

## 정렬법을 이용한 범주형 자료의 근사 측정동일성 분석\*

주 영 신

성균관대학교 심리학과 / 박사과정

장 승 민†

성균관대학교 심리학과 / 교수

요인모형을 이용한 측정동일성 분석에는 다집단 확인적 요인분석이 주로 사용되어왔다. 전통적인 관점에서 측정동일성이 성립한다고 결론을 내리기 위해 비교 집단이나 시점 사이에 요인구조와 측정모수가 차이가 없이 완전히 동일하다는 근거가 필요하다. 그러나 집단의 수가 많은 경우 이와 같은 조건은 지나치게 엄격하거나 비현실적이 되기도 한다. 근사 측정동일성(approximate measurement invariance)은 비교 집단의 개수가 많을 때 이들 사이의 측정동일성을 현실적으로 검토하기 위해 제안된 완화된 형태의 측정동일성을 의미한다. 근사 측정동일성을 분석하는 주요 절차인 다집단 확인적 요인분석의 정렬법은 주로 연속형 측정 변수에 대한 요인모형을 다수 국가 간에 비교하는 데 사용되어 왔다. 반면 연령에 따른 측정동일성의 분석은 주로 다집단 확인적 요인분석을 이용하여 둘 또는 셋으로 구분된 연령 집단을 비교하는 방식으로 수행되어 왔다. 그러나 연령에 따른 측정동일성의 분석에도 연령 집단을 세분한 다수의 연령 집단에 대해 근사 측정동일성을 적용할 수 있다. 본 연구는 근사 측정동일성의 개념과 필요성, 분석 모형에 대해 논의하고 정렬법을 이용하여 심리검사의 연령에 따른 근사 측정동일성을 분석하는 방법과 절차를 제시하였다. 특히 범주형 자료에 대한 측정동일성 및 근사 측정동일성 분석에서의 개념적 유사점과 차이점을 논의하였으며 다집단 확인적 요인분석과 정렬법을 이용하여 인지기능 평가를 위한 검사인 K-MMSE 검사점수의 26개 연령 집단에 대한 측정동일성 및 근사 측정동일성을 분석하였다. 이를 통해 전통적인 측정동일성과 비교하여 근사 측정동일성 분석의 유용성을 강조하고 범주형 자료의 근사 측정동일성 분석과 관련된 논의사항들을 정리하였다.

주요어 : 범주형 변수의 요인분석, 다집단 확인적 요인분석, 근사 측정동일성, 정렬법, K-MMSE

\* 본 논문은 2022년 4월 한국심리측정평가학회에서 발표되었음.

† 교신저자: 장승민, 성균관대학교 심리학과, 서울특별시 종로구 성균관로 25-2, Tel: 02-760-0692

E-mail: jahngs@skku.edu



Copyright © 2023, The Korean Psychological Association. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial Licenses(<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution and reproduction in any medium, provided the original work is properly cited.

측정동일성(measurement invariance)은 같은 수준의 속성을 가지는 대상을 측정한 결과가 이 속성과 무관한 조건으로 인해 달라지지 않는 것을 의미한다(Meredith, 1993; Millsap, 2011). 예를 들어 어떤 척도를 이용한 우울의 측정의 성별에 대해 측정동일성을 가지면 우울 수준이 같은 남성과 여성은 이 척도에서 동일한 검사점수를 얻을 것이다. 서로 다른 조건 사이의 측정동일성은 해당 조건에서 얻은 측정 자료를 비교·분석하여 평가할 수 있다. 일반적으로 심리학 연구에서 측정동일성의 분석은 서로 다른 집단이나 시점에서 동일한 도구에 의해 측정된 검사점수 또는 잠재변수를 같은 의미로 해석하고 비교하는 것이 타당한지를 판단하기 위해 수행된다. 따라서 측정변수나 잠재변수에 대한 분석 결과를 집단 간에 비교할 때는 이들 사이에 측정동일성이 성립하는지를 먼저 검토하는 것이 중요하다(Vandenberg & Lance, 2000).

측정동일성의 분석에는 같은 측정지표들(예, 척도 문항 점수)에 대한 확인적 요인분석(confirmatory factor analysis, CFA)을 여러 집단에 동시에 적용하여 비교하는 다집단 CFA(Jöreskog, 1971)가 널리 사용된다. 다집단 CFA를 사용하면 측정지표와 요인 관계의 동일성(요인동일성, factorial invariance)이 집단 간에 성립하는지를 확인할 수 있다. 측정지표와 요인의 관계는 요인구조와 측정모수들(예, 요인부하량, 절편 등)에 의해 결정되는데 이 요소들의 동일성을 일정한 순서에 따라 단계적으로 평가하여 요인동일성의 수준을 결정할 수 있다(Meredith, 1993; Vandenberg & Lance, 2000).

전통적인 의미에서 요인부하량이나 절편과 같은 측정모수의 동일성은 각 지표변수의 측정모수의 값이 집단 간에 완전히 일치하는 것

을 의미한다. 표본 자료를 이용한 분석에서 완전한 측정동일성(exact measurement invariance)은 모든 집단에서 측정모수가 동일하다는 제약이 부여된 모형의 카이제곱( $\chi^2$ )과 이러한 제약이 부여되지 않은 모형의 카이제곱의 차이가 통계적으로 유의하지 않을 때 지지된다. 그러나 카이제곱 검정은 표본 크기에 민감하기 때문에 집단 사이에 모수 추정값의 차이가 매우 작더라도 표본 크기가 크면 집단 간에 차이가 없다는 영가설을 쉽게 기각하며 이러한 경향은 추정모수의 개수가 많을수록 더욱 강하게 나타난다. 따라서 지표변수가 많을수록 측정모수의 개수도 비례하여 증가하는 다집단 CFA에서 높은 수준의 완전한 측정동일성을 지지하는 결과를 얻기는 쉽지 않다(Cheung & Rensvold, 2002; Herzog et al., 2007; Kenny & McCoach, 2003).

이러한 이유로 표본 크기가 큰 자료에 대해 측정동일성을 분석할 때는 카이제곱 대신에 CFI나 RMSEA와 같은 근사적합도지수의 차이에 근거하여 동일성의 수준을 결정하는 방식을 사용하기도 하며(Chen, 2007; Cheung & Rensvold, 2002), 일부 문항의 측정모수를 동일성 제약 없이 자유롭게 추정하도록 허용하는 부분 측정동일성(partial measurement invariance)을 적용하기도 한다(Byrne et al., 1989). 그러나 측정동일성 분석 시 다양한 상황에 공통적으로 적용할 수 있는 근사적합도지수를 활용한 단일한 규칙은 없으며 지표변수의 속성에 따라 판단 기준이 달라질 수도 있다(Sass et al., 2014). 또한 부분 측정동일성은 분석 절차와 순서가 명료하지 않고 임의적이어서 측정동일성을 평가하는 절차로 한계가 있다. 특히 비교 집단의 수가 많아지면 추정모수의 개수가 크게 증가하게 되며 이 경우 측정동일성을 평

가하는 일반적 절차나 그에 대한 보완적 절차의 한계점이 더 두드러진다.

Asparouhov와 Muthén(2014)은 비교 집단의 수가 많아 측정모수의 개수가 많고 표본 크기가 큰 등의 이유로 완전한 측정동일성을 충족하기 어려운 경우에 적용할 수 있는 덜 엄격한 의미의 측정동일성으로 근사 측정동일성(approximate measurement invariance)의 개념을 제안하였다.<sup>1)</sup> 전통적인 의미의 측정동일성과 달리 근사 측정동일성은 측정모수 간의 근소한 차이를 허용한다. 다집단 CFA에 대한 정렬법(alignment method)은 이러한 근사 측정동일성 모형을 추정하고 분석하기 위해 사용되는 절차이다. 정렬법은 먼저 형태 동일성만 가정된 다집단 CFA에서 각 집단의 측정모수를 추정한 후 이들이 서로 유사한 값을 갖도록 변환한다. 이는 자료를 설명하는 모형의 본질은 동일하게 유지한 채 잠재변수의 척도를 조정하는 재모수화를 통해 측정모수들의 차이를 줄이는 방법이다. 근사 측정동일성과 정렬법은 주로 대규모 서베이 자료를 국제적으로 수집하여 국가 간 특성을 비교하는 연구에 많이 적용되어왔다(Lomazzi, 2018; Marsh et al., 2018; Munck et al., 2018; Son et al., 2019).

정렬법은 국가 간 비교 외에도 다른 분야에 유용하게 적용될 수 있는데 그중 하나가 발달적 변화와 관련된 심리검사의 연령에 따른 측정동일성 분석이다. 심리검사나 심리척도의

연령에 따른 측정동일성 분석은 임의의 연령 기준으로 구분된 둘 또는 셋 정도의 연령 집단에 대해 다집단 CFA를 수행함으로써 이루어지는 것이 일반적이다. 그러나 이런 방식의 연령 집단 구분은 측정 도구와 대상의 연령에 따른 측정동일성에 관한 논리적·이론적 근거에 의한 것이라기보다는 표본 크기가 충분치 않거나 분석 절차를 단순화할 수 있다는 등의 현실적 이유 또는 편의적인 이유에 의한 경우가 많다. 따라서 표본 크기가 충분하고 이론상이나 분석상 다른 이유가 없다면 연령 집단을 세분하여 다수의 연령 집단에 걸쳐 촘촘하게 측정동일성을 검토한다면 연령에 따른 검사점수 또는 요인평균의 차이에 대한 해석을 더욱 명확하게 할 수 있다.

정렬법을 이용한 근사 측정동일성 분석은 지표변수가 연속형 변수이고 요인과 선형적 관계를 맺는다고 가정된 맥락에서 주로 적용되어왔다. 그러나 심리학을 비롯한 행동과학 분야에서 잠재변수를 측정하기 위해 사용되는 측정변수는 등간의 속성을 가지는 연속형 변수가 아니고 순서가 있으면서도 범주적으로 구분되는 순서 범주형(ordered-category) 변수 또는 이분형(dichotomous) 변수인 경우가 흔하다. 지표변수가 범주형인 경우에도 측정동일성 분석은 다집단 CFA를 이용하거나 정렬법을 사용할 수 있다. 그러나 범주형 변수에 대한 정렬법을 이해하기 위해서는 범주형 변수에 대한 요인모형과 다집단 CFA에 대한 이해가 전제되어야 한다. 범주형 변수에 대한 요인모형은 식별을 위한 척도 설정 방식, 측정동일성의 요소, 추정법과 분석 절차 등에서 연속형 변수에 대한 요인모형과 다른 점이 많다.

본 연구는 많은 집단 사이의 측정동일성을 평가하기 위해 제안된 근사 측정동일성의 개

1) 요인동일성은 측정동일성을 정의하는 한 가지 방법으로 엄밀한 의미에서 둘은 동의어가 아니다(Meredith, 1993). 그러나 근사 측정동일성에 관한 문헌에서는 근사 요인동일성이라는 용어가 사용되고 있지 않다. 따라서 이 글에서도 근사 요인동일성이라는 용어가 더 적절한 맥락에서도 이를 근사 측정동일성으로 통합하여 사용하였다.

념과 필요성을 논의하고 다집단 CFA의 정렬법이 어떻게 근사 측정동일성을 모델링하고 평가하는지를 소개하였다. 특히 주로 국가 간 비교 연구 등에 활용되는 정렬법을 연령에 따른 측정동일성을 분석하고 평가하는 데에 확장하여 적용하였다. 또한 연속형 지표변수의 측정모형을 정렬하는 절차에 대한 논의를 범주형 자료의 측정모형으로 확장하고 실제 분석 예시를 통해 이 절차가 어떻게 적용될 수 있는지를 보였다.

다음 절에서는 먼저 측정동일성의 개념과 다집단 CFA를 이용한 측정동일성 분석 절차를 요약하고 이 절차의 제한점을 논의하였다. 이어서 다집단 CFA의 해가 정렬을 통해 어떻게 근사 측정동일성을 최대화하고 분석하는지, 정렬법의 원리와 절차를 개략하였다. 다음으로 범주형 변수에 대한 다집단 CFA의 특징을 요약하고 정렬법이 이러한 자료에 어떻게 적용될 수 있는지를 논의하였다. 끝으로 6,500명 이상의 장노년을 대상으로 수집한 패널자료에 포함된 심리검사(K-MMSE)의 문항 점수를 사용하여 범주형 변수에 대한 다집단 CFA와 정렬법이 연령에 따른 근사 측정동일성 분석에 어떻게 적용될 수 있는지 예시하였다.

### 측정동일성 분석과 근사 측정동일성

다집단 CFA를 이용한 측정동일성 분석과 근사 측정동일성

다집단 CFA(Jöreskog, 1971)는 여러 집단에 동일한 요인구조를 가진 측정모형을 설정하여 모수를 추정하고 모형 적합도를 평가한다. 지표변수가 연속형 변수이고 잠재변수와 선형적

관계를 가정할 때 다집단 CFA 모형은 식 (1)과 같이 표현한다.

$$y_{ipg} = \nu_{pg} + \lambda_{pg}\eta_{ig} + \epsilon_{ipg} \quad (1)$$

집단  $g$ 에 속하는 개인  $i$ 에 대해, 변수  $p$ 의 관찰값  $y_{ipg}$ 는 절편  $\nu_{pg}$ , 요인부하량  $\lambda_{pg}$ 를 가중치로 하는 요인점수  $\eta_{ig}$ , 그리고 오차  $\epsilon_{ipg}$ 의 합으로 결정된다. 다집단 CFA의 기본 모형은 각 문항의 측정모수인 절편, 요인부하량, 오차분산( $\sigma_{pg}^2$ )을 집단별로 추정한다.

다집단 CFA에서 지표변수와 요인의 관계를 결정하는 요소들의 집단 간 동일성은 전형적으로 (1) 요인구조의 동일성, (2) 요인부하량의 동일성, (3) 절편의 동일성의 순서로 평가한다.<sup>2)</sup> 먼저 동일한 요인구조를 가지는 식 (1)의 다집단 CFA 모형이 분석 자료에 부합하면, 즉 모형 카이제곱 검정에 의해 완전 적합 가설이 기각되지 않거나 근사적합도지수가 기준에 부합하면 이들 사이에는 가장 낮은 수준의 동일성인 형태 동일성이 충족된다. 다집단 CFA를 이용하여 더 높은 수준의 측정동일성을 확인할 때에는 동일성 제약의 수준이 낮은 모형과 높은 모형을 비교한다. 이때 두 모형의 카이제곱의 차이( $\Delta\chi^2$ )가 통계적으로 유의하지 않으면 더 높은 수준의 동일성이 지지되고 유의

2) 이상의 세 가지 요소의 동일성은 집단 간에 요인의 분산, 공분산, 평균이나 지표변수의 평균을 비교할 때 요구되는 동일성과 관련된다. 이외에도 오차 분산/공분산의 동일성을 추가로 평가할 수 있으나 이것의 중요성과 필요성에 대해서는 학자들마다 이견이 있다(Gregorch, 2006; Little, 2013). 이 글에서는 잠재변수의 분석과 관련된 요인동일성을 다루며 오차 분산의 동일성에 대한 논의는 포함하지 않았다.

하면 지지되지 않는다. 요인구조의 동일성과 더불어 요인부하량의 동일성이 충족되면 더 높은 수준의 동일성인 약한 동일성이 성립한다. 여기에 절편의 동일성이 추가로 충족되면 더 높은 수준인 강한 동일성이 성립한다. 다집단 CFA에서 약한 동일성 이상의 요인 동일성이 성립하면 잠재변수의 분산과 공분산의 집단 간 차이, 강한 동일성 이상의 요인동일성이 성립하면 잠재변수의 평균의 집단 간 차이를 타당하게 분석할 수 있다 (Brown, 2015).

측정동일성의 수준을 확인하기 위한 카이제곱 차이 검정은 표본 크기가 클 때 영가설을 쉽게 기각하여 측정동일성의 수준을 더 낮게 결론을 내리게 하는 경향이 있다(Cheung & Rensvold, 2002). 이러한 이유로 측정동일성의 수준을 결정할 때에는 모형 간 카이제곱의 차이가 아니라 CFI나 RMSEA와 같은 근사적합도 지수의 차이를 참조하는 방법이 많이 사용된다. 대규모 표본에서는 근사적합도지수가 나빠지는 정도인  $\Delta CFI$ 가 .01보다 작거나(Cheung & Rensvold, 2002),  $\Delta RMSEA$ 가 .015 이하(Chen, 2007)이면 더 높은 수준의 측정동일성 모형을 채택하는 것이 권장된다. 그러나 표본 크기와 집단 간 표본 크기의 균형, 비동일성의 패턴 등에 따라 요구되는 기준이 다르고 특히 표본 크기가 작을 때는 더 엄격한 기준이 요구된다(예,  $\Delta CFI < .005$ ,  $\Delta RMSEA < .010$ , Chen, 2007).

측정동일성을 집단 사이에 측정모수 값의 완전한 일치로 정의하는 방식은 집단의 수가 많을 때에는 적용이 어렵다. 예를 들어 다수 집단의 측정동일성 분석이 대표적으로 적용되는 국가 간 비교의 경우, 국가마다 가지고 있는 고유한 문화적, 사회적 특징으로 인해 심

리적 구성개념을 측정하는 검사도구의 측정동일성이 수많은 국가 간에 완전하게 일치하기는 매우 어려우며 측정동일성이 집단 간 잠재평균의 비교가 가능한 수준까지 확립되는 사례는 매우 드물다(Muthén & Asparouhov, 2018; Van de Schoot et al., 2015).

이러한 이유로 분석집단이 많은 경우 비교를 위해 요구되는 측정동일성을 덜 엄격하게 정의하는 근사 측정동일성(approximate measurement invariance)의 개념이 제안되었다(Asparouhov & Muthén, 2014; Muthén & Asparouhov, 2014). 근사 측정동일성은 측정모수의 완전한 일치를 동일성의 기준으로 요구하지 않고 집단 간 측정모수들의 차이가 수용 가능한 수준이라면 근사적으로 동일하다고 간주한다. 근사 측정동일성을 평가하는 절차로 가장 널리 사용되는 방법이 정렬법이다(Asparouhov & Muthén, 2014).

#### 정렬법을 이용한 근사 측정동일성 분석

정렬법은 측정모수에 대한 동일성 제약 없이 다집단 CFA의 형태 동일성 모형에서 추정된 각 집단의 측정모수의 차이가 최소화되도록 변환하고 정렬하는 절차이다. 정렬 후의 모형은 정렬 전의 모형에서 요인분산과 요인평균의 척도를 조정한 모형으로 두 모형은 수학적으로 동일하다. 같은 모형에서 요인의 척도를 조정하면 각 지표변수의 측정모수인 요인부하량과 절편도 그에 따라 달라진다. 정렬법에서 수행되는 변환은 집단 간 측정모수의 차이의 총량이 최소가 되도록 각 집단의 요인 척도를 조정한 재모수화를 의미한다. 이는 탐색적 요인분석에서 요인부하량의 형태가 단순 구조에 가깝도록 요인을 회전하는 것과 유사

하다(Asparouhov & Muthén, 2014).

정렬법의 과정을 좀 더 구체적으로 설명하면 다음과 같다. 먼저 각 집단의 요인평균( $\alpha_g$ )과 요인분산( $\psi_g$ )을 각각 0과 1로 고정하고 모든 지표변수의 요인부하량과 절편을 자유롭게 추정하는 다집단 CFA의 기저 모형(M0)을 추정한다. M0에서 추정된 집단  $g$ 의 변수  $p$ 에 대한 절편과 요인부하량을 각각  $\nu_{pg,0}$ ,  $\lambda_{pg,0}$ 이라고 할 때, 요인평균과 요인분산을 각각  $\alpha_g$ 와  $\psi_g$ 로 변환하면 이 절편과 요인부하량은 각각 아래와 같이 변환된다.

$$\lambda_{pg,1} = \frac{\lambda_{pg,0}}{\sqrt{\psi_g}} \quad (2)$$

$$\nu_{pg,1} = \nu_{pg,0} - \alpha_g \frac{\lambda_{pg,0}}{\sqrt{\psi_g}} \quad (3)$$

각 지표변수에 대해 모든 집단 쌍의 요인부하량의 비동일성과 절편의 비동일성을 더하는 과정을 모든 지표변수에 대해 수행하면 측정모수의 불일치의 총량을 수량화할 수 있다. 이를 식으로 나타내면 아래 식 (4)와 같다.

$$F = \sum_p \sum_{g_1 < g_2} w_{g_1, g_2} f(\lambda_{pg_1,1} - \lambda_{pg_2,1}) + \sum_p \sum_{g_1 < g_2} w_{g_1, g_2} f(\nu_{pg_1,1} - \nu_{pg_2,1}) \quad (4)$$

총 손실함수  $F$ 가 최소화될 때까지  $\alpha_g$ 와  $\psi_g$ 를 반복적으로 변화시키면서 측정모수를 정렬한다. 이때 모든 집단의 측정모수에 대해 성분손실함수(component loss function, CLF)  $f$ 가 먼저 수행되며, 그 후  $F$ 를 통해 집단 간 측정

모수의 차이가 줄어든다(Asparouhov & Muthén, 2014).  $w$ 는  $f$ 의 가중치로 집단의 표본 크기가 클수록  $F$ 에 더 많이 기여한다.

식 (4)의 최소화를 통해 첫 번째 집단을 제외한 나머지 집단에서  $\alpha_g$ 와  $\psi_g$ 가 추정되면 집단 간에  $\alpha_g$ ,  $\psi_g$ 의 비교가 가능해진다. 모형의 식별을 위해 모든 집단의 요인분산의 곱은 1이 되도록 제약한다. 이때 참조집단의 요인평균은 0이 되도록 제약하거나(FIXED) 제약하지 않을(FREE) 수 있다.

각 집단의 측정모수들이 정렬된 후에는 집단에 따라 어떤 모수가 근사적으로 동일한지 혹은 그렇지 않은지 판단하기 위해 각 집단의 모수의 차이를 통계적으로 평가하는 분석이 수행된다. 먼저 근사적으로 동일한 집단을 묶기 위해 초기 세트(집단 쌍)를 선택한 후 이 세트의 평균과 각 집단의 모수의 차이에 대해 통계적 유의성을 판단하여 동일성 집단의 세트를 갱신·확장한다. 이 과정은 더 이상 근사적으로 동일한 집단이 늘어나지 않을 때까지 반복된다(Asparouhov & Muthén, 2014).

정렬법의 분석 결과를 통해 측정동일성을 다각적으로 평가할 수 있다. 첫째, Muthén과 Asparouhov(2014)는 정렬법 분석 결과에서 근사적으로 동일하지 않은 집단의 비율이 25% 이하인 경우에는 정렬 과정이 적절히 이루어진 것이며 결과를 신뢰하고 사용할 수 있다고 하였다. 이를 통해 분석의 전반적인 타당성을 검토할 수 있다. 둘째, 정렬 합치 함수(fit function)를 이용하여 측정모수 및 지표변수별로 부적합(misfit)의 크기를 구할 수 있다. 정렬 합치 함수는 정렬 후 남아 있는 부적합의 정도를 뜻하여 0에서 멀수록 비동일성이 큰 것을 의미한다. 마지막으로  $R^2$ 은 M0에서 모수

의 변산성이 정렬 과정에서 변화하는 요인평균 및 요인분산의 변산성에 의해 흡수되는 정도를 의미한다(Asparouhov & Muthén, 2014; Muthén & Asparouhov, 2018).  $R^2$ 이 클수록 정렬이 효율적이었다고 해석할 수 있다.

### 범주형 자료의 측정동일성 분석

#### 범주형 변수를 위한 확인적 요인분석

표준적인 CFA에서 지표변수가 연속형 변수라면 잠재변수와 지표변수의 관계가 선형적으로 가정되며 지표변수 사이의 선형적 관련성을 나타내는 공분산행렬이 분석에 사용된다. 최대우도법은 표본공분산행렬  $S$ 와 모형에 의해 예측되는 예측공분산행렬  $\hat{\Sigma}$  간의 차이를 나타내는 합치 함수(fit function)를 가장 작게 만드는 모수들의 조합을 추정한다(Kline, 2016).

지표변수가 범주형일 때 CFA에서의 측정동일성은 각 지표변수의 특정 응답 범주를 선택할 확률이 집단마다 다르지 않고 서로 동일함을 의미한다(Millsap, 2011). 따라서 범주형 변수의 측정모형에서 지표-요인 관계는 선형적으로 표현되지 않고 비선형적으로 표현된다(Edwards et al., 2012). 또한 지표변수가 범주형일 경우 이 변수는 정규분포할 수 없다.

다변량 정규성을 가정할 수 없는 범주형 지표변수의 경우 표본 공분산행렬은 이들 사이의 관계를 적절히 반영하지 못한다. 이러한 문제를 해결하기 위해 범주형 지표변수  $y$ 의 반응 범주의 기저에 연속형 변수  $y^*$ 가 존재한다고 가정하고 CFA 모형을 추정할 수 있다.  $y^*$ 는 잠재반응변수(latent response variable,

Muthén, 1984)라고 부르며 해당 지표변수의 특정 범주에 응답하도록 하는 잠재적 특성의 양(amount)으로 정의된다. 이렇게 하면 범주형 지표변수  $y$  대신 연속형 잠재반응변수  $y^*$  간의 공분산행렬을 추정하고 이에 대한 요인 분석을 수행할 수 있다. 이때 잠재반응변수 사이의 상관은 지표변수  $y$ 가 이분형이라면 사분상관으로, 다분형이라면 다분상관으로 추정한다. 이러한 과정을 통해 요인과 범주형 지표변수  $y$ 간의 비선형적 관계를 연속형 잠재반응변수  $y^*$ 와 요인의 선형적 관계로 대체한다.

범주형 지표변수  $y$ 와 잠재반응변수  $y^*$  간의 관계는 응답 반응의 비율을 고려하여 분계점  $\tau$ 에 의해 매개된다. 예를 들어, 세 개의 응답 범주 1, 2, 3을 가진 문항  $y$ 가 있을 때  $y^*$ 가  $\tau_1$ 을 넘지 않는다면  $y$ 의 값은 1이 되며,  $y^*$ 가  $\tau_1$ 보다 크지만  $\tau_2$ 과 같거나 작을 경우에는 2가 되고,  $\tau_2$ 보다 큰 경우는 3이 된다. 이처럼 분계점은 응답 범주가 구분되는 잠재반응변수 상의 경계를 의미하며 어떤 문항의 응답 범주의 수가  $m$ 개일 경우  $m-1$ 개가 추정된다.

이처럼 연속형 변수의 공분산행렬이나 범주형 변수의 잠재반응변수의 상관행렬을 이용하여 모형을 추정하는 방법을 제한정보 추정법이라고 부른다. 범주형 CFA에서는 모형의 식별을 위해 잠재반응변수에도 척도를 설정해야 한다. 잠재반응변수의 척도설정 방법은 델타 척도화(Delta scaling)와 세타 척도화(Theta scaling) 두 가지가 있다. 델타 척도화는 모든 잠재반응변수의 분산을 고정(예, 1)하고 세타 척도화는 각 잠재반응변수의 오차분산을 고정(예, 1)한다(Kline, 2016).

잠재반응변수를 사용하는 범주형 CFA의 추정치는 자료의 분포 특성을 반영하는 가중행렬을 사용하는 가중최소제곱(weighted least squares, WLS)을 사용할 수 있다. 그러나 WLS의 안정적인 추정을 위해서는 가중행렬의 계산을 위해 매우 큰 표본 크기가 필요하며 (Muthén & Kaplan, 1992), 카이제곱 통계량이 과대추정되거나 표준오차가 과소추정되는 문제가 있다(Dolan, 1994). 이에 대한 대안으로 평균 및 분산 조정 가중 최소제곱(mean-and variance-adjusted least squares, WLSMV) 추정이 널리 사용되는데 이 추정법은 WLS에 사용하는 가중행렬의 대각 요소만 사용하는 Diagonal WLS의 한 형태로 표본 크기가 충분히 크다면 다양한 종류의 범주형 변수에서 비교적 정확한 추정 결과를 제공하는 것으로 보고되었다 (Muthén & Kaplan, 1985, 1992).

한편 공분산행렬이나 상관행렬과 같은 요약 자료 대신 전체 자료를 이용하여 CFA를 추정하는 경우 이를 완전정보 추정이라고 한다. 범주형 자료 CFA의 완전정보 추정에서는 제한정보 추정과는 달리 잠재변수와 지표변수 사이의 비선형적 관계를 직접적으로 설정한다. 이때에는 요약자료를 사용하지 않기 때문에 표본 공분산행렬과 모형에서 예측된 공분산행렬 간의 차이를 비교하는 절차가 없기 때문에 모형 적합도 통계량이나 근사적합도지수를 산출할 수 없다. 측정동일성 분석에서는 모형 적합도에 대한 평가를 할 수 있다는 점에서 일반적으로 완전정보 추정보다 제한정보 추정이 선호되지만 잠재변수의 조절효과나 요인 부하량에 대한 상호작용 등을 포함하는 특수한 모형의 경우에는 제한정보 추정이 불가하고 완전정보 추정만을 사용해야 하는 경우도 있다.

## 범주형 자료의 측정동일성 분석과 정렬법

범주형 자료의 CFA를 다집단 분석으로 확장하여 측정동일성을 분석할 때는 연속형 변수에서의 절차와 유사한 방식으로 수행한다. 형태 동일성 모형부터 강한 동일성 모형까지 요인구조 및 측정모수에 동일성 제약을 부여한 후 분석하여 각 모형의 적합도 통계량 및 근사적합도지수가 수용 가능한지 확인 후 모형 간의 비교를 통해 요인모형의 측정동일성 수준을 판단한다. 다만 범주형 자료에서의 강한 동일성 모형은 절편이 아닌 분계점의 동일성을 의미한다는 점에서 차이가 있다.

범주형 자료의 다집단 CFA를 사용한 측정동일성의 평가도 연속형 자료와 마찬가지로 카이제곱 통계량과 근사적합도지수를 이용할 수 있는데 이를 위해서는 제한정보 추정이 필요하다. 지표변수의 정규분포 가정이 타당하지 않은 경우 CFA에 대한 카이제곱 검증에는 Satorra와 Bentler(2001)의 보정 계수(correction factor)를 사용하여 보정된  $\chi^2$ 과  $\Delta\chi^2$  검증을 적용할 수 있다. 그러나 WLSMV 추정에서는 공분산행렬에 대한 보정까지 필요하기 때문에  $\Delta\chi^2$  검증에 대한 계산이 매우 복잡하여 Mplus와 같은 프로그램에서는 이를 위해 별도의 옵션(DIFFTEST)을 제공한다.

그러나 WLSMV 추정에서의  $\Delta\chi^2$  검증도 대표본에서 영가설 기각에 민감하기 때문에 (French & Finch, 2006; Sass et al., 2014) 연속형 자료의 다집단 CFA의 평가 방법과 마찬가지로 근사적합도지수의 차이를 이용하여 측정동일성의 수준을 평가하는 방법이 제안되었다. Rutkowski와 Svetina(2017)는 범주형 CFA의 측정동일성 분석에서 근사적합도지수의 절단점을



제안하였다. 구체적으로는  $\Delta\chi^2$ 과 함께 고려하되 형태 동일성과 약한 동일성 모형의 RMSEA 차이값( $\Delta RMSEA$ )이 .05보다 작을 경우에는 약한 동일성 모형 채택, 그리고 약한 동일성과 강한 동일성 모형의  $\Delta RMSEA$ 가 .01보다 작을 때에는 강한 동일성 모형을 채택할 수 있다고 하였다.  $\Delta CFI$ 의 절단값 또한 함께 제시하였으나, 선행 연구에서 CFI는 여러 가지 조건에서 예상 밖의 범위가 나타나거나 모형을 오탐지한 경우가 많아 권장되지 않는다.

범주형 자료의 정렬법 분석에는 제한정보 추정법 대신 완전정보 추정법이 사용된다. 주로 MLR(robust maximum likelihood) 또는 MLF(maximum likelihood with standard error approximation using the first-order derivative) 추정이 사용된다. 범주형 자료를 위한 정렬법에서는 식 (4)의 총 손실함수  $F$ 에 절편( $\nu_p$ ) 대신 변수  $p$ 의  $c$ 번째 분계점( $\tau_{pc}$ )이 들어가는 아래의 식 (5)를 최소화하도록 요인부하량과 분계점을 정렬한다.

$$F = \sum_p \sum_{g_1 < g_2} w_{g_1, g_2} f(\lambda_{pg_1, 1} - \lambda_{pg_2, 1}) + \sum_p \sum_c \sum_{g_1 < g_2} w_{g_1, g_2} f(\tau_{pcg_1, 1} - \tau_{pcg_2, 1}) \quad (5)$$

범주형 자료에서 완전정보를 사용하는 추정 은 곧 문항 단위의 관찰값을 지표변수로 하는 비선형 요인분석 모형을 뜻하므로 IRT 모형과 같은 의미를 가진다(Asparouhov & Muthén, 2022). IRT 모형을 다집단 분석으로 확장하면 집단에 따라 문항의 난이도 및 변별도 모수에 차이가 존재하는지 확인할 수 있다. 이는 구조방정식모형의 측정동일성 분석에 상응하는

개념이며, 만약 집단 사이에 난이도나 변별도 모수가 다르다면 차별문항기능(Differential Item Functioning, DIF)이 있다고 해석한다(Woods, 2009; Woods & Grimm, 2011).

## 범주형 자료에 대한 근사 측정동일성 분석의 예시

### 분석자료

근사 측정동일성 분석의 예시를 위하여 고령화연구패널조사(KLoSA)의 2018년 수집 자료 중 K-MMSE 검사의 문항 점수를 사용하였다. 전체 표본 6,940명 중 393명의 K-MMSE 문항 응답에서 결측이 확인되었다. 이들 중 모든 검사 문항에 응답하지 않은 386명의 자료를 제외한 6,554명의 자료가 측정동일성 분석에 사용되었다. 이 중 여성은 3,792명(57.9%)이었고 연령의 범위는 55세에서 102세, 평균은 69.4세(SD=9.9)였다.

표 1. K-MMSE의 측정 영역 및 점수 범위

영역	점수 범위	
	수정 전	수정 후*
기억등록	0-3	0-2
시간지남력	0-5	0-2
장소지남력	0-5	0-2
주의집중 및 계산	0-5	0-4
기억회상	0-3	0-3
언어	0-8	0-4
시각적 구성	0-1	0-1

\*분석을 위해 응답자가 매우 적은 일부 점수 범주를 통합

K-MMSE(Kang et al., 1997)는 MMSE (Mini-Mental State Examination, Folstein et al., 1975)의 한국어판으로 초기 단계의 치매 환자들을 선별하기 위해 지남력, 기억력, 주의집중력, 언어능력을 포함한 인지기능의 여러 영역을 짧은 시간에 측정하는 검사이다. 각 영역 점수 범위는 0-1부터 0-8까지 다양하며 영역 점수를 합산한 총점으로 검사점수를 산출한다. K-MMSE의 측정 영역과 영역별 점수 범위는 표 1과 같다.

노화는 일반적으로 인지기능을 저하시킨다고 알려져 있으며 장년기 이후에는 연령이 높을수록 MMSE의 검사점수가 낮게 나타난다(Crum et al., 1993; Kim et al., 1997). 그러나 연령 증가에 따른 MMSE 점수의 감소를 노화에 따른 인지기능의 저하를 보여주는 명백한 경험적 근거로 해석하기 위해서는 MMSE 점수가 서로 다른 연령에 대해 측정동일성을 가진다는 것이 전제되어야 한다.

K-MMSE의 검사점수가 다양한 연령에서 측정동일성을 가지는지를 확인하기 위해서 20개 이상의 가급적 많은 수의 연령 집단을 구성하여 분석을 수행하였다. 먼저 출생 연도별로 하나의 연령 집단을 구성하되 다집단 CFA가 적절히 수행될 수 있도록 집단 표본의 크기가 150보다 작은 경우에는 둘 이상의 연령 집단을 하나로 묶었다(Kline, 2005; Yoon & Lai, 2018). 하지만 이렇게 구성한 일부 연령 집단에서 일부 영역의 점수 범주에서는 사례 수가 너무 작아 분석이 수행될 수 없었다. 예를 들어 330명으로 구성된 55세 연령 집단에서 시간지남력 점수의 분포는 0점-0명, 1점-0명, 2점-3명, 3점-9명, 4점-6명, 5점-293명이었고 장소지남력 점수의 분포는 0점-0명, 1점-1명, 2점-1명, 3점-1명, 4점-16명, 5점-292명이었다. 이런

경우 해당 연령 집단에서 모든 점수 범주 사이의 분계점을 추정할 수 없다. 이러한 경우에는 분석이 수행될 수 있도록 모든 연령 집단에서 각 점수 범주의 사례 수가 최소 3이 되도록 일부 영역 점수의 범주를 통합하였다. 예를 들어 시간지남력과 장소지남력은 원래 0점에서 5점까지 여섯 개의 범주로 구분되었던 것을 0점에서 3점을 하나로 묶어 0점으로, 4점은 1점으로, 5점은 2점으로 변환하여 세 개의 범주를 가지도록 수정하였다. 마찬가지로 기억등록과 주의집중 및 계산은 0점과 1점, 언어는 0점에서 4점까지를 하나의 범주로 묶었다. 이렇게 수정된 각 영역의 점수 범위는 표 1에서 확인할 수 있다.<sup>3)</sup>

이렇게 각 영역의 점수 범주를 수정한 후에도 특정 연령 집단에서 두 개의 영역 점수가 교차하는 범주의 빈도가 0으로 나타나는 경우(예, 시간지남력과 장소지남력이 모두 0점인 경우)가 있었다. 즉, 빈도가 0인 교차 범주의 조합에 따라 다분상관이나 분계점이 추정되지 않는 연령 집단이 있었다. 이러한 경우에는 빈도가 0인 교차 범주의 발생을 줄이기 위해 두 개의 연령 집단을 하나로 통합하였다(예, 55-56세). 자료 변환의 기준을 정리하면 (1) 각 집단의 최소 표본은 150명 이상이 되게 하고, (2) 이렇게 구성한 연령 집단에서 영역별 각 응답범주의 최소 빈도는 3 이상이 되도록 빈도가 작은 범주는 통합하였다. 그리고 (3) 각 영역 간 교차범주에서 빈도가 0인 경우가 있으면 추가적으로 인접 연령과 통합하였다. 이러한 과정을 거쳐 표 1, 표 2와 같이 분석자

3) 범주형 자료에 대한 CFA에서는 응답 범주의 교차빈도가 0 또는 매우 작은 수인 경우 일부 범주를 통합하는 방법이 흔히 사용된다(Rutkowski et al., 2019; Tennant et al., 2004).

표 2. 연령 단위 집단 구성 및 표본 수

연령 집단	N	연령 집단	N	연령 집단	N	연령 집단	N
55-56세	633	68세	230	75세	193	82세	154
57-58세	518	69세	209	76세	219	83-84세	251
59-60세	502	70세	214	77세	220	85-86세	161
61-62세	483	71세	205	78세	182	87-88세	156
63-64세	451	72세	211	79세	191	89세 이상	228
65-66세	431	73세	189	80세	179		
67세	170	74세	178	81세	182		

료를 구성하였다.

#### 분석절차

K-MMSE의 영역 점수를 나타내는 7개의 지표변수에 대한 단일 요인 모형에 대해 이 요인이 서로 다른 연령 집단에서 측정동일성을 가지는지 확인하기 위해 다집단 CFA와 정렬법을 각각 실시하였다. 요인모형의 지표변수로는 7개의 영역별 점수를 사용하였다. 응답범주 통합 후 시각적 구성은 두 개, 기억등록, 시간지남력, 장소지남력은 세 개, 기억등록은 네 개, 주의집중 및 계산과 언어는 다섯 개의 응답범주를 가지기 때문에 순서범주형 지표변수로 다루었다.

다집단 CFA 모형은 제한정보법인 WLSMV로 추정하여 모형 적합도의 비교를 통해 측정동일성 수준을 평가하였다. 요인의 척도는 참조집단(첫 번째 연령 집단)의 요인분산을 1로 고정하여 설정하였으며, 나머지 연령 집단의 요인분산을 추정하기 위해 기억등록(참조지표)의 요인부하량이 모든 집단에서 같도록 제약하였다. 각 지표변수의 잠재반응변수의 척도는 참조집단에서 잠재반응변수의 잔차분산을

1로 고정하여 설정하였다(세타 척도화). 나머지 집단에서는 잔차분산이 자유롭게 추정될 수 있도록 각 지표변수의 첫 번째 분계점이 모든 집단에서 동일하도록 제약하였다. 그리고 참조집단의 요인평균은 0으로 고정하고 나머지 집단의 요인평균이 추정될 수 있도록 기억등록의 두 번째 분계점이 모든 집단에서 같게 제약하였다. 이러한 제약은 순서범주형 변수에 대한 다집단 CFA 모형의 식별을 위한 최소한의 제약으로 이를 통해 형태 동일성 모형을 설정하였다. 형태 동일성 제약에 각 요인부하량이 전체 연령 집단에서 동일성을 가지는 제약을 추가하여 약한 동일성 모형을 설정하였다. 마지막으로 약한 동일성 제약에 각 분계점이 전체 연령 집단에서 동일성을 가지는 제약을 추가하여 강한 동일성 모형을 설정하였다.

순서범주형 변수의 요인모형에 대한 정렬법 분석은 제한정보법을 사용할 수 없어 완전정보법인 MLR로 추정하였다(Muthén & Muthén, 1998-2020). 먼저 모든 집단의 요인 평균을 고정하지 않고 측정모수를 정렬하는 자유(FREE)정렬을 실시하였다. 이 결과에서 요인평균이 가장 0에 가까웠던 18번째 집단을 참조집단으

표 3. 다집단 CFA를 이용한 측정동일성 평가(WLSMV 추정)

모형	$\chi^2$	df	CFI	SRMR	RMSEA	$\Delta\chi^2*$	$\Delta df$	$\Delta RMSEA$
형태동일성	1256.4***	364	.970	.065	.099 [.093, .105]			
약한동일성	1487.9***	514	.968	.070	.087 [.082, .092]	344.3***	150	-.012
강한동일성	1847.4***	764	.964	.075	.075 [.071, .079]	418.6***	250	-.012

\*WLSMV 추정 사용 시 Mplus에서 제공하는 DIFFTEST를 이용하여 얻은 값

로 재설정하여 이 집단의 요인평균을 0, 요인 분산을 1로 두고 측정모수를 정렬하는 고정(FIXED) 정렬을 다시 실시하였다. 모든 분석은 Mplus 8.4로 수행되었다.

#### 다집단 CFA 분석결과

단일 요인 모형에 대한 26개 연령 집단의 형태 동일성 모형은  $\chi^2$  검정에 의한 완전 적합 가설은 기각되었으나 근사적합도지수 CFI (.970)와 SRMR(.065)은 적합한 수치를 보였다. 다집단 분석에서 집단이 20개 이상 많을 때 RMSEA(.099)는 큰 값을 나타내는 경향이 있다는 점을 고려하여 완화된 기준(예, .10 이하)을 적용하면 형태 동일성을 충족한 것으로 판단할 수 있다(Rutkowski & Svetina, 2014).

형태 동일성 모형과 약한 동일성 모형의 비교( $\Delta\chi^2(150)=344.3***, p<.001$ )와 약한 동일성 모형과 강한 동일성 모형의 비교( $\Delta\chi^2(250)=418.6***, p<.001$ )는 요인부하량과 절편의 완전한 측정동일성을 지지하지 않았다. 근사적합도지수의 비교에 의한 동일성 평가는 Rutkowski와 Svetina(2017)가 제안한 범주형 CFA의 기준을 참고하였다.  $\Delta CFI$ 의 경우 두 비

교 모두에서 .01보다 작은 감소를 보였다.  $\Delta RMSEA$ 의 경우에는 오히려 제약이 큰 모형이 제약이 작은 모형에 비해 낮은 수치를 보였다. 따라서 근사적합도지수의 변화량을 기준으로 할 때는 다집단 CFA에서 강한 동일성이 성립한 결과로 해석할 수 있다.

#### 정렬법 분석결과

Muthén과 Asparouhov(2014)의 경험적 규칙에 근거하면 표 4에서 모든 요인부하량은 비동일하다고 식별된 모수가 없었으며 분계점은 주의집중 및 계산, 언어, 시각적 구성 영역에서 비동일성이 발생하였다. 이들의 평균 비동일성 비율은 7.75%로 기준인 25%보다 작으므로 정렬 결과는 신뢰할 수 있었다.

모수별 비동일성의 비율은 측정모수 또는 지표변수 단위의 정렬 합치 함수(fit function contribution)와 함께 해석할 수 있다. 지표변수 단위의 총 정렬 합치 함수는 정렬 후 해당 변수가 가진 전반적인 비동일성의 정도를 의미한다. 표 4에서 주의집중 및 계산, 언어 영역의 총 정렬 합치 함수는 각각 -834.44, -700.759로 다른 영역에 비해 상대적으로 컸다.

표 4. 정렬법을 이용한 측정동일성 평가(MLR 추정)

변수	비동일 집단의 비율(%)	정렬 합치 함수	$R^2$	정렬 전 분산	정렬 후 분산
총 정렬 합치 함수: -494.959					
시간					
TIME	0	-195.892	.174	.797	.381
지남력					
TIMES\$1	0	-166.814	.551	3.515	.295
TIMES\$2	0	-132.253	.741	2.793	.061
총 정렬 합치 함수: -474.558					
공간					
PLACE	0	-156.77	.234	.123	.056
지남력					
PLACE\$1	0	-168.572	.94	1.634	.125
PLACE\$2	0	-149.216	.949	1.156	.069
총 정렬 합치 함수: -428.466					
기억등록					
REGIS	0	-152.905	0	.101	.077
REGIS\$1	0	-132.876	.951	2.095	.055
REGIS\$2	0	-142.685	.946	1.686	.076
총 정렬 합치 함수: -834.44					
주의집중 및 계산					
ATTEN	0	-150.672	.408	.153	.048
ATTEN\$1	15(15,18,21,23)	-201.202	.899	2.060	.289
ATTEN\$2	8(15,16)	-185.233	.927	1.743	.199
ATTEN\$3	0	-147.585	.975	1.299	.059
ATTEN\$4	0	-149.748	.974	1.329	.065
총 정렬 합치 함수: -590.192					
기억회상					
RECAL	0	-149.551	.162	.115	.065
RECAL\$1	0	-168.656	.86	2.930	.213
RECAL\$2	0	-125.736	.971	2.049	.031
RECAL\$3	0	-146.249	.958	2.056	.087
총 정렬 합치 함수: -700.759					
언어					
LAN	0	-155.194	.194	.767	.353
LAN\$1	0	-159.427	.814	7.239	.769
LAN\$2	0	-131.275	.877	5.142	.197
LAN\$3	0	-118.046	.936	4.909	.081
LAN\$4	4(22)	-136.817	.902	4.140	.265
총 정렬 합치 함수: -324.146					
시각적 구성					
DRAW	0	-167.403	.34	.415	.152
DRAW\$1	4(21)	-156.743	.909	3.578	.169

주. \$ 기호가 없는 모수는 요인부하량을 의미하며 \$ 뒤의 숫자는 각 지표변수의 분계점을 나타냄.

그리고 측정모수 단위의 정렬 합치 함수는 주의집중 및 계산 영역의 첫 번째, 두 번째 분계점(ATTEN\$1, ATTEN\$2; -201.202, -185.233), 언어 영역의 네 번째 분계점(LAN\$4; -136.817), 시각적 구성 영역의 분계점(DRAW\$1; -156.743)과 같이 비동일한 집단이 식별된 모수들의 값이 상대적으로 컸다.

한편 요인부하량의  $R^2$ 은 0에서 .4 사이의 범위를 보였으며 분계점에서는 대부분 1에 가까웠다. 또한 요인부하량은 전반적으로 정렬 전후의 분산의 차이가 크지 않았으나 분계점은 상대적으로 전후 차이가 컸다. 따라서 요인부하량은 정렬 전후 변산성의 차이가 작았으나, 분계점은 연령 집단 사이의 변산성이 정렬 과정을 통해 많이 감소하였다.

한편 비동일한 집단의 비율, 정렬 합치 함수,  $R^2$ 의 관계가 불일치하게 나타나는 모수들도 있었다. 언어 영역의 첫 번째 분계점(LAN\$1)은 비동일한 집단이 없었으나 정렬 합치 함수의 값은 비동일한 집단이 존재하는 언어 영역의 네 번째 분계점(LAN\$4)보다 크게 나타났다. 이와 같은 현상이 발생하는 원인은 다음과 같이 유추할 수 있다. 먼저 정렬 후 후속 분석 중 동일한 집단을 분류하는 과정에

서 통계적 유의성이 탐지될 수 있는 충분한 표본 크기에 도달하지 못했을 가능성이 있다 (Byrne & van de Vijver, 2017). 다른 이유로는 소수의 비동일한 집단이 식별되는 추정 원리상 집단에 따라 전반적으로 비슷한 수준의 비동일성이 존재할 경우 상대적으로 큰 수준의 비동일성을 가진 집단은 잘 탐지되지 않았을 수 있다. 위 내용을 종합하면 요인부하량은 본래 집단에 따라 비동일성이 크지 않았을 것으로 예상되며, 분계점은 집단에 따라 차이가 컸으나 정렬 과정을 통해 비동일성이 상당 수준 개선되었다고 할 수 있다. 그림 1은 다집단 CFA의 강한 동일성 모형과 정렬법에서 추정된 각 연령 집단의 요인평균을 산점도로 제시한 것으로, 두 분석에서 추정된 요인평균 간 상관은 .99로 거의 흡사하였다.

## 논 의

본 논문에서는 다수 집단을 비교할 때 적용할 수 있는 완전한 동일성의 완화된 형태인 근사 측정동일성의 개념을 소개하고, 다수의 연령 집단에 대해 정렬법을 적용하여 범주형 자료의 분석절차를 예시하였다. 완전한 측정동일성을 평가하는 범주형 다집단 CFA에서는 측정동일성의 수준에 따라 동일성 제약을 상이하게 부여한 모형 간의  $\Delta\chi^2$  검증과 근사 적합도지수를 이용한 비교를 통해 측정동일성의 수준을 결정한다. 반면 정렬법에서는 측정모수의 정렬을 통해 집단 간 비동일성이 개선되었으며 분석결과를 통해 비동일성의 개선 정도와 정렬 과정의 효율성을 평가할 수 있다. 따라서 정렬 후 얻어진 요인모수를 이용해 집단 간 요인평균, 요인분산의 비교가 가능하며

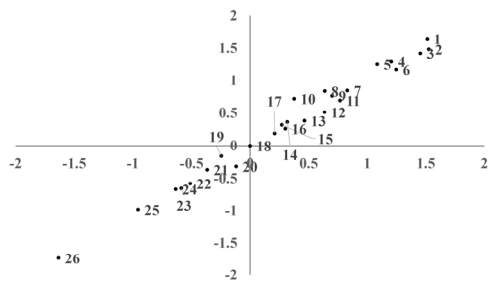


그림 1. 다집단 CFA의 강한 동일성 모형과 정렬법으로 추정된 집단별 요인평균의 산점도. 숫자는 표 2의 연령 집단의 순서를 나타냄

동시에 일부 집단에서 근사적으로 동일하지 않은 모수를 확인할 수 있다.

근사 측정동일성에서는 측정동일성 성립 여부에 대한 결론을 내리기 위한 통계적 검증의 측면보다는 집단에 따라 유사한 모수 세트를 얻기 위해 모수를 제약하는 모형화(modeling)의 측면을 강조한다. 분석과정에서도 다집단 CFA는 측정동일성의 수준에 따른 모형을 분석 후 비교하여 측정동일성의 수준을 평가하지만 정렬법에서는 각 집단의 모수들이 근사적으로 동일해지도록 형태 동일성 모형에서 요인부하량, 절편(또는 분계점)의 정렬이 동시에 수행된다. 따라서 근사적으로 동일하다는 가정 하에 잠재변수 간의 구조적 관계를 분석하거나 잠재변수의 요인평균 및 분산을 비교하는 목적으로 활용될 수 있다. 한편 범주형 자료의 정렬법은 IRT 모형에 상응하기 때문에 문항 단위에서 동일성을 위반하는 모수를 찾는 탐색적인 도구가 될 수 있다(Luong & Flake, 2022). 그러므로 완전한 동일성이 확립되기 어려운 자료에서 비동일성에 대한 파악과 근사 동일성을 가진 모형을 동시에 추정할 수 있으며, 자료에서 비동일성을 유발하는 원인에 대한 예상이나 정보가 있는 경우 특히 도움이 될 것이다. 또한 정렬법에서의 다수 집단은 주로 국가 비교 연구에 많이 적용되어 왔으나 다수 기업을 비교하거나 본 논문과 같이 연령 집단을 비교하는 맥락에도 사용할 수 있다. 특히 연령 집단을 비교할 때 종단자료를 적절히 활용한다면 연령에 따른 발달적 변화를 정확하게 추정하는 데에도 유용할 것이다.

대규모 자료에 적합한 측정동일성 분석에 대해서는 오랜 시간동안 논의되어 왔다. 최근 이러한 문제를 다룰 수 있는 베이지안 구조방정식 모형, 베이지안 다층 확인적 요인분석

모형 등 새로운 방법이 제안되었다(Asparouhov & Muthén, 2014; Muthén & Asparouhov, 2013, 2018). 그러나 이를 적용한 연구들은 대부분 연속형 자료를 다루고 있다(Byrne & van de Vijver, 2017; Ciecuch et al., 2018; Lomazzi, 2018; Munck et al., 2018). 범주형 다집단 CFA를 응용한 근사 측정동일성 분석의 소개는 범주형 지표변수의 특성을 적절히 반영하면서도 비교 집단이 많은 대규모 자료에서 근사 측정동일성 분석의 유용성을 확대할 수 있을 것이다.

본 연구에서는 분석절차를 예시하는 과정에서 패널 자료를 일부 재구성하여 사용하였다. 따라서 분석결과는 응답범주를 재구성했을 때의 결과이며 원 범주에 대한 분석결과와는 다를 수 있다. 그러나 예시분석이 범주형 자료에서 근사 측정동일성 분석의 표준적인 사례임을 고려했을 때 결과를 이해하고 해석하는데에는 문제가 없다. 따라서 본 연구를 바탕으로 추후 연구자들이 정렬법을 이용한 근사 측정동일성 분석 시 참고자료로 활용할 수 있을 것이다.

## 참고문헌

- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 495-508.  
<http://doi.org/10.1080/10705511.2014.919210>
- Asparouhov, T., & Muthén, B. (2022). Multiple group alignment for exploratory and structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 30(2), 1-23.  
<http://doi.org/10.1080/10705511.2022.2127100>

- Brown, T. A. (2015). *Confirmatory factor analysis for applied research (2nd ed.)*. New York, NY: Guilford Publications.  
<http://www.guilford.com/books/Confirmatory-Factor-Analysis-for-Applied-Research/Timothy-Brown/9781462515363>
- Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for the equivalence of factor covariance and mean structures. *Psychological Bulletin*, 105(3), 456-466.  
<http://doi.org/10.1037/0033-2909.105.3.456>
- Byrne, B. M., & van de Vijver, F. V. D. (2017). The maximum likelihood alignment approach to testing for approximate measurement invariance: A paradigmatic cross-cultural application. *Psicothema*, 29(4), 539-551.  
<http://doi.org/10.7334/psicothema2017.178>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464-504.  
<http://doi.org/10.1080/10705510701301834>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233-255.  
[http://doi.org/10.1207/S15328007SEM0902\\_5](http://doi.org/10.1207/S15328007SEM0902_5)
- Cieciuch, J., Davidov, E., Algesheimer, R., & Schmidt, P. (2018). Testing for approximate measurement invariance of human values in the European Social Survey. *Sociological Methods & Research*, 47(4), 665-686.  
<http://doi.org/10.1177/0049124117701478>
- Crum, R. M., Anthony, J. C., Bassett, S. S., & Folstein, M. F. (1993). Population-based norms for the Mini-Mental State Examination by age and educational level. *Journal of the American Medical Association*, 269(18), 2386-2391.  
<http://doi.org/10.1001/jama.1993.03500180078038>
- Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology*, 47(2), 309-326.  
<http://doi.org/10.1111/j.2044-8317.1994.tb01039.x>
- Edwards, M. C., Wirth, R. J., Houts, C. R., & Xi, N. (2012). Categorical data in the structural equation modeling framework. In R. Hoyle (Eds.), *Handbook of structural equation modeling* (pp. 195-208). New York, NY: The Guilford Press.  
<http://www.guilford.com/books/Handbook-of-Structural-Equation-Modeling/Rick-Hoyle/9781462544646>
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189-198.  
[http://doi.org/10.1016/0022-3956\(75\)90026-6](http://doi.org/10.1016/0022-3956(75)90026-6)
- French, B. F., & Finch, W. H. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural Equation Modeling*, 13(3), 378-402.  
[http://doi.org/10.1207/s15328007sem1303\\_3](http://doi.org/10.1207/s15328007sem1303_3)



- Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor framework. *Medical Care*, 44(11 Suppl 3), S78-S94.  
<http://doi.org/10.1097/01.mlr.0000245454.12228.8f>
- Herzog, W., Boomsma, A., & Reinecke, S. (2007). The model-size effect on traditional and modified tests of covariance structures. *Structural Equation Modeling*, 14(3), 361-390.  
<http://doi.org/10.1080/10705510701301602>
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36(4), 409-426.  
<http://doi.org/10.1007/BF02291366>
- Kang, Y., Na, D. L., & Hahn, S. (1997). A validity study on the Korean Mini-Mental State Examination (K-MMSE) in dementia patients. *Journal of the Korean Neurological Association*, 15(2), 300-308.  
<http://www.jkna.org/journal/view.php?number=4088>
- Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, 10(3), 333-351.  
[http://doi.org/10.1207/S15328007SEM1003\\_1](http://doi.org/10.1207/S15328007SEM1003_1)
- Kim, H., Kim, E., & Na, D. (1997). Naming deficits in patients with dementia of the Alzheimer type: Error analysis of Korean version-Boston naming test. *Journal of Korean Neurology*, 15(5), 1012-1021.  
<http://www.jkna.org/journal/view.php?number=4172>
- Kline, R. B. (2005). *Psychological testing: A practical approach to design and evaluation*. California: SAGE publications.  
<http://methods.sagepub.com/book/psychological-testing>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling (4th ed.)*. New York, NY: Guilford publications.  
<http://www.guilford.com/books/Principles-and-Practice-of-Structural-Equation-Modeling/Rex-Kline/9781462551910>
- Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford publications.  
<http://www.guilford.com/books/Longitudinal-Structural-Equation-Modeling/Todd-Little/9781462510160>
- Lomazzi, V. (2018). Using alignment optimization to test the measurement invariance of gender role attitudes in 59 countries. *Methods, Data, Analyses: A Journal for Quantitative Methods and Survey Methodology*, 12(1), 77-103.  
<http://doi.org/10.12758/mda.2017.09>
- Luong & Flake (2022). Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting. *Psychological Methods*, Advance online publication.  
<http://doi.org/10.1037/met0000441>
- Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2018). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of

- latent means across many groups. *Psychological Methods*, 23(3), 524-545.  
<http://doi.org/10.1037/met0000113>
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58(4), 525-543.  
<http://doi.org/10.1007/BF02294825>
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Taylor & Francis Group.  
<http://www.taylorfrancis.com/books/mono/10.4324/9780203821961/statistical-approaches-measurement-invariance-roger-millsap>
- Munck, I., Barber, C., & Torney-Purta, J. (2018). Measurement invariance in comparing attitudes toward immigrants among youth across Europe in 1999 and 2009: The alignment method applied to IEA CIVED and ICCS. *Sociological Methods & Research*, 47(4), 687-728.  
<http://doi.org/10.1177/0049124117729691>
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1), 115-132.  
<http://doi.org/10.1007/BF02294210>
- Muthén, B., & Asparouhov, T. (2013). BSEM measurement invariance. *Mplus web notes*, 17, 1-48.  
<http://www.statmodel.com/examples/webnotes/webnote17.pdf>
- Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: the alignment method. *Frontiers in Psychology*, 5, 978.  
<http://doi.org/10.3389/fpsyg.2014.00978>
- 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.  
<http://doi.org/10.1177/0049124117701488>
- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2), 171-189.  
<http://doi.org/10.1111/j.2044-8317.1985.tb00832.x>
- Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45(1), 19-30.  
<http://doi.org/10.1111/j.2044-8317.1985.tb00832.x>
- Muthén, L., & Muthén, b. (1998-2020). *Mplus: Statistical analysis with latent variables user's guide 8.0*. Los Angeles, CA: Muthén & Muthén.  
[http://www.statmodel.com/HTML\\_UG/introV8.htm](http://www.statmodel.com/HTML_UG/introV8.htm)
- 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.  
<http://doi.org/10.1177/0013164413498257>
- Rutkowski, L., & Svetina, D. (2017). Measurement invariance in international surveys: Categorical indicators and fit measure performance. *Applied Measurement in Education*, 30(1), 39-51.

- <http://doi.org/10.1080/08957347.2016.1243540>  
 Rutkowski, L., Svetina, D., & Liaw, Y. L. (2019). Collapsing variables and measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(5), 790-802.  
<http://doi.org/10.1080/10705511.2018.1547640>
- Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling*, 21(2), 167-180.  
<http://doi.org/10.1080/10705511.2014.882658>
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507-514.  
<http://doi.org/10.1007/BF02296192>
- Son, S., Kim, H., & Hong, S. (2019). Measurement invariance testing using random effects model for many groups: Multilevel confirmatory factor analysis (ML CFA) and multilevel factor mixture modeling (ML FMM). *Korean Journal of Psychology: General*, 38(2), 185-218.  
<http://dx.doi.org/10.22257/kjp.2019.6.38.2.185>
- Tennant, A., Penta, M., Tesio, L., Grimby, G., Thonnard, J. L., Slade, A., Lawton, G., Simone, A., Carter, J., Lundgren-Nilsson, A., Tripolski, M., Ring, H., Biering-Sørensen, F., Marincek, C., Burger, H., & Phillips, S. (2004). Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: The PRO-ESOR project. *Medical Care*, 42(1), I37-I48.  
<http://doi.org/10.1097/01.mlr.0000103529.63132.77>
- Van de Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology*, 6, 1064.  
<http://doi.org/10.3389/fpsyg.2015.01064>
- 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.  
<http://doi.org/10.1097/01.mlr.0000103529.63132.77>
- Woods, C. M. (2009). Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research*, 44(1), 1-27.  
<http://doi.org/10.1080/00273170802620121>
- Woods, C. M., & Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement*, 35(5), 339-361.  
<http://doi.org/10.1177/0146621611405984>
- Yoon, M., & Lai, M. H. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(2), 201-213.  
<http://doi.org/10.1080/10705511.2017.1387859>

1차원고접수 : 2023. 03. 07.

최종게재결정 : 2023. 04. 06.

## Analysis of Approximate Measurement Invariance of Categorical Indicators Using Alignment

Youngshin Ju

Seungmin Jahng

Department of Psychology, Sungkyunkwan University

Multiple-group confirmatory factor analysis (CFA) has been widely used to analyze measurement equivalence using factor models. Traditionally, in order to conclude that measurement equivalence exists, it is necessary to have evidence that the factor structure and measurement parameters, such as factor loadings are completely identical, with no differences across target groups or measurement occasions. However, when the number of groups is large, these conditions can become overly stringent or impractical. Approximate measurement invariance is a relaxed form of measurement invariance and proposed to practically examine measurement invariance between many groups. Alignment is a widely used procedure for analyzing approximate measurement invariance and has been used to compare measurement models with continuous indicator between many countries. On the other hand, the analysis of measurement invariance of test scores across ages reflecting developmental changes in psychological attributes has been mainly conducted by comparing two or three small age groups using multiple-group CFA. However, in the analysis of measurement invariance across ages, approximate measurement invariance can be applied to ages that are divided into a large number of age groups. The current study discusses the idea, necessity, and analytical model of approximate measurement invariance and illustrates the analysis of the approximate measurement invariance of psychological tests over time using the alignment method. In particular, we discussed the conceptual similarities and differences between measurement invariance and approximate measurement invariance for factor models with categorical data, and introduced the analytical process of measurement invariance and approximate measurement invariance for 26 age groups of K-MMSE test scores, a test for assessing cognitive function, using multiple-group CFA and the alignment method. Based on the results of the analysis, the implications of approximate measurement invariance versus traditional measurement invariance were explained, and discussions related to analysis of approximate measurement invariance with categorical data were summarized.

*Key words* : categorical confirmatory factor analysis, multiple group confirmatory factor analysis, approximate measurement invariance, Alignment, K-MMSE

## 부 록

### (1) 다집단 확인적 요인분석을 위한 Mplus 코드 - 강한 동일성 모형

TITLE: MG CFA

DATA: FILE = mmse2018.csv;

VARIABLE: NAMES = REGIS RECAL DRAW TIME PLACE ATTEN LAN group;  
USEVARIABLE = REGIS RECAL DRAW TIME PLACE ATTEN LAN group;  
CATEGORICAL = REGIS RECAL DRAW TIME PLACE ATTEN LAN;  
GROUOPING = group (1=A 2=B 3=C 4=D 5=E 6=F 7=G 8=H 9=I 10=J  
11=K 12=L 13=M 14=N 15=O 16=P 17=Q 18=R 19=S 20=T  
21=U 22=V 23=W 24=X 25=Y 26=Z);  
MISSING = REGIS RECAL DRAW TIME PLACE ATTEN LAN(999);

ANALYSIS: ESTIMATOR = WLSMV;  
PARAMETERIZATION = THETA;  
DIFFTEST = WEAK\_THETA.DAT;  
STARTS = 20;

MODEL: f1 BY REGIS\* (a)  
RECAL DRAW TIME PLACE ATTEN LAN;  
[REGIS\$1] (b);  
[REGIS\$2] (c);

MODEL A: {f1@0}; f1@1; ! 참조집단의 요인척도설정  
REGIS@1 RECAL@1 DRAW@1 TIME@1 PLACE@1 ATTEN@1 LAN@1;  
! 잔차분산 고정

한국심리학회지: 일반

(2) 범주형 자료의 정렬법을 위한 Mplus 코드

TITLE: Alignment

DATA: FILE = mmse2018.csv;

VARIABLE: NAMES = REGIS RECAL DRAW TIME PLACE ATTEN LAN group;  
USEVARIABLE = REGIS RECAL DRAW TIME PLACE ATTEN LAN group;  
CATEGORICAL = REGIS RECAL DRAW TIME PLACE ATTEN LAN;  
CLASSES = C(26);  
KNOWNCLASS IS C(group = 1-26);  
MISSING = REGIS RECAL DRAW TIME PLACE ATTEN LAN(999);

ANALYSIS: TYPE = MIXTURE;  
ESTIMATOR = MLR;  
ALIGNMENT = FIXED(18); ! 요인평균이 0에 가까운 18번 집단의 요인평균 고정  
! 요인분산만 제약할 경우 FREE로 기재  
ALGORITHM = INTEGRATION;

MODEL: %OVERALL%  
f1 BY REGIS RECAL DRAW TIME PLACE ATTEN LAN;

OUTPUT: ALIGN;