하여 김한조(2009)는 비균일적 차별문항기능(부분 측정단위동일성)에 대한 연구에서 차별기능의 비율이 높은 경우 자유-기저 전략이, 차별기능의 비율이 낮은 경우 제약-기저 전략이 유리했음을 보고한 바 있다.

오류 탐지율

각 조건에 대해 오류 탐지율을 계산한 결과는 표 5와 같다. 정확 탐지율과 마찬가지로 일부 중간 조건(작은 집단효과, 표본크기 비율 3:2)에 대한 결과는 생략하였다.

표 5에서 볼 수 있는 바와 같이, 오류 탐지율에서는 전반적으로 탐지 전략에 따른 차이가 두드러진다. 모형과 관계없이 제약-기저 전략을 사용한 경우 오류 탐지율이 최대 .95(큰 집단효과, 큰 측정변수 신뢰도 차이, 표본크기 1000, 표본크기 비율 1:1, 큰 차별기능 조건)까지 증가하는 것을 볼 수 있다. 차별문항기능의 정도가 높거나 표본크기가 큰 경우의 오류 탐지율이 특히 더 크게 나타나는데, 이는 제약-기저 전략에서 모든 문항의 측정원점

에 동일화제약을 부여한 모형을 $\Delta\chi^2$ 검증의 기저모형으로 사용하기 때문인 것으로 판단된다.

그림 5에 제시하였듯, 표본 크기가 동일한 경우에는 MACS 모형과 MIMIC 모형의 차이보다는 제약-기저 전략과 자유-기저 전략의 차이가 두드러졌다. 자유-기저 전략을 사용한 경우 조건과 관계없이 MACS 모형과 MIMIC 모형 모두에서 .03 수준의 오류 탐지율을 보이는 것을 알 수 있다. 이는 일반적으로 설정하는 유의수준인 .05보다 낮은 수준으로, 연구문항 외에 균일적 차별기능문항이 존재하더라도 연구문항을 균일적 차별기능문항으로 잘못 탐지하는 경우가 드물 것으로 예상할 수 있다. 그러나 제약-기저 전략을 사용한 경우, 연구문항 외의 문항에 균일적으로 차별적인 문항기능이 존재하면 연구문항을 균일적 차별기능문항으로 잘못 탐지할 가능성이 높을 것이다.

표본크기가 다른 경우 자유-기저 전략을 사용한 MACS 모형은 .02-.04, MIMIC 모형은 .02-.06 수준의 오류 탐지율을 보였다. 일부 조건에서 MIMIC 모형의 오류 탐지율이 .06까지 상승하는 것을 볼 수 있으나, 현실적으로 크게 무리가 없는 수준이며, Bonferroni 방법 등을 통해 교정할 여지가 있다. 반면 제약-기저 전략을 사용한 경우에는 MACS 모형의 오류 탐지율이 .03-.65, MIMIC 모형의 오류 탐지율이 .03-.87로 수용하기 어려운 수준으로 나타났다. 이 경우 MIMIC 모형의 오류 탐지율이 MACS 모형보다 높게 나타났다.

표 5. 오류 탐지률

오류 형태	표본크기 비율 33				표본크기 비율 43			
	사탐기능 함양	자본 사탐기능	중간 사탐기능	문 사탐기능	사탐기능 함양	자본 사탐기능	중간 사탐기능	문 사탐기능
관성 정답 정답의 오류 충성률수 크기 신뢰도	MACS (23) (23) (23) (23)	MACS (23) (23) (23) (23)	MACS (23) (23) (23) (23)	MACS (23) (23) (23) (23)	MACS (23) (23) (23) (23)	MACS (23) (23) (23) (23)	MACS (23) (23) (23) (23)	MACS (23) (23) (23) (23)
결함	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300
중간	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300
결함	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300
중간	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300
결함	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300
중간	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300
결함	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300
중간	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300	300 300 300 300

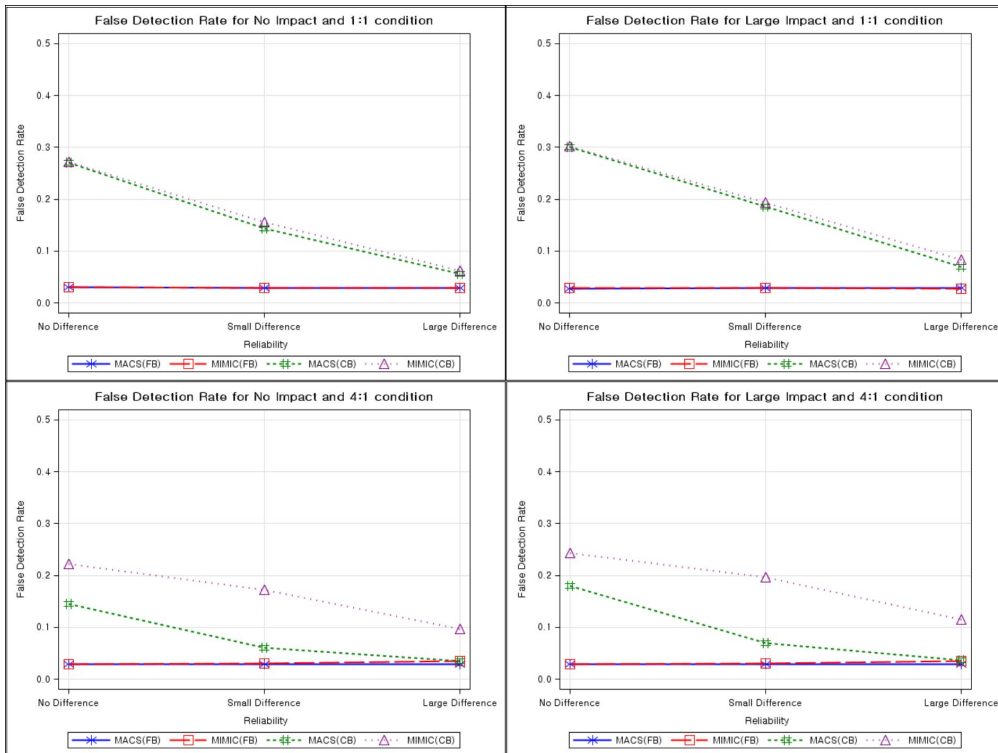


그림 5. 조건별 오류 탐지율

논 의

본 연구에서는 MACS 모형과 MIMIC 모형의 균일적 차별문항기능 탐지율을 비교하기 위해 몬테 카를로 모의실험을 실시하였다. 이를 위해 집단효과 크기, 측정변수 신뢰도의 차이, 표본크기의 비율, 표본크기, 차별문항기능의 크기를 독립변수로 설정하였다. 또한 두 모형에 각각 두 가지 차별문항기능 탐지 전략을 적용하여 탐지 전략에 따른 차이를 관찰하였다. 먼저 표본크기가 집단 간에 동일한 경우, 집단효과가 존재하고 집단 간에 측정변수 신뢰도의 차이가 크면 MIMIC 모형의 정확 탐지율이 저하되는 것으로 나타났다. 그러나 이러한 효과가 주목할 만큼 크게 나타나지는 않았

기 때문에, MIMIC 모형이 가정 위배에 대해 강내성을 가지는 것으로 해석된다. 반면 집단 효과가 존재하지 않는 경우에는 MIMIC 모형의 요인계수 추정치가 왜곡됨에도 불구하고 정확 탐지율이 저하되지 않았다.

표본크기가 집단 간에 다른 경우에는 MIMIC 모형의 정확 탐지율이 전반적으로 우수하였다. 특히 측정변수 신뢰도가 집단 간에 크게 다른 경우에도 MIMIC 모형의 정확 탐지율은 크게 저하되지 않았다. 반면 MACS 모형의 정확 탐지율은 상대적으로 낮았으며, 특히 측정변수 신뢰도가 집단 간에 다른 경우 정확 탐지율이 현저하게 저하되었다.

오류 탐지율에서는 모형에 따른 차이보다 탐지 전략에 따른 차이가 현저하게 나타났다.

표본크기가 동일한 경우와 그렇지 않은 경우 모두 자유-기저 전략의 오류 탐지율이 낮게 나타났으며, 일반적으로 사용하는 1종 오류 확률의 기준인 .05과 유사하거나 낮은 수준이었다. 그러나 제약-기저 전략을 사용한 경우 MIMIC 모형과 MACS 모형 모두 높은 수준의 오류 탐지율을 보였으며, 특히 제약-기저 전략을 사용한 MIMIC 모형은 가장 높은 수준의 오류 탐지율을 보였다.

두 모형 모두 제약-기저 전략을 사용한 경우의 정확 탐지율이 높았으나, 이 경우 오류 탐지율이 높게 나타났기 때문에 제약-기저 전략의 실용성을 재고할 필요가 있다. 올바른 통계적 의사결정을 위해서는 정확 탐지율이 높을 뿐 아니라 오류 탐지율이 낮아야 하기 때문이다. 자유-기저 전략을 사용한 MIMIC 모형은 이 두 가지 조건을 모두 충족하므로, 균일적 차별문항기능 탐지 시 가장 권장할 만한 방법으로 볼 수 있다. 표본크기가 집단 간에 다르고 측정변수 신뢰도의 차이가 클 경우 자유-기저 전략을 사용한 MIMIC 모형의 오류 탐지율이 다소 높을 수 있지만 그 크기가 .06 수준으로 크게 우려할 정도는 아니며, 오류 탐지율이 우려되는 상황이라면 대안으로서 Bonferroni 방법 등을 통해 오류 탐지율을 교정할 수도 있다⁵⁾. 다만 MIMIC 모형은 비균일적

5) 실제로 추가분석을 실시한 결과, Bonferroni 방법을 사용했을 때 자유-기저 전략을 사용한 MIMIC 모형의 오류 탐지율은 모든 조건을 통틀어 최댓값이 .03에 불과할 정도로 감소하였다. 이때의 정확 탐지율은 최악의 조건인 '큰 집단효과, 큰 측정변수 신뢰도 차이, 표본크기 비율 4:1' 조건에서 교정 전에 비해 .07 정도 하락하였다. 반면 자유-기저 전략을 사용한 MACS 모형의 경우 동일 조건에서 정확 탐지율이 .10 가까이 하락하는 등 Bonferroni 교정으로 인한 정확 탐지율 손실이

차별문항기능이 존재하지 않는 상황을 가정하고 있으므로, 비균일적 차별문항기능이 의심되거나 발견된 상황에서는 자유-기저 전략을 사용한 MACS 모형을 권장한다.

본 연구는 의의는 다음과 같다. 첫째, 다양한 조건 하에서 구조방정식모형에 기반한 균일적 차별문항기능 탐지 방법들을 경험적으로 비교하였다. 그동안 MIMIC 모형이 추가적으로 가지는 세 가지 가정(측정단위 동일성, 측정오차분산 동일성, 요인분산 동일성)에 대해 여러 문헌에서 지적하였으나(예: Brown, 2006; Hancock et al., 2000), 동일한 조건 하에서 MACS 모형과의 수행을 비교한 연구는 없었다. 또한 집단 간 표본크기의 차이에 따른 각 모형의 수행은 종종 연구된 바 있으나(예: Kaplan & George, 1995; Gelin & Zumbo, 2007), 두 모형의 차이에 대해서는 체계적으로 비교되지 않았다. 둘째, 실제 연구에서 관찰 가능한 표본 크기를 고려하여 현실적인 참조점을 제시하였다. 특히 심리학 연구에서 자주 발견되는 매우 작은 표본(N=20)을 고려하였기 때문에 다양한 연구 상황에 적용이 가능하다. 본 연구에서는 충분한 크기(1,000회)의 반복시행을 실시하였으므로, 실제 연구에서 정확 탐지율과 오류 탐지율을 예측하는데 본 연구의 결과(표 4, 표 5)를 활용할 수 있다. 마지막으로, 차별문항기능을 탐지하는데 사용되는 두 가지 전략(제약-기저, 자유-기저)을 두 모형에 모두 적용함으로써 효과적인 방안을 도출하였다.

본 연구의 제한점 및 추후 연구에 대한 제언은 다음과 같다. 첫째, MACS 모형과 MIMIC 모형 간의 비교에 초점을 두었기 때문에, 비균일적 차별문항기능이 존재하지 않는 상황을

MIMIC 모형에 비해 크게 나타났다.

가정하였다. 일반적으로 구조방정식모형을 이용할 경우 측정단위 동일성 검증이 측정원점 동일성에 선행하지만, MIMIC 모형에서는 이를 검증할 수 없다. 따라서 비균일적 차별기능이 존재하는 상황에서 MIMIC 모형의 수행을 관찰할 필요가 있다. 둘째, 본 연구에서는 6개의 문항 중 1개의 문항에만 균일적 차별기능문항이 존재하는 상황을 가정하였다. 실제 자료에서는 여러 개의 차별기능문항이 공존할 수 있기 때문에, 차별기능문항의 비율을 체계적으로 조작하여 그 효과를 검토할 필요가 있다. 셋째, 본 연구에서는 모든 측정변수가 표준정규분포를 따르는 것으로 설정되었다. 현실적으로 개별 문항이 정규분포를 따르는 경우는 거의 없으므로, 문항들이 서열변수(예: 5점척도)인 경우에 대한 모의실험 연구가 필요하다. 마지막으로, 본 연구에서는 1개 요인만을 상정하였는데, 실제로 심리검사는 다요인 구조인 경우가 많다. 따라서 복수의 요인이 존재하는 상황에 대한 후속연구가 필요할 것이다.

참고문헌

김성훈 (1995). 문항차별기능의 인지심리학적 해석 가능성 검토. *교육평가연구*, 8(2), 39-62.

김한조 (2010). 다집단 분석에서의 부분동일성 검증 전략들: 적절한 전략을 찾기 위한 시뮬레이션 연구. 성균관대학교 석사학위논문.

성태제 (1993). 차별기능(편파성) 문항 추출을 위한 Raju 방법과 MH 방법의 비교 연구. *교육평가연구*, 6(1), 91-12.

이순목, 금은희, 이찬순 (2010). 다집단 분석의

문제: 평균구조분석에서의 측정원점 동일성 검증 필요 여부. *교육평가연구*, 23(2), 391-416.

Ackerman, T. A. (1992). A Didactic Explanation of Item Bias, Item Impact, and Item Validity from a Multidimensional Perspective. *Journal of Educational Measurement*, 29(1), 67-91.

Barendse, M. T., Oort, F. J., & Garst, G. J. A. (2010). Using restricted factor analysis with latent moderated structures to detect uniform and nonuniform measurement bias: a simulation study. *Advanced Statistical Analysis*, 94, 117-127.

Bentler, P. M. & Bonett, D. G. (1980). Significant tests and goodness of fit in the analysis of covariance structure. *Psychological Bulletin*, 88, 588-606.

Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.

Byrne, B. M., Shavelson, R. J., & Muthen, M. (1989). Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance. *Psychological Bulletin*, 105(3), 456-466.

Camilli, G. (1979). A Critique of the chi-square method assessing item bias. Unpublished paper, Laboratory of Educational Research, University of Colorado at Boulder.

Chan, D. (2000). Detection of Differential Item Functioning on the Kirton Adaptation-Innovation Inventory Using Multiple-Group Mean and Covariance Structure Analyses. *Multivariate Behavioral Research*, 35(2), 169-199.

- Cohen, J. (1988). *Statistical Power Analysis for Behavioral Sciences*(2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.
- Dragow, F. (1984). Scrutinizing Psychological Tests: Measurement Equivalence and Equivalent Relations with External Variables. *Psychological Bulletin*, 95, 134-135.
- Gelin, M. N. & Zumbo, B. D. (2007). Operating characteristics of the DIF MIMIC approach using Jöreskog's covariance matrix with ML and WLS estimation for short scales. *Journal of Modern Applied Statistical Methods*, 6(2), 573-588.
- Hancock, G. R., Lawrence, F. R., & Nevitt, J. (2000). Type I error and power of latent mean methods and MANOVA in factorially invariant and noninvariant latent variable systems. *Structural Equation Modeling*, 7(4), 534-556.
- Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in Exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement*, 65(2), 202-226.
- Holland, P. W. & Thayer, D. T. (1988). Differential Item Functioning and the Mantel-Haenszel Procedure. In Wainer, H. & Braun, H. I. (Eds.), *Test Validity*(pp.129-145). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.
- Jöreskog, K. G. (1971). Statistical Analysis of sets of congeneric tests. *Psychometrika*, 36(2), 109-133.
- Jöreskog, K. G. & Goldberger, A. S. (1975). Estimation of a Model with Multiple Indicator and Multiple Causes of a Single Latent Variable. *Journal of American Statistical Association*, 70(351), 631-639.
- Jöreskog, K. G. & Sörbom, D. (1989). *LISREL 7: A Guide to the Program and Applications*. Chicago: SPSS Publications.
- Kaplan, D. & George, R. (1995). A study of the power associated with testing factor mean differences under violations of factorial invariance. *Structural Equation Modeling*, 2(2), 101-118.
- Lee, J. H. (2009). *Type I Error and Power of the Mean and Covariance Structure Confirmatory Factor Analysis for Differential Item Functioning Detection: Methodological Issues and Resolutions*. Unpublished Doctoral Dissertation. The University of Kansas.
- Little, T. D. (1997). Mean and Covariance Structures(MACS) Analyses of Cross-Cultural Data: Practical and Theoretical Issues. *Multivariate Behavioral Research*, 32(1), 53-76.
- Lord, F. M. (1980). *Application of item response theory to practical testing problems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.
- Lubke, G. H. & Dolan, C. V. (2003). Can unequal residual variances across groups mask differences in residual means in the common factor model?. *Structural Equation Modeling*, 10(2), 175-192.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84-99.
- Marsh, H. W., Balla, J. R., & McDonald, R. P.

- (1988). Goodness-of-fit indices in confirmatory factor analysis: the effect of sample size. *Psychological Bulletin*, 103, 391-41.
- Maydeu-Olivares, A. & Cai, L. (2006). A cautionary note on using G2(dif) to assess relative model fit in categorical data analysis. *Multivariate Behavioral Research*, 41, 55-64.
- Meade, A. W. & Lautenschlager, G. J. (2004). A Comparison of Item Response Theory and Confirmatory Factor Analytic Methodologies for Establishing Measurement Equivalence/Invariance. *Organizational Research Methods*, 7, 361-388.
- Mellenberg, G. J. (1982). Contingency Table Models for Assessing Item Bias. *Journal of Educational Statistics*, 7(2), 105-118.
- Meredith, W. (1993). Measurement Invariance, Factor Analysis and Factorial Invariance. *Psychometrika*, 58(4), 525-543.
- Meredith, W. & Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika*, 57(2), 289-311.
- Miller, T. R. & Spray, J. A. (1993). Logistic Discriminant Function Analysis for DIF Identification of Polytomously Scored Items. *Journal of Educational Measurement*, 30(2), 107-122.
- Millsap, R. E. & Everson, H. T. (1993) Methodological Review: Statistical Approaches for Assessing Measurement Bias. *Applied Psychological Measurement*, 17(4), 297-334.
- Muthén, B. O. (1979). A structural probit model with latent variables, *Journal of the American Statistical Association*, 74, 807-811.
- Muthén, B. O. (1983). Latent variable structural equation modeling with categorical data, *Journal of Econometrics*, 22, 43-65.
- Muthén, L. K. and Muthén, B. O. (1998-2010). *Mplus User's Guide. Sixth Edition*. Los Angeles, CA: Muthén & Muthén.
- Oort, F. J. (1998). Simulation study of item bias detection with restricted factor analysis. *Structural Equation Modeling*, 5(2), 107-124.
- Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo Experiments: Design and Implementation. *Structural Equation Modeling*, 8(2), 287-312.
- Pennell, R. (1968). The influence of communality and N on the sampling distributions of factor loadings. *Psychometrika*, 33(4), 423-439.
- Raju, N. S., Raffitte, L. J., & Byrne, B. M. (2002). Measurement Equivalence: A Comparison of Methods Based on Confirmatory Factor Analysis and Item Response Theory. *Journal of Applied Psychology*, 87(3), 517-529.
- Rock, D. A., Werts, C. E., & Flaughner, R. L. (1978). The use of analysis of covariance structure for comparing the psychometric properties of multiple variables across populations, *Multivariate Behavioral Research*, 13, 408-418.
- SAS Institution Inc. (2011). *SAS/STAT® 9.3 User's Guide*. Cary, NC: SAS Institute Inc.
- Schmitt, N. & Kuljanin, G. (2006). Measurement Invariance: Review of practice and implications. *Human Resource Management Review*, 18, 210-222.
- Shealy, R. & Stout, W. (1993). A Model-based

- Standardization Approach that separates true bias/DIF from Group ability differences and Detects Test Bias/DTF as well as Item Bias/DIF. *Psychometrika*, 58(2), 159-194.
- Smith, L. L. (2002). On the Usefulness of Item Bias Analysis to Personality Psychology. *Personality and Social Psychology Bulletin*, 28, 754-763.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, 27, 229-239.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting Differential Item Functioning With Confirmatory Factor Analysis and Item Response Theory: Toward a Unified Strategy. *Journal of Applied Psychology*, 91(6), 1292-1306.
- Steenkamp, J. E. M. & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research*, 25, 78-9.
- Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In Wainer, H. & Braun, H. I. (Eds.), *Test Validity*(pp. 147-172). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.
- Vandenberg, R. J. & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, 3(1), 4-7.
- Yoon, M. & Millsap, R. E. (2007). Detecting Violations of Factorial Invariance Using Data-Based Specification Searches: A Monte Carlo Study. *Structural Equation Modeling*, 14(3), 435-463.

1차 원고접수 : 2013. 09. 30.
수정원고접수 : 2013. 12. 14.
최종게재결정 : 2013. 12. 19.

**Detection of differential item functioning
using structural equation modeling:
A comparison of MACS and MIMIC**

Soocheol Yun

Language Education Institute

Soonmook Lee

Sungkyunkwan Univ.

Two models, MACS and MIMIC, can be used to detect Differential Item Functioning(DIF) in a Structural Equation Modeling framework. Although these two models can be considered as special cases of general Structural Equation Models, they may perform differently in various research contexts due to differences in statistical assumptions and the way in which each model uses data. In particular, since MIMIC model requires some additional assumptions, its performance may decline when those assumptions are not satisfied. Furthermore, the performance of MIMIC model will be superior to that of MACS model when sample sizes vary among groups because the former uses a single dataset including group variable(s), unlike the latter. Although many articles have commented on these predictions, no systematic research comparing the performance of the two models under these circumstances had yet to be conducted. Thus, we investigated the differences in performance of these two models under various conditions, specifically the size of impact, differences in measurement variable reliability, sample size ratio, total sample size, the size of differential item functioning, and the strategy for detecting DIF through a Monte Carlo simulation study. We found that the performance of MIMIC model in detecting uniform DIF did not decline significantly, although one of its additional assumptions was violated. Moreover, MIMIC model was superior to MACS model when sample sizes differed between two groups. Finally, we emphasize the importance of employing appropriate strategies for effective use of the two models to detect uniform DIF.

Key words : *Differential Item Functioning, Structural Equation Modeling, MACS, MIMIC*

부 록 I. 관심집단의 측정오차분산 계산 과정

I-1. 6번째 측정변수의 신뢰도가 높은(0.8) 경우

$$\text{측정변수 신뢰도 공식}^* \rho_6 = \frac{\lambda_6^2 \phi}{\lambda_6^2 \phi + \theta_6} \text{에서 } \lambda_6 = 0.8, \phi = 1, \rho_6 = 0.8 \text{이므로 } \theta_6 = 0.16$$

I-2. 6번째 측정변수의 신뢰도가 중간(0.5)인 경우

$$\text{측정변수 신뢰도 공식}^* \rho_6 = \frac{\lambda_6^2 \phi}{\lambda_6^2 \phi + \theta_6} \text{에서 } \lambda_6 = 0.8, \phi = 1, \rho_6 = 0.5 \text{이므로 } \theta_6 = 0.64$$

I-3. 6번째 측정변수의 신뢰도가 낮은(0.3) 경우

$$\text{측정변수 신뢰도 공식}^* \rho_6 = \frac{\lambda_6^2 \phi}{\lambda_6^2 \phi + \theta_6} \text{에서 } \lambda_6 = 0.8, \phi = 1, \rho_6 = 0.3 \text{이므로 } \theta_6 = 1.493$$

*측정변수 신뢰도 공식은 각주 1)을 참고

부 록 II. 측정오차분산과 표본크기에 따른 균일적 차별문항기능의 크기

균일적 차별문항기능의 크기는 식 (II-1)과 같이 정의된다.

$$d = \frac{\tau_6^{(0)} - \tau_6^{(1)}}{\sigma_{pooled}} \quad (\text{II-1})$$

통합표준편차 σ_{pooled} 는 식 (II-2)와 같이 정의된다.

$$\sigma_{pooled} = \frac{(n^{(0)} - 1)\sigma_6^{(0)} + (n^{(1)} - 1)\sigma_6^{(1)}}{n^{(0)} + n^{(1)} - 2} \quad (\text{II-2})$$

여기서 $\sigma_6^{(0)} = \sqrt{\lambda_6^{2(0)}\phi + \theta_6^{(0)}}$ = 참조집단의 6번째 측정변수 표준편차

$\sigma_6^{(1)} = \sqrt{\lambda_6^{2(1)}\phi + \theta_6^{(1)}}$ = 관심집단의 6번째 측정변수 표준편차

$n^{(0)}$ = 참조집단의 표본크기

$n^{(1)}$ = 관심집단의 표본크기

6번째 측정변수 신뢰도의 차이에 따라 관심집단의 측정오차분산인 $\theta_6^{(1)}$ 가 달라지며, $n^{(0)}$ 과 $n^{(1)}$ 도 조건에 따라 달라지기 때문에, 식 (II-2)에 의해 통합표준편차가 각 조건에 따라 다른 값을 가지게 된다. 따라서 식 (II-1)과 식 (II-2)에 의해 관심집단에서 6번째 측정변수의 측정원점은 d 가 동일하더라도 조건에 따라 다른 값을 가질 수 있다. 예를 들어 작은 차별문항기능 조건 ($d = 0.2$)에서 d 값이 동일함에도 불구하고 관심집단에서 6번째 측정변수의 측정원점은 다음과 같이 달라질 수 있다.

표 II-1. 표본크기와 측정오차분산에 따른 관심집단의 측정원점 차이 예시

표본크기		측정오차분산		측정변수 표준편차		통합	관심집단의
참조집단	관심집단	참조집단	관심집단	참조집단	관심집단	표준편차	측정원점
60	40	0.16	0.16	0.894	0.894	0.894	1.821
60	40	0.16	1.493	0.894	1.460	1.153	1.769

표 II-2는 d 의 크기에 따른 관심집단의 측정원점 크기를 나타낸다.

표 II-2. d 의 크기에 따른 관심집단의 측정원점 크기

d	관심집단의 측정원점 크기			
	M	SD	Min	Max
0	2.000	-	2.000	2.000
0.2	1.789	0.037	1.717	1.821
0.5	1.471	0.092	1.293	1.553
0.8	1.154	0.147	0.869	1.284

부 록 III. 모의자료 생성 및 분석용 Mplus 명령문

III-1. 모의자료 생성에 사용한 Mplus 명령문

title: Data generation for condtion (조건 번호)

montecarlo:

names are y1-y6;

nobservations=(참조집단의 표본크기) (관심집단의 표본크기);

nreps=1000;

```
ngroups=2;
repsave=all;
save=rep*.txt;
results=result.out;
model population:
  f1 by y1*1.0 y2*0.6 y3*0.8 y4*0.6 y5*1.0 y6*0.8;
  f1*1;
  y1*0.43 y2*1.44 y3*0.64 y4*0.84 y5*1.0 y6*0.16;
  [y1*2.0 y2*1.5 y3*2.5 y4*1.0 y5*3.0 y6*2.0];
  [f1*0];
model population-g2:
  y6*(관심집단의 측정오차분산);
  [f1*(집단효과)];
  [y6*(관심집단의 측정원점)];
output: tech9;
```

III-2. MACS 모형에 사용한 Mplus 명령문 예시(자유-기저 전략의 기저모형)

```
title: Free-Baseline MACS - null model
data: type=montecarlo;
      file is replist.txt;
      variable: names are y1-y6 g;
      grouping is g(1=ref 2=foc);
model: f by y1@1 y2-y6*;
      f*;
      [f@0];
      [y1-y6];
model foc:
      y1-y6*;
      f*;
      [f];
      [y2-y6];
savedata: results are macsfb1_cond.par;
```

※ 연구모형 진행 시 “model foc:”의 [y2-y6] 부분 수정

III-3. MIMIC 모형에 사용한 Mplus 명령문 예시(제약-기저 전략의 기저모형)

```
title: Constrained-Baseline MIMIC - null model
data: type=montecarlo;
      file is replist.txt;
      variable: names are y1-y6 g;
model: f by y1@1 y2-y6;
      f on g;
      f*;
savedata: results are mimiccb1_cond.par;
```

※ 연구모형 진행 시 “model:” 부분에 “y2 on g;” 등의 경로 추가