

## 다수 집단의 측정동일성 검정을 위한 임의효과 모형: 다층 확인적 요인분석(ML CFA)과 다층 요인혼합모형(ML FMM)의 비교

손 수 경                      김 효 진                      홍 세 희<sup>†</sup>

고려대학교 교육학과

집단 비교 연구 시 측정동일성의 성립 여부는 집단 간 의미 있는 비교를 하기 위한 필수 요건으로 제시되고 있다. 이를 위해 일반적으로 다집단 확인적요인분석(MG CFA)이 널리 사용되어 왔으나, MG CFA는 비교집단이 소수일 경우에 적합한 것으로, 많은 집단을 비교하게 되는 국가 비교 연구에서는 그 한계가 제기된다. 따라서 본 연구에서는 10개 이상의 많은 수의 국가(혹은 집단)를 비교하기에 적절한 방법론인 다층 확인적요인분석(ML CFA)과 다층 요인혼합모형(ML FMM)을 이용한 분석 방법론을 기술하였다. ML CFA는 절편만을 임의효과로 추정하는 임의절편모형과 요인계수도 임의효과로 추정하는 임의절편 및 임의요인계수모형으로 구분하여 장단점을 기술하였다. 구체적으로 각 방법론에 대한 이론적 모형과 측정동일성 검정 절차를 제시하고, 기존의 MG CFA에 비해 지니는 이점 및 적용 시 유의해야 할 사항을 서술하였다. 또한 이러한 방법론을 적용한 예시로서, PISA 2015 자료를 활용하여 학생이 인식한 과학의 도구적 동기 및 즐거움에 대해 국가별 측정동일성 검정 절차를 분석하고 국가별 잠재평균을 추정하였다. 마지막으로 본 연구의 향후 연구 및 의의에 대해 논의하였다.

주요어 : 국가비교, 측정동일성, 잠재평균 비교, 다층 확인적요인분석, 다층 요인혼합모형

<sup>†</sup> 교신저자: 홍세희, 고려대학교 교육학과, 서울시 성북구 안암로 145, E-mail: [seehong@korea.ac.kr](mailto:seehong@korea.ac.kr)

## 서 론

최근 20년 동안 국가 및 문화비교 연구(cross-cultural research)는 방법론적인 발전과 함께 심리학을 비롯한 교육학, 사회학 등의 여러 분야에서 활발히 수행되고 있다. 예를 들어, 심리학에서는 국가 간 심리학적 구인에 대한 비교가 수행된 바 있으며(예: 공감 및 관점의 수용 비교 연구: Chopik, O'Brien, & Konrath, 2017; 정신건강에 대한 국가 비교 연구: Žemojtel-Piotrowska et al, 2018), 교육학에서는 OECD 학업성취도 국제비교 연구(Programme for International Student Assessment: PISA), 수학·과학 성취도 추이변화 국제비교 연구(the Trends in International Mathematics and Science Study: TIMSS) 등을 통해 국가 간 학생들의 학업성취 및 교수·학습의 비교 연구가 수행되고 있다(예: 학업적 자기개념의 비교: Marsh, & Hau, 2003; TIMSS 수학 평가의 비교: 박찬호, 2017).

이 같은 국가 비교 연구에서 중요한 방법론적 이슈는 측정동일성(Measurement Invariance)에 관한 것이다(Byrne, & van de Vijver, 2010). 측정동일성은 개인이 어느 집단에 속하는지에 관계없이 해당 검사의 구인(construct)을 같은 방식으로 측정하는가에 대한 검정을 말하며, 특정 집단에 속한 개인이 서로 다른 방식으로 검사 문항에 응답하는 것은 측정동일성이 위배되는 것으로 볼 수 있다(Mellenbergh, 1989; Meredith, 1993; Millsap, & Everson, 1991; Vandenberg, & Lance, 2000). 예를 들어, 자아탄력성에 대한 성차를 비교하고자 할 때, 문항에서 주어지는 스트레스 상황이나 대응 전략은 남성과 여성에게 동일한 수준으로 이해되는 것이 아니라면(즉, 측정동일성이 성립하지

않는다면) 단순히 문항 총점 혹은 평균만을 가지고 성별 비교를 하는 것은 의미가 없을 수 있다. 집단 간 잠재평균의 비교 혹은 구조 모형의 계수 비교 등을 실시할 때 측정동일성의 성립은 필수적으로 선행되어야 하는 요건이며, 측정동일성이 성립하지 않으면 의미 있는 비교를 할 수 없을 뿐 아니라 잘못된 결과를 도출할 수 있다. 따라서 집단의 비교 연구에서 측정동일성의 성립은 매우 중요한 이슈가 된다.

그러나 기존에 널리 사용되어온 방식인 다 집단 확인적요인분석(Multiple Group Confirmatory Factor Analysis: MG CFA, Jöreskog, 1971)은 많은 수의 국가를 비교할 때는 적절하지 않은 것으로 보고되고 있다(Jang et al., 2017; Kim, Joo, Lee, Wang, & Stark, 2016; Selig, Card, & Little, 2008). MG CFA는 비교 집단을 고정된 것으로 가정하여 모든 집단별로 측정모형을 모델링하고 추정하므로, 주로 소수 집단의 비교(예: 성별, 학교급별 비교)에서 유용한 방법으로 알려져 있다. 그러나 MG CFA 모형을 활용하여 다수의 집단을 대상으로 비교할 시에는 모든 집단별로 측정모형을 만들고 비교해야 하므로 쌍으로 비교하는 수가 매우 크게 늘게 되고, 이는 번거로운 절차를 요구할 뿐 아니라 측정동일성의 결과를 잘못 내릴 가능성이 높아지게 된다(Asparouhov, & Muthén, 2014; Rutkowski, & Svetina, 2014). 특히 많은 수의 국가를 비교하는 경우, 국가 간에는 매우 다른 국가적 배경과 특성이 자리하고 있기 때문에, 모든 나라가 동시에 측정동일성을 만족하기는 상당히 어렵다(Muthén, & Asparouhov, 2018). 이에 따라 방법론 연구자들은 많은 수의 집단을 비교할 때 적절한 측정동일성 검정 방법에 대해 연구하여 왔다. 대표적으로 기존

의 MG CFA는 국가별로 모형을 각기 추정하는 고정효과 모형으로 볼 수 있는 반면, 다수의 집단을 비교할 때는 비교집단들에 대한 평균과 분산을 추정함으로써 추정 모수의 수를 낮추는 임의효과 모형이 제시되었다(Asparouhov, & Muthén, 2014; Fox, 2010). 임의효과를 추정하는 모형에는 자료의 다층적 구조를 모델링하여(즉, 다층 확인적요인분석(Multilevel Confirmatory Factor Analysis: ML CFA), 문항의 절편을 임의로 추정하는 임의절편 모형이 있으며(Jak, Oort, & Dolan, 2013; Davidov, Dülmer, Schlüter, Schmidt, & Meuleman, 2012), 절편뿐 아니라 문항을 설명하는 요인의 계수(요인계수)도 임의효과로 추정하는 임의절편 및 임의요인계수 모형이 있다(De Boeck, 2008, Frederickx et al., 2010; Verhagen, & Fox, 2013; Asparouhov, & Muthén, 2015). 한편, 비교하고자 하는 집단의 수가 많은 경우에는 모든 집단을 동일선상에서 비교하기보다는 관찰되지 않는 이질성으로 인해 잠재계층을 형성할 가능성이 있다(Allua, Stapleton, & Beretvas, 2008). 이에 따라 Kim 외(2017)는 잠재계층모형과 확인적 요인분석이 합쳐진 다층 요인혼합모형(Multilevel Factor Mixture Modeling: ML FMM)을 활용하여 비교집단을 잠재계층으로 구분하고 잠재계층 간 측정동일성을 검정하는 방법을 제시해오고 있다(Kim, & Wang, 2018).

이들 방법론은 집단을 고정효과로 추정하는 것 대신 임의효과로 추정하는 방법인 ML CFA 중 임의절편모형과 ML FMM의 방법은 기존의 MG CFA에서와 같이 비교집단 간 측정동일성을 검정할 때 요인계수와 절편이 동일함을 가정하는 정확(exact) 검정을 가정한다. 정확 검정은 비교집단 간 해당 모수(즉, 요인계수, 절편)를 동일한 값으로 제약하여 요인계수동일

성, 절편동일성 모형을 순차적으로 검정하는 방법을 말한다. 다만, MG CFA는 비교집단이 두 개 혹은 세 개로 작을 경우에 집단의 잠재평균을 고정효과로 추정함으로써 분산을 허용하지 않는 반면, 임의절편모형과 ML FMM은 비교집단이 많을 경우에 다층 모형을 사용하여 집단의 잠재평균을 임의 효과로 추정하게 된다. 기존의 일반적인 다집단 분석은 이러한 정확 검정 방법의 접근법을 보편적으로 사용하여 왔으므로(예: 이종희, 김기연, 김수진, 2011; Chen, 2007; Horn, & McArdle, 1992), 본 연구에서는 이를 확장하여 비교집단이 많은 경우에 대한 측정동일성 검정 방법을 고찰하고자 하였다. 한편, 비교집단이 많은 경우 요인계수동일성 모형이나 절편동일성 모형은 지지 되지 않는 경우가 많다(Muthén, Asparouhov, 2018). 그런 경우에는 정확히 하나의 값으로 제약하기보다는 약간의 차이는 허용하는 근사 동일성을 가정하는 것이 적절할 수 있다(Asparouhov, & Muthén, 2015; Muthén, & Asparouhov, 2018). 즉, 모수에서의 차이가 크지 않고 모형의 적합도가 적절하다면 근사 검정을 통해 평균 비교가 가능하다.

최근 이들 방법론을 적용한 연구가 학술지에서 발표되고 있다. 예를 들면, Jang과 그의 동료들(2017)은 MG CFA와 ML CFA의 임의절편모형 그리고 근사 측정동일성을 가정하는 정렬 최적화(Alignment Optimization) 모형으로 26개 국가 간 삶의 만족도를 추정한 결과를 비교하였으며, MG CFA의 한계점을 제시하면서 ML CFA의 임의절편모형과 정렬 최적화 방법의 사용을 권고하였다. 또한, Muthén과 Asparouhov(2018)은 26개 유럽 국가 조사 자료를 활용하여 전통-순응(Tradition-conformity)의 문항에 대해 MG CFA, ML CFA의 임의절편모

형과 임의절편 및 임의요인계수모형 그리고 정렬 최적화 모형으로 추정된 결과를 비교하였으며, MG CFA를 제외한 나머지 방법들이 유사한 결과를 나타내었음을 보고하였다. 또한 Da Costa와 Dias(2014)는 유럽연합 15개국을 대상으로 빈곤의 속성을 인식하는데 3가지 유형의 잠재계층(개인주의, 사회주의, 운명주의)이 있음을 검증하였으며 이들 유형에 대해 비교하였다.

이에 본 연구에서는 ML CFA의 임의절편모형과 임의절편 및 임의요인계수 모형, 그리고 ML FMM를 중심으로 이들 방법론에 대해서 상세히 설명하고 장단점을 분석함과 더불어 경험적 연구에 적용할 때 유의해야 할 점을 정리하였다. 또한, 실제 자료를 가지고 적용 사례를 제시함으로써 적용 연구자들에게 유용한 정보를 제공하고자 한다. 본 연구는 다수의 집단을 비교할 수 있는 측정동일성 분석 절차에 대한 최근 연구 성과를 소개하고, 후행 연구자들이 해당 방법을 보다 쉽게 적용할 수 있도록 구체적인 예시를 제공하는 것을 연구목적으로 한다.

### 고정효과 모형을 이용한 다수 집단의 측정동일성 검정

다집단 확인적요인분석(Multiple Group Confirmatory Factor Analysis, MG CFA)

#### MG CFA 모형식

측정동일성은 수학적으로는 능력(ability)이 소속집단(group membership)과 독립적일 때 관측된 점수를 얻을 조건부 확률로 정의된다

(Mellenbergh, 1989; Meredith, & Millsap, 1992; Yoon, & Millsap, 2007). 이러한 정의는 다음 (식 1)과 같이 표현할 수 있으며, 이때  $Y$ 는 어떤 구인에 대해 관측된 점수,  $W$ 는 그 구인을 측정하는 잠재 요인(latent construct),  $G$ 는 집단을 나타낸다. 이는 어떤 구인에 대해 동일한 능력을 가진 개인들이 집단에 관계없이 관측된 점수를 얻을 확률이 같을 때 측정동일성이 성립한다고 할 수 있다.

$$P(Y|W, G) = P(Y|W) \quad (\text{식 1})$$

비교집단 간에 측정동일성을 검정하기 위한 가장 주된 방법론은 확인적요인분석(CFA) 모형을 이용하는 것이다. MG CFA에서의 집단별 측정 모형은  $p$ 개의 관찰변수와  $m$ 개의 공통요인 간에 선형 관계를 가지며 다음 (식 2)과 같이 표현된다.

$$y = \tau + \Lambda\eta + \epsilon \quad (\text{식 2})$$

여기서  $y$ 는 관측 점수의  $p \times 1$  벡터,  $\tau$ 는 측정절편의  $p \times 1$  벡터,  $\Lambda$ 은  $p \times m$ 의 요인계수행렬,  $\eta$ 는 요인 점수의  $m \times 1$  벡터이며  $\epsilon$ 는 잔차 요인에 대한  $p \times 1$  벡터이다. 다집단일 경우, (식 2)는 (식 3)과 같이 소속집단을 가리키는  $g$ 가 추가된 측정모형이 된다.

$$y_g = \tau_g + \Lambda_g\eta_g + \epsilon_g \quad (\text{식 3})$$

각 집단 내에서 잠재 점수( $\eta$ )와 잔차 점수( $\epsilon$ )는 상관관계가 없는 것으로 가정하며(즉,  $COV(\eta_g, \epsilon_g) = 0$ ), 집단  $g$ 에서  $y$ 의 공분산 구조는 (식 4)와 같다.

$$\Sigma_g = \Lambda_g \Phi_g \Lambda_g' + \Theta_g \quad (\text{식 4})$$

여기서  $\Theta_g$ 는 잔차 분산의 대각선 행렬이고,  $\Phi_g$ 는 소속집단  $g$ 에서의 요인 공분산 행렬이다. 또한, (식 3)에서 잔차 요인이 평균이 0임을 가정하면, 각 소속집단에서  $\mathbf{y}$ 의 기댓값은 (식 5)와 같다. 이때,  $g$ 집단에서의 잠재평균( $\eta_g$ )은  $\alpha_g$ 이다.

$$E(\mathbf{y}) = \tau_g + \Lambda_g \alpha_g \quad (\text{식 5})$$

### MG CFA의 측정동일성 검정 절차

측정동일성의 검정은 형태동일성(Configural invariance), 요인계수동일성(Metric invariance 또는 Weak invariance), 절편동일성(Scalar invariance 또는 Strong invariance), 요인분산동일성(Strict invariance)의 순서로 위계적으로 검정한다(Widaman, & Reise, 1997). 일반적으로 집단 간 평균 비교는 절편동일성까지 만족하면 의미 있는 비교가 가능한 것으로 제시되고 있다(Little, 1997, 2000; Meredith, 1993). 따라서 본 연구에서 요인분산동일성은 절차의 설명과 논의에서 제외하였다.

형태동일성은 요인계수  $\Lambda_g$ 의 패턴(0과 0이 아닌 요인계수)이 비교집단 간에 동일한 것을 의미한다(Steenkamp, & Baumgartner, 1998). 형태동일성은 측정동일성 검정 절차에서 기저모형(baseline model)이 되며, 형태동일성이 만족해야 요인계수동일성 검정이 가능하다. 요인계수동일성은 비교집단 간에 동일한 요인계수 행렬( $\Lambda_g = \Lambda$ )을 갖는 것으로, 요인의 절편과 분산은 다를 수 있다(Horn, & McArdle, 1992; Millsap, 1997). 요인계수동일성이 지지되면, 절편동일성 검정이 가능하다. 절편동일성은 비

교집단 간에 동일한 요인계수 행렬( $\Lambda_g = \Lambda$ )을 가질 뿐만 아니라 절편도 동일한 것( $\tau_g = \tau$ )으로, 절편동일성까지 만족해야 집단평균 비교가 의미를 갖게 된다(Meredith, 1993).

### MG CFA의 한계점

MG CFA 모형은 국가 비교와 같이 비교집단이 많은 경우에는 다음과 같은 점에서 유의해야 할 필요가 있다. 첫째, MG CFA는 집단을 고정 분류(fixed classification)한 것으로 가정하여 검정하는 방법이다. 고정 분류는 모집단에서 모든 가능한 집단을 고려한 것을 의미하는데 예를 들어, 성별에 따른 잠재평균을 비교하고자 할 때, 비교 가능한 집단은 남학생과 여학생뿐이므로 고정효과로 추정하는 것이 적절하다. 그러나 비교집단이 많은 경우에 집단은 고정 분류(fixed classification)보다 임의 분류(random classification)로 가정하는 것이 더 적절하다(Jak, Oort, & Dolan, 2013). 집단을 고정효과로 추정하는 경우에는 각 집단별로 측정모형을 만들고 집단별로 식별된 모형에서 추정한 모수들을 비교한다. 따라서 비교집단이 많을수록 추정 모수는 상당히 증가하며, 그만큼 상당히 많은 표본 크기가 요구된다(Asparouhov, & Muthén, 2014). 예를 들어, Kim 외(2017)의 연구에 따르면, 집단 수가 50개 일 때, MG CFA는 형태동일성 모형에서 900개의 자유모수를 추정한 반면, ML CFA 모형은 30개의 자유모수를 추정한 것으로 나타났다. 적은 수의 모수를 추정하는 것이 항상 유리한 것은 아니나 많은 모수의 추정은 표본 크기가 충분히 크지 않은 경우 문제가 발생할 가능성이 높다. 이에 따라 비교집단이 많은 경우에 기존의 MG CFA는 권장되지 않는 경향이 있

다(Jang et al., 2017; Kim et al., 2017).

둘째, 모형 평가의 측면에서 CFA 모형에서의 적합도(즉,  $\chi^2$ , CFI, RMSEA)의 기준이 다수의 집단이 있는 모형에서는 적절하지 않을 수 있다. 일반적으로 CFA 모형의 적합도는 CFI와 TLI가 .90이상, RMSEA는 .08미만이면 적합도가 괜찮은 것으로 해석한다(홍세희, 2000; Browne, & Cudeck, 1993). 그러나 다수의 집단이 있는 경우 이 기준을 만족하기는 쉽지 않으며, 모형 수정이 상당히 이루어진 후에 기준값에 도달할 수 있고 이는 측정동일성의 결과를 잘못 내릴 가능성이 높아지는 문제가 발생한다(Asparouhov, & Muthén, 2014). 따라서 다수 집단의 비교 시에는 일반적인 적합도 기준과는 다른 기준을 고려해야 한다. 예를 들어, Rutowski와 Svetina(2014)는 비교집단의 수가 많을 경우, RMSEA는 실제 모형 적합도와 관계없이 기준값보다 커지는 경우가 많다고 보고한 바 있다. 이들은 시뮬레이션 연구를 통해 비교집단이 10개 혹은 20개 이상일 경우, 기준값을 보다 완화한 .10으로 할 것을 권고하였다. 또한, 더 많은 수의 집단 비교 시에는 명확한 가이드라인을 제시하지 않았지만 그 기준이 더욱 완화될 수 있음을 제안하였다. MG CFA에서 모형 비교 시(예: 형태동일성 모형과 요인계수동일성 모형의 비교)에는 내재된 모형의 비교이므로  $\chi^2$  검정이 이루어지나, 이는 표본 크기에 민감하므로 많은 연구자들이 CFA와 RMSEA의 차이값을 더 많이 고려하고 있다.  $\Delta CFA$ 는 .01 이하(Cheung & Rensvold, 2002),  $\Delta RMSEA$ 는 .015 이하(Chen, 2007)의 기준을 적용한다. 또한, Rutowski와 Svetina(2014)는 비교집단이 많은 경우, 절편동일성 검정에서는 위의 기준치가 적절하나, 요인계수동일성에서는  $\Delta CFA$ 는 .02 이하,  $\Delta RMSEA$ 는 .03 이하의

표 1. MG CFA 모형 비교에 대한 적합도 지수 기준(Rutowski와 Svetina, 2014)

| 동일성 검정            | CFI               | RMSEA              |
|-------------------|-------------------|--------------------|
| 형태동일성             | .90 이상            | .10 이하             |
| 형태동일성 vs. 요인계수동일성 | $\Delta \leq .01$ | $\Delta \leq .015$ |
| 요인계수동일성 vs. 절편동일성 | $\Delta \leq .02$ | $\Delta \leq .03$  |

기준이 필요함을 제안하였다(표 1 참고).

### 임의효과 모형을 이용한 다수 집단의 측정동일성 검정

다층 확인적요인분석(Multilevel Confirmatory Factor Analysis, ML CFA)

국가 비교 연구에서 개인은 국가에 내재된 자료이므로 다층모형과 구조방정식모형의 결합 모형인 다층 확인적요인분석(ML CFA)을 활용할 수 있다. 이때 1수준은 국가에 속한 개인, 2수준은 비교하고자 하는 국가가 되며, 국가는 모집단으로부터 무작위로 추출한 변수로 가정하여 2수준에서 임의효과로 모델링한다. 따라서 1수준과 2수준 모두에서 측정모형을 설정할 수 있다.

비교집단이 많은 경우에 ML CFA는 MG CFA에 비해 여러 가지 장점을 지닌다. 먼저, MG CFA는 모든 집단별로 측정모형을 만들어야 하는 반면, ML CFA는 무선 집단들 간의 평균적인 모형을 나타내는 하나의 측정모형을 만들고 집단 간 분산을 평가하므로 MG CFA에 비해 상대적으로 간단하다. 또한 ML CFA

는 집단을 임의효과로 처리하므로 집단 수가 많은 것은 문제가 되지 않으며 오히려 집단 수가 많을수록 수행력이 좋은 것으로 보고된다(Kim et al., 2017). 선행연구에서 다층적 집단은 2수준 분산의 정확한 추정을 위해 최소 30개 집단은 필요하다고 제시하고 있다(Kreft, & De Leeuw, 1998; Maas & Hox, 2005). 한편, ML CFA는 집단 간 분산이 작을 경우 모형이 수렴되지 않는 문제가 발생할 수 있으며, 측정모형은 MG CFA와 같이 각 비교집단 별로 만들어지는 것이 아니라 각 수준(집단 내, 집단 간)에서 만들어지므로 각 집단에 대한 모형 추정은 가능하지 않다. 만약 연구자가 MI 성립이 아닌 집단별로 모수를 추정하는데 관심이 있다면 MG CFA를 고려해야 할 것이다.

ML CFA는 임의효과로 추정하는 모수의 범위에 따라 두 가지 모형으로 나눌 수 있다. 첫 번째 모형은 2수준에서의 절편만을 임의효과로 추정하는 임의절편 모형(random intercept model; Jak, Oort, & Dolan, 2013; 2014)이며, 두 번째 모형은 2수준에서 절편뿐 아니라 요인 부하량도 임의효과로 추정하는 임의절편 및 임의요인계수 모형(random intercept, random loading model; Muthén, & Asparahov, 2018)이다. 두 모형에 대한 모형식과 측정동일성 검정 방법을 다음에 상세히 기술하였다.

#### 임의절편모형(random intercept model)의 모형식

ML CFA에서 총 분산은 전체평균에서 기인하는 집단 간 분산과 집단평균에서 기인하는 집단 내 분산의 합으로 표현된다(Kaplan, Kim, & Kim, 2009; Muthén, 1994). 즉, 공분산행렬( $\Sigma_T$ )은 집단 간 공분산 행렬( $\Sigma_B$ )과 집단 내 공분산 행렬( $\Sigma_W$ )로 분해할 수 있으며 (식 6)

과 같다.

$$\Sigma_T = \Sigma_B + \Sigma_W \quad (\text{식 6})$$

이에 따라 집단 내 수준의 측정모형( $Y_{Wij}$ )은 집단 내 분산으로 구성되고 집단 간 수준의 측정모형( $Y_{Bj}$ )은 집단 간 분산으로 구성되기 때문에 각각 다음과 같이 표현될 수 있다.

$$Y_{Wij} = \tau_W + \Lambda_W \eta_{Wij} + \epsilon_{Wij}, \quad (\text{식 7})$$

$$\epsilon_{Wij} \sim N(0, \Theta_W),$$

$$Y_{Bj} = \tau_B + \Lambda_B \eta_{Bj} + \epsilon_{Bj}, \quad (\text{식 8})$$

$$\epsilon_{Bj} \sim N(0, \Theta_B)$$

여기서 집단 내 개인의 점수는 집단 평균과 개인점수의 차이로 설명될 수 있으므로 집단 내 절편  $\tau_W$ 은 0으로 고정한다. ML CFA의 통합식을 표현하면 다음과 같다.

$$Y_{ij} = \tau_B + \Lambda_B \eta_{Bj} + \Lambda_W \eta_{Wij} + \epsilon_{Bj} + \epsilon_{Wij} \quad (\text{식 9})$$

(식 9)는 (식 10)과 같이 임의절편모형의 식으로 동일하게 표현할 수 있다.

$$\text{Level 1: } Y_{ij} = \tau_j + \Lambda_W \eta_{Wij} + \epsilon_{Wij},$$

$$\epsilon_{Wij} \sim N(0, \Theta_W),$$

$$\text{Level 2: } \tau_j = \tau_B + \Lambda_B \eta_{Bj} + \epsilon_{Bj},$$

$$\epsilon_{Bj} \sim N(0, \Theta_B) \quad (\text{식 10})$$

### 임의절편모형의 측정동일성 검정 절차

본 연구에서 다루고 있는 다층 모형은 집단 내 개인들의 평균적인 특성을 집단 간 수준에서 나타내고 있으며, 이는 비교하고자 하는 구인이 집단 내와 집단 간에 동일한 요인구조를 가짐을 의미한다(Stapleton, Yang, & Hancock, 2016). 예를 들어, 각 국가에 속한 학생들의 과학에 대한 흥미가 1수준에서 5개의 변인으로 측정되고, 2수준에서는 동일한 잠재변인이 각 나라의 과학에 대한 흥미의 평균을 나타낸다면 1수준과 2수준은 모두 하나의 잠재변인을 가지는 동일한 요인구조를 가지게 된다. 다층 모형은 경우에 따라서 1수준과 2수준의 요인구조가 다른 경우가 있을 수 있으나(예: 자아존중감 요인은 개인수준에서는 정적요인과 부적요인의 두 요인인 반면, 학교수준에서는 하나의 요인을 가짐(Zimprich, Perren, & Hornung, 2005)), 본 연구는 국가 및 문화 비교 연구(cross-cultural research)를 수행함에 있어 1수준과 2수준 간 요인구조가 같은 상황을 가정한다. 이에 대해 Ryu와 Mehta (2017)은 특별기고(Teacher's corner)에서 수준 간 측정 동일성(metric invariance across levels)을 강조한 바 있다. 수준 간 측정 동일성은 다음의 세 가지 조건을 만족해야 한다. 1) 각 수준에서 나타내는 요인은 개념적으로 같아야 하고, 2) 각 수준 간 요인의 형태가 같아야 하며(즉, 교차 수준 간에 요인의 형태동일성이 만족해야 함), 3) 각 수준 간에 요인계수가 같아야 한다(즉, 교차 수준 간 요인계수 동일성이 만족해야 함). 교차 수준 간 요인계수의 동일성이 성립한다는 것은 특정 잠재변인이 집단 내와 집단 간에 같은 의미를 갖는 것으로 해석할 수 있으며, 이는 분산이 각 수준으로 분해될 수 있음을 의미한다(Zyphur, Kaplan, & Christian,

2008; van de Vijver, & Poortinga, 2002). 각 수준에서 분산이 추정되어야 집단 내 상관계수인 ICC도 계산과 해석이 가능하다(Mehra, & Neale, 2005). 이에 대한 보다 자세한 사항은 Ryu와 Mehta(2017, 941p)를 참조하기 바란다.

따라서 임의절편모형의 측정동일성 검정은 교차 수준 간에 같은 요인구조가 동일하게 적용하는 형태동일성 모형, 요인계수가 집단에 따라 동일하다면 집단 내 수준과 집단 간 수준에서의 요인계수 역시 동일할 것이라고 가정하는 요인계수동일성 모형, 그리고 모든 집단 간에 요인절편이 동일한 절편동일성 모형을 순차적으로 적합하여 모형 비교를 할 수 있다.

형태동일성 모형은 교차 수준 간 요인계수와 잔차분산을 자유롭게 추정한다(즉,  $\Lambda_W \neq \Lambda_B$ ,  $V(\epsilon_{Bj}) \neq 0$ ). 모형의 지지는 일반적인 CFA 적합도 지수인  $\chi^2$ , CFI, RMSEA로 평가한다. 형태동일성이 만족되면, 비교집단 간에 교차수준(집단 내 수준과 집단 간 수준, cross-level)에 대한 요인계수를 동일하게 제약함으로써, 요인계수동일성을 검정할 수 있다(Jak, Oort, & Dolan, 2013; Tay et al., 2014). 즉, 교차 수준 간에 동일한 요인구조를 가지는 경우에 요인계수가 모든 집단에서 동일하다면( $\Lambda_j = \Lambda$ ), 집단 내 CFA 모형의 요인계수는 집단 간 CFA 모형의 요인계수와 같아야 한다( $\Lambda_W = \Lambda_B$ ,  $V(\epsilon_{Bj}) \neq 0$ ). 또한, 요인계수동일성 모형에서 집단 간 수준의 잔차 분산들을 모두 0으로 제약함으로써 절편동일성을 검정할 수 있다( $V(\epsilon_{Bj})=0$ ). 집단 수준에서 요인계수뿐 아니라 절편도 모든 집단에서도 동일하다면( $\tau_j = \tau$ ), 집단 간의 절편에 대한 분산이 0이 되기 때문이다. 즉, 집단 간 수준의 잔차

분산이 0이 된다. 만약 절편동일성이 지지되지 않으면 수정 지수를 바탕으로 0으로 제한한 잔차 분산을 하나씩 풀어주어 부분 절편동일성으로 진행할 수 있다.

#### 임의절편 및 임의요인계수 모형(random intercept and random loading model)의 모형식

임의절편 및 임의요인계수 모형은 엄격한 정확 동일성 가정에서 흔히 발생하는 낮은 적합도의 문제를 해결하기 위해 베이지안을 이용하여 검정할 수 있도록 제안된 방법으로, 이 모형은 2수준의 절편뿐 아니라 요인계수도 임의효과로 추정한다. 집단의 수가 많을 경우, 절편동일성은 성립하지 않는 경우가 흔히 발생하며(Muthén, & Asparouhov, 2018), 이는 집단 간 비교를 가능하지 않도록 하기 때문에 정확히 같은 값으로 추정하도록 동일하게 제약하는 방법보다는 작은 값의 차이는 허용하도록 하여 집단 간 평균 비교가 가능하도록 하는 것이 대안적인 방법이 될 수 있다. 임의절편 및 임의요인계수 모형은 측정 모수(즉, 절편과 요인계수)가 집단에 따라 변화하지만 공통의 평균과 분산을 가지는 것으로 정의된다. 이는 정확히 같은 값으로 제약함으로써 측정동일성을 검정하는 정확검정(exact testing)과 달리, 근사(approximate) 측정동일성을 가정한다(Asparouhov, & Muthén, 2015; Muthén, & Asparouhov, 2018). 임의절편 및 임의요인계수 모형식은 다음 (식 11)로 표현할 수 있다.

$$\begin{aligned} \text{Level 1: } Y_{ij} &= \tau_j + \Lambda_j \eta_{wij} + \epsilon_{wij}, \\ \text{Level 2: } \tau_j &= \tau_B + \Lambda \eta_{Bj} + \epsilon_{Bj}, \\ \Lambda_j &= \Lambda + \Lambda \eta_{\psi j} + \epsilon_{\Lambda j}. \quad (\text{식 11}) \end{aligned}$$

여기서  $\eta_{wij} \sim N(0,1)$ ,  $\epsilon_{wij} \sim N(0,\theta)$ ,  $\epsilon_{Bj} \sim N(0,\sigma_B^2)$ ,  $\epsilon_{\Lambda j} \sim N(0,\sigma_A^2)$ ,  $\eta_{Bj} \sim N(0,\psi)$ ,  $\eta_{\psi j} \sim N(0,\sigma^2)$ 이다. 절편에서의 분산은  $\sigma_B^2$ , 요인계수에서의 분산은  $\sigma_A^2$ , 요인평균에서의 분산은  $\psi$  그리고 요인 분산에서의 분산은  $\sigma^2$ 로 추정된다. 즉, 모형에서 절편( $\tau_j$ )과 요인계수( $\Lambda_j$ )의 평균과 분산이 추정되며 이를 토대로 동일성 정도를 판단할 수 있게 된다. 예를 들어, 각 집단의 요인계수는 서로 다른 값을 가지지만 평균이 1이고( $\mu_\lambda = 1$ ), 분산이 0.01( $\sigma_\lambda^2 = 0.01$ )로 추정되면, 집단들의 요인계수는 95% 신용구간에서 0.804~1.196의 범위가 되어, 근사 측정동일성이 지지되는 것으로 해석할 수 있다(Muthén, & Asparouhov, 2018).

임의절편 및 임의요인계수 모형의 추정은 베이지안 추정이 요구된다. 일반적으로 모수 추정에 활용하는 최대우도법을 이용하여 요인계수를 임의효과로 추정하게 되면 많은 차원은 수치적분이 필요로 하게 되는데 현재 개발된 프로그램으로는 사실한 불가능하기 때문이다(Asparouhov, & Muthén, 2015).

또한 이 모형에서는 근사 측정동일성을 가정하고 있으므로 정확 검정에서처럼 형태동일성, 요인계수동일성, 절편동일성의 모형을 비교하는 방식을 취하지 않는다. 절편과 요인계수의 평균과 분산을 추정하여 분산이 크다면 동일하지 않는 정도가 큰 것으로 해석할 수 있다.

#### 다층 요인혼합모형(Multilevel Factor Mixture Modeling, ML FMM)

요인혼합모형(Factor Mixture Modeling, FMM)은 요인(잠재 연속 변수)과 잠재계층(잠재 범주

형 변수)을 동시에 통합하여 확인적요인분석과 잠재계층분석이 합쳐진 모형으로(Lubke, & Muthén, 2005), 관찰되지 않은 집단(unobserved group)에 대해 측정동일성을 검정할 수 있다(Buzick, 2010; Vermunt, 2007). 기존의 다집단 분석(multiple group analysis)은 관찰된 집단(예: 성별)만을 대상으로 비교가 가능한 반면, FMM은 관찰되지 않는 집단에서의 이질성을 잠재계층으로 구분하고, 잠재계층 간 측정동일성 검정 및 평균 비교가 가능하다는 점에서 전통적인 다집단 분석모형에 비해 유연한 분석들을 제공한다(Kim, Joo, Lee, Wang, & Stark, 2016). 초기에 관찰되지 않은 집단을 잠재계층으로 구분하여 국가 비교 연구를 수행한 연구들은 주로 문항반응이론(Item response theory, IRT)에서부터 연구되어 왔다. 예를 들어 Cheung과 Rensvold(2000)은 문항에 대한 서로 다른 반응 양식이 극단적 반응 양식(Extreme response styles, ERS)과 묵인 반응 양식(acquiescence response styles, ARS)과 같은 관찰되지 않는 집단을 형성할 수 있으며, 이는 문화적 차이의 중요한 원인이 될 수 있음을 연구하였고, De Jong과 Steenkamp(2010)은 국가 비교 연구에서 다층 다차원 순서형 IRT 모형(mixture multilevel multidimensional ordinal IRT models)을 제안하며 각 잠재계층에 속한 국가들은 서로 다른 측정모수를 가짐을 연구하였다.

FMM은 국가(혹은 문화) 간 비교와 같이 비교하는 집단이 많을 경우 특히 유용할 수 있다. 전통적인 측정동일성 방법뿐 아니라 ML CFA 방법은 여러 비교집단(group)이 모두 한꺼번에 동질적인지를 검정하고 있으며, 만약 대부분의 집단에서 동일성이 만족되더라도 일부의 집단에서 동일성이 만족되지 않으면 영가

설을 기각하여 집단 간 비교가 어렵다는 한계가 있다. 반면, FMM은 비교집단들 중에는 일부 이질적인 잠재계층(heterogeneous latent class)이 있음을 가정한다. 비교집단들 중에서는 측정동일성이 성립되는 집단들이 군집하여 잠재계층을 형성할 수 있으며, 각 잠재계층 내에 어떠한 집단이 속했는지를 확인할 수 있다. 현실적으로 비교집단이 많을 경우 모든 집단의 측정 동일성이 만족되기란 쉽지 않다는 점에서 국가 비교 연구의 경우 FMM은 매우 유용한 접근법이 된다(Kim et al., 2017; Kim, & Wang, 2018).

국가 비교 연구는 개인 혹은 집단이 국가에 내재된 다층 자료이므로 자료 구조에 따라서 다층 요인혼합모형(Multilevel Factor Mixture Modeling, 이하 ML FMM)을 이용하여 분석하는 것이 적절할 수 있다(Allua, Stapleton, & Beretvas, 2008; Kim, Cao, Wang, & Nguyen, 2017; Kim, & Wang, 2018).

#### ML FMM 모형식

다수 집단에 대한 ML FMM의 모형식은 2수준 잠재계층  $c$ 에 대한 조건부 측정모형으로, 측정모형은 1수준에서 만들어지며 잠재계층별로 구성된다. 1수준의 모형식은 (식 12)과 같다. 여기서, 집단  $j$ 의 잠재계층( $C_j$ )이  $c$ 에 속할 때의 조건부 측정모형은 집단  $j$ 에 속한 개인  $i$ 의 관찰점수  $Y_{ij}$ 와 잠재변수  $\eta_{ijc}$ 와의 관계를 나타낸다(Kim et al., 2017; Kim, & Wang, 2018).

$$[Y_{ij}|c] = \tau_{jc} + \Lambda_c^{(W)} \eta_{ijc}^{(W)} + \epsilon_{ijc}^{(W)},$$

$$\epsilon_{ijc}^{(W)} \sim N(0, \Theta_c^{(W)}) \quad (\text{식 12})$$

여기서, (W)는 집단 내, 즉 1수준을 나타내며,  $\tau_{jc}$ ,  $\Lambda_c^{(W)}$ 와  $\epsilon_{ijc}^{(W)}$ 는 각각 잠재계층  $c$ 에 속하는 집단  $j$ 의 절편, 요인계수 행렬, 개인  $i$ 의 잔차를 의미한다. 관찰점수인  $Y_{ij}$ 는 다변량 정규분포를 가정한다. 2수준 모형은 임의 절편  $\tau_{jc}$ 에 대해 (식 13)과 같이 정리할 수 있다.

$$\tau_{jc} = \mu_c + \epsilon_{jc}^{(B)}, \quad \epsilon_{jc}^{(B)} \sim N(0, \Theta_c^{(B)}) \quad (\text{식 13})$$

여기서 (B)는 집단 간, 즉 2수준을 나타내며,  $\mu_c$ 와  $\epsilon_{jc}^{(B)}$ 는 각각 잠재계층  $c$ 의 절편, 집단  $j$ 의 잔차를 의미한다. 각 수준에서의 잔차는 평균이 0이고, 분산이  $\Theta_c^{(W)}$ ,  $\Theta_c^{(B)}$ 인 정규분포를 따른다. ML FMM에서는 잠재계층별로 모수 추정이 가능하므로 2수준에서 측정동일성이 성립하지 않는 모수를 찾아낼 수 있다. 다시 말해, 요인계수( $\Lambda_c^{(W)}$ ), 절편( $\mu_c$ ), 잔차분산( $\Theta_c^{(W)}$ ,  $\Theta_c^{(B)}$ )은 잠재계층별로 이질성을 나타낼 수 있다. 잠재계층 간 이질성의 원인은 비교집단 간 측정동일성이 성립하지 않기 때문이기도 하지만 잠재평균과 같은 구조 모수에서의 이질성으로도 나타난다.

집단  $j$ 가 속하는 잠재계층을 추정하기 위해서는 참조 잠재계층  $k$ 에 대비하여 잠재계층  $c$ 에 속할 로그 확률을 계산하는 다항회귀분석 모형이 사용된다.

$$\ln \left[ \frac{P(c_j = c)}{P(c_j = k)} \right] = \nu_c \quad (\text{식 14})$$

여기서  $\nu_c$ 는 잠재계층  $c$ 의 절편을 말하며,

모형식별을 위해 참조계층은 0으로 제약된다. 본 연구에서는 측정동일성 검정을 위한 측정 모형에만 관심을 두고 있기 때문에 예측변수를 고려하지 않았으나, 각 국가가 속하는 잠재계층을 예측하기 위한 설명변수를 2수준에서 고려할 수 있다. 이를 모형에 포함할 경우, 집단 간(between) 수준의 관측변수인  $X_j$ 로 모델링되며, (식 15)와 같게 된다. 이때,  $\Gamma_c$ 는 잠재계층  $c$ 에 대한 회귀계수를 의미한다.

$$\ln \left[ \frac{P(c_j = c|X_j)}{P(c_j = k|X_j)} \right] = \nu_c + \Gamma_c X_j \quad (\text{식 15})$$

#### ML FMM의 측정동일성 검정 절차

ML FMM에서는 여러 비교집단에서 측정동일성이 성립하는 잠재계층이 몇 개인지에 대해 먼저 탐색적 접근이 필요하다. 많은 집단 비교를 하기 위한 여러 방법론의 수행력을 비교한 Kim 외(2017) 연구에서는 잠재계층의 수와 측정동일성 성립의 수준을 미리 알 수 없기 때문에 잠재계층의 수를 1부터 늘려감과 동시에 여러 측정동일성 수준의 모형을 만들어 비교하는 것을 제안하고 있다. 즉, 잠재계층이 한 개인 CFA 모형(One-class model), 잠재계층이 두 개인 형태동일성 모형(Two-class configural model), 잠재계층이 두 개인 요인계수 동일성 모형(Two-class metric model), 잠재계층이 두 개인 절편동일성 모형(Two-class intercept model), 잠재계층이 세 개인 형태동일성 모형 등이다. 한편, 다층자료에서 ML FMM을 이용한 측정동일성 검정의 효율성을 시뮬레이션한 Kim 외(2016) 연구에서는 잠재계층의 수와 측정동일성 성립의 수준에 대해 모든 가능한 조합으로 모형을 비교하기보다 형태동일성의 모형에서 잠재계층의 수를 1부터 늘려가면서 모

형을 비교하여 잠재계층 수를 결정한 후, 해당 잠재계층 수의 모형에서 측정동일성 수준을 검정하는 2단계 방식을 제안하고 있다.

본 연구에서는 한 번에 많은 모형을 탐색적으로 비교하는 것보다 잠재계층의 수를 먼저 결정하고 해당 잠재계층의 수에서 순차적인 측정동일성 검정을 수행하는 방식에 초점을 두었다.

이에 대한 상세한 설명은 다음과 같다. 먼저, 형태동일성의 모형에서 잠재계층의 수를 1부터 늘려가면서 모형을 비교하여 잠재계층 수를 결정해야 한다. 이때 모형 비교는 내재된 모형이 아니므로 일반 잠재계층분석에서와 같이 정보지수인 AIC(Akaike information criterion; Akaike, 1974), BIC(Bayesian information criterion; Schwarz, 1978), saBIC(sample-size-adjusted BIC; Sclove, 1987)와 LRT 검정 방법인 LMR-LRT (Lo-Mendel l-Rubin test; Lo, Mendell, & Rubin, 2001), BLRT(bootstrap LRT; McLachlan, & Peel, 2000), 분류지수인 Entropy (Ramaswamy, Desarbo, Reibstein, & Robinson, 1993) 등이 사용될 수 있다. 두 번째 단계로, 잠재계층 수를 결정한 후에는 기존의 다집단 분석에서와 같이 잠재계층 간에 모수(요인계수, 절편)를 제약하여 동일성을 검정하는 절차를 수행한다. 이때 만약 1개의 잠재계층 모형이 지지된다면, 이는 모든 비교집단이 동일한 형태를 가진다는 것을 의미하며 모든 집단에 대해서 요인계수동일성, 절편동일성 절차로 나아갈 수 있다. 절편동일성 모형까지 만족이 된다면 모든 비교집단은 하나의 동일한 잠재평균을 갖는다는 해석이 가능하다. 그러나 만약 두 개(혹은 그 이상의) 잠재계층 모형이 지지된다면, 이는 많은 비교하고자 하는 집단 간에 이질적인 잠재계층이 두 개가 있다는 것으로 두 잠재계층 간에 요

인계수 동일성 및 절편 동일성 모형을 순차적으로 검증하여 두 잠재계층 간에 잠재평균을 비교할 수 있다.

형태동일성 검정에서 잠재계층의 수가 결정되면, 1수준에서 잠재계층 간에 요인계수를 동일하게 제약하여 요인계수동일성을 검정할 수 있다. 이때 1수준에서의 잠재평균 및 분산과 2수준에서의 측정변수의 절편은 자유롭게 추정한다. 요인계수동일성 모형의 지지 여부는 앞서 언급한 정보지수와 더불어 요인계수동일성 모형은 형태동일성에 일부 모수가 제약된 내재된 모형이므로 LRT(likelihood ratio test) 차이검정을 실시할 수 있다. 그러나 다층모형과 같이 모형 추정에 MLR(maximum likelihood estimation with robust standard errors)을 이용하는 경우에는 Satorra와 Bentler(2001)가 제안한 SB LRT(Satorra-Bentler scaled likelihood ratio test)가 일반 LRT보다 더 적절한 것으로 제시되고 있다(Kim et al., 2016).

1수준에서의 잠재계층 간 요인계수와 2수준에서의 절편이 잠재계층 간에 동일하게 제약된다. 또한 모형을 식별하고 잠재계층 간 비교가능한 척도를 부여하기 위해 하나의 잠재계층에 대한 잠재평균과 분산은 각각 0과 1로 고정하고, 다른 잠재계층은 해당 모수를 자유롭게 추정하여 잠재계층 간 잠재평균의 차이를 추정할 수 있다. 요인계수동일성 모형에서와 마찬가지로 절편동일성 모형의 지지 여부는 정보지수 및 SB LRT 검정을 통해 확인할 수 있다. 또한 다른 접근법에서와 마찬가지로 절편동일성이 지지되지 않을 경우 부분 절편동일성 모형으로 진행할 수 있다.

### 임의효과 모형에 대한 장단점 분석

지금까지의 내용을 바탕으로 ML CFA(임의절편모형, 임의절편 및 임의요인계수 모형)와 ML FMM에 대한 장단점을 정리해 보고자 한다.

첫째, ML CFA(임의절편모형, 임의절편 및 임의요인계수 모형)와 ML FMM을 이용한 분석은 기존의 다집단 확인적요인분석(MG CFA)에 비해 다음의 측면에서 이점을 지닌다. MG CFA는 집단을 고정 분류(fixed classification)로 가정하여 단일 수준에서 검정하는 반면, ML CFA의 두 모형과 ML FMM은 모두 다층자료의 속성을 반영하여 집단을 임의효과로 가정하여 다층 모형으로 검정한다. 따라서 두 방법론 모두 비교하는 집단의 수가 많을 경우 MG CFA에 비해 이점이 지닌다.

둘째, ML CFA의 두 모형과 ML FMM을 비교하여 측정동일성 검정절차와 그 차이점에 대해 기술하면 다음과 같다. ML CFA의 임의절편모형은 집단 간 수준에서 하나의 평균적인 측정모형을 만들고 집단 간 수준이 집단 내 수준에서의 모형과 동일함을 보여줌으로써 측정동일성 수준을 검정한다. 즉, 집단 간과 집단 내 수준이 동일한 요인구조를 가지고 있는 것으로 평가되면 형태동일성, 이에 더하여 두 수준 간에 동일한 요인계수를 가지고 있는 것으로 지지되면 요인계수동일성, 집단 간 수준에서의 잔차 분산이 0이면 절편동일성이 지지된 것으로 평가한다. ML CFA의 임의절편 및 임의요인계수 모형은 임의절편모형과 같이 집단 간 수준에서 하나의 평균적인 측정모형을 만들어 검정하나, 분산을 허용하는 근사 측정동일성을 가정하여 분산의 정도를 검토함으로써 측정동일성의 성립 여부를 판단한다.

반면, ML FMM은 집단 내 수준에서 측정모형을 만들고, 동일한 측정모형을 공유하는 집단들로 이루어진 계층(cluster)을 찾는다. 이 계층은 명시적으로 나누어진 집단(예: 성별, 학교 유형 등)이 아닌 잠재적인 원인으로 인해 이질성을 가진 집단으로 볼 수 있다. ML CFA와 달리 ML FMM은 집단 간 수준에서의 측정모형은 만들지 않으며, 계층을 나눈 후에 각 계층 간에 집단 내 수준에서 측정동일성 수준을 검정한다.

셋째, 각 방법론의 이점과 유의해야 할 사항을 기술하면 다음과 같다. ML CFA의 임의절편모형은 많은 수의 비교집단에 대해 하나의 평균적인 측정모형으로 동일성 수준을 검정하므로 각 집단마다 잠재평균을 도출할 수 있는 장점이 있는 반면, 비교집단이 많을수록 측정동일성이 성립할 수 있는 가능성은 낮아지게 될 것이므로 측정동일성 가정이 성립하지 않을 가능성이 크다. 그런 경우에 모수에서 약간의 차이는 허용하는 ML CFA의 임의절편 및 임의요인계수모형은 유용하게 활용될 수 있다. 그러나 이 모형은 베이지안 추정을 필요로 하기 때문에 분석에 상당히 많은 시간이 소요된다. 또한 베이지안 추정의 경우 ML 추정과 다르게 적합도 지수가 다양하지 않다. 일반적으로 베이지안 추정의 적합도는 PPP(posterior predictive p-value)의 값으로 판단하나(0.05 혹은 0.01보다 작으면 좋지 않은 적합도임, Asparouhov, & Muthén, 2012), 다층 임의효과모형 추정 시에는 PPP값이 산출되지 않는다. 한편, ML FMM은 집단 간에 어떠한 이유로든지 발생할 수 있는 이질성을 고려할 수 있으며, 이를 토대로 계층을 분리하고 계층 간에 측정동일성을 검정할 수 있는 장점이 있다. 그러나 ML FMM은 계층별로 동일한 모수

를 추정하므로 같은 계층에 속한 집단들은 요인계수, 절편뿐 아니라 잠재평균에서도 하나의 값으로 추정된다. 또한 임의절편모형에서는 측정동일성 검정 절차에서 요구되는 내재된 모형의 비교만 수행하는 반면, ML FMM에서는 측정동일성 검정 절차 이전에 먼저 계층의 수를 확인하기 위해 여러 모형을 만들어 비교하는 작업이 필요하다. 이때 일반적으로 계층을 늘려갈수록 모형을 추정하는 데 상당히 많은 시간이 소요되며, local maxima 등의 문제로 수렴이 잘 되지 않는 경우가 흔히 발생한다(Lubke, & Muthén, 2005). 더욱이 각 개인의 소속 계층은 각 측정동일성 검정 모형마다 달라질 수 있다(Kim et al., 2016). 예를 들어, 요인계수동일성에서는 계층 1에 속하였으나 절편동일성에서는 계층 2에 속하게 되는 경우이다. 또한 서로 다른 측정동일성 검정 모형 간에는 계층 순서에 변동이 일어나는 라벨 스위칭(label switching)이 일어날 수 있어 주의가 필요하다. 즉 요인계수동일성에서는 동일한 계층이 1로 표시되고, 절편동일성에서는 2로 표시되는 경우이다(Redner, & Walker, 1984).

### 분석 방법의 적용 예시

본 장에서는 PISA(Programme for International Student Assessment, OECD) 2015의 자료를 활용하여 MG CFA, ML CFA(임의절편모형), ML CFA(임의절편 및 임의요인계수 모형) 그리고 ML FMM의 분석방법을 적용한 예시를 제시하였다. 연구자들이 실제로 분석할 때 경험할 수 있는 여러 상황을 보여주기 위해 예시는 두 가지로 제시하였다. 첫 번째 예시에서는

ML CFA의 임의절편모형에서 절편동일성 모형이 지지된 것으로, 해당 모형에서의 평균과 MG CFA, ML FMM의 평균을 비교하였다. 그러나 일반적으로 다수의 집단비교에서 절편동일성 모형은 충족되지 않는 경향이 있다. 이에 따라 두 번째 예시에서는 ML CFA의 임의절편 및 임의요인계수 모형을 적용하여 요인평균을 비교할 수 있는 사례를 제시하였다. 이 장에서는 다수의 집단을 비교할 때 검정해야 하는 측정동일성의 단계를 상세하게 기술하였으며, 이를 통해 많은 연구자들의 이해를 돕고 활용을 도모하는 것을 목표로 한다.

### 적용 예시 1: 과학의 도구적 동기에 대한 국가별 측정동일성 검정 및 평균 비교

#### 자료

첫 번째 예시에서 사용한 자료는 PISA 2015 학생 설문지에 포함된 과학에 관한 도구적 동기이다. 이 검사는 ‘과학교과를 잘하기 위해 노력하는 것은 미래에 내가 원하는 일을 하기 위함이다’, ‘과학교과에서 배운 것은 나의 미래에 필요하기 때문에 중요하다’, ‘과학교과는 장래에 나의 전문성을 높이기 위해 가치있다’, ‘내가 학교에서 배운 과학지식은 장래에 내가 직업을 구하는데 도움이 될 것이다’의 4개의 문항으로 구성되어 있다. 분석 대상은 PISA 2015에 참여한 72개 나라 중 지역적으로 조사된 나라들(예: 미국 메사추세츠)을 제외한 55개국으로 411,531명의 학생 자료로 구성되어 있다.

#### 연구방법

먼저, 각 분석 방법별로 측정동일성을 검정하기 이전에 요인을 구성하는 문항에 대

하여 평균, 표준편차, 왜도와 첨도를 포함한 기초통계분석을 진행하였으며, ICC(Intra-Class Correlation, 문항의 전체 분산 중 2수준 분산이 차지하는 비율로 집단 내 상관계를 나타냄)와 신뢰도를 분석하였다. 두 번째로는 기존에 널리 사용되어 왔던 다집단 확인적요인분석(MG CFA)을 활용하여 측정동일성이 성립하는지 분석하였다. MG CFA에서 형태동일성모형, 요인계수동일성모형, 절편동일성모형을 비교하기 위해  $\chi^2$  차이검정, CFI, RMSEA, SRMR의 적합도 지수를 통해 모형을 결정하였다. 적합도 지수의 적용은 표 1에서 제시한 바와 같이 Rutowski와 Svetina(2014) 연구에서 제안한 기준치를 적용하였다. 세 번째로는 다층 확인적요인분석(ML CFA)의 임의절편모형을 이용하여 측정동일성을 검정하였다. MG CFA에서와 마찬가지로  $\chi^2$  차이검정, CFI, RMSEA, SRMR의 적합도 지수를 확인하였다. 그리고 MG CFA와 ML CFA의 두 방법을 이용하여 분석한 국가별 잠재평균 추정치를 비교하였다. 마지막으로 ML FMM을 통하여 55개의 국가가 관찰되지 않는 잠재계층으로 분류될 수 있는지 알아보고 결과를 검토하였다. 분석 프로그램은 Mplus 7.4를 사용하였다.

#### 기초통계분석 결과

각 문항에 대한 기술통계는 표 2와 같다. 과학의 도구적 동기의 모든 문항에 대한 ICC

는 약 .04로 나타났다. 교육학 및 심리학에서의 ICC는 일반적으로 같은 다른 분야에서 보다 작은 편이며(Hox, & Maas, 2001), .05보다 작은 값도 흔히 관찰되는 경향이 있다(Bogler, & Nir, 2012; Peugh, 2010). 또한 본 자료는 학생이 학교에 속한 위계적 속성을 나타내고 있으므로 분석모형은 자료에 맞게 모델링할 필요가 있으며, 집단 내 표본의 크기가 크기 때문에 다층모형을 적용하지 않은 경우 제 1종 오류가 증가할 수 있다(홍세희, 2015; Barcikowski, 1981). 따라서 본 연구에서 다층 모형으로 적용한 것은 타당하다고 볼 수 있다.

문항의 Cronbach  $\alpha$  전체 신뢰도는 .92로 좋은 신뢰도를 나타냈다. 또한, 자료는 국가에 학생이 내재된 다층자료이므로 1수준과 2수준의 신뢰도를 구분하여 산출하여 보았다(Geldhof, Preacher, & Zyphur, 2014). 그 결과 1수준과 2수준의 신뢰도는 각각 .92, .99로 매우 양호한 것으로 나타났다.

#### MG CFA의 분석 결과

MG CFA는 ML 추정을 디폴트로 하며, 국가는 집단변인으로 간주된다. 형태동일성 모형의 식별을 위해서 첫 번째 문항의 요인계수는 1로 고정되며, 해당 문항의 절편은 모든 나라에 대하여 같게 제약된다. 또한, 모든 문항의 잔차공분산은 0으로 고정된다.

MG CFA의 결과는 표 3에 제시되어 있다.

표 2. 과학에 대한 도구적동기 기술통계

| 문항 | N       | M    | SD  | 왜도  | 첨도   | ICC |
|----|---------|------|-----|-----|------|-----|
| 1  | 364,402 | 2.06 | .87 | .59 | -.23 | .04 |
| 2  | 363,016 | 2.15 | .88 | .44 | -.47 | .04 |
| 3  | 362,291 | 2.10 | .86 | .51 | -.30 | .04 |
| 4  | 361,638 | 2.21 | .89 | .38 | -.55 | .04 |

표 3. MG CFA의 측정동일성 검정 결과

| 모형      | $\chi^2$ | df  | p    | CFI  | RMSEA | SRMR | $\Delta\chi^2$ | $\Delta df$ | $p(\Delta\chi^2)$ |
|---------|----------|-----|------|------|-------|------|----------------|-------------|-------------------|
| 형태동일성   | 21951.06 | 112 | .000 | .982 | .166  | .018 | -              | -           | -                 |
| 요인계수동일성 | 27326.31 | 277 | .000 | .977 | .117  | .046 | 5375.25        | 165         | .000              |
| 절편동일성   | 35800.60 | 442 | .000 | .970 | .106  | .048 | 8474.288       | 165         | .000              |

형태동일성모형의 적합도는  $\chi^2(112)=21951.06$ ,  $p<.001$ , CFI=.982, RMSEA=.166으로 나타났다. CFI에서는 좋은 적합도를 보였으며 RMSEA도 만족스러운 것으로 판단하였다. 이는 Rutowski와 Svetina(2014)의 제안에 따르면 10개 정도의 비교집단이 있을 경우 .10을 제안하였고, 26개 나라를 비교한 선행연구(Jang et al., 2017)에서는 .15로 기준값을 제시하였으므로, 본 연구에서의 비교집단이 55개임을 감안하면 만족스러운 적합도로 해석할 수 있을 것이다. 따라서 해당 변인에 대한 형태동일성이 만족하는 것으로 보았다. 두 번째로, 요인계수동일성에 대한 검정을 시행하였다. 그 결과,  $\chi^2(277)=27326.31$ ,  $p<.001$ , CFI=.977, TLI=.972, RMSEA=.117의 적합도를 보였다. 형태동일성모형과 비교하면,  $\Delta\chi^2(165)=5375.25$ ,  $p<.001$ ,  $\Delta CFI=-.005$ ,  $\Delta RMSEA=-.049$ 로 나타나  $\chi^2$  차이검정은 통계적으로 유의했지만  $\Delta CFI$ 가 .02를 넘지 않고  $\Delta RMSEA$ 는 오히려 작아졌으므로 상대적 적합도 지수의 결과를 통해 요인계수동일성을 지지하였다. 세 번째로, 절편동일

성 검정 결과는  $\chi^2(458)=35800.60$ ,  $p<.001$ , CFI=.970, TLI=.977, RMSEA=.106의 적합도를 보였다. 요인계수동일성모형과 절편동일성모형의 비교를 통해  $\Delta\chi^2(165)=8474.288$ ,  $p<.001$ ,  $\Delta CFI=-.007$ ,  $\Delta RMSEA=-.012$ 로 이전 모형비교와 마찬가지로  $\chi^2$  차이검정은 통계적으로 유의했지만, 상대적 적합도 지수의 결과( $\Delta CFI$ 가 .01 이하)를 통해 절편동일성모형을 지지하였다.

**ML CFA(임의절편모형)의 분석 결과**

본 연구 자료의 ICC는 .04 정도로 각 문항에 국가 간 분산이 어느 정도 존재한다는 것을 알 수 있다. 따라서 ML CFA 분석은 적절한 것으로 볼 수 있다. 임의절편모형에 대한 측정동일성 검정을 위해 형태동일성, 요인계수동일성 그리고 절편동일성을 순서대로 검정하였다. 국가는 집단 변수(cluster variable)로 간주된다. 분석 결과는 표 4에 제시하였다.

먼저, 1수준과 2수준의 요인계수와 2수준 잠재 분산을 자유롭게 추정하는 형태동일성

표 4. ML CFA(임의절편모형)의 측정동일성 검정 결과

| 모형      | $\chi^2$ | df | p    | CFI  | RMSEA | SRMR-W | SRMR-B | $\Delta\chi^2$ | $\Delta df$ | $p(\Delta\chi^2)$ |
|---------|----------|----|------|------|-------|--------|--------|----------------|-------------|-------------------|
| 형태동일성   | 21707.43 | 4  | .000 | .982 | .117  | .020   | .006   |                |             |                   |
| 요인계수동일성 | 21714.46 | 7  | .000 | .982 | .088  | .020   | .007   | 7.03           | 3           | .709              |
| 절편동일성   | 23032.53 | 12 | .000 | .981 | .073  | .020   | .031   | 1318.07        | 5           | .000              |

모형을 검정한 결과,  $\chi^2(4)=21707.43$ ,  $p<.001$ , CFI=.982, RMSEA=.117의 적합도를 보였다. RMSEA는 기준을 만족하지 않았지만 CFI가 좋은 적합도를 보임을 알 수 있다. 다음으로, 교차수준 간 요인계수는 같게 고정하고 2수준의 잠재분산은 자유롭게 추정하는 요인계수동일성 모형을 검정하여 보았다. 그 결과,  $\chi^2(7)=21714.46$ ,  $p<.001$ , CFI=.982, RMSEA=.088의 적합도를 보였다. 형태동일성 모형과 비교하였을 때,  $\Delta\chi^2$ 가 7.03이고, 차이값이 유의하지 않았으므로 요인계수동일성 모형을 지지하는 것으로 나타났다. CFI의 차이값은 0에 가까우며, RMSEA는 작아졌으므로 두 지수 또한 요인계수동일성 모형을 지지하는 것으로 나타났다. 마지막으로 수준 간 요인계수와 2수준의 잠재분산을 동일하게 고정한 절편동일성 모형의 검정 결과를 제시하였다. 그 결과,  $\Delta\chi^2$ 에 대한  $p$ 값은 유의하여 요인계수동일성 모형을 지지하였지만, CFA와 RMSEA의 두 적합도 지수가 절편동일성을 지지하였다. 따라서 종합적으로 절편동일성 모형을 지지할 수 있다. 절편동일성 모형 분석 시 *Mplus* 명령어에 `save = fscores` 구문을 이용하면 국가별 잠재평균을 추정치를 얻을 수 있으며, 그 결과는 표 8에 제시하였다.

### ML FMM의 분석 결과

55개의 나라가 과학의 도구적 동기에 대해

관찰되지 않은 이질적인 잠재계층으로 분류되는지 그리고 분류된 잠재계층 간 평균비교가 가능한지 알아보기 위해 ML FMM을 이용하여 측정동일성 검정을 실시하였다.

먼저, 잠재계층 수를 결정하기 위해 집단 내 모형에 측정모형을 설정하고 각 문항의 요인계수를 자유롭게 추정하도록 하였다. 분석 결과는 표 5에 제시하였다. 잠재계층을 1개부터 3개까지 점차적으로 잠재계층 수를 늘려가며 정보지수(AIC, BIC, saBIC)와 LRT 검정(LLRT, BLRT) 결과를 살펴보았다. 그 결과, 계층 수가 3개일 때의 AIC, BIC, saBIC 값이 가장 낮았다. BLRT의 결과는 2개 계층을 지지하였지만, 본 연구에서는 정보지수의 고려를 통해 계층 수 3개의 모형을 선택하였다.

잠재계층 수를 3개로 하여 형태동일성, 요인계수동일성, 절편동일성의 절차에 따라 측정동일성 검정을 수행하였고, 그 결과는 표 6과 같다. 먼저, AIC, BIC, saBIC의 값은 요인계수동일성에서 가장 작은 값을 보임을 알 수 있었다. 또한 앞서 언급하였던 대로, 추정방법으로 MLR을 이용한 모형에서의 내재된 모형 비교를 위하여 SB LRT를 고려하였다.

형태동일성과 요인계수동일성 모형의 비교에서는 SB LRT의  $p$ -value가 .000으로 형태동일성을 지지하였으나 세 가지 정보지수에서는 모두 요인계수동일성을 지지하는 것으로 나타나 요인계수동일성을 지지하는 것으로 보았다.

표 5. ML FMM의 잠재계층 수 결정

| 잠재 계층 수 | AIC       | BIC       | SABIC     | LLRT | BLRT  | Entropy |
|---------|-----------|-----------|-----------|------|-------|---------|
| 1       | 2874049.8 | 2874049.8 | 2874142.3 |      |       |         |
| 2       | 2854174.9 | 2854447.3 | 2854367.8 | .240 | .000  | 1.000   |
| 3       | 2853701.5 | 2854115.4 | 2853994.6 | .240 | 1.000 | 1.000   |

표 6. ML FMM의 측정동일성 검정 결과

| 모형      | AIC       | BIC       | SABIC     | SB LRT |
|---------|-----------|-----------|-----------|--------|
| 형태동일성   | 2853701.5 | 2854115.4 | 2853994.6 |        |
| 요인계수동일성 | 2848586.8 | 2848935.3 | 2848833.6 | .000   |
| 절편동일성   | 2849099.2 | 2849382.4 | 2849299.8 | .000   |
| 부분절편동일성 | 2844642.9 | 2844947.8 | 2844858.8 | .000   |

요인계수동일성과 절편동일성의 비교에서는 세 정보지수가 모두 요인계수동일성 모형에서 더 낮고 SB LRT 결과에서도 p-value가 .000으로 나타나 종합적으로 요인계수동일성을 지지하는 것으로 나타났다. 따라서 절편동일성의 모형에서 잠재계층 간 측정문항의 절편을 동일하게 제약하였던 것을 하나씩 풀어주는 부분 절편동일성 모형의 검정을 수행하였다. 그 결과 첫 번째 문항의 절편을 자유롭게 추정하였더니 정보지수가 요인계수동일성 모형보다 감소하였다.

위의 측정동일성 검정을 종합한 결과, 세 잠재계층은 부분 측정동일성이 성립하는 것으로 볼 수 있다. 즉, 과학에 대한 도구적 동기는 55개의 나라가 관찰되지 않는 변인으로 인해 세 잠재계층으로 분류될 수 있으며, 잠재계층 간에는 동일한 요인계수를 가지고 문항 1을 제외한 나머지 문항에서는 동일한 측정변수의 절편을 가짐을 알 수 있다. 최종모형인 부분 절편동일성 모형에서 분류된 세 잠재계

층의 잠재평균은 표 7에 제시하였다. *Mplus*에서 ML FMM 분석 결과에는 잠재계층별로 요인계수, 잠재평균 등의 추정치만을 제시하고 있으므로, 각 국가에 어느 잠재계층에 속하는지 파악하기 위해서는 *save=cprob* 명령어를 추가로 기술해 주면 된다. 또한 ML FMM은 잠재계층 간 측정동일성 검정을 하여 잠재계층 간 잠재평균이 의미가 있으나, 본 연구에서는 추가적으로 ML CFA에서와 마찬가지로 국가별 잠재평균도 산출하여 표 8에 제시하여 보았다. 그러나 ML FMM에서 국가별 잠재평균을 제시한 것은 단지 이해를 돕기 위한 것이며, 이 모형에서는 잠재계층 내 집단 간에는 측정동일성 검정이 이루어지지 않았다는 점을 유의해야 한다. 마지막으로, 본 연구에서 세 잠재계층으로 분류된 원인은 이 분석 결과로는 알 수 없으며 향후 국가 간 과학에 대한 도구적 동기와 관련이 높을 것으로 예상할 수 있는 공변인을 투입하여 추가적인 분석이 가능할 것이다.

표 7. ML FMM에 대한 잠재계층 분석 결과

|      | 로짓(확률) <sup>a</sup> | 개수 | 비율(%) | 잠재평균   |
|------|---------------------|----|-------|--------|
| 계층 1 | .45 (61.1)          | 21 | 38.2  | 0.000  |
| 계층 2 | .36 (58.8)          | 20 | 36.4  | -0.334 |
| 계층 3 | -                   | 14 | 25.5  | -0.293 |

주. a. 잠재계층 3 (참조집단) 대비 각 잠재계층에 속할 로짓을 확률로 계산하여 제시함.

표 8. MG CFA, ML CFA, ML FMM의 분석 방법에 따른 국가별 도구적 동기의 잠재평균

| 순위 | MG CFA |        | ML CFA 임의절편모형 |        | ML FMM |        |      |
|----|--------|--------|---------------|--------|--------|--------|------|
|    | 국가     | 잠재평균   | 국가            | 잠재평균   | 국가     | 잠재평균   | 잠재계층 |
| 1  | 독일     | 0.324  | 독일            | 0.321  | 독일     | 0.393  | 1    |
| 2  | 오스트리아  | 0.302  | 오스트리아         | 0.303  | 오스트리아  | 0.371  | 1    |
| 3  | 네덜란드   | 0.296  | 네덜란드          | 0.297  | 네덜란드   | 0.364  | 1    |
| 4  | 스위스    | 0.279  | 스위스           | 0.284  | 스위스    | 0.344  | 1    |
| 5  | 체코     | 0.192  | 체코            | 0.193  | 체코     | 0.243  | 1    |
| 6  | 헝가리    | 0.157  | 헝가리           | 0.155  | 헝가리    | 0.200  | 1    |
| 7  | 룩셈부르크  | 0.150  | 룩셈부르크         | 0.152  | 룩셈부르크  | 0.196  | 1    |
| 8  | 일본     | 0.131  | 일본            | 0.140  | 일본     | 0.178  | 1    |
| 9  | 벨기에    | 0.131  | 벨기에           | 0.134  | 벨기에    | 0.174  | 1    |
| 10 | 프랑스    | 0.112  | 프랑스           | 0.122  | 프랑스    | 0.156  | 1    |
| 11 | 한국     | 0.096  | 한국            | 0.091  | 슬로바키아  | 0.119  | 1    |
| 12 | 슬로바키아  | 0.083  | 슬로바키아         | 0.085  | 한국     | 0.110  | 3    |
| 13 | 슬로베니아  | 0.070  | 슬로베니아         | 0.072  | 슬로베니아  | 0.099  | 1    |
| 14 | 덴마크    | 0.062  | 덴마크           | 0.067  | 덴마크    | 0.089  | 1    |
| 15 | 라트비아   | 0.039  | 라트비아          | 0.041  | 라트비아   | 0.067  | 1    |
| 16 | 이탈리아   | 0.031  | 이탈리아          | 0.033  | 이탈리아   | 0.058  | 1    |
| 17 | 노르웨이   | 0.024  | 노르웨이          | 0.026  | 노르웨이   | 0.050  | 1    |
| 18 | 폴란드    | 0.007  | 폴란드           | 0.009  | 폴란드    | 0.032  | 1    |
| 19 | 호주     | 0.000  | 호주            | 0.000  | 크로아티아  | 0.014  | 1    |
| 20 | 크로아티아  | -0.009 | 크로아티아         | -0.005 | 호주     | 0.000  | 3    |
| 21 | 핀란드    | -0.015 | 핀란드           | -0.012 | 불가리아   | -0.013 | 1    |
| 22 | 불가리아   | -0.031 | 불가리아          | -0.029 | 핀란드    | -0.018 | 3    |
| 23 | 에스토니아  | -0.055 | 에스토니아         | -0.048 | 에스토니아  | -0.066 | 1    |
| 24 | 마카오    | -0.057 | 마카오           | -0.056 | 아이슬란드  | -0.067 | 2    |
| 25 | 아이슬란드  | -0.058 | 아이슬란드         | -0.057 | 마카오    | -0.068 | 3    |
| 26 | 홍콩     | -0.074 | 홍콩            | -0.072 | 스페인    | -0.080 | 2    |
| 27 | 스페인    | -0.089 | 스웨덴           | -0.088 | 홍콩     | -0.088 | 1    |
| 28 | 스웨덴    | -0.091 | 스페인           | -0.090 | 이스라엘   | -0.094 | 3    |
| 29 | 대만     | -0.093 | 대만            | -0.091 | 스웨덴    | -0.103 | 1    |

표 8. MG CFA, ML CFA, ML FMM의 분석 방법에 따른 국가별 도구적 동기의 잠재평균 (계속)

| 순위 | MG CFA  |        | ML CFA 임의절편모형 |        | ML FMM  |        |      |
|----|---------|--------|---------------|--------|---------|--------|------|
|    | 국가      | 잠재평균   | 국가            | 잠재평균   | 국가      | 잠재평균   | 잠재계층 |
| 30 | 러시아     | -0.095 | 러시아           | -0.091 | 대만      | -0.106 | 2    |
| 31 | 이스라엘    | -0.099 | 이스라엘          | -0.102 | 러시아     | -0.110 | 3    |
| 32 | 그리스     | -0.115 | 그리스           | -0.109 | 그리스     | -0.132 | 2    |
| 33 | 우루과이    | -0.117 | 우루과이          | -0.119 | 우루과이    | -0.138 | 2    |
| 34 | 미국      | -0.146 | 미국            | -0.147 | 미국      | -0.171 | 2    |
| 35 | 포르투갈    | -0.152 | 포르투갈          | -0.149 | 포르투갈    | -0.174 | 3    |
| 36 | 칠레      | -0.166 | 칠레            | -0.164 | 칠레      | -0.188 | 3    |
| 37 | 콜롬비아    | -0.172 | 콜롬비아          | -0.167 | 콜롬비아    | -0.194 | 2    |
| 38 | 아일랜드    | -0.178 | 아일랜드          | -0.177 | 아일랜드    | -0.198 | 2    |
| 39 | 몬테네그로   | -0.180 | 몬테네그로         | -0.180 | 영국      | -0.204 | 2    |
| 40 | 뉴질랜드    | -0.184 | 뉴질랜드          | -0.184 | 몬테네그로   | -0.207 | 2    |
| 41 | 영국      | -0.185 | 영국            | -0.185 | 뉴질랜드    | -0.214 | 2    |
| 42 | 터키      | -0.198 | 터키            | -0.186 | 터키      | -0.216 | 3    |
| 43 | 리투아니아   | -0.208 | 리투아니아         | -0.202 | 리투아니아   | -0.234 | 2    |
| 44 | 코스타리카   | -0.218 | 코스타리카         | -0.218 | 코스타리카   | -0.247 | 2    |
| 45 | 캐나다     | -0.257 | 캐나다           | -0.257 | 캐나다     | -0.299 | 2    |
| 46 | 브라질     | -0.271 | 브라질           | -0.268 | 브라질     | -0.306 | 3    |
| 47 | 페루      | -0.304 | 페루            | -0.297 | 페루      | -0.343 | 2    |
| 48 | 싱가포르    | -0.304 | 싱가포르          | -0.301 | 카타르     | -0.345 | 2    |
| 49 | 카타르     | -0.305 | 카타르           | -0.305 | 싱가포르    | -0.354 | 2    |
| 50 | 멕시코     | -0.306 | 멕시코           | -0.305 | 멕시코     | -0.356 | 3    |
| 51 | 중국      | -0.310 | 중국            | -0.309 | 중국      | -0.362 | 3    |
| 52 | 태국      | -0.315 | 태국            | -0.313 | 아랍에미리트  | -0.366 | 3    |
| 53 | 아랍에미리트  | -0.324 | 아랍에미리트        | -0.324 | 태국      | -0.366 | 2    |
| 54 | 도미니카공화국 | -0.349 | 도미니카공화국       | -0.348 | 도미니카공화국 | -0.394 | 3    |
| 55 | 튀니지     | -0.382 | 튀니지           | -0.373 | 튀니지     | -0.428 | 2    |

주. 분석 방법에 따라 잠재평균 순위에 차이가 있는 국가에 색 표시를 하였음.

### 국가별 잠재평균 추정

MG CFA는 참조집단의 잠재평균을 0으로 설정하여 그 외 집단의 평균을 추정한다. 반면, ML CFA와 ML FMM은 참조집단의 설정이 없이 각 나라별로 잠재평균을 추정하므로, 본 연구에서는 MG CFA의 추정치와 비교하기 위해 참조집단인 호주와의 차이값을 계산하여 표 8에 제시하였다.

MG CFA, ML CFA와 ML FMM을 통해 추정한 55개국의 잠재평균 결과는 다음과 같다. 과학의 도구적 동기는 세 방법 모두에서 독일이 가장 높은 잠재평균 점수를 나타냈고, 튀니지가 가장 낮은 점수를 보였다. ML FMM 결과에서는 잠재평균이 높은 나라들이 대부분 잠재계층 1에 속하는 것으로 나타나고 있고, 잠재계층 2와 3은 서로 섞여 있는 것을 확인할 수 있다.

MG CFA와 ML CFA가 추정한 잠재평균과 국가별 순위는 거의 유사하나, 일부 나라에서 순위가 다른 것으로 나타났다. 즉, MG CFA에서는 스페인의 잠재평균이 스웨덴보다 다소 높은 것으로 나타났으나, ML CFA에서는 반대의 결과가 나타났다. 그러나 ML FMM과 MG CFA, ML CFA의 두 분석 결과는 상당히 다른 결과를 나타냈다. 이는 ML FMM에서 잠재평균이 유사한 국가들이 서로 다른 잠재계층에 속하면서 순서가 달라진 것으로 추측해볼 수 있다. 예를 들어, 한국과 슬로바키아 두 국가는 MG CFA와 ML CFA에서는 한국의 잠재평균이 다소 높은 것으로 나타났지만 ML FMM에서는 반대의 결과를 나타냈으며 속한 잠재계층도 두 국가 간에 다른 것으로 나타났다.

적용 예시 2: 과학의 즐거움에 대한 국가별 측정동일성 검정 및 평균 비교

### 자료

두 번째 예시에서 사용한 자료는 PISA 2015 학생 설문지에 포함된 과학에 대한 즐거움이 다. 이 검사는 '나는 과학에 대한 주제를 배울 때 재미를 느끼는 편이다', '나는 과학에 관련된 책을 읽는 것을 좋아한다', '나는 과학관련 문제를 풀 때 행복하다', '나는 과학에 관련된 새로운 지식을 찾는 것이 즐겁다', '나는 과학에 대하여 배우는 것이 흥미롭다'의 5개 문항으로 구성되어 있다. 분석 대상은 예시 1과 마찬가지로 55개국의 411,531명의 학생 자료이다.

### 연구방법

두 번째 예시는 측정동일성이 만족하지 않는 경우(즉, 절편동일성 모형이 지지되지 않는 경우)에는 임의절편 및 임의요인계수 모형을 통해 다수 집단의 평균 비교가 가능함을 보여주기 위한 것이다. 따라서 MG CFA, ML CFA의 임의절편모형 그리고 ML CFA의 임의절편 및 임의요인계수모형 방법 비교에 초점을 두었다. 예시 1에서와 마찬가지로 분석프로그램은 *Mplus 7.4*를 사용하였다.

### 기초통계분석 결과

각 문항에 대한 기술통계는 표 9와 같다. 모든 문항에 대한 ICC는 약 .04로 나타났으며, 문항의 Cronbach  $\alpha$  전체 신뢰도는 .94로 좋은 신뢰도를 나타냈다.

### MG CFA의 분석 결과

먼저 MG CFA를 이용하여 과학에 대한 즐거움의 형태동일성, 요인계수동일성, 절편동일성 모형을 순차적으로 검증하였다(표 10). 형태동일성모형은  $\chi^2(290)=20287.794, p<.001$ ,

표 9. 과학에 대한 즐거움 기술통계

| 문항 | N      | M    | SD   | 왜도    | 첨도    | ICC  |
|----|--------|------|------|-------|-------|------|
| 1  | 405611 | 2.72 | .894 | -.465 | -.475 | .035 |
| 2  | 404138 | 2.58 | .891 | -.207 | -.692 | .043 |
| 3  | 403078 | 2.62 | .877 | -.282 | -.598 | .049 |
| 4  | 403287 | 2.79 | .871 | -.546 | -.266 | .044 |
| 5  | 402923 | 2.78 | .887 | -.486 | -.410 | .045 |

표 10. MG CFA의 측정동일성 검정 결과

| 모형      | $\chi^2$  | df  | p    | CFI  | RMSEA | SRMR | $\Delta\chi^2$ | $\Delta df$ | $p(\Delta\chi^2)$ |
|---------|-----------|-----|------|------|-------|------|----------------|-------------|-------------------|
| 형태동일성   | 20287.794 | 290 | .000 | .988 | .013  | .013 |                |             |                   |
| 요인계수동일성 | 25078.316 | 518 | .000 | .986 | .082  | .042 | 4790.522       | 228         | .000              |
| 절편동일성   | 62582.173 | 746 | .000 | .964 | .109  | .055 | 37503.857      | 228         | .000              |

CFI=.988, TLI=.977, RMSEA=.013으로 좋은 적합도를 보였으므로 해당 변인에 대한 형태동일성이 만족한다고 할 수 있다. 두 번째로 요인계수동일성 모형을 검증한 결과,  $\chi^2(518)=25078.316, p<.001, CFI=.986, TLI=.984, RMSEA=.082$ 으로 나타났다. 형태동일성모형과의 비교 결과  $\Delta\chi^2(228)=4790.522, p<.001, \Delta CFI=-.002, \Delta RMSEA=.069$ 으로  $\chi^2$ 통계량이 통계적으로 유의했으며, 상대적 적합도 지수의 결과에서는 CFI의 차이는 기준치인 .01보다 낮았지만 RMSEA에서 기준치인 .015를 넘어서 증가했으므로 요인계수동일성이 지지된다고 볼 수 없다. 또한 절편동일성모형에서도 CFI와 RMSEA의 두 지수가 기준치보다 더 안좋아졌으므로 MG CFA 모형을 이용한 측정동일성 검정에서는 형태동일성 모형만이 지지된 것으로 볼 수 있다.

**ML CFA(임의절편모형)의 분석 결과**

과학에 대한 즐거움 요인에 대한 ML CFA의 임의절편모형에 대한 분석 결과는 표 11과 같다. 예시 1에서와 같이 형태동일성, 요인계수동일성, 절편동일성 모형을 순차적으로 검정하였다. 먼저, 1수준과 2수준의 요인계수와 2수준 잠재 분산을 자유롭게 추정하는 형태동일성모형을 검정한 결과,  $\chi^2(10)=716.52, p<.001, CFI=.984, RMSEA=.013$ 로 나타나 만족스러운 적합도를 보임을 알 수 있다. 다음으로, 교차수준 간 요인계수는 같게 고정하고 2수준의 잠재분산은 자유롭게 추정하여 요인계수동일성 모형을 검정하였다. 형태동일성모형과의 비교 결과,  $\Delta\chi^2$ 가 263.41으로 유의하게 나타나 형태동일성모형을 지지하는 것으로 나타났다,  $\Delta CFI$ 은 .006,  $\Delta RMSEA$ 은 0에 가까우므로 두 지수에서는 요인계수동일성 모형을 지지하는 것으로 나타났다. 마지막으로 교차

표 11. ML CFA 임의절편모형의 측정동일성 검정 결과

| 모형      | $\chi^2$ | df | p    | CFI  | RMSEA | SRMR<br>-W | SRMR<br>-B | $\Delta\chi^2$ | $\Delta df$ | $p(\Delta\chi^2)$ |
|---------|----------|----|------|------|-------|------------|------------|----------------|-------------|-------------------|
| 형태동일성   | 716.52   | 10 | .000 | .984 | .013  | .012       | .012       |                |             |                   |
| 요인계수동일성 | 979.93   | 14 | .000 | .978 | .013  | .012       | .015       | 263.41         | 4           | .000              |
| 절편동일성   | 19769.43 | 19 | .000 | .988 | .051  | .012       | .115       | 18789.5        | 5           | .000              |

수준 간 요인계수와 2수준의 잠재분산을 동일하게 고정한 절편동일성 모형의 검정 결과를 제시하였다. 그 결과,  $\Delta\chi^2$ 에 대한 p값은 유의하여 요인계수동일성 모형을 지지하였지만,  $\Delta CFI$ 은 .01로 기준치인 .02보다 낮지만  $\Delta RMSEA$ 은 .038으로 기준치인 .03보다 더 컸으므로 절편동일성을 지지하는 것으로 볼 수 없다. 따라서 임의절편모형의 측정동일성 검정 결과는 요인계수동일성 모형을 지지한다고 할 수 있다.

**ML CFA(임의절편 및 임의요인계수 모형)의 분석 결과**

본 예시에서는 ML CFA의 임의절편모형을 통한 측정동일성 검정 결과, 절편동일성 모형

을 지지하지 못하였으므로, 모형의 2수준에서 절편뿐 아니라 요인계수도 임의효과로 추정하는 임의절편 및 임의요인계수 모형을 검정하였다. 앞서 기술하였듯이 이 모형은 베이지안 추정이 요구되며 베이지안 추정에서의 추정치는 신용구간이 0을 포함하지 않으면 유의한 것으로 해석한다. 분석 결과, 표 12에서와 같이 절편과 요인계수의 평균 및 분산 추정치는 모두 유의한 것으로 나타났다. 또한 분산의 추정치는 모두 0.01보다 작은 값으로 나타나 근사 동일성이 지지되는 것으로 해석할 수 있다. 2수준에서의 요인 평균값은 표 13에 제시하였다.

표 12. ML CFA 임의절편 및 임의요인계수 모형의 절편과 요인계수의 분산 추정치

| 문항<br>번호 | 절편    |                |      |              | 요인계수 |              |      |              |
|----------|-------|----------------|------|--------------|------|--------------|------|--------------|
|          | 평균    |                | 분산   |              | 평균   |              | 분산   |              |
|          | 추정치   | 신용구간           | 추정치  | 신용구간         | 추정치  | 신용구간         | 추정치  | 신용구간         |
| 문항 1     | 2.717 | [2.662, 2.770] | .001 | [.000, .001] | .730 | [.707, .753] | .007 | [.004, .010] |
| 문항 2     | 2.577 | [2.524, 2.631] | .000 | [.000, .001] | .730 | [.708, .752] | .006 | [.004, .009] |
| 문항 3     | 2.603 | [2.548, 2.659] | .001 | [.000, .001] | .740 | [.718, .763] | .007 | [.004, .011] |
| 문항 4     | 2.793 | [2.739, 2.846] | .000 | [.000, .001] | .751 | [.728, .774] | .003 | [.002, .005] |
| 문항 5     | 2.766 | [2.710, 2.823] | .000 | [.000, .000] | .762 | [.740, .786] | .004 | [.002, .007] |

표 13. ML CFA 임의절편 및 임의요인계수 모형의 국가별 즐거움의 잠재평균

| 순위 | 국가명    | 잠재평균   | 순위 | 국가명   | 잠재평균   |
|----|--------|--------|----|-------|--------|
| 1  | 튀니지    | 0.425  | 29 | 아이슬란드 | -0.007 |
| 2  | 싱가포르   | 0.400  | 30 | 노르웨이  | -0.026 |
| 3  | 도미니카   | 0.371  | 31 | 몬테네그로 | -0.028 |
| 4  | 태국     | 0.343  | 32 | 이스라엘  | -0.033 |
| 5  | 멕시코    | 0.320  | 33 | 룩셈부르크 | -0.034 |
| 6  | 아랍에미리트 | 0.310  | 34 | 호주    | -0.046 |
| 7  | 중국     | 0.272  | 35 | 폴란드   | -0.052 |
| 8  | 페루     | 0.271  | 36 | 러시아   | -0.063 |
| 9  | 캐나다    | 0.210  | 37 | 스페인   | -0.064 |
| 10 | 코스타리카  | 0.205  | 38 | 스웨덴   | -0.089 |
| 11 | 카타르    | 0.202  | 39 | 프랑스   | -0.108 |
| 12 | 포르투갈   | 0.187  | 40 | 이탈리아  | -0.119 |
| 13 | 리투아니아  | 0.175  | 41 | 스위스   | -0.120 |
| 14 | 콜롬비아   | 0.169  | 42 | 벨기에   | -0.160 |
| 15 | 브라질    | 0.147  | 43 | 대만    | -0.164 |
| 16 | 홍콩     | 0.145  | 44 | 크로아티아 | -0.182 |
| 17 | 불가리아   | 0.140  | 45 | 우루과이  | -0.186 |
| 18 | 미국     | 0.123  | 46 | 핀란드   | -0.189 |
| 19 | 마카오    | 0.116  | 47 | 한국    | -0.244 |
| 20 | 아일랜드   | 0.108  | 48 | 독일    | -0.298 |
| 21 | 뉴질랜드   | 0.085  | 49 | 헝가리   | -0.305 |
| 22 | 에스토니아  | 0.079  | 50 | 슬로바키아 | -0.324 |
| 23 | 그리스    | 0.047  | 51 | 체코    | -0.340 |
| 24 | 영국     | 0.029  | 52 | 일본    | -0.409 |
| 25 | 라트비아   | 0.006  | 53 | 오스트리아 | -0.455 |
| 26 | 터키     | 0.000  | 54 | 슬로베니아 | -0.495 |
| 27 | 칠레     | -0.001 | 55 | 네덜란드  | -0.571 |
| 28 | 덴마크    | -0.006 |    |       |        |

## 논 의

집단 비교 연구 시 측정동일성의 성립 여부를 집단 간 의미 있는 비교를 하기 위한 필수 요건으로 제시되고 있다. 이를 위해 일반적으로 다집단 확인적요인분석(multiple group confirmatory factor analysis, MG CFA)이 널리 사용되어 왔다. 그러나 국제 비교 연구에서는 많은 수의 국가(혹은 집단)를 비교하는 경우가 흔히 발생한다. 기존에 널리 사용되었던 다집단분석(multiple group confirmatory factor analysis, MG CFA)에서는 일반적으로 그 대상이 두 개 정도이었으나, 10개 이상의 많은 수의 국가나 문화를 비교하기 위해서는 보다 적절한 통계분석 방법이 요구된다(Kim, Cao, Wang, & Nguyen, 2017).

이에 따라 본 연구에서는 많은 수의 집단을 비교하기 위해 모수를 임의효과로 추정하는 방법론을 다층 확인적요인분석(ML CFA)의 임의절편모형과 임의절편 및 임의요인계수모형, 그리고 다층 요인혼합모형(ML FMM)으로 정리하였다. 또한 이러한 방법론을 적용한 예시로서 PISA 2015 자료를 활용하여 학생들이 지각한 과학의 도구적 동기 및 과학의 즐거움의 구인에 대해 국가별 측정동일성 검정을 실시하고 잠재평균을 분석하였다.

다수의 집단을 비교하기 위해 임의효과모형으로 추정하는 이들 방법론은 공변인을 모형에 포함하여 비동일성의 원인을 조사하는 연구로 확대할 수 있다. ML CFA에 공변량이 포함된 모형은 MIMIC(multiple-indicators, multiple-causes) 모형으로 불리며, 요인에 대한 공변인을 포함함으로써 집단 간의 차이를 설명할 수 있다(Jöreskog, & Goldberger, 1975). 이때 공변인은 범주형, 연속형 변수 모두 가능

하다. 예를 들어, 국가 간 차이를 특정 제도의 시행 여부(범주형 변수) 혹은 GDP(연속형 변수)로 설명하고자 한다면, 이를 요인에 대한 설명변수로 포함할 수 있으며, 여러 개의 공변인을 동시에 모델링할 수 있다. 또한 측정동일성이 성립하지 않았다면 문항에서의 비동일성을 탐색하기 위해 공변인을 각 문항에 회귀시켜 유의한지 확인할 수 있다. 그 외 다층 MIMIC 모형에 대한 자세한 사항은 본 연구의 범위를 벗어나므로 더 자세한 사항은 다음의 논문을 참고할 수 있다(예: Jak, Oort, & Dolan, 2013; 2014; Kim, & Cao, 2015; Kim, Yoon, Wen, & Kwok, 2015). 마찬가지로 ML FMM에서는 소속 계층(class membership)을 설명하는 공변인을 모형에 포함하여 비동일성의 원인을 조사할 수 있으며(Lubke, & Muthén, 2005), ML CFA에서와 같이 이러한 공변인은 여러 개가 투입될 수 있다.

본 연구에서 다루었던 방법론 이외에도 다수의 집단을 비교하는 방법으로 정렬 최적화(Alignment Optimization; Asparouhov, & Muthén, 2014) 방법이 있다. 정렬 최적화는 모든 측정모수에 대한 함수를 정의하고 가장 눈에 띄는 비동일성 문항을 제외한 사소한 차이는 허용함으로써 이 함수가 최적이 되도록 셋(set)을 찾는 것이다. 심화된 내용은 본 연구의 범위를 벗어나므로 논의에서 제외하였다. 정리하자면, 연구자들은 다수의 집단을 비교하고자 할 때, ML CFA, ML FMM과 더불어 정렬 최적화 방법도 적용 가능하므로, 연구자들이 이를 고려한다면 자신의 연구주제에 더 적합한 방법론을 적용하는데 도움이 될 것이다.

마지막으로 본 연구는 국내 연구자들에게 국가비교 연구와 같이 자료가 다층 구조이면서, 비교하고자 하는 집단이 다수일 때 잠재

평균을 분석할 수 있는 방법론을 상세히 소개하였다는 점에서 의의가 있다. Kim 외(2017)는 다수 집단에 대한 다섯 가지의 측정동일성 방법(즉, MG CFA, ML CFA 모형과 ML FMM의 정확검정 방법과 베이지안 근사 검정과 정렬 최적화 방법의 근사검정 방법)에 대해 소개하고 시뮬레이션을 통해 효과성을 비교한 바 있다. 그러나 해당 연구에서의 ML CFA 모형은 임의절편모형만 다루고 있으며, 측정동일성 검정 절차가 상세하지 않은 측면이 있다. 본 연구에서는 ML CFA에 대해 임의절편모형뿐 아니라 임의절편 및 임의요인계수 모형으로 확대하여 측정동일성 검정 절차를 상세히 소개하였으며, 방법론적 인지도가 낮은 ML FMM의 방법에 대하여 보다 구체적으로 소개하고 검정 절차를 상세히 보여주었다는 점에서 의의가 있다. 마지막으로 국내 연구자들이 국외의 시뮬레이션 연구를 통해 적용연구를 한다는 것이 쉽지 않다는 측면에서, PISA 2015의 실제 자료를 통해 과학 동기 구인 중 도구적 동기와 즐거움에 대한 국가별 잠재평균분석의 연구방법, 분석결과, *Mplus* syntax를 예시로 제시함으로써 향후 국내 연구자들에게 활용 가능성을 제시하였다는 점에서 의의가 있다.

## 참고문헌

- 박찬호 (2017). A Comparative Study on TIMSS Mathematics Assessment of Korea, Japan, and the USA: A Differential Item Functioning Approach. *비교교육연구*, 27, 1-19.
- 이종희, 김기연, & 김수진 (2011). 수학 학업성취도가 높은 국가의 수학-정의적 영역 요인 분석 및 측정 동일성 검증. *학교수학*, 13(2), 307-321.
- 홍세희 (2000). 구조 방정식 모형의 적합도 지수 선정기준과 그 근거. *한국심리학회지: 임상*, 19(1), 161-177.
- 홍세희 (2015). *위계적자료 분석을 위한 횡단 다층모형*. 서울: 에스앤엠리서치그룹.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.
- Allua, S., Stapleton, L. M., & Beretvas, S. N. (2008). Testing latent mean differences between observed and unobserved groups using multilevel factor mixture models. *Educational and Psychological Measurement*, 68(3), 357-378.
- Asparouhov, T., & Muthén, B. (2012, July). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. In annual meeting of the National Council on Measurement in Education, Vancouver, British Columbia.
- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 495-508.
- Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational and Behavioral Statistics*, 6(3), 267-285.
- Bogler, R., & Nir, A. E. (2012). The importance of teachers' perceived organizational support to job satisfaction: What's empowerment got to do with it?. *Journal of Educational Administration*, 50(3), 287-306.
- Browne, M. W., & Cudeck, R. (1993). Alternative

- ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Buzick, H. M. (2010). Testing for heterogeneous factor loadings using mixtures of confirmatory factor analysis models. *Frontiers in psychology, 1*, 1-9.
- Byrne, B. M., & Van de Vijver, F. J. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing, 10*(2), 107-132.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464-504.
- Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology, 31*, 187-212.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.
- Chopik, W. J., O'Brien, E., & Konrath, S. H. (2017). Differences in empathic concern and perspective taking across 63 countries. *Journal of Cross-Cultural Psychology, 48*(1), 23-38.
- Da Costa, L. P., & Dias, J. G. (2014). Perceptions of poverty attributions in Europe: A multilevel mixture model approach. *Quality & Quantity, 48*(3), 1409-1419.
- Davidov, E., Dülmer, H., Schlüter, E., Schmidt, P., & Meuleman, B. (2012). Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. *Journal of Cross-Cultural Psychology, 43*(4), 558-575.
- De Boeck, P. (2008). Random Item IRT Models. *Psychometrika, 73*, 533-559.
- De Jong, M. G., & Steenkamp, J. B. E. (2010). Finite mixture multilevel multidimensional ordinal IRT models for large scale cross-cultural research. *Psychometrika, 75*(1), 3-32.
- Frederickx, S., F. Tuerlinckx, P. De Boeck, and D. Magis. (2010). RIM: A Random Item Mixture Model to Detect Differential Item Functioning. *Journal of Educational Measurement, 47*, 432-457.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*(1), 72-91.
- Horn, J. L., & McArdle, J. J. (1992). A practical guide to measurement invariance in research on aging. *Experimental Aging Research, 18*, 117-144.
- Hox, J. J., & Maas, C. J. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural equation modeling, 8*(2), 157-174.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal, 6*(1), 1-55.
- Jak, S., Oort, F. J., & Dolan, C. V. (2013). A test for cluster bias: Detecting violations of

- measurement invariance across clusters in multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(2), 265-282.
- Jak, S., Oort, F. J., & Dolan, C. V. (2014). Measurement bias in multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(1), 31-39.
- Jang, S., Kim, E. S., Cao, C., Allen, T. D., Cooper, C. L., Lapierre, L. M., ... & Abarca, N. (2017). Measurement invariance of the satisfaction with life scale across 26 countries. *Journal of Cross-Cultural Psychology*, 48(4), 560-576.
- 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 indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351a), 631-639.
- Kaplan, D., Kim, J.-S., & Kim, S.-Y. (2009). Multilevel latent variable modeling: Current research and recent developments. In R. Millsap & A. Maydeu-Olivares (Eds.), *Sage handbook of quantitative methods in psychology* (pp. 592-612). Thousand Oaks, CA: Sage.
- Kim, E. S., & Cao, C. (2015). Testing group mean differences of latent variables in multilevel data using multiple-group multilevel CFA and multilevel MIMIC modeling. *Multivariate behavioral research*, 50(4), 436-456.
- Kim, E. S., & Wang, Y. (2018). Investigating Sources of Heterogeneity with Three-Step Multilevel Factor Mixture Modeling: Beyond Testing Measurement Invariance in Cross-National Studies. *Structural Equation Modeling: A Multidisciplinary Journal*, 1-17.
- Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: A comparison of five approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(4), 524-544.
- Kim, E. S., Joo, S. H., Lee, P., Wang, Y., & Stark, S. (2016). Measurement invariance testing across between-level latent classes using multilevel factor mixture modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(6), 870-887.
- Kim, E. S., Yoon, M., Wen, Y., Luo, W., & Kwok, O. M. (2015). Within-level group factorial invariance with multilevel data: multilevel factor mixture and multilevel mimic models. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(4), 603-616.
- Kreft, I. G. G., & De Leeuw, J. (1998). *Introducing multilevel modeling*. Newbury Park, CA: Sage.
- 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.
- Little, T. D. (2000). On the comparability of constructs in cross-cultural research: A critique of Cheung and Rensvold. *Journal of Cross-cultural Psychology*, 31(2), 213-219.
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88, 767-778.
- Lubke, G. H., & Muthén, B. (2005). Investigating

- population heterogeneity with factor mixture models. *Psychological Methods*, 10(1), 21-39.
- Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86-92.
- Marsh, H. W., & Hau, K. T. (2003). Big-Fish--Little-Pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American psychologist*, 58(5), 364.
- McLachlan, G., & Peel, D. (2000). Finite mixture models. Hoboken, NJ: Wiley.
- Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological methods*, 10(3), 259-284.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13(2), 127-143.
- 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.
- Millsap, R. E. (1997). Invariance in measurement and prediction: Their relationship in the single-factor case. *Psychological Methods*, 2(3), 248-260.
- Millsap, R. E., & Everson, H. (1991). Confirmatory measurement model comparisons using latent means. *Multivariate Behavioral Research*, 26(3), 479-497.
- Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological methods & Research*, 22(3), 376-398.
- Muthén, B. O., and Asparouhov, T. (2013). *BSEM Measurement Invariance Analysis*. Mplus Web Notes: No. 17. Available online at: www.statmodel.com
- Muthén, B., & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups: alignment and random effects. *Sociological Methods & Research*, 47(4), 637-664.
- Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of school psychology*, 48(1), 85-112.
- Ramaswamy, V., Desarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science*, 12(1), 103-124.
- Redner, R. A., & Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26, 195-239.
- Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31-57.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507-514.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461-464.
- Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 333-343.
- Selig, J. P., Card, N. A., & Little, T. D. (2008).

- Latent variable structural equation modelling in crosscultural research: Multigroup and multilevel approaches. In F. J. R. van de Vijver, D. A., van Hemert, & Y. H. Poortinga (Eds.), *Multilevel analysis of individuals and cultures* (pp. 93-119). Mahwah, NJ: Erlbaum.
- Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics, 41*(5), 481-520.
- Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78-90.
- Tay, L., Woo, S. E., & Vermunt, J. K. (2014). A conceptual and methodological framework for psychometric isomorphism: Validation of multilevel construct measures. *Organizational Research Methods, 17*(1), 77-106.
- Van de Vijver, F. J., & Poortinga, Y. H. (2002). Structural equivalence in multilevel research. *Journal of Cross-Cultural Psychology, 33*(2), 141-156.
- 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), 4-70.
- Verhagen, A. J. and J. P. Fox. (2013). Bayesian Tests of Measurement Invariance. *The British Journal of Mathematical and Statistical Psychology, 66*, 383-401.
- Vermunt, J. K. (2007, August). Multilevel mixture item response theory models: An application in education testing. Paper presented at the 56th Session of the International Statistical Institute, Lisbon, Portugal.
- 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.
- Żemojtel Piotrowska, M., Piotrowski, J. P., Osin, E. N., Ciecuch, J., Adams, B. G., Ardi, R., ... & de Clunie, G. T. (2018). The mental health continuum short form: The structure and application for cross cultural studies-A 38 nation study. *Journal of clinical psychology, 74*(6), 1034-1052.
- Zyphur, M. J., Kaplan, S. A., & Christian, M. S. (2008). Assumptions of cross-level measurement and structural invariance in the analysis of multilevel data: Problems and solutions. *Group Dynamics: Theory, Research, and Practice, 12*(2), 127-140.
- 1차원고접수 : 2018. 10. 30.  
2차원고접수 : 2019. 03. 20.  
3차원고접수 : 2019. 06. 10.  
최종게재결정 : 2019. 06. 27.

**Measurement Invariance Testing Using Random Effects model  
for Many groups: Multilevel Confirmatory Factor Analysis (ML CFA)  
and Multilevel Factor Mixture Modeling (ML FMM)**

**Sookyung Son**

**Hyojin Kim**

**Sehee Hong**

Korea University, Department of Education

In multi-group analysis, measurement invariance is a requirement for meaningful comparisons between groups. Multi-group confirmatory factor analysis (MG CFA) has been widely used for group comparisons. However, MG CFA is appropriate for comparisons with a small number of groups and is limited for a large number of groups, in particular, in cross-cultural comparative studies. To overcome the limitation of MG CFA, this study described alternative approaches: multilevel confirmatory factor analysis (ML CFA) and multilevel factor mixture modeling (ML FMM), which are effective for comparing more than 10 groups. In ML CFA, its advantages and disadvantages were described by introducing two models: random intercept models that estimate only intercepts as random effects and random intercept and loading models that estimate intercepts and factor loadings as random effects. Specifically, this study presented theoretical models for the two methods and procedures for testing measurement invariances. In addition, this study discussed advantages of ML CFA, relative to those of MG CFA, and several points that should be considered when applying ML CFA. And, as an example of applying ML CFA, this study conducted latent means analysis across countries for instrumental motivation of science and enjoyment perceived by students using the PISA 2015 data. Finally, implications of this study and future research directions were discussed.

*Key words* : many group comparison, measurement invariance, latent means comparison, multiple group analysis, multilevel confirmatory analysis, multilevel factor mixture modeling

**부 록**

1. MG CFA

Data: file is data.dat;  
 Variable: names are cnt Imot1-Imot4;  
 Usevariables are Imot1-Imot4;  
 Grouping is cnt (36=AUS 40=AUT ... **중략**  
 970=QCH 971=QES);  
 Missing=all(-999);  
 Analysis: estimator = ml;  
 Model = configural metric scalar **!형태동일성,**  
**요인계수동일성, 절편동일성 모형 검정**  
 Model: instrument by Imot1-Imot4;

2. ML CFA(임의절편모형)

1) **형태동일성모형**

Data: file is data.dat;  
 Variable: names are cnt Imot1-Imot4;  
 Usevariables are Imot1-Imot4;  
 Missing=all(-999);  
 Cluster = cnt;  
 Analysis: type = twolevel;  
 Model:  
 %within%  
 fw by Imot1-Imot4;  
 %between%  
 fb by Imot1-Imot4;

2) **요인계수동일성모형**

(Data~Analysis는 상동)  
 model:  
 %within%  
 fw by Imot1

Imot2(1)  
 Imot3(2)  
 Imot4(3);  
 %between%  
 fb by Imot1  
 Imot2(1)  
 Imot3(2)  
 Imot4(3);  
 Savedata: FILE = metric\_output.dat;  
 Save = fscores;

3) **절편동일성모형**

(Data~Analysis는 상동)

model:  
 %within%  
 fw by Imot1  
 Imot2(1)  
 Imot3(2)  
 Imot4(3);  
 %between%  
 fb by Imot1  
 Imot2(1)  
 Imot3(2)  
 Imot4(3);  
 Imot1-Imot4@0;  
 Savedata: FILE = scalar\_output.dat;  
 Save = fscores;

3. ML CFA(임의절편 및 임의요인계수모형)

Data: file is data.dat;  
 Variable: names are cnt enjsci1-enjsci5;  
 Usevariables are enjsci1-enjsci5;  
 Missing=all(-999);  
 Cluster = cnt;

```

Analysis: type=twolevel; proc=2; type=random;          %cb#2%
estimator=bayes; fbiter=5000;                          f by Imot1-Imot3;
Model:                                                  Imot1-Imot4;
%within%                                              {f} ; f;
s1-s5 | fw BY enjsci1-enjsci5; fw@1;                  %cb#3%
%between%                                             f by Imot1-Imot3;
enjsci1-enjsci5 s1-s5;                                Imot1-Imot4;
{s1-s5*1} (p1-p5);                                   {f} ; f;
fb by enjsci1-enjsci5*1 (p1-p5);                     %between%
enj by s1-s5*1 (p1-p5);                              %overall%
fb enj;                                               {Imot4}(int);
Output: tech1 tech8;                                  {Imot1-Imot3};
Plot: type = plot3;                                   %cb#2%
factors = fb(100);                                   {Imot4}(int);
savedata: file = bayes_output.dat;                   %cb#3%
save = fscores (100);                                {Imot4}(int);
    
```

#### 4. ML FMM

##### 1) 형태동일성모형

```

Data: file is data.dat;
Variable: names are cnt Imot1-Imot4;
Usevariables are Imot1-Imot4;
Missing=all(-999);
Classes = cb(3);
Between = cb;
Cluster = cnt;
Analysis: type = twolevel mixture;
Algorithm = integration;
Processors = 4;
Model:
%within%
%overall%
f by Imot1-Imot4*;
{f@0}; f@1;
    
```

##### 2) 요인계수동일성모형

(Data~Analysis는 상동)

```

Model:
%within%
%overall%
f by Imot1-Imot4*;
{f@0}; f@1;
%cb#2%
Imot1-Imot4;
{f} ; f;
%cb#3%
Imot1-Imot4;
{f} ; f;
%between%
%overall%
{Imot4}(int);
{Imot1-Imot3};
    
```

### 3) 절편동일성모형

(Data~Analysis는 상동)

Model:

%within%

%overall%

f by Imot1-Imot4\*;

{f@0}; f@1;

%cb#2%

Imot1-Imot4;

{f} ; f;

%cb#3%

Imot1-Imot4;

{f} ; f;

%between%

%overall%

{Imot1-Imot4}(int1-int4);

savdata: File is class3\_scalar.txt;

save=cprob fscores;