교육평가연구
Journal of Educational Evaluation
2018, 제31권 제3호, 583-608
http://dx.doi.org/10.31158/JEEV.2018.31.3.583

다차원 2모수 IRT 모형에 대한 대칭적 및 비대칭적 특성함수 척도연계 방법의 기능 비교

김 아 름 김 성 훈*
한양대학교

본 연구는 다차원 문항반응이론(MIRT) 척도연계 방법 중 우수성이 보고된 문항특성함수(ICF) 방법과 검사특성함수(TCF) 방법에 초점을 두고 각 방법의 대칭적 및 비대칭적 버전의 통계적 특성과 상대적 기능의 차이를 탐구하고자 하였다. 이를 위해, 다차원 2모수 로지스틱(M2PL) 모형에 대해 기존의 비대칭성을 수정한 대칭적 ICF 방법 및 TCF 방법을 제시하였으며, 모의실험을 통해 네 방법(비대칭적 ICF 및 TCF 방법, 대칭적 ICF 및 TCF 방법)의 기능을 분석하였다. 모의실험 요인으로 공통문항의 비율, 준거함수에 대한 가중치 유형과 능력 지점의 수 등을 고려하였으며, 네 척도연계 방법이 모의실험 조건에서 능력분포의 모수와 문항모수를 얼마나 정확하고 안정적으로 추정하는지를 탐색하였다. 또한 모의실험 자료를 이용하여 각 척도연계 방법의 대칭성을 확인하였다. 주요 결과는 다음과 같다. 첫째, 능력분포 모수의 복원에 있어서 대칭적 ICF 방법이 가장 작은 척도연계의 오차를 산출하였다. 둘째, 문항모수의 복원에 있어서 비대칭적 ICF 방법이 가장 작은 척도연계의 오차를 산출하였다. 셋째, 네 방법에 대하여 대칭성을 조사한 결과, 대칭적 ICF 및 TCF 방법은 모두 척도연계의 대칭성을 만족하였지만, 비대칭적 ICF 및 TCF 방법은 그렇지 않았다.

주제어 : 다차원 문항반응이론(MIRT), 척도연계, 대칭성, 문항/검사 특성함수 방법

* 교신저자 : 김성훈, 한양대학교 사범대학 교육학과, seonghoonkim@hanyang.ac.kr

Ⅰ. 서 론

현대 검사이론인 문항반응이론(item response theory: IRT)은 검사자료의 분석을 위한 일종의 통계적 모형이다. 교육 및 심리 분야에서는 IRT를 사용하여 피험자가 지닌 잠재적 특성(구인) 혹은 능력을 측정하는 데 관심을 가진다. IRT의 관심 모수는 피험자의 특성 혹은 능력의 수준을 나타내는 피험자 능력모수와 문항의 특성을 나타내는 문항모수이며, 두 모수는 수리적으로 모형화되어 피험자 능력척도 상에 표현된다. 하나의 능력척도 상에서 피험자와 문항의 특성을 모두 설명할 수 있다는 장점으로 인해 IRT는 문항 및 검사 제작, 검사 동등화, 수직 척도화, 문항은행의 구축, 컴퓨터 적응검사, 차별적 문항기능 분석 등 여러 분야에서 활용되고 있다(de Ayala, 2008; Kolen & Brennan, 2014; Yen & Fitzpatrick, 2006).

IRT를 위와 같이 다양한 분야에 활용하기 위해서는 척도 미결정성(scale indeterminacy) 문제를 먼저 해결해야 한다. 하나의 구인을 측정하는 일차원 IRT(unidimensional IRT: UIRT) 모형의 경우 능력척도는 원점과 단위에 있어서 미결정성을 지닌다. 두 개 이상의 구인을 측정하는 다차원 IRT(multidimensional IRT: MIRT) 모형을 이용할 경우, 능력 공간(ability space) 상에 문항모수와 능력모수를 표현하기 때문에 능력척도는 원점과 단위뿐만 아니라 좌표축의 방향(orientation)에 있어서 미결정성을 지니게 된다(Reckase, 2009). 이러한 척도 미결정성으로 인해 서로 다른 검사에서 IRT 모수 추정을 통해 얻어진 두 능력척도는 일반적으로 대등하지 않으며, 두 검사에서 얻어진 문항모수 및 능력모수 추정치들을 비교하기 위해서는 두 능력척도 중 하나를 기준 척도로 삼아 모든 IRT 모수 추정치들을 선형 변환하여 그 공통의 기준 척도 상에 놓을 필요가 있다. 이와 같이 서로 다른 검사에서 정의된 두 능력척도 간의 선형 변환 관계를 찾아 두 능력척도를 연계하는 작업을 IRT 척도연계(scale linking)라 부르며, 척도연계의 목적은 공통 능력척도의 개발에 있다(Kim, Harris, & Kolen, 2010; Kolen & Brennan, 2014).

이론상 IRT 척도연계는 공통의 문항모수 혹은 공통의 피험자 능력모수를 이용하여 실시할 수 있으나, 현실적으로 검사의 보안 문제 및 연습효과로 인한 검사점수의 왜곡이 발생할 우려가 있기 때문에 공통문항 기반 척도연계가 주로 사용된다(Kolen & Brennan, 2014). 이러한 점에 비추어 본 연구에서는 공통문항을 기반으로 한 척도연계를 다루었으며, 이하에서 언급하는 척도연계는 모두 공통문항 기반 척도연계를 가리킨다.

지금까지 여러 UIRT 및 MIRT 모형에 대해 척도연계 방법들이 개발되어 왔다. 대표적인 UIRT 척도연계 방법으로는 (a) 문항모수 추정치의 적률을 이용한 방법(Loyd & Hoover, 1980; Marco, 1977), (b) 문항 혹은 검사의 특성함수(characteristic function)를 이용한 방법(Haebara, 1980; Stocking & Lord, 1983), (c) 문항모수 추정치 간의 차이를 최소화하는 방법(Divgi, 1985;

Ogasawara, 2001) 등을 들 수 있다. 이러한 척도연계 방법들은 다차원 2모수 혹은 3모수 로지스틱(multidimensional two- or three-parameter logistic: M2PL/M3PL) 모형(Davey, Oshima, & Lee, 1996; Li & Lissitz, 2000; Min, 2003; Oshima, Davey, & Lee, 2000)과 다차원 등급반응(graded response) 모형(김성훈, 2018) 등으로 확장되어 왔다. UIRT 척도연계 방법과 달리, MIRT 척도연계 방법은 능력 좌표축의 방향에 대한 미결정성을 해결하는 과정에서 직교 회전(orthogonal rotation)을 사용하는지, 사각 회전(oblique rotation)을 사용하는지에 따라 크게 두 가지 방법으로 나뉜다. 여기서, 직교 회전은 능력 변수들 간의 상관이 존재하지 않음을 가정하기 때문에 사각 회전의 특수한 경우라고 할 수 있다. 사각 회전을 사용하는 방법에는 Oshima et al.(2000)이 제안한 Direct/최소제곱 방법, 문항특성함수(item characteristic function: ICF) 방법, 검사특성함수(test characteristic function: TCF) 방법 등이 포함된다. 이와 달리, 직교 회전을 사용하는 척도연계 방법에는 Schönemann(1966)의 직교 프로크루스테스(orthogonal Procrustes) 해법에 토대를 둔, Li & Lissitz(2000)와 Min(2003)의 방법을 들 수 있다.

이러한 UIRT 및 MIRT 척도연계 방법들에 대하여 대부분의 선행 연구들(김성훈, 2018; Davey et al., 1996; Hanson & Béguin, 2002; Kim & Cohen, 1998; Min, 2003, 2007; Oshima et al., 2000; Simon, 2008)은 방법들 간 기능의 상대적 우수성을 검토하는 데 주된 관심을 두었으며, 특성함수 방법이 다른 방법들보다 정확하고 안정적인 척도연계 상수를 산출하였다고 보고하였다. 나아가, 일부 연구들은 척도연계 방법의 상대적 특성으로서 대칭성(symmetry)을 검토하였다(김성훈, 민경석, 강동희, 2016; Kim & Kolen, 2007). 두 검사형 N과 O의 능력척도에 대해 “N에서 O로” 그리고 “O에서 N으로” 척도연계를 실시한다고 하면, 대칭적 척도연계 방법을 통해 산출된 “N에서 O로”의 척도연계 함수는 “O에서 N으로”의 척도연계 함수의 역(inverse)이어야 한다. 결과적으로 대칭적 척도연계 방법은 N의 능력척도와 O의 능력척도 간에 고유한 일대일 대응관계를 밝혀준다. 이와 같은 척도연계의 대칭성은 검사 동등화에서의 대칭성과 동일 개념이며, 검사 동등화를 다루는 많은 문헌에서는 대칭성을 성공적인 동등화를 실시하기 위한 필요 요건 중 하나로 꼽고 있다(Dorans & Holland, 2000; Holland & Dorans, 2006; Kim et al., 2010; Kolen & Brennan, 2014; Lord, 1980). 특히, Kolen & Brennan(2014)은 대칭성이 성립하지 않는 방법을 사용하여 척도연계를 수행할 경우 체계적 오차(systematic error)가 발생할 수 있음을 강조하였다.

검사 동등화에서 그러하듯이, IRT 척도연계에서도 대칭성이 중요하게 다루어져야 할 특성임에도 불구하고 지금까지 대부분의 MIRT 척도연계 방법들은 비대칭적으로 제시되어 왔다. 더욱이 선행 연구자들은 그들이 제시한 척도연계 방법이 비대칭적 방법임을 인식하고 비대칭적 방법을 대칭적 방법으로 수정하거나 새로운 대칭적 척도연계 방법을 제시하는 일에 별다른 관심을 기울이지 않았다. 이 때문에 비대칭적 MIRT 척도연계 방법을 어떻게 대칭적

방법으로 수정할 수 있으며, 대칭적 방법은 비대칭적 방법에 비추어 어떠한 기능의 차이를 보이는지에 대한 연구가 이루어지지 않았다. IRT의 여러 활용 분야에서 척도연계가 차지하는 역할과 중요성을 감안할 때, 대칭적 척도연계 방법의 개발과 그 특성에 대한 연구는 필히 이루어져야 할 것으로 보인다.

이러한 연구의 필요성과 관련하여, 본 연구의 목적은 지금까지 개발된 MIRT 척도연계 방법 중 우수성이 확인된 ICF 방법과 TCF 방법 각각에 대하여 대칭적 방식을 제시하고 모의실험을 통해 기존의 비대칭적 ICF 방법 및 TCF 방법과 비교하여 그 기능적 차이를 비교 분석하는 데 있다. 특히, ICF 및 TCF 방법의 최소화 준거함수와 관련한 요인들(가중치 분포, 능력 지점의 수 등)에 대해 그 요인들의 수준을 변화시켰을 때 척도연계의 정확성이 어떻게 달라지는지를 집중 탐구하였다. 이를 통해 척도연계의 실제에서 각 요인에 대한 최적의 수준은 어떻게 지정되어야 할지를 탐색하였다. 또한 제시된 대칭적 ICF 및 TCF 척도연계 방법이 대칭성 조건에 부합하는 척도연계 상수를 산출하는지를 모의실험 자료를 사용하여 실증적으로 확인하고자 하였다.

Ⅱ. 이론적 배경

1. MIRT 모형

UIRT 모형에서 확장된 MIRT 모형은 각 차원에서의 능력 수준이 다른 차원에서의 능력 수준에 의하여 상호 보완되는지에 따라 보완 모형(compensatory model)과 비보완/부분보완 모형(noncompensatory or partially compensatory model)으로 크게 나눌 수 있다(Reckase, 2009). 두 모형 중 모수 추정이 더 용이한 보완 모형에 대해 연구가 많이 이루어져 왔으며, 대표적인 보완 모형으로는 M2PL 모형과 M3PL 모형(Reckase, 1985, 2009)이 있다. 본 연구에서는 M2PL 모형에 대한 대칭적 및 비대칭적 척도연계 방법을 다룬다. m-차원 능력 공간에서 M2PL 모형은 이분 채점 문항 i에 대한 피험자 j의 정답 확률을 다음과 같이 나타낸다.

$$P(u_{ij}=1|\boldsymbol{\theta}_j, \mathrm{a}_i, d_i) = \frac{1}{1+\exp[-D(\mathrm{a}_i{}'\boldsymbol{\theta}_j + d_i)]}, \tag{1}$$

여기서 u_{ij}는 문항 i에 대한 피험자 j의 점수(문항을 맞혔을 때 u_{ij}=1, 틀렸을 때 u_{ij}=0)이고; $\mathrm{a}_i = (a_{i1}, a_{i2}, ..., a_{im})'$로서 문항 i에 대한 $m \times 1$ 변별도 모수 벡터이며; d_i는 문항 i의 곤

란도 관련 스칼라 모수이고; $\boldsymbol{\theta}_j = (\theta_{j1}, \theta_{j2}, \ldots, \theta_{jm})'$로서 각 차원에 대한 피험자 j의 능력수준을 나타내는 $m \times 1$ 벡터이다. 척도상수 D는 정규 오자이브(normal ogive) 모형에서의 척도 단위와 최대한 일치하도록 1.7 혹은 1.702로 지정되곤 한다.

일차원 2PL 모형에 상응하도록, M2PL 모형의 변별도 모수 벡터와 곤란도 관련 모수는 다차원 변별도(multidimensional discrimination: MDISC) 모수와 다차원 곤란도(multidimensional difficulty: MDIFF) 모수로 다음과 같이 재정의할 수 있다(Reckase, 2009).

$$\mathrm{MDISC}_i = \sqrt{\mathbf{a}_i{}'\mathbf{a}_i} \;\; ; \;\mathrm{MDIFF}_i = \frac{-d_i}{\sqrt{\mathbf{a}_i{}'\mathbf{a}_i}} \tag{2}$$

2. 척도연계와 대칭성

공통문항을 이용한 척도연계를 서술하기 위하여 서로 다른 피험자 집단이 동형검사 N과 O에 응시하여 얻은 문항반응 자료를 분석하여 N형 검사의 모수들을 O형 검사의 능력척도 상에 연계한다고 하자. 그리고 두 검사형은 동일한 m개의 구인을 측정하고, n개의 공통문항을 가지며, N형 검사와 O형 검사의 능력척도는 각각 $\boldsymbol{\theta}_N$과 $\boldsymbol{\theta}_O$로 표기한다고 하자. IRT 모수 불변성에 따르면, 공통문항 i에 대해 식 (1)의 정답 유목에 대한 반응함수의 값(즉, 확률)은 N형 검사의 능력척도와 문항모수를 사용하든 O형 검사의 능력척도와 문항모수를 사용하든 동일한 값을 가진다. 이를 형식적으로 표현하면(능력벡터에서 특정 피험자를 나타내는 하위첨자 j를 생략),

$$P(\boldsymbol{\theta}_N; \mathbf{a}_{Ni}, d_{Ni}) = P(\boldsymbol{\theta}_O; \mathbf{a}_{Oi}, d_{Oi}). \tag{3}$$

위 식 (3)이 성립하기 위해서는 두 능력척도와 문항모수는 다음의 선형 관계를 지녀야 한다.

$$\boldsymbol{\theta}_O = \boldsymbol{\theta}_N^* = \mathrm{A}\,\boldsymbol{\theta}_N + \boldsymbol{\beta}, \tag{4}$$

$$\mathbf{a}'_{Oi} = (\mathbf{a}_{Ni}^*)' = \mathbf{a}'_{Ni}\mathrm{A}^{-1}, \tag{5}$$

$$d_{Oi} = d_{Ni}^* = d_{Ni} - \mathbf{a}'_{Ni}\mathrm{A}^{-1}\boldsymbol{\beta}, \tag{6}$$

여기서 별표(*)가 달린 기호들은 모두 θ_O 척도 상으로 변환된 척도 혹은 문항모수를 나타내고, A는 두 척도의 단위와 방향을 조정하는 $m \times m$ 회전행렬(rotation matrix)이며, β는 두 척도의 원점의 위치를 조정하는 $m \times 1$ 이동벡터(translation vector)이다(Davey et al., 1996). A와 β는 통틀어 척도연계 상수라 불린다.

위 식 (4)~(6)은 θ_O 척도를 기준으로 하여 두 능력척도와 문항모수 간의 선형 관계를 표현한 것이다. 이와 반대로, 식 (3)은 θ_N 척도를 기준으로 하여 두 능력척도와 문항모수 간의 관계를 다음과 같이 표현할 수 있음을 의미한다.

$$\theta_N = \theta_O^{\#} = \mathrm{A}^{-1}\theta_O + (-\mathrm{A}^{-1}\beta), \tag{7}$$

$$\mathrm{a}'_{Ni} = (\mathrm{a}_{Oi}^{\#})' = \mathrm{a}'_{Oi}\mathrm{A}, \tag{8}$$

$$d_{Ni} = d_{Oi}^{\#} = d_{Oi} + \mathrm{a}'_{Oi}\beta, \tag{9}$$

여기서 샤프(#)가 달린 기호들은 모두 θ_N 척도 상으로 변환된 척도 혹은 문항모수를 나타낸다. 식 (4)에 비추어 볼 때, 식 (7)에서의 A^{-1}과 $-\mathrm{A}^{-1}\beta$은 θ_O 척도를 θ_N 척도로 변환할 때의 회전행렬과 이동벡터에 해당한다.

위 식 (5)와 (6) 그리고 식 (8)과 (9)는 n개의 모든 공통문항에 대해 성립하는 유일한 척도연계 상수 A와 β가 존재함을 가정한다. 이러한 존재성은 IRT 모형이 자료에 완전히 부합하고, 문항모수가 표집의 오차 없이 추정될 때(즉, 문항모수가 모집단 모수일 때) 성립한다. 그러나 실제에서 IRT 모형이 검사 자료에 완전히 부합할 가능성은 거의 없으며, 문항모수는 표본 자료를 통해 추정되는 것이 일반적이다. 따라서 공통문항에 대한 기준 척도 상의 문항모수(추정치)와 변환된 문항모수(추정치)는 완전히 일치하지 않고 얼마간의 차이(즉, 척도연계의 오차)가 존재할 수 있다. 서론에서 언급한 공통문항 기반 척도연계 방법들은 기본적으로 공통문항 전체에서 발생하는 척도연계의 오차를 최소화 하는 목적으로 척도연계 상수 A와 β를 추정한다(김성훈 외, 2016). 여러 척도연계 방법 중 본 연구에서 관심을 두고 있는 ICF 방법과 TCF 방법은 척도연계의 오차를 문항/검사 특성함수를 통해 포착한다는 특징을 지닌다.

서론에서 서술하였듯이, 척도연계의 대칭성은 연계될 두 능력척도 간에 고유한 일대일 대응관계가 성립함을 의미한다. 척도연계의 대칭성을 구체적으로 논의하기 위해, "θ_N 척도에서 θ_O 척도로"의 변환에 사용되는 척도연계 상수를 A_N와 β_N로 나타내고, 반대로 "θ_O 척도

에서 θ_N 척도로"의 변환에 사용되는 척도연계 상수를 A_O와 β_O로 나타내자. 식 (4)에 따르면, A_N=A, β_N=β이고, 식 (7)에 따르면, $\mathrm{A}_O=\mathrm{A}^{-1}$, $\beta_O=-\mathrm{A}^{-1}\beta$이다. 그런데 식 (7)은 식 (4)의 역함수로서, 두 식은 두 능력척도 간의 고유한 일대일 대응관계를 유일한 A와 β를 사용하여 나타내고 있다. 따라서 "대칭적" 척도연계 방법은 다음 관계에 있는 척도연계 상수 세트를 산출하는 방법을 의미한다(아래에서 I는 항등행렬이고, 0은 영벡터임).

$$\mathrm{A}_O = \mathrm{A}_N^{-1};\ \mathrm{A}_N\mathrm{A}_O = \mathrm{I} \tag{10}$$

$$\beta_O = -\mathrm{A}_N^{-1}\beta_N;\quad \mathrm{A}_N^{-1}\beta_N + \beta_O = 0 \tag{11}$$

3. MIRT ICF 및 TCF 방법

척도연계의 대칭성 요건이 중요함에도 불구하고, ICF 방법과 TCF 방법은 비대칭적 방식으로 제시되어 왔다(김성훈 외, 2016). 이 절에서는, 식 (4)에서처럼, "θ_N 척도에서 θ_O 척도로의" 연계에 필요한 척도연계 상수 A와 β를 추정하는 것을 전제로 각 특성함수 방법에 대해 비대칭적 방식을 수정한 대칭적 방식을 서술한다. 서술의 편의상, ICF/TCF 방법의 대칭적 및 비대칭적 방식을 "대칭적 ICF/TCF 방법"과 "비대칭적 ICF/TCF 방법"으로 부른다. 대칭적 및 비대칭적 ICF 방법과 TCF 방법은 컴퓨터 프로그램 STMIRT(Kim, 2017)에서 지원된다.

1) 대칭적 및 비대칭적 ICF 방법

ICF 방법은 두 능력척도 상에 표현된 공통문항의 문항특성함수(ICF)의 차이를 최소화하는 척도연계 상수를 추정하는 방법으로, UIRT의 Haebara 방법(Haebara, 1980)을 MIRT로 확장시킨 방법이다(Oshima et al., 2000). Kim & Kolen(2007)이 지적하였듯이, Haebara(1980)는 "대칭적" 척도연계 상수 A와 β를 구하기 위해 N형 검사와 O형 검사의 두 능력척도 모두를 사용하여 최소화 준거함수를 정의하였다. 이와 동일하게, MIRT 대칭적(S) ICF 방법을 위한 최소화 준거함수(f_{ICF-S})는 θ_O와 θ_N 척도 모두를 사용하여 다음과 같이 정의할 수 있다.

$$f_{ICF-S}(\mathrm{A},\beta) = I_O + I_N\ , \tag{12}$$

$$I_O = \frac{1}{nq_O}\sum_{\boldsymbol{\theta}_O}\sum_{i=1}^{n}\left[P\left(\boldsymbol{\theta}_O;\mathrm{a}_{Oi},d_{Oi}\right)-P\left(\boldsymbol{\theta}_O;\mathrm{a}_{Ni}^{*},d_{Ni}^{*}\right)\right]^2 W_{\boldsymbol{\theta}_O}, \tag{13}$$

$$I_N = \frac{1}{nq_N} \sum_{\theta_N} \sum_{i=1}^{n} \left[P(\theta_N; \mathrm{a}_{Ni}, d_{Ni}) - P(\theta_N; \mathrm{a}^{\#}_{Oi}, d^{\#}_{Oi}) \right]^2 W_{\theta_N}, \tag{14}$$

여기서 I_O은 θ_O 척도 상에서 포착되는 척도연계 오차를 반영하고, I_N은 θ_N 척도 상에서 포착되는 척도연계 오차를 반영하며; W_{θ_O}와 W_{θ_N}는 특정 θ(θ_O 혹은 θ_N) 지점에서의 O형과 N형 검사 각각에 응시한 피험자 집단의 능력분포의 밀도를 반영하는 가중치를 나타내고; q_O와 q_N은 각 능력척도에서 모든 θ 지점에 대한 가중치의 합을 나타낸다. 위 식에서 명시적으로 드러나지 않는 부분은 각 능력척도에서의 θ 지점의 수이다. 각 능력차원에서의 능력 지점의 수를 동일하게 K라 하면, 전체 θ 지점의 수는 K^m이다. 이와 달리 MIRT 비대칭적(NS) ICF 방법을 위한 최소화 준거함수(f_{ICF-NS})는, Oshima et al.(2000)이 제시하였듯이, 위 식 (13)의 I_O만을 사용하여 다음과 같이 정의할 수 있다.

$$f_{ICF-NS}(\mathrm{A}, \beta) = I_O \tag{15}$$

위 식 (12)와 식 (15)에서 정의된 대칭적 및 비대칭적 최소화 준거함수는 A와 β에 대한 비선형적 함수로서 그 최소화 해를 직접적으로 구할 수 없다. ICF 방법의 A와 β에 대한 최소화 해를 구하기 위해서는 다변인 quasi-Newton 방법과 같은 최적화 기법(Dennis & Schnabel, 1996)이 필요하다(김성훈, 2018). STMIRT 프로그램은 다변인 quasi-Newton 기법을 사용하여 대칭적 및 비대칭적 ICF 방법의 척도연계 상수를 구한다.

2) 대칭적 및 비대칭적 TCF 방법

TCF 방법은 두 능력척도 상에 표현된, n개 공통문항에 대한 검사특성함수(TCF)의 차이를 최소화하는 척도연계 상수를 추정하는 방법으로, UIRT의 Stocking-Lord 방법(Stocking & Lord, 1983)을 MIRT로 확장한 방법이다(Oshima et al., 2000). 이를 위해 ICF 방법에서처럼 최소화 준거함수가 사용된다. MIRT 대칭적 TCF 방법에 대한 최소화 준거함수(f_{TCF-S})는 다음과 같이 정의할 수 있다.

$$f_{TCF-S}(\mathrm{A}, \beta) = T_O + T_N, \tag{16}$$

$$T_O = \frac{1}{q_O} \sum_{\theta_O} \left[T(\theta_O) - T^*(\theta_O) \right]^2 W_{\theta_O}, \tag{17}$$

$$T_N = \frac{1}{q_N} \sum_{\boldsymbol{\theta}_N} \left[T(\boldsymbol{\theta}_N) - T^{\#}(\boldsymbol{\theta}_N) \right]^2 W_{\boldsymbol{\theta}_N}, \tag{18}$$

여기서 $T(\boldsymbol{\theta}_O) = \sum_{i=1}^{n} P(\boldsymbol{\theta}_O ; \mathrm{a}_{Oi}, d_{Oi})$, $T^{*}(\boldsymbol{\theta}_O) = \sum_{i=1}^{n} P(\boldsymbol{\theta}_O ; \mathrm{a}^{*}_{Ni}, d^{*}_{Ni})$, $T(\boldsymbol{\theta}_N) = \sum_{i=1}^{n} P(\boldsymbol{\theta}_N ; \mathrm{a}_{Ni}, d_{Ni})$, 그리고 $T^{\#}(\boldsymbol{\theta}_N) = \sum_{i=1}^{n} P(\boldsymbol{\theta}_N ; \mathrm{a}^{\#}_{Oi}, d^{\#}_{Oi})$는 공통문항 기반 검사특성함수를 나타내고, 나머지 기호들은 식 (13)과 (14)에서 정의된 것과 동일하다. 이와 달리, MIRT 비대칭적 TCF 방법에 대한 최소화 준거함수(f_{TCF-NS})는, Oshima et al.(2000)이 제시하였듯이, 식 (17)의 T_O만으로 다음과 같이 정의할 수 있다.

$$f_{TCF-NS}(\mathrm{A}, \beta) = T_O \tag{19}$$

ICF 방법에서처럼, 위 식 (16)과 (19)에서 정의된 대칭적 및 비대칭적 TCF 방법의 최소화 준거함수는 A와 β에 대한 비선형적 함수로서 최소화 해를 직접적으로 구할 수 없으며, 그 해를 구하기 위해서는 다변인 최적화 기법이 필요하다. 두 TCF 방법의 척도연계 상수를 구하기 위해 STMIRT 프로그램을 사용할 수 있다.

Ⅲ. 모의실험 연구

앞서 서술한 대칭적 및 비대칭적 ICF 방법과 TCF 방법의 특성과 상대적 기능을 검토하기 위해 다양한 척도연계 조건을 형성하여 모의실험을 실시하였다. 김성훈 외(2016)에서처럼, 각 척도연계 조건에서 “네 가지” 척도연계 방법(대칭적 및 비대칭적 ICF 방법, 대칭적 및 비대칭적 TCF 방법)을 100번 반복 실시하여 각 방법이 척도연계(표본 자료의 능력척도 θ_N에서 모의실험용 능력척도 θ_O으로의 연계)를 통해 모의실험용 검사의 문항모수와 이 검사에 응시한 피험자 집단의 다차원 능력분포를 얼마나 잘 복원하는지를 분석하였다. 또한 모의실험 자료를 사용하여 각 척도연계 방법의 척도연계 수행에서의 대칭성 혹은 비대칭성을 확인하였다. 모의실험 연구의 구체적인 설계와 분석은 다음과 같았다.

1. 검사의 구성

모의실험에 사용한 검사는 두 개의 구인을 측정하되, 모의실험 요인 중 공통문항의 비율

에 따른 척도연계 결과의 차이를 확인함에 있어서 전체 문항의 수(즉, 검사의 길이)가 충분하도록 총 50개의 이분 채점 문항으로 구성하였다. M2PL 모형 기반 척도연계를 수행하기 위해 각 문항의 변별도 모수(a_1, a_2)와 곤란도 관련 모수(d)를 생성하였다. 연구의 편의상 전체 문항을 10 문항씩 다섯 개의 문항군으로 나누고 각 문항군이 통계적으로 최대한 유사한 특성을 가지도록 문항모수를 생성하였다. 그 중 첫 두 개의 문항군을 공통 문항으로 사용하였으며, 각 문항군에 동일하게 적용한 M2PL 문항모수 생성 절차는 다음과 같았다.

기하학적으로, M2PL 모형의 변별도 모수는 이차원 능력 공간 (θ_1, θ_2) 상에서 원점과의 거리를 나타내는 MDISC와 방향코사인(direction cosine) 값에 의해 결정되며, 곤란도 관련 모수는 MDISC와 MDIFF에 의해 결정된다[식 (2) 참고]. θ_1 축에 대한 방향코사인 값을 $\cos\alpha(0° < \alpha < 90°)$라 할 때, MDISC, MDIFF, α를 먼저 생성한 후 다음의 두 식을 이용하여 M2PL 문항모수를 생성하였다(이차원이므로 θ_2 축에 대한 방향코사인 값은 $\sin\alpha$).

$$a_{i1} = \mathrm{MDISC}_i \times \cos\alpha_i \ ; \ a_{i2} = \mathrm{MDISC}_i \times \sin\alpha_i \tag{20}$$

$$d_i = -\mathrm{MDIFF}_i \times \mathrm{MDISC}_i \tag{21}$$

구체적으로, 각 문항군에 대해 10개의 문항이 능력 공간에 고르게 퍼지도록[즉, 혼합구조(mixed structure)를 띠도록] θ_1 축의 양의 방향과 이루는 각의 크기를 기준으로 하여 능력 공간을 (0°, 18°), [18°, 36°), [36°, 54°), [54°, 72°), [72°, 90°)의 다섯 구간으로 나눈 후, 각 구간을 균일분포의 구간으로 하여 α를 2회씩 랜덤 추출하였다. MDISC의 경우, 모의실험용 검사가 현실성을 지니면서 동시에 문항모수 추정이 적절히 이루어지게끔 균일분포 U(0.6, 1.8)에서 랜덤 추출하였다. UIRT의 곤란도 모수에 대응하는 MDIFF는 표준 정규분포 N(0,1)에서 랜덤 추출하였다. 이상과 같이 생성된 모의실험용 검사의 M2PL 모형 문항모수들에 대한 기술통계를 공통 문항과 전체 문항에 대해 요약하여 제시하면 <표 1>과 같다.

2. 모의실험 요인

모의실험 요인은 공통문항의 비율, 최소화 준거함수를 정의하는 데 필요한 가중치의 분포 및 각 능력차원에 대한 능력 지점의 수 등을 고려하였으며, 각 요인의 수준은 다음과 같이 설정하였다. 첫째, 공통문항의 비율(proportion of common items: PCI)은 전체 검사 길이의 20%와 40%의 두 수준으로 설정하였다. PCI 20%는 공통문항 동등화 설계에서 권고되는 최소한

〈표 1〉 모의실험용 검사의 공통 문항과 전체 문항에 대한 M2PL 문항모수의 기술 통계

문항군		a_1	a_2	d	MDISC	MDIFF	α
1~10	M	0.81	0.79	-0.05	1.25	-0.02	44.72
	SD	0.40	0.41	1.20	0.23	0.95	26.06
	MIN	0.14	0.04	-2.43	0.82	-1.69	2.34
	MAX	1.35	1.34	2.28	1.59	1.76	81.96
11~20	M	0.80	0.76	-0.07	1.24	0.00	43.45
	SD	0.43	0.45	1.22	0.27	0.94	26.73
	MIN	0.05	0.01	-2.45	0.84	-1.33	0.36
	MAX	1.44	1.45	1.49	1.74	1.41	87.90
1~50	M	0.79	0.77	-0.03	1.22	0.00	45.08
	SD	0.44	0.40	1.12	0.29	0.95	26.43
	MIN	0.01	0.00	-2.45	0.66	-1.85	0.07
	MAX	1.48	1.45	2.28	1.74	1.76	89.50

주. M=평균; SD=표준편차; MIN=최솟값; MAX=최댓값.

의 길이다(Angoff, 1971; Kolen & Brennan, 2014). PCI 40%는 공통문항 비율의 증가에 따른 척도연계 결과의 변화를 검토하기 위해 선정하였다.

둘째, 최소화 준거함수를 정의하는 데 필요한 가중치(weight: WGT)는 능력분포의 반영 여부에 따라 이변량 정규(bivariate normal: BVN) 가중치와 균일(uniform: UNF) 가중치의 두 수준으로 정하였다. Oshima et al.(2000)의 연구를 비롯한 기존의 MIRT 척도연계 연구들은 ICF 방법 혹은 TCF 방법을 사용함에 있어서 동일한 가중치를 사용하여 척도연계를 수행하였다. 그러나 사용하는 가중치에 따라 다른 척도연계 결과가 산출될 수 있기(Kim & Kolen, 2007) 때문에 균일분포 가중치를 사용하는 경우와 능력분포를 반영한 가중치를 사용하는 경우를 비교하여 가중치의 유형에 따른 척도연계 결과의 차이를 살펴보고자 하였다.

셋째, 최소화 준거함수의 정의를 위한 각 능력차원의 능력 지점의 수(number of ability points: NP)는 5, 9, 15의 세 수준으로 설정하였다. 각 능력차원에 대해, MIRT 척도연계 프로그램인 IPLINK(Lee & Oshima, 1996)는 11개의 능력 지점을, STMIRT는 7개의 능력 지점을 디폴트로 사용한다. 본 연구에서는 두 프로그램의 디폴트 값의 평균인 9개를 기준으로 하여 7개보다 작은 5개와 11개보다 큰 15개를 추가적으로 선정함으로써 능력 지점의 수의 변화에 따른 척도연계 결과의 차이를 검토하였다.

3. 모의실험 절차

모의실험용 검사의 M2PL 문항모수가 θ_O 척도 상에 표현되어 있고, 이 검사의 표본 자료가 얻어졌다고 하자. MIRT에서는 척도 미결정성을 해결하기 위하여 표본 피험자의 능력분포가 표준 BVN 분포(평균=0, 분산=1, 공분산=0)라고 임의로 가정한 채 문항모수를 추정하는 것이 일반적이다(김성훈 2018; 김성훈 외, 2016). 이렇게 임의로 추정된 0－I 척도를 θ_N 척도라 하면, 이 척도는 θ_O 척도와 일반적으로 대등하지 않으며, 따라서 “θ_N 척도에서 θ_O 척도로”의 척도연계의 필요성이 발생한다. 이 점을 감안하여, 모의실험용 검사에 대해 “표본 자료의 능력척도(θ_N)에서 모의실험용 능력척도(θ_O)로”의 척도연계를 수행하기 위해 피험자 모집단의 기저 능력분포를 BVN(μ_1=0.5, μ_2=0.5, σ_1^2=1.25, σ_2^2=0.8, σ_{12}=0.4) 분포로 설정하였다. 구체적으로, 두 능력변수의 분산을 분산=1과 비교하여 약 0.1 표준편차 다른 1.25와 0.8로 설정하였으며, 구인 간 상관이 높을수록 검사자료는 일차원성을 띠기 때문에(Ackerman, 1989), 두 구인의 상관이 적절히 존재하되 높지 않도록 상관계수를 0.4로 설정하였다(두 분산의 곱=1이므로 상관계수=공분산). 또한, 연계되는 두 능력척도의 평균이 서로 다르도록 모집단의 평균 벡터를 표준화 평균=0보다 0.5 표준편차만큼 높게 설정하였다.

이와 같이 설정한 피험자 모집단에서 피험자 표본을 추출하여 표본 검사자료를 생성한 후, 이 자료를 사용하여 θ_N 척도 상에서 추정한 문항모수들을 피험자 모집단의 능력척도 θ_O 상으로 연계하는 작업을 100번 반복하는 모의실험을 수행하였다. 이 때, 표본의 크기인 피험자 수는 비교적 정확한 MIRT 문항모수를 추정하기 위하여 요구되는 최소 요건인 2,000으로 선정하였다(Ackerman, 1994). 표본 자료 생성, 문항모수 추정 및 척도연계의 구체적 방법은 다음과 같았다.

1) 표본 검사자료의 생성

표본 검사자료는 2,000명의 피험자가 50-문항 모의실험용 검사에 응시하였을 때 산출되는 2,000×50 문항점수들의 행렬이 되도록 모의생성 하였다. 검사자료 행렬에서 피험자 j의 문항 i에 대한 점수(u_{ij}=0 혹은 1)는 다음과 같이 산출하였다. 먼저 모집단 능력분포인 BVN (μ_1=0.5, μ_2=0.5, σ_1^2=1.25, σ_2^2=0.8, σ_{12}=0.4) 분포에서 피험자 j의 능력벡터 $\boldsymbol{\theta}_j$를 랜덤 추출하였다. 그 다음 균일분포 U(0,1)에서 난수 R을 생성하여 $P(\boldsymbol{\theta}_j;\mathrm{a}_i,d_i) \ge R$ 이면 u_{ij}에 1을, $P(\boldsymbol{\theta}_j;\mathrm{a}_i,d_i) < R$이면 u_{ij}에 0을 할당하였다. 표본 검사자료 생성에 필요한 모든 연산은 R 프로그램(R Development Core Team, 2018)을 사용하여 수행하였다.

2) 모수 추정 및 척도연계

표본 검사자료의 문항모수를 추정하기 위하여 TESTFACT 프로그램(Wood et al., 2003)을 사용하였다. 본 연구는 임의의 능력척도 상에 놓인 문항모수 추정치에 대하여 척도연계를 수행하는 것에 주안점을 두고 있으므로, 김성훈 외(2016)에서처럼, 좌표계 회전을 실시하지 않은 초기 문항모수 추정치를 표본 검사자료의 문항모수로 선택하였다.

추정된 문항모수는 임의의 0－I 다차원 능력척도(즉, 평균=0, 분산=1, 공분산=0) 상에 놓이기 때문에 그 문항모수를 모집단의 능력척도로 연계하기 위하여 MIRT 척도연계를 수행하였다. 모의실험 요인인 공통문항 비율, 가중치 유형 및 능력 지점의 수를 교차한 2×2×3=12개의 척도연계 조건 각각에서 대칭적 및 비대칭적 ICF 방법과 대칭적 및 비대칭적 TCF 방법의 네 가지 척도연계 방법을 사용하여 "θ_N 척도에서 θ_O 척도로"의 척도연계를 수행하였다. 이 척도연계에서 20%의 공통문항은 1~10번이었으며, 40%의 공통문항은 1~20번 문항이었다. 그리고 최소화 준거함수의 정의와 관련하여, UNF 가중치의 경우 －3부터 3까지 동일 간격을 가지는 5, 9, 혹은 15개의 능력 지점을 설정하였으며, BVN 가중치의 경우 Gauss-Hermite 기법을 사용하여 능력 지점과 가중치를 설정하였다. 이 모든 척도연계 작업은 STMIRT 프로그램을 사용하여 수행하였다.

4. 분석 및 평가

1) 기저 능력분포의 복원

MIRT 척도연계의 핵심은 두 능력척도를 연계하는 회전행렬 A와 이동벡터 β의 추정이다. 그러나 다차원 능력 공간에서 발생하는 회전에 대한 미결정성으로 인해 추정된 회전행렬을 그 자체로 분석하는 것은 적절하지 않다(김성훈, 2018; Oshima et al., 2000). 이 문제를 해결하기 위해, 김성훈(2018)의 연구에서처럼, A와 β의 복원 정도를 직접 분석하는 대신 다음 식 (22)에 기초하여 모집단 능력분포의 분산-공분산 행렬(Σ_{θ_O})과 평균 벡터(μ_{θ_O})의 복원 정도를 분석하였다.

$$\mu_{\theta_O} = \mathrm{A}\mu_{\theta_N} + \beta = \beta \;\; ; \;\; \Sigma_{\theta_O} = \mathrm{A}\Sigma_{\theta_N}\mathrm{A}' = \mathrm{AA}', \tag{22}$$

여기서 0－I 추정에 의하여 $\mu_{\theta_N} = 0$이고 $\Sigma_{\theta_N} = \mathrm{I}$이다. 각 척도연계 방법에 대해, 모집단 능력분포의 복원을 평가하기 위하여 각 척도연계 조건에서 산출된 100개의 능력변수의 평균(μ_1, μ_2), 분산(σ_1^2, σ_2^2) 및 공분산(σ_{12})에 대한 추정치들의 BIAS와 RMSE를 계산하였다. 구체적으로,

μ_1로 예를 들면, 능력분포 모수의 복원을 검토하기 위한 BIAS와 RMSE는 다음과 같이 산출하였다(아래 식에서, $\hat{\mu}_{1r}=r$번째 반복시행에서 얻어진 μ_1에 대한 추정치).

$$\mathrm{BIAS}(\hat{\mu}_1)=\sum_{r=1}^{100}\frac{1}{100}\left(\hat{\mu}_{1r}-\mu_1\right);\quad \mathrm{RMSE}(\hat{\mu}_1)=\sqrt{\sum_{r=1}^{100}\frac{1}{100}\left(\hat{\mu}_{1r}-\mu_1\right)^2} \tag{23}$$

2) **문항모수의 복원**

기저 능력분포의 복원과 더불어 척도연계를 통해 문항모수가 얼마나 잘 복원되었는지를 평가하였다. 모든 척도연계 조건에서 동일한 문항군에 대해 문항모수의 복원 정도를 평가하기 위해 21번~50번의 30개의 비공통 문항들의 모수를 평가 대상으로 하였다. 문항모수의 복원은 김성훈(2018)의 연구에서처럼 30개 문항 전체에 대한 절대 BIAS와 RMSE의 요약통계량인 RASB(root of average squared bias)와 RAMSE(root of average MSE)를 계산하여 평가하였다. 다음의 식은 문항모수 a_1에 대한 RASB와 RAMSE를 계산한 공식이며, 나머지 두 모수 a_2와 d에도 동일하게 적용하였다.

$$\mathrm{RASB}(\hat{a}_1^*)=\sqrt{\frac{1}{30}\sum_{i=21}^{50}\left[\sum_{r=1}^{100}\frac{1}{100}\left(\hat{a}_{i1r}^*-a_{i1}\right)\right]^2};\ \mathrm{RAMSE}(\hat{a}_1^*)=\sqrt{\frac{1}{30}\sum_{i=21}^{50}\sum_{r=1}^{100}\frac{1}{100}\left(\hat{a}_{i1r}^*-a_{i1}\right)^2}, \tag{24}$$

여기서 $\hat{a}_{1r}^*$은 r번째 반복 시행에서 얻어진 θ_O 척도 상의 a_{i1}^*의 추정치이다. 위 식에서 알 수 있듯이, RASB와 RAMSE는 30개의 비공통 문항에 대하여 계산한 평균적인 절대 BIAS와 RMSE이다.

3) **척도연계의 대칭성 분석**

능력분포 및 문항모수의 복원에 대한 평가와 별도로, 네 가지 척도연계 방법(대칭적 및 비대칭적 ICF 방법과 대칭적 및 비대칭적 TCF 방법) 각각이 대칭적인 척도연계 상수를 산출하는지 비대칭적 척도연계 상수를 산출하는지를 분석하였다. 이 분석을 위해, 각 방법에 대해, 모의실험의 각 척도연계 조건에서 첫 번째 반복시행 자료에 대해 “$\boldsymbol{\theta}_O$ 척도에서 $\boldsymbol{\theta}_N$ 척도로”의 척도연계를 추가적으로 실시하여 회전행렬 A_O와 이동벡터 β_O를 구하였다. 이미 실시된 “$\boldsymbol{\theta}_N$ 척도에서 $\boldsymbol{\theta}_O$ 척도로”의 척도연계에서 산출된 회전행렬과 이동벡터를 A_N과 β_N라 하면, 각 척도연계 방법에 대해 척도연계의 대칭성, 즉 식 (10)에서의 $\mathrm{A}_N\mathrm{A}_O=\mathrm{I}$와 식 (11)에서의 $\mathrm{A}_N^{-1}\beta_N+\beta_O=\mathbf{0}$이 성립하는지를 평가하였다.

Ⅳ. 연구결과

모의실험의 결과는 모집단의 능력분포 모수 및 문항모수의 복원과 척도연계의 대칭성 분석의 두 측면으로 나누어 정리하였다. 능력분포 모수의 복원 결과는 <표 2>와 <표 3>에, 문항모수의 복원 결과는 <표 4>에, 척도연계의 대칭성 분석 결과는 <표 5>에 제시하였다. 각 표에서 대칭적 ICF 및 TCF 방법은 ICF-S와 TCF-S로 표기하였고, 비대칭적 방법은 ICF-NS와 TCF-NS로 표기하였다.

1. 능력분포 모수 및 문항모수의 복원

1) 능력분포 모수의 복원 결과

능력분포 모수에 대한 BIAS 결과를 살펴보면, <표 2>에서 보듯이, 일부 조건을 제외하고, 두 평균(μ_1, μ_2)에 대한 BIAS가 0.05 이내로 매우 작았다. 이는 모든 방법에서 이동벡터가 정확하게 추정되었음을 의미한다. 이와 달리 공분산(σ_{12})의 BIAS는 0.1 이내이고, 두 분산(σ_1^2, σ_2^2)의 BIAS는 0.2 이내였다. 이것은 회전행렬의 추정 오차는 이동벡터보다 다소 큼을 나타낸다. 또한, 두 분산과 공분산의 BIAS의 부호는 모두 음수로 다소 과소 복원되었다.

모의실험 요인과 관련하여, 공통문항의 비율이 증가함에 따라 첫 번째 분산의 BIAS의 크기는 크게 증가하는 경향을 보였으며, 나머지 네 모수의 BIAS는 비슷한 수준을 유지하였다. 가중치의 유형과 관련해서는 UNF 가중치를 사용할 때 각 모수의 복원에 대한 네 방법은 유사한 패턴을 보인 반면, BVN 가중치를 사용할 때, 두 ICF 방법과 두 TCF 방법의 패턴이 다르게 나타났다. 두 ICF 방법의 경우 두 분산의 BIAS가 고르게 나타난 반면, 두 TCF 방법의 경우 첫 번째 분산의 BIAS는 작으나 두 번째 분산의 BIAS는 매우 크게 나타났다. 능력 지점의 수에 따른 BIAS를 살펴보면, 두 ICF 방법은 능력 지점의 수가 증가함에 따라 BIAS가 감소하지만 전반적인 수준에는 큰 차이가 없었다. 반면, 두 TCF 방법의 BIAS는 능력 지점의 수에 따라 다소 큰 차이를 보였는데, 특히 BVN 분포를 사용하는 경우에 능력 지점의 수에 따라 BIAS는 큰 차이를 보였다. 또한, 대칭성과 관련하여 복원 결과를 살펴보면, 전반적으로 BVN 가중치를 사용하는 경우에는 대칭적 방법이 더 작은 BIAS를 산출했으나, UNF 가중치를 사용하는 경우 대칭성에 따른 방법 간 BIAS의 차이는 뚜렷한 패턴을 찾기 힘들었다.

능력분포 모수에 대한 RMSE를 살펴보면, <표 3>에서 보듯이, 전반적인 RMSE의 크기는 BIAS의 크기와 비례하였다. 또한 BIAS의 결과에서처럼, 두 ICF 방법 간 그리고 두 TCF 방법 간에 유사한 RMSE 결과를 산출하였다.

〈표 2〉 능력변수의 평균과 분산-공분산 복원 결과 (BIAS)

조건	방법	PCI=20%					PCI=40%				
		μ_1	μ_2	σ_1^2	σ_2^2	σ_{12}	μ_1	μ_2	σ_1^2	σ_2^2	σ_{12}
WGT=UNF											
NP=5	ICF-NS	.009	.003	-.094	-.151	-.090	.016	.008	-.137	-.131	-.086
	ICF-S	.013	.002	-.096	-.147	-.085	.016	.005	-.150	-.131	-.080
	TCF-NS	.032	-.029	-.098	-.156	-.095	.019	-.008	-.136	-.136	-.084
	TCF-S	.035	-.036	-.103	-.158	-.093	.021	-.015	-.147	-.139	-.080
NP=9	ICF-NS	.011	.005	-.092	-.148	-.089	.014	.009	-.134	-.132	-.085
	ICF-S	.012	.006	-.090	-.143	-.083	.016	.007	-.141	-.129	-.077
	TCF-NS	.033	-.033	-.101	-.160	-.094	.021	-.012	-.129	-.142	-.084
	TCF-S	.031	-.033	-.102	-.161	-.089	.020	-.014	-.134	-.142	-.079
NP=15	ICF-NS	.011	.005	-.089	-.148	-.089	.014	.008	-.133	-.132	-.085
	ICF-S	.012	.006	-.087	-.143	-.082	.016	.007	-.139	-.129	-.076
	TCF-NS	.035	-.036	-.096	-.163	-.095	.022	-.015	-.123	-.145	-.085
	TCF-S	.031	-.034	-.097	-.163	-.089	.019	-.014	-.128	-.144	-.079
WGT=BVN											
NP=5	ICF-NS	.017	-.011	-.090	-.121	-.084	.015	-.003	-.129	-.107	-.073
	ICF-S	.016	-.003	-.096	-.111	-.074	.017	.000	-.126	-.103	-.067
	TCF-NS	.066	-.083	-.022	-.180	-.114	.043	-.046	-.038	-.176	-.091
	TCF-S	.036	-.043	-.050	-.163	-.094	.024	-.021	-.056	-.152	-.084
NP=9	ICF-NS	.017	-.010	-.090	-.123	-.082	.016	-.003	-.121	-.107	-.072
	ICF-S	.016	.001	-.090	-.107	-.069	.017	.002	-.117	-.099	-.063
	TCF-NS	.042	-.055	-.061	-.175	-.092	.036	-.039	-.055	-.165	-.087
	TCF-S	.026	-.029	-.051	-.163	-.087	.017	-.012	-.046	-.152	-.082
NP=15	ICF-NS	.017	-.010	-.086	-.123	-.083	.016	-.003	-.119	-.108	-.073
	ICF-S	.016	.001	-.086	-.107	-.070	.017	.002	-.116	-.099	-.063
	TCF-NS	.038	-.049	-.056	-.173	-.094	.032	-.035	-.049	-.164	-.089
	TCF-S	.026	-.029	-.045	-.165	-.089	.015	-.010	-.034	-.157	-.085

주. PCI = 공통문항의 비율; WGT=가중치 분포; NP=능력 지점의 수.

〈표 3〉 능력변수의 평균과 분산-공분산 복원 결과 (RMSE)

		PCI=20%					PCI=40%				
조건	방법	μ_1	μ_2	σ_1^2	σ_2^2	σ_{12}	μ_1	μ_2	σ_1^2	σ_2^2	σ_{12}
WGT=UNF											
NP=5	ICF-NS	.046	.044	.167	.165	.101	.041	.035	.154	.140	.092
	ICF-S	.038	.037	.154	.161	.096	.036	.029	.162	.140	.085
	TCF-NS	.072	.086	.146	.174	.106	.061	.071	.160	.149	.091
	TCF-S	.066	.083	.148	.173	.103	.053	.063	.165	.149	.086
NP=9	ICF-NS	.037	.037	.158	.162	.100	.036	.031	.149	.141	.091
	ICF-S	.035	.034	.142	.157	.094	.034	.027	.152	.138	.083
	TCF-NS	.064	.077	.151	.176	.105	.057	.068	.154	.153	.091
	TCF-S	.061	.076	.147	.176	.100	.053	.063	.154	.153	.085
NP=15	ICF-NS	.036	.035	.152	.163	.100	.036	.030	.147	.141	.091
	ICF-S	.035	.033	.137	.157	.093	.034	.026	.150	.137	.082
	TCF-NS	.065	.079	.145	.179	.105	.058	.070	.150	.156	.092
	TCF-S	.061	.076	.141	.178	.100	.053	.064	.149	.155	.085
WGT=BVN											
NP=5	ICF-NS	.034	.032	.116	.136	.095	.033	.029	.139	.118	.079
	ICF-S	.033	.028	.120	.124	.086	.033	.027	.135	.112	.073
	TCF-NS	.094	.125	.094	.196	.126	.084	.105	.100	.188	.099
	TCF-S	.064	.085	.099	.178	.104	.063	.079	.106	.162	.091
NP=9	ICF-NS	.034	.030	.116	.136	.093	.033	.027	.130	.117	.078
	ICF-S	.033	.028	.114	.120	.081	.033	.026	.126	.108	.070
	TCF-NS	.069	.093	.103	.190	.103	.070	.090	.106	.176	.094
	TCF-S	.058	.075	.097	.177	.098	.054	.068	.099	.162	.090
NP=15	ICF-NS	.034	.030	.114	.136	.094	.033	.026	.128	.117	.079
	ICF-S	.033	.028	.110	.120	.083	.033	.026	.124	.107	.070
	TCF-NS	.064	.085	.102	.187	.104	.064	.082	.110	.175	.097
	TCF-S	.056	.072	.095	.179	.099	.052	.065	.097	.167	.092

주. PCI = 공통문항의 비율; WGT=가중치 분포; NP=능력 지점의 수. 각 열에 해당하는 능력분포 모수별 산출된 가장 작은 RMSE 값에 밑줄 침.

공통문항의 비율 요인과 관련하여, 첫 번째 분산을 제외하면 나머지 네 모수에 대하여 모든 척도연계 조건에서 공통문항의 비율이 증가함에 따라 RMSE의 크기가 감소하였다. 첫 번째 분산의 경우, 공통문항의 비율이 증가할 때, UNF 가중치를 사용한 두 ICF 방법을 제외하고는 모두 RMSE가 증가하였다. 가중치의 유형에 따른 결과를 살펴보면, 다섯 능력분포 모수의 복원에 있어서 두 ICF 방법은 BVN 가중치를 사용할 때 더 작은 RMSE를 산출하였다. 이와 달리, 두 TCF 방법은 가중치와 관련한 뚜렷한 패턴을 보이지 않았다. 또한 전반적으로 능력 지점의 수가 증가할수록 더 작은 RMSE가 산출되었으며, 공통문항의 비율이 낮을수록 그 차이가 큰 경향을 보였다.

각 척도연계 방법에 대하여, 다섯 능력분포 모수의 복원에서 가장 작은 RMSE를 산출한 경우의 수를 열별로(10개 열) 세면, 비대칭적 ICF 방법은 1번, 대칭적 ICF 방법은 7번, 비대칭적 TCF 방법은 1번, 대칭적 TCF 방법은 1번으로 대칭적 ICF 방법이 능력분포 모수의 복원에 가장 우수함을 보였다. 또한, 방법별 산출된 가장 작은 RMSE 값은 모두 BVN 가중치를 사용할 때 얻어졌으며, 공통문항의 비율이 낮은 경우에는 능력지점의 수를 5 혹은 9로 정할 때, 공통문항의 비율이 높을 때는 9 혹은 15개로 정할 때 얻어졌다.

2) 문항모수의 복원 결과

문항모수에 대한 RASB와 RAMSE 결과를 차례대로 살펴보면 다음과 같았다. <표 4>의 RASB 열에서 보듯이, 세 유형의 문항모수의 RASB는 공통문항의 비율이 증가함에 따라 서로 다른 변화 양상을 보였다. 먼저, a_1 모수의 경우, 공통문항의 비율이 증가와 관계없이 비슷한 수준의 RASB가 산출되었으며, a_2 모수의 RASB는 공통문항의 비율이 증가함에 따라 감소하는 경향을 보였다. 반면, d 모수의 경우 공통문항의 비율이 증가함에 따라 RASB가 증가하는 경향을 보였다. 가중치의 유형과 관련하여, 전반적으로 두 변별도 모수의 경우 UNF 가중치를 사용할 때의 RASB가, d 모수의 경우 BVN 가중치를 사용할 때의 RASB가 더 작게 산출되었다. 네 방법 모두 능력 지점의 수에 따른 RASB의 차이는 거의 없었다.

<표 4>의 RAMSE 열에서 보듯이, 두 TCF 방법을 사용했을 때의 d 모수에 대한 복원 결과를 제외하면 공통문항의 비율이 증가할수록 더 작은 RAMSE가 산출되었다. 또한, 세 유형의 문항모수에 대해, 두 TCF 방법보다 두 ICF 방법을 사용하였을 때 산출된 RAMSE가 더 작았다. 능력 지점의 수와 관련하여, 각 방법에 대해 RAMSE는 능력 지점의 수에 따른 차이가 거의 없었으며, 일부 조건의 경우, 능력 지점의 수가 증가함에 따라 오히려 RAMSE가 증가하는 경향을 보였다.

각 방법에 대하여, 세 유형의 문항모수의 복원에서 가장 작은 RAMSE를 산출한 경우의 수를 열별로(6개의 열) 세면, 비대칭적 ICF 방법은 4번, 대칭적 ICF 방법은 2번, TCF 방법은 0

〈표 4〉 문항모수의 복원 결과

		RASB						RAMSE					
		PCI=20%			PCI=40%			PCI=20%			PCI=40%		
조건	방법	a_1	a_2	d	a_1	a_2	d	a_1	a_2	d	a_1	a_2	d
WGT=UNF													
NP=5	ICF-NS	.036	.036	.055	.036	.032	.058	.090	.083	.086	.084	.073	.084
	ICF-S	.036	.037	.054	.036	.030	.056	.089	.080	.084	.084	.072	.082
	TCF-NS	.036	.037	.048	.036	.032	.051	.092	.079	.087	.086	.074	.085
	TCF-S	.036	.036	.047	.036	.030	.049	.091	.078	.084	.086	.072	.082
NP=9	ICF-NS	.036	.037	.056	.036	.032	.058	.090	.082	.087	.084	.073	.084
	ICF-S	.037	.039	.054	.036	.032	.056	.089	.080	.084	.084	.072	.082
	TCF-NS	.036	.036	.048	.036	.032	.050	.092	.079	.084	.086	.074	.084
	TCF-S	.037	.035	.046	.036	.032	.049	.092	.078	.082	.086	.073	.082
NP=15	ICF-NS	.036	.038	.055	.036	.032	.057	.090	.082	.086	.084	.073	.083
	ICF-S	.037	.040	.054	.036	.032	.056	.089	.080	.084	.084	.072	.082
	TCF-NS	.037	.037	.047	.036	.033	.050	.092	.079	.084	.086	.074	.084
	TCF-S	.037	.036	.045	.036	.033	.048	.092	.078	.082	.086	.073	.082
WGT=BVN													
NP=5	ICF-NS	.039	.041	.045	.040	.036	.047	.087	.077	.075	.084	.073	.074
	ICF-S	.041	.041	.046	.041	.037	.047	.087	.077	.075	.084	.073	.074
	TCF-NS	.038	.051	.043	.040	.048	.043	.092	.083	.082	.090	.080	.086
	TCF-S	.038	.046	.042	.039	.046	.045	.091	.079	.079	.087	.079	.083
NP=9	ICF-NS	.039	.041	.045	.040	.038	.046	.087	.077	.075	.084	.073	.074
	ICF-S	.042	.044	.046	.042	.040	.046	.087	.078	.075	.084	.074	.074
	TCF-NS	.039	.043	.042	.039	.045	.043	.093	.078	.079	.088	.079	.083
	TCF-S	.039	.046	.044	.040	.048	.047	.091	.079	.079	.087	.080	.082
NP=15	ICF-NS	.039	.042	.046	.040	.038	.046	.087	.078	.075	.084	.074	.074
	ICF-S	.042	.045	.046	.042	.041	.046	.087	.078	.075	.085	.075	.074
	TCF-NS	.039	.044	.042	.039	.046	.044	.092	.079	.078	.088	.080	.082
	TCF-S	.040	.047	.044	.040	.050	.048	.091	.080	.079	.087	.081	.083

주. PCI = 공통문항의 비율; WGT=가중치 분포; NP=능력 지점의 수. 각 열에 해당하는 문항모수별 산출된 가장 작은 RAMSE 값에 밑줄 침.

번이었다. 이러한 결과는, 전반적으로 볼 때, 문항모수의 복원에 있어 비대칭적 ICF 방법이 가장 작은 오차를 산출하였음을 의미한다. 또한, 능력분포 모수의 복원과 마찬가지로 UNF 가중치를 사용하는 것보다 BVN 가중치를 사용할 때 보다 작은 척도연계 오차가 발생하였다. 능력 지점의 수에 따른 문항모수의 복원에는 큰 차이는 없었으나, 두 ICF 방법에 대하여, 대체로 5 혹은 9개의 능력 지점을 사용할 때 척도연계 오차가 더 작게 산출되는 경향을 보였다.

2. 척도연계의 대칭성 분석

<표 5>는 각 척도연계 조건에서 첫 번째 반복시행에서 얻은 자료를 사용하여 네 척도연계 방법의 대칭성을 분석할 결과를 제시하고 있다. 이 표에서 보듯이, 기대한 대로, 두 대칭적 방법은 $\mathrm{A}_N\mathrm{A}_O=\mathrm{I}$, $\mathrm{A}_N^{-1}\beta_N+\beta_O=0$의 결과를 산출하였다. 이와 달리 두 비대칭적 방법은 모든 척도연계 조건에서 $\mathrm{A}_N\mathrm{A}_O\neq\mathrm{I}$, $\mathrm{A}_N^{-1}\beta_N+\beta_O\neq 0$의 결과를 보이는 비대칭적 척도연계 상수를 산출하였다. 각 비대칭적 방법은 공통문항의 비율과 가중치에 따라 비대칭의 정도에서 약간의 차이를 보였지만, 각 요인별 뚜렷한 차이 패턴은 없었다. 주목할 만한 결과로, TCF 방법의 경우 BVN 가중치를 사용할 때 $\mathrm{A}_N^{-1}\beta_N+\beta_O$의 값이 0.1~0.2로 다소 컸다.

Ⅴ. 논의 및 제언

본 연구는 MIRT 척도연계 방법 중 ICF 방법과 TCF 방법을 대칭적으로 수정・제시한 후 기존의 비대칭적인 방법과 수정된 대칭적 방법의 기능적 차이를 모의실험을 통해 살펴보았다. 네 척도연계 방법(대칭적 및 비대칭적 ICF 방법, 대칭적 및 비대칭적 TCF 방법)에 대하여 능력분포 모수 및 문항모수를 복원하여 그 상대적 기능을 비교하였으며, 네 방법에 대한 대칭성을 분석하였다.

모의실험의 결과, 모집단의 능력분포 모수를 추정함에 있어 대칭적 ICF 방법이 가장 우수하며, 문항모수의 추정에는 대칭성과 관련하여 방법 간 차이는 거의 존재하지 않지만, ICF 방법이 TCF 방법보다 우수함을 확인할 수 있었다. 특히, ICF 방법에 대한 준거함수의 가중치로는 능력분포를 반영한 BVN 가중치를 사용할 때 척도연계의 정확성과 안정성을 보다 높일 수 있음을 확인하였다. 이러한 결과는 MIRT 척도연계에서 균일 가중치를 사용한 ICF 방법과 TCF 방법을 다룬 선행연구에서 찾아 볼 수 없는 것으로, MIRT ICF 및 TCF 방법에

〈표 5〉 척도연계의 대칭성 분석 결과

조건	방법	PCI=20% $\mathrm{A}_N\mathrm{A}_O$	PCI=20% $\mathrm{A}_N^{-1}\beta_N+\beta_O$	PCI=40% $\mathrm{A}_N\mathrm{A}_O$	PCI=40% $\mathrm{A}_N^{-1}\beta_N+\beta_O$
아래의 모든 조건	ICF-S TCF-S	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$
WGT=UNF					
NP=5	ICF-NS	$\begin{bmatrix}1.001 & -0.008\\-0.016 & 1.004\end{bmatrix}$	$\begin{bmatrix}0.021\\0.073\end{bmatrix}$	$\begin{bmatrix}1.037 & -0.020\\-0.029 & 1.023\end{bmatrix}$	$\begin{bmatrix}0.012\\0.058\end{bmatrix}$
	TCF-NS	$\begin{bmatrix}1.004 & -0.005\\-0.011 & 1.007\end{bmatrix}$	$\begin{bmatrix}0.025\\0.079\end{bmatrix}$	$\begin{bmatrix}1.012 & -0.017\\-0.010 & 1.022\end{bmatrix}$	$\begin{bmatrix}0.012\\0.085\end{bmatrix}$
NP=9	ICF-NS	$\begin{bmatrix}0.996 & -0.006\\-0.012 & 0.989\end{bmatrix}$	$\begin{bmatrix}0.017\\0.044\end{bmatrix}$	$\begin{bmatrix}1.019 & -0.013\\-0.030 & 1.013\end{bmatrix}$	$\begin{bmatrix}0.007\\0.046\end{bmatrix}$
	TCF-NS	$\begin{bmatrix}1.021 & -0.035\\-0.030 & 1.031\end{bmatrix}$	$\begin{bmatrix}0.022\\0.014\end{bmatrix}$	$\begin{bmatrix}1.010 & -0.025\\-0.017 & 1.024\end{bmatrix}$	$\begin{bmatrix}0.010\\-0.009\end{bmatrix}$
NP=15	ICF-NS	$\begin{bmatrix}0.995 & -0.005\\-0.011 & 0.986\end{bmatrix}$	$\begin{bmatrix}0.016\\0.041\end{bmatrix}$	$\begin{bmatrix}1.018 & -0.013\\-0.029 & 1.011\end{bmatrix}$	$\begin{bmatrix}0.007\\0.043\end{bmatrix}$
	TCF-NS	$\begin{bmatrix}1.024 & -0.039\\-0.034 & 1.032\end{bmatrix}$	$\begin{bmatrix}0.021\\-0.006\end{bmatrix}$	$\begin{bmatrix}1.011 & -0.029\\-0.019 & 1.027\end{bmatrix}$	$\begin{bmatrix}0.009\\-0.030\end{bmatrix}$
WGT=BVN					
NP=5	ICF-NS	$\begin{bmatrix}0.996 & 0.011\\-0.008 & 0.960\end{bmatrix}$	$\begin{bmatrix}0.007\\0.007\end{bmatrix}$	$\begin{bmatrix}1.000 & -0.005\\-0.002 & 0.982\end{bmatrix}$	$\begin{bmatrix}0.003\\0.007\end{bmatrix}$
	TCF-NS	$\begin{bmatrix}1.039 & -0.018\\-0.061 & 0.991\end{bmatrix}$	$\begin{bmatrix}-0.003\\-0.178\end{bmatrix}$	$\begin{bmatrix}0.997 & 0.005\\-0.003 & 0.972\end{bmatrix}$	$\begin{bmatrix}-0.001\\-0.097\end{bmatrix}$
NP=9	ICF-NS	$\begin{bmatrix}0.999 & 0.010\\-0.012 & 0.955\end{bmatrix}$	$\begin{bmatrix}0.005\\0.009\end{bmatrix}$	$\begin{bmatrix}0.999 & -0.005\\-0.012 & 0.982\end{bmatrix}$	$\begin{bmatrix}0.004\\0.007\end{bmatrix}$
	TCF-NS	$\begin{bmatrix}0.990 & -0.014\\-0.008 & 0.986\end{bmatrix}$	$\begin{bmatrix}0.000\\-0.132\end{bmatrix}$	$\begin{bmatrix}0.980 & -0.022\\0.004 & 0.997\end{bmatrix}$	$\begin{bmatrix}-0.001\\-0.178\end{bmatrix}$
NP=15	ICF-NS	$\begin{bmatrix}1.001 & 0.008\\-0.015 & 0.955\end{bmatrix}$	$\begin{bmatrix}0.005\\0.010\end{bmatrix}$	$\begin{bmatrix}0.997 & -0.003\\-0.016 & 0.983\end{bmatrix}$	$\begin{bmatrix}0.003\\0.008\end{bmatrix}$
	TCF-NS	$\begin{bmatrix}0.999 & -0.020\\-0.015 & 0.993\end{bmatrix}$	$\begin{bmatrix}0.000\\-0.110\end{bmatrix}$	$\begin{bmatrix}0.977 & -0.030\\0.005 & 1.004\end{bmatrix}$	$\begin{bmatrix}-0.001\\-0.196\end{bmatrix}$

주. PCI = 공통문항의 비율; WGT=가중치 분포; NP=능력 지점의 수.

의미 있는 시사점을 제공한다고 볼 수 있다. 또한, 비대칭적 방법만을 고려할 때, ICF 방법이 TCF 방법보다 능력분포 모수 및 문항모수의 복원에서 우수함을 확인할 수 있었다. 이러한 결과는 Simon(2008)과 김성훈(2018)의 연구와 일치하는 결과이다.

한편, 네 방법에 대하여 실제 대칭성이 성립하는지 확인해본 결과 대칭적으로 수정된 ICF 방법과 TCF 방법은 대칭성이 성립하였지만, 기존의 비대칭적 ICF 방법과 TCF 방법에는 대칭성이 성립하지 않았다. 김성훈 외(2016)에서 논의되었듯이 비대칭적인 척도연계를 수행하는 경우 두 능력척도 간 일대일 대응관계가 성립하지 않기 때문에 척도연계의 방향에 따라 문항모수와 능력모수는 왜곡 추정될 수 있다. 이러한 척도연계의 비대칭성으로 인해, Kolen & Brennan(2014)이 지적하였듯이, 척도연계와 동등화에 있어 체계적 오차가 발생할 수 있다. 따라서 이러한 체계적 오차를 줄이기 위해서는 비대칭적 방법보다는 대칭적 방법을 사용하는 것이 바람직할 것이다. 또한, 문항모수의 복원에서 비대칭적 방법이 대칭적 방법보다 작은 척도연계의 오차를 산출하였으나, 두 방법 간의 차이가 미미하였고 능력분포의 복원 및 두 능력척도 간 대칭성을 고려하면 대칭적 방법, 특히 대칭적 ICF 방법을 사용하여 척도연계를 수행하는 것이 바람직할 것이다. 물론 이러한 진술은 검사동등화의 관점에서 그러하다. 김성훈 외(2016)에서 논의되었듯이, 관심 공통의 척도가 기준 척도 하나이고, 척도연계의 목적이 그 공통 척도를 유지하는 데 있다면 문항모수의 복원에 우수성을 보인 비대칭적 방법을 사용해도 무방할 것이다. 사실 기존 연구의 대부분이 비대칭적 방법만을 제시한 것은 이런 이유와 관련되어 있을 것이다.

마지막으로 후속연구를 위한 제언을 하면 다음과 같다. 첫째, ICF 방법과 TCF 방법은 최소화 준거함수의 수정으로 대칭적 방법으로 확장될 수 있는데, 이는 M2PL 모형뿐만 아니라 다른 MIRT 모형에 대해서도 적용할 수 있다. 또한, MIRT 모형의 특수한 형태인 Bifactor 모형에 대해서도 동일한 방법을 사용하여 대칭적 ICF 및 TCF 방법을 제시할 수 있다. 이러한 여러 MIRT 모형에 대해 대칭적 방법과 비대칭적 방법이 어떤 기능적 차이를 보이는지 살펴볼 필요가 있을 것이다. 둘째, 본 연구에서 다루지 않은 MIRT 척도연계 방법인 최소제곱 방법, Li & Lissitz(2000)의 방법, Min(2003)의 방법은 모두 성격상 일반화 선형(회귀) 방법에 기초한 비대칭적인 방법으로 볼 수 있다. 이 세 가지 방법들에 대하여 대칭적 방법으로의 수정 가능성에 대하여 모색할 수 있다. 김성훈(2018)은 최소제곱 방법은 대칭적으로 수정될 수 있다고 하였으나, Li & Lissitz와 Min의 방법에 대해서는 그 가능성을 언급하지 않았다. 비대칭적 최소제곱 방법을 대칭적으로 어떻게 수정할지 그리고 Li & Lissitz와 Min의 방법을 대칭적 방법으로 수정 가능한지, 나아가 수정된 방법과 기존의 방법 간에는 어떠한 기능적 차이가 있는지 등에 대한 연구가 이루어질 필요가 있을 것이다. 셋째, 일반적으로 IRT 척도연계 연구는 기준 검사와 그 동형검사 각각에서 개별 추정된 공통 문항의 모수를 이용하여 수행된다. 이 경우, 두 검사의 모수 추정치는 회전의 미결정성으로 인해 모두 임의의 능력척도에 놓이게 되기 때문에 기준 검사의 모수 추정치를 기준 척도로 연계한 후, 동형검사의 모수 추정치를 기준 검사 추정치의 척도로 연계하는 작업이 뒤따를 필요가 있다. 즉, 척도연계를

기준 검사에서 한 번 그리고 동형검사에서 한 번 더 수행해야 하는 문제가 발생한다. 이러한 상황에서, 각 척도연계 방법을 두 번 사용한다고 할 때, 대칭적 방법과 비대칭적 방법 간에는 어떠한 기능적 차이가 있는지에 대한 연구가 이루어질 필요가 있을 것이다.

참고문헌

김성훈(2018). MIRT 등급반응 모형을 위한 공통-문항 척도연계 방법. **교육평가연구**, 31(2), 387-409.

김성훈, 민경석, 강동희(2016). 모수 복원 상황에서 MIRT 척도연계 방법의 기능 진단. **교육평가연구**, 29(4), 669-696.

Ackerman, T. A. (1989). Unidimensional IRT calibration of compensatory and noncompensatory multidimensional items. *Applied Psychological Measurement*, *13*, 113-127.

Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, *7*, 255-278.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, DC: American Council on Education.

Davey, T., Oshima, T. C., & Lee, K. (1996). Linking multidimensional item calibrations. *Applied Psychological Measurement*, *20*, 405-416.

de Ayala, R. J. (2008). *The theory and practice of item response theory*. New York, NY: Guilford Press.

Dennis, J. E., & Schnabel, R. B. (1996). *Numerical methods for unconstrained optimization and nonlinear equations*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Divgi, D. R. (1985). A minimum chi-square method for developing a common metric in item response theory. *Applied Psychological Measurement*, *9*, 413-415.

Dorans, N. J., & Holland, P. W. (2000). Population invariance and the equatability of tests: Basic theory and the linear case. *Journal of Educational Measurement*, *37*, 281-306.

Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, *22*, 144-149.

Hanson, B. A., & Béguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, *26*, 3-24.

Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 187-220). Westport, CT: American Council on Education and Praeger.

Kim, S. (2017). STMIRT: A computer program for scale transformation using multidimensional item response theory models (Version 1.0) [Computer software]. Author.

Kim, S., Harris, D. J., & Kolen, M. J. (2010). Equating with polytomous item response models. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 257-291). New York, NY: Routledge.

Kim, S., & Kolen, M. J. (2007). Effects on scale linking of different definitions of criterion functions for the IRT characteristic curve methods. *Journal of Educational and Behavioral Statistics*, *32*, 371-397.

Kim, S.-H., & Cohen, A. S. (1998). A Comparison of linking and concurrent calibration under item response theory. *Applied Psychological Measurement*, *22*, 131-143.

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York, NY: Springer.

Lee, K., & Oshima, T. C. (1996). IPLINK: Multidimensional and unidimensional item parameter linking in item response theory. *Applied Psychological Measurement*, *20*, 230.

Li, Y. H., & Lissitz, R. W. (2000). An evaluation of the accuracy of multidimensional IRT linking. *Applied Psychological Measurement*, *24*, 115-138.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Loyd, B. H., & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, *17*, 179-193.

Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, *14*, 139-160.

Min, K.-S. (2003). *The impact of scale dilation on the quality of the linking of multidimensional item response theory calibration* (Unpublished doctoral dissertation). Michigan State University.

Min, K.-S. (2007). Evaluation of linking methods for multidimensional IRT calibrations. *Asia Pacific Education Review*, *8*, 41-55.

Ogasawara, H. (2001). Least squares estimation of item response theory linking coefficients. *Applied Psychological Measurement*, *25*, 373-383.

Oshima, T. C., Davey, T. C., & Lee, K. (2000). Multidimensional linking: Four practical approaches. *Journal of Educational Measurement*, *37*, 357-373.

R Development Core Team. (2018). R: A language and environment for statistical computing (Version 3.4.4). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, *9*, 401-412.

Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer.

Schönemann, P. H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, *31*, 1-10.

Simon, M. K. (2008). *Comparison of concurrent and separate multidimensional IRT linking of item parameters* (Unpublished doctoral dissertation). University of Minnesota.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, *7*, 201-210.

Wood, R., Wilson, D. T., Gibbons, R., Schilling, S. G., Muraki, E., & Bock, R. D. (2003). TESTFACT: Test scoring, item statistics, and item factor analysis (Version 4) [Computer software]. Chicago, IL: Scientific Software International.

Yen, W. M., & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 111-153). Westport, CT: American Council on Education and Praeger.

◎ 논문접수: 2018. 7. 31 / 수정본 접수: 2018. 9. 13 / 게재승인: 2018. 9. 15

저 자 소 개

・김아름 : 한양대학교 교육학과 학사 및 석사과정 졸업. 교육측정 및 평가 전공. 관심 연구 분야는 검사동등화, 문항반응이론, 대규모 자료 분석 등임. areumkim4369@gmail.com

・김성훈 : 본 논문의 교신저자. The University of Iowa에서 교육측정 및 통계 전공으로 박사 학위 취득. 한양대학교 교육학과 교수로 재직. 관심 연구 분야는 검사동등화 및 척도연계, 문항반응이론, 신뢰도, 기준설정, 실험 설계 및 분석 등임. seonghoonkim@hanyang.ac.kr

〈ABSTRACT〉

Comparison of Symmetric Versus Nonsymmetric Characteristic-Function Scale Linking Methods for a Multidimensional Two-Parameter IRT Model

Areum Kim Seonghoon Kim
Hanyang University

In multidimensional item response theory (MIRT) scale linking, it has been shown that the item characteristic function (ICF) and test characteristic function (TCF) methods should perform better than other methods. The purpose of this study is to investigate the statistical properties and performances of the symmetric and nonsymmetric versions for each of the ICF and TCF methods. For this purpose, the symmetric versions of the ICF and TCF methods, which had been presented as nonsymmetric ones in previous studies, were first presented for use with the multidimensional two-parameter logistic (M2PL) model. Then, the relative performances of the four methods, symmetric and nonsymmetric ICF methods and symmetric and nonsymmetric TCF methods, were investigated through computer simulations under various scale linking conditions. The simulation study included three critical factors of (1) proportion of common items, (2) weighting scheme for the criterion functions, and (3) number of ability points in each dimension for the criterion functions. The performances of the four methods were evaluated in terms of accurate recovery of the parameters of the underlying population ability distribution and the M2PL item parameters for a 50-item multiple-choice test. Also, the four methods were analyzed as to whether they produced symmetric solutions or not. Main results were as follows. First, in the recovery of the underlying population distribution, the symmetric ICF method yielded the smallest linking error. Second, in the recovery of the item parameters, the nonsymmetric ICF method resulted in the smallest linking error. Third, as expected, symmetric ICF and TCF methods did produce symmetric linking coefficients, but nonsymmetric ICF and TCF methods did not.

Keywords : Multidimensional item response theory (MIRT), scale linking, symmetry, item characteristic function (ICF) method, test characteristic function (TCF) method