

번안 심리검사 타당화 작업에 대한 체계적 검토: 검수와 보고 관행에 대한 검토와 제언

김 미 립* 임 예 지

고려대학교

본 연구는 한국 심리검사 번안 타당화 연구를 체계적으로 검토하여 국내 심리검사 번안의 검수와 보고에 대해 제언하는 것을 목표로하였다. 교육 및 심리검사의 표준과 ITC 검수 지침에 기반하여 2017년부터 2023년까지 출간한 총 107개 번안검사의 타당화 관행을 검토하였으며, 번안검사의 심리측정적 속성에 대한 정보가 충분히 보고되고 있는지 또한 각 연구가 타당도 확보를 위해 적합한 분석을 실시했는지 조사하였다. 검토 결과, 두 편을 제외하고 모든 번안검사 타당화 연구는 신뢰도를 보고하였고 요인분석을 거의 필수적으로 사용하여 구성 타당도를 검증하였음을 알 수 있었다. 따라서 해당 요인분석의 시행과 보고 관행에 대해 보다 면밀히 살펴보고 몇 가지 제언을 함으로써 응용 연구자들에게 도움이 되고자 하였다. 또한 같은 개념에 대하여 연구마다 각기 다른 용어의 사용이 빈번함을 지적하며, 심리학 연구에서 사용하는 학술 용어의 통합을 제언하였다. 체계적 검토 결과에 따른 본 연구의 제언은 번안검사 뿐만 아니라 새로운 검사의 개발 및 타당화를 목표로 하는 연구자들에게도 도움이 될 것으로 기대한다.

주요어 : 번안검사 타당화, 검수 지침, 체계적 검토, 심리측정적 속성, 요인분석

† 교신저자: 김미림, 고려대학교 4단계 BK21 심리학교육연구단, (02841) 서울시 성북구 안암로 145, Tel: 02-3290-2558, E-mail: mirimkim@korea.ac.kr



Copyright © 2024, 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.

대다수 사회과학 연구가 잠재적 특성을 수리화하는 작업, 즉 측정을 통하여 양적 연구를 실시한다. 연구의 타당성을 위해서는 각 연구에 알맞은 측정 도구의 사용이 필수적이며, 이에 따라 적합한 측정 도구를 개발하고 도구의 적합성 및 타당성을 검증하는 작업 또한 중요하다. 타당화가 충분히 이루어진 검사 도구는 후속 연구에서도 계속해서 측정 도구로 활용될 수 있으며, 만약 원검사가 외국어로 개발된 것일 경우에는 한국어로 번역 및 국내 문화와 정서에 알맞게 번안하고 해당 번안검사를 타당화한다.

번안검사의 타당화는 검사를 단순히 번역하는 것 이상의 작업이라고 볼 수 있다. 이는 원검사의 언어를 해당 국가의 언어로 번역하는 작업뿐만 아니라 현지 문화권 상황에 더욱 적합하게 하기 위한 노력이 추가적으로 필요하기 때문이다. 부적절한 번안은 검사의 타당도를 저해하고, 부적합한 도구의 사용은 연구 결과를 오도할 수도 있으므로 원검사와 마찬가지로 번안검사의 타당화 작업 역시 필요하다(Clark & Watson, 2019). 이에 따라 국제검사 위원회(ITC: International Test Commission)에서는 번역 및 번안 검사를 위한 지침을 제공함으로써 번안 작업의 중요성을 강조하고 올바른 타당화 작업을 소개하고 있다.

서동기와 이순목이 번역한 한국어판 ITC 지침(ITC Guidelines for Translating and Adapting Tests; Gregoire, 2018) 크게 여섯 가지 지침으로 이루어져 있다. 첫 번째 지침은 번안 이전에 필요한 선행조건과 관련이 있으며, 원검사에 대한 번안 허가의 필요성을 이야기하고, 원검사와의 언어적/문화적 차이에 따라 발생할 수 있는 여러 상황에 적합한 절차를 설명한다. 두 번째 지침은 언어와 문화를 고려한 번역

설계와 절차에 기반한 검사개발 지침으로 대상 언어와 문화, 그리고 검사에 대한 지식이 풍부한 전문가와 대상 모집단에게 번안 검사 문항, 지침, 그리고 집행을 유사하게 하기 위한 근거 제공을 강조한다. 또한 본격적인 대규모 검사 이전에 파일럿 검사와 그에 대한 문항 분석을 실시한다면 본 검사의 시행에 참고할 수 있는 유용한 정보를 미리 파악하고 추후 확인적 증거로도 활용할 수 있음을 언급하기도 하였다. 세 번째 지침은 검수 방법에 대한 것으로, 타당화 과정에 쓰이는 실제적 자료 분석법(empirical data analysis)과 관련이 있다. 심리측정적 속성(psychometric properties)에 기반한 다수의 분석법을 설명하고, 적합한 신뢰도 및 타당도 확보를 위한 방법을 안내한다. 번역에서의 이슈나 문화 차이에서 발생할 수 있는 문항 비동등성을 나타내는 문항차별기능(differential item functioning)을 살펴보는 분석 또한 언급하고 있으며, 요인분석 등을 통한 검사 구조 동등성 검증을 강조하고 있다. 네 번째와 다섯 번째 지침은 각각 검사의 실시와 해석 단계에서 언어/문화적 차이에서 비롯된 결과의 차이가 발생할 수 있음을 알리고 있으며, 그 차이를 최소화하기 위한 방안을 소개한다. 마지막 지침으로, ITC는 문서화를 강조한다. 번안검사 연구자는 번안 검사와 원검사의 공통점 및 차이점에 대해서 상세하게 문서화하고, 번안검사를 사용하는 실무자들을 위하여 검사 매뉴얼을 제공해야 할 의무가 있음을 강조한다. 번안검사 개발과 관계된 기술 보고서와 검사 사용 지침에 대한 설명이 충분해야 추후 번안검사 사용의 타당도를 저해하지 않을 수 있다. 앞서 설명한 전반적인 지침 외에 구체적인 검수와 검사 실시 상황을 알아보고 관련 지침을 파악하고자 하는 연구자들

은 ITC 지침을 필수적으로 참고할 것을 제안한다. 특히 세 번째 지침에서 설명하고 있는 심리측정적 속성과 구체적인 분석법들은 국내 번안검사 연구에 큰 도움이 될 것으로 기대한다.

최근에는 국외에서 개발된 측정도구 및 검사에 쉽게 접근할 수 있기 때문에, 연구자들은 새로운 도구를 빠르게 접하고 이를 번안하여 연구에 이용할 수 있다. 일례로 한국심리학회 산하 학술지에서 출간한 번안 연구의 동향을 살펴보면, 2017년에는 아홉 개의 연구로 비교적 적은 편이었으나 2023년에는 21건으로 증가하여 번안검사에 대한 수요가 증가함을 보였다. 따라서 적합한 번안검사 타당화는 더욱 중요해졌으며 이에 본 연구는 번안검사를 이용하는 연구자들의 타당화 연구 현황을 알아보고 그 적합성을 살펴보는 것을 주요 목표로 하였다. 특별히 국내에 ITC 지침이 소개된 2017년 이후의 번안 심리검사 연구를 대상으로 Thoemmes와 Kim(2011)의 방법을 참고한 체계적 검토(systematic review)를 실시하였고, ITC 검수 지침에 기반하여 번안검사의 정보와 타당화 과정을 살펴보았다. 본 연구는 검토 결과에 대한 독자들의 올바른 이해를 위하여, 번안과 검수 지침을 먼저 소개하고, 기반이 되는 이론적 배경을 살펴보고자 한다. 후반부에는 체계적 검토의 결과를 종합하여 번안 심리검사 연구의 타당화 현황에 대하여 논의하였다. 즉, 번안검사의 심리측정적 속성에 대한 경험적 증거가 충분한지 그리고 증거에 기반한 적합한 판단이 이루어졌는지 살펴보았다. 본 연구는 ITC 지침 중 양적 방법에 의한 검수와 문서화에 대해서 강조하고 있으므로 검토 결과 및 제언은 번안검사 연구자들에게만 국한될 필요가 없으며 새로운 검사의 개발 및

타당화를 목표로 하는 연구자들에게도 도움이 될 것으로 기대한다.

번안

연구자별로 적합한 번안 과정에 대해 다양하게 정의하고 있으나 본 연구는 Beaton 등(2000)과 Wild 등(2005)이 설명한 구조로 이를 설명하고자 한다. 첫째, 일차 번역은 원검사와 번안검사에 해당하는 언어와 문화에 대해 충분한 지식과 이해를 가지고 있는 두 명의 이중 언어 사용자가 실시한다. 둘째, 각 번역 결과가 종합되도록 조율을 한다. 셋째, 번역본을 다시 원검사의 언어로 번역하는 과정, 즉 역번역 과정을 거쳐 더욱 자연스러운 번안 결과를 기획한다. 이때 역번역은 원검사의 언어를 모국어로 사용하는 번역자가 담당하도록 한다. 넷째, 문화적으로 자연스러운 번안을 위해 제삼의 번역자 혹은 전문가 패널을 활용하여 검사 간의 동등성을 확인한다(Brislin, 1970; Gregoire, 2018). 이는 언어적 동등성뿐만 아니라 개념적인 동등성을 확보하기 위함이다. 마지막으로, 예비 검사를 실시하여 검사에 대한 응답자의 이해도를 파악하고, 추가로 내용 타당도에 대한 근거를 확보한다.

검수 지침

연구자는 다양한 경험적 분석을 통해서 번안검사가 원검사와 크게 다르지 않음을 검수하여 검사 동등성을 확립할 필요가 있다. 이와 관련하여, ITC 지침은 연구자가 고려해야 하는 첫 번째 사항으로 충분한 표본크기를 이야기한다. 번안검사의 표본은 원검사와 마찬가지로 검사 대상의 모집단을 대표할 수 있어

야 하며, 분석 모형을 안정적으로 추정할 수 있을 만큼 충분한 크기여야 하기 때문이다. 적합한 표본을 확보했다면, 원검사와의 동등성을 검증하고, 변안검사의 심리측정적 속성을 살펴본다.

검사 목적에 따라, 원검사와 변안검사가 확립해야 하는 동립성의 종류는 상이할 수 있다. ITC 지침은 OECD 주관의 국제 학업성취도 평가인 PISA(programme for international student assessment)와 우울척도를 예시로 들고 있다(Gregoire, 2018). PISA 검사는 OECD 국가를 포함한 다양한 국가들의 학생 역량을 평가하고 비교하는 것을 목표로하므로, 원검사와 변안검사 간의 검사 양식 동등성을 검수할 필요가 있다. 또한 국가 간 검사점수의 비교를 목적으로 척도체계의 동등화(equating) 및 연계(linking) 작업을 실시한다. 반면, 우울척도는 언어체계가 다른 임상군 간의 우울점수 비교보다는 변안검사가 우울의 측정에 타당한 도구인지, 그리고 측정된 우울 점수가 적은 오차를 갖는지의 여부가 더 중요하다. 따라서 원검사와 변안검사의 척도체계 비교보다는 언어, 혹은 문화에 따른 검사 구조의 동등성을 확인하고 변안검사의 타당성을 확보하기 위한 통계적 접근을 고려한다.

원검사가 신뢰롭고 타당하다는 근거가 충분할지라도, 해당 검사의 개정판에 대해서는 별도의 타당화 작업이 필요하다(AERA, APA, NCME, 2014; Gregoire, 2018). 같은 논리로 변안검사 또한 추가적인 타당화 작업을 실행하여 변안검사 사용에 대한 타당성을 뒷받침할 수 있어야 하며, 연구자는 타당화 과정에서 근거한 검사 정보를 제공할 연구적 책무가 있다.

신뢰도

신뢰도는 같은 검사를 이용하여 동일한 피험자의 구성개념을 반복 측정할 때 검사의 결과가 일관된 정도를 나타낸다. 즉 신뢰도가 낮은 검사는 반복 측정에 따라 상이한 결과가 발생할 수 있기 때문에(강태훈, 김명연, 2023; 성태제, 2002), 연구자들은 변칙적인 상황에도 일정한 검사 점수를 확보하고자 신뢰도가 높은 검사를 사용하고자 한다. 또한 American Educational Research Association(AERA), American Psychological Association(APA), 그리고 National Council on Measurement in Education(NCME) (2014)에서 출간한 교육 및 심리검사를 위한 표준(Standards for Educational and Psychological, Testing)에 따르면 연구자는 검사의 시행마다 신뢰도를 보고해야 하며, 변안 검사의 경우에는 원검사와 변안검사의 신뢰도를 모두 보고할 필요가 있다고 하였다. 이에 대하여, 본 연구는 검사 이론에 기반한 신뢰도의 정의를 소개함으로써 검사의 신뢰도가 거듭 강조되는 이유를 살펴보고자 한다.

연구에서 항상 언급되는 신뢰도의 개념과 신뢰도 지수는 고전검사이론(classical test theory)에 기반하고 있다. 고전검사이론은 관찰점수 X 가 진점수(true score) T 와 측정오차(measurement error) E 로 구성된 확률변수(random variable)라는 개념을 바탕으로 신뢰도의 개념을 정의하였다(Crocker & Algina, 1986). 측정에 오차가 전혀 없다면 검사점수는 항상 일정하겠지만, 현실적으로 측정오차가 없는 검사를 개발하기란 어려운 일이다. 따라서 검사를 개인에게 무한 번 시행할 경우, 검사점수는 시행마다 근소한 차이를 가질 수 있으며, k 번째 검사를 통해 관찰된 점수 X_k 에 대해

확률 p_k 를 갖는 확률분포를 따르게 된다. 확률변수의 기댓값은 확률변수의 평균값으로 표현할 수 있는데, 이를 이용하여 고전검사 이론에서는 확률변수 X 의 기댓값을 진점수 ($T = \sum_{k=1}^K X_k p_k$)라고 정의하고, 관찰점수의 평균값으로 진점수를 표현한다. 또한 관찰점수와 진점수의 차이값, 즉 관찰점수가 진점수에서 벗어난 정도를 측정오차로 정의한다. 식 (1)은 이와 같은 세 가지 개념의 관계를 나타낸다.

$$X = T + E \quad (1)$$

또한 진점수와 측정오차 간에 상관이 없다는 가정에 기반하여 관찰점수의 분산(ρ_X^2)을 진점수 분산(ρ_T^2)과 측정오차 분산(ρ_E^2)의 합으로 나타낼 수 있으며, 이는 식 (2)와 그림 1로 표현할 수 있다. 이때 신뢰도는 관찰점수의 분산에 대해 진점수 분산이 차지하는 비율로 정의한다($\rho_{XX} = \sigma_T^2 / \sigma_X^2$).

$$\rho_X^2 = \rho_T^2 + \rho_E^2 \quad (2)$$

그림 1을 통해 관찰점수 분산 내에서 진점수 분산의 차지하는 비율이 클수록 측정오차의 비율이 작아지는 것을 쉽게 파악할 수 있다. 앞서 설명한 검사점수의 일관도와 연결해 보면, 신뢰도가 높은 검사일수록 측정오차보

다 진점수가 차지하는 비율이 높아 일관된 결과를 제공하는 경향이 있다고 볼 수 있다. 따라서 측정오차의 영향력이 적은 것을 증명하기 위해서라도 신뢰도는 계속해서 검증되고 강조되어야 할 검사의 속성이 될 수밖에 없다.

그러나 검사의 높은 신뢰도가 항상 검사의 적합성으로 귀결하는 것은 아니므로 주의가 필요하다(Raykov & Marcoulides, 2011). 신뢰도 지수의 한 가지인 Cronbach's α 계수는 검사를 이루는 문항 간 공분산을 이용하여 산출되기 때문에 검사를 이루는 문항 간의 일관성, 즉 내적 일관성을 나타낸다. α 계수는 측정 도구의 적합성을 나타내기 위한 특성으로 보고되고 있으나, 검사를 이루는 문항 간의 상관성이 높을수록 그리고 문항의 개수가 많아질수록 그 크기가 커지는 경향이 있기 때문에, 높은 신뢰도가 단순히 좋은 검사를 나타내는 것은 아님을 알 수 있다. 단적인 예로, 우울을 측정하기 위하여 지능검사를 사용한 경우를 생각해 보자. 모든 연구자가 지능검사를 통해 우울을 측정하는 것이 적합하지 않음을 알고 있지만, 지능검사를 구성하는 문항 간 상관성이 크고, 문항의 개수가 충분히 많다면 연구 목표와 관계없이 연구 도구의 신뢰도는 높게 추정된다. 다시 말해 검사가 연구에 적합한지와는 무관하게, 검사의 신뢰도는 높게 나타날 수 있다. 이러한 예는 신뢰도 이외의 추가적인 근거에 기반하여 검사를 평가할 필요가 있음을 시사한다(Crocker & Algina, 1986). 검사 타당도가 중요한 이유가 바로 여기에 있다.

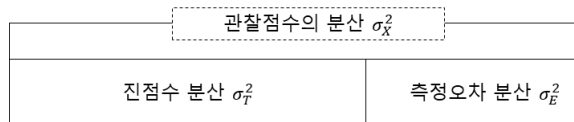


그림 1. 관찰점수, 진점수, 그리고 측정오차 분산의 관계

타당도

최근까지 검사 타당도의 개념은 변화하고 있으며, 이에 따라 검사가 근거해야 하는 타당도의 세부 유형 또한 다양해지고 있다(강태훈, 김명연, 2023). 교육 및 심리검사의 표준(AERA, APA, & NCME, 2014)의 정의에 따르면, 타당도는 검사를 통한 의사결정이 측정(measurement) 혹은 예측(prediction)과 같은 검사의 설계 목적을 충족하는 정도를 의미하며, 그 유형을 크게 내용 타당도(content validity), 준거 타당도(criterion validity), 그리고 구성 타당도(construct validity)로 분류할 수 있다. 그러나 연구 분야에 따라 세 가지 타당도의 하위 유형이 조금씩 상이하기 때문에 본 연구에서는 미국심리학회(APA)의 정의에 따라 세부 유형을 논의하고자 한다. 따라서 예측 타당도(predictive validity)와 공존 타당도(concurrent validity)는 준거 타당도(criterion validity)로, 그리고 수렴 타당도(convergent validity)와 변별 타당도(discriminant validity)는 구성 타당도(construct validity)의 하위 유형으로 분류하였다.

내용 타당도는 타당도 확보를 위한 질적인 근거로, 이를 이용하여 검사의 내용이 측정하고자 하는 구성개념 및 속성과 관계가 있는지에 대해 주관적으로 판단한다(신진아 외, 2021). 간단한 예로, 도박중독을 측정하고자 하는 검사는 중독의 잠재 속성과 관계된 내용으로 구성되어야 한다. 그러나 특별히 도박중독을 나타내는 행동을 기술함으로써 다른 중독(예: 약물중독)과는 차별화할 수 있는 행위를 측정하고, 이에 해당하는 잠재점수를 수량화할 수 있도록 설계해야 한다. 이렇게 심리검사 문항은 대상 구성개념을 나타내는 특수한 행위 및 생각 또한 기술할 수 있어야 하므

로, 해당 분야 전문가들의 의견을 수렴하고 검사 문항을 설계함으로써 내용 타당도를 확보할 수 있다.

준거 타당도는 검사와 외부 준거변수와의 상관관계를 살펴봄으로써 검사의 타당도를 검증하는 방식이며, 예측 타당도와 공존 타당도가 이에 포함된다. 예측 타당도가 중요한 상황으로는 선발 현장을 예로 들 수 있다. 인적성 검사에 기반하여 응시자의 채용 여부를 결정할 경우, 검사의 타당도는 검사 점수가 채용 이후 응시자의 업무성과(준거변수)를 얼마나 예측할 수 있는지와 관계가 있을 것이다(Jenkins & Griffith, 2004). 검사 점수와 업무능력 간의 상관관계가 강할수록 높은 예측 타당도를 의미하며, 검사 점수가 높은 사람들을 우선적으로 선발하는 것에 대한 근거가 된다. 한편, 공존 타당도를 알아보기 위해서는 기존에 널리 쓰이는 검사를 준거변수로 하여 제작 검사와의 상관을 살펴보면, 예측 타당도와 마찬가지로 준거 검사와의 상관관계가 강할수록 공존 타당도가 높다고 판단한다(신진아 외, 2021). 널리 쓰이고 있는 검사는 반복적으로 타당화 작업을 거친 검사를 의미하기 때문에, 공신력 높은 검사와의 상관관계가 높은 검사 역시 타당한 검사로 공인할 수 있다는 논리다.

검사는 예측뿐만 아니라 측정을 주요 목적으로 하여 실시할 수 있다. 측정은 구성개념에 대한 응답자의 상태(status)와 위치(standing)에 숫자를 부여하여 수치적으로 추정하는 것을 의미하며(Raykov & Marcoulides, 2011), 구성 타당도는 이러한 과정이 적합한지에 대한 근거가 된다. 구성개념은 실재를 관찰할 수 없는 잠재적 속성이기 때문에, 임의의 척도 없이는 구성개념의 차원(dimension)에 대한 개인의 위치를 판단할 수 없다. 따라서 측정을 통

해 개개인의 구성개념에 숫자를 부여함으로써 잠재 속성의 위치를 가능하고 상대적인 정도를 비교한다. 구성 타당도의 근거는 검사 문항과 구성개념 간의 관계에 기반하며, 검사 점수가 나타내는 것이 실제 측정하고자 하는 대상 구성개념인지, 또한 검사가 구성개념에 최대한 근접한 점수를 추정하였는지를 판단한다(AERA, APA, & NCME, 2014).

구성 타당도는 단 한 가지 통계분석을 통해 평가할 수 있는 개념이 아니며 다양한 근거에 기반하여 종합적으로 평가하게 된다. 구성 타당도를 평가하는 접근 방식은 크게 두 가지로 요약할 수 있는데, 검사를 이루는 문항과 구성개념 간의 관계성을 검증함으로써 타당화하는 것을 내적 접근(internal approach), 외부/준거 검사와의 관계성에 기반하여 구성 타당도를 평가하는 것을 외적 접근(external approach)이라고 한다(Clark & Watson, 2019; Loevinger, 1957; Messic, 1995; Raykov & Marcoulides, 2011). 앞서 강조하였듯이 구성 타당도의 확보를 위해 검사의 다양한 속성을 평가하는 것이 바람직하므로, 두 가지 접근 방식에 모두 근거하여 구성 타당도 개념을 확보할 것을 권고한다.

본 연구는 구성 타당도를 살펴보기 위한 방법으로 요인분석을 다루고자 한다. 요인분석 모형은 구성 타당도를 살펴보기에 적합한 모형 중 하나라고 볼 수 있으며, 국내 심리검사 타당화 연구를 검토한 결과 실제로 가장 빈번하게 사용되고 있는 모형임을 알 수 있었다. 요인분석의 측정모형에 주목하는 경우, 연구자는 요인-문항 간의 관계를 살펴봄으로써 각 문항이 요인(구성개념)을 적절히 측정하는지 평가한다(내적 접근). 이때 문항과 요인 간의 관계가 충분히 커서 잠재점수가 관찰점수를 설명하는 정도가 높을수록 검사 구성 타당도

의 적합한 근거가 될 수 있다. 또한 요인분석 모형을 통해 대상 검사와 외부 검사에 해당하는 요인 간 관계를 살펴봄으로써 구성 타당도를 논의할 수 있으며, 이처럼 준거와의 관계에 근거하는 구성 타당도를 특별히 수렴 및 변별 타당도라고 일컫는다(외적 접근). 검사가 대상 구성개념을 잘 측정했다면, 유사한 구성개념을 측정하는 준거검사와 연관성이 높아 수렴하는 경향이 커야 한다. 두 검사가 측정하는 요인 간의 상관이 높을수록, 두 검사의 동질성에 기반한 수렴 타당도의 근거가 된다. 연구자는 수렴 타당도를 확보함과 동시에 변별 타당도에 대해서도 고려할 필요가 있다. 대상 검사와 준거 검사가 측정하는 구성개념이 서로 다를 경우, 두 검사는 서로 연관성이 낮고 서로 변별하는 경향이 커야 한다. 요인 상관이 낮을수록 각 검사가 측정하는 요인 간의 이질성에 기반하여 변별 타당도의 근거를 확보할 수 있다.

심리학 연구에서 언급되는 또 다른 타당도로는 증분 타당도(incremental validity)를 들 수 있다. 증분 타당도는 Sechrest(1954)가 고안한 개념으로, 대상 검사가 기존 검사와 변별이 잘 되는지 그리고 기존 검사와 비교했을 때 종속변수에 대해 추가로 예측하는 면모가 있는지 살핀다.

요인분석 모형

요인분석은 구성개념의 차원성과 문항의 특성을 함께 평가할 수 있기 때문에 심리측정적 속성을 확인하는 주요 분석 방법으로 사용되고 있으며(Brown, 2015), ITC 지침 또한 요인분석 모형을 비중 있게 다루고 있다. 따라서 본 연구는 국내 변안검사 연구에서의 요인분석

관행을 ITC 지침 및 선행 연구로부터의 권고 사항과 비교하고 타당화 작업에 유용한 사항을 논의하고자 하였다. 비교 결과를 기술하기 전에 요인분석 모형에 대한 이론적 배경을 아래에 설명하였다.

탐색적 요인분석과 확인적 요인분석

심리학 연구를 포함한 대다수의 사회과학 연구에서 활용하는 요인분석 모형은 공통요인 분석(common factor analysis)으로 탐색적 요인분석 혹은 확인적 요인분석 모형을 의미하며(이순목, 1994; Brown, 2015), 식 (3)과 같은 선형 식으로 나타낼 수 있다.

$$x_{ij} = \lambda_{j1}\xi_{j1} + \lambda_{j2}\xi_{j2} + \dots + \lambda_{jr}\xi_{jr} + \delta_{ij} \quad (3)$$

x_{ij} 는 사람 i 가 문항 j 에 대해 응답한 문항 점수를 의미하며, 이는 첫 번째 공통요인 ξ_{j1} 부터 r 번째 공통요인 ξ_{jr} , 그리고 고유요인 δ_{ij} 와의 선형관계로 표현할 수 있다. 고전검사이론에서 관찰점수의 분산을 진점수와 측정오차 분산의 구성으로 정의했듯이, 공통요인분석 모형은 문항점수의 분산을 공통분산(communality)과 고유분산(unique variance)으로 표현한다(Kline, 2013). 공통분산은 공통요인이 문항점수를 설명하는 정도를 나타내며 고유분산은 공통요인이 설명하지 못하는 문항 특유의 특성이 차지하는 정도를 의미한다. 이때 요인분석은 다수 문항의 공통분산에 관여하는 요인을 추출하고 그 속성을 살펴보는 것을 목표로 한다(Brown, 2015).

요인구조에 대한 가설의 차이

탐색적 요인분석과 확인적 요인분석의 가장 큰 차이는 요인의 개수와 요인-문항 간 관계에 대한 가설 유무에 있다. 탐색적 요인분석은 요인-문항 간의 불명확한 관계에 기반하여 공통분산을 설명하는 적합한 개수의 요인과 요인구조를 탐색하는 것을 목표로 하지만, 확인적 요인분석의 경우 구조에 대한 명확한 가설을 설정하고 이를 검증한다(Brown, 2015). 두 모형의 요인구조 차이는 그림 2를 통하여 살펴볼 수 있다.

그림 2는 여섯 개의 문항과 두 개 요인 간의 관계를 나타내고 있다. 탐색적 요인분석을 나타내는 (a)의 경우, 각 요인이 모든 문항을 설명하고 각 문항이 다수의 요인으로 회귀함으로써 교차 요인부하(crossed-factor loading)를 추정하게 된다. 이는 검사 문항과 요인의 관계에 대한 명확한 가설이 없기 때문에 모든

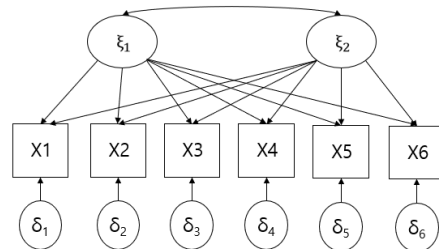
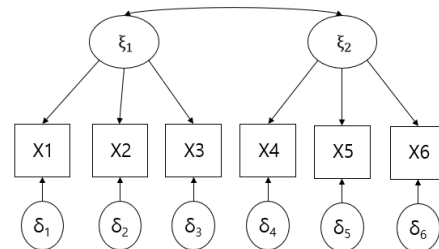


그림 2. (a) 탐색적 요인분석 모형



(b) 확인적 요인분석 모형

요인-문항 관계를 살펴보기 위함이다. 그림 (a)의 예시는 두 개의 요인에 기반한 요인구조를 나타내고 있으나, 실제 탐색적 요인분석은 요인 개수에 대한 모든 경우의 수(예: 여섯 개의 문항에 대해서는 단일 요인부터 6요인까지)에 기반하여 가능한 요인-문항 관계를 살펴보게 된다. 가설이 부재한 경우에 유용하기 때문에 검사 제작의 초기 단계에 실시하여 요인구조를 탐색한다(Flora & Flake, 2017). 반면 확인적 요인분석 모형의 경우, 요인 개수와 요인-문항 관계에 대한 구체적인 가설에 기반하여 모형을 설정하며 그 결과 그림 (b)와 같이 각 문항은 특정 요인에만 회귀한다. 따라서 요인과 문항 간의 유의미한 관계를 예측할 수 있는 경우에 대해서는 해당 요인부하를 추정하고, 그렇지 않은 관계에 대해서는 요인부하를 0으로 고정한다. 추정해야 하는 모수의 수가 적어짐으로써 탐색적 요인분석 모형보다 추정이 용이하다는 장점이 있으며, 확인적 요인분석은 가설에 기반하므로, 요인구조에 대한 탐색 이후에 최종 모형을 확인 및 타당화 하기 위하여 쓰일 수 있다. 즉, 탐색적 요인분석을 통하여 요인구조에 대한 가설을 설정하고, 이를 확인적 요인분석을 통하여 교차 타당화하는 식으로 검사를 타당화하는 것이 일반적이다(Brown, 2015; Flora & Flake, 2017).

요인분석 모형의 설정, 평가 및 수정

요인구조에 대한 두 요인분석 모형의 차이는 모형의 설정(specification), 식별(identification), 그리고 분석 방법에도 차이를 갖게 한다. 이에 대하여 본 연구는 탐색적 요인분석부터 확인적 요인분석의 순으로 각 모형의 설정과 평가를 간략히 설명하고자 하였으며, 표 1에 이

를 정리하여 연구자들에게 도움이 되고자 하였다.

탐색적 요인분석은 대략 (1) 요인의 추출, (2) 요인개수 결정, (3) 그리고 요인구조의 회전을 걸쳐 모수를 추정한다. 요인 추출은 추출법에 따라 차이를 보일 수 있는데, 추정법의 예로는 최대우도법(maximum likelihood estimation)과 주축분해법(principal axis factoring)을 들 수 있다(이순목, 1994; De Winter & Dodou, 2012). 통계 소프트웨어에 따라 추출법에 대한 선택의 폭은 상이할 수 있다. 이때 한 가지 주의할 것은 일부 소프트웨어에서 이용할 수 있는 주성분 분석(principal component analysis)의 활용이다. 특히 SPSS에서는 요인 추출법으로 주성분 분석을 선택할 수 있는데, 주성분 분석은 공통요인분석의 맥락보다는 자료 축소를 연구 목적으로 하는 요인분석 모형이기 때문에 연구자의 주의가 필요하다(이순목, 1995; Brown, 2015). 공통된 요인의 추출과 그 해석을 목표로 하는 경우에는 주성분 분석 이외의 추정법이 적절하다고 볼 수 있다(Gorsuch, 1990).

탐색적 요인분석은 각 요인개수와 문항-요인 간 관계에 대한 명확한 가설이 없는 채로 수행되기 때문에 두 번째, 요인의 개수 결정 단계에서 연구자의 주관적 판단이 필요하다. 요인 개수의 판단, 즉 모형 결정에 참고할 수 있는 평가지표로는 스크리 도표(Cattell, 1966), 평행분석(parallel analysis; Horn, 1965), MAP test (Velicer, 1976) 등을 들 수 있으며, 소프트웨어에 따라 이용할 수 있는 평가지표에 차이가 있을 수 있다. 소프트웨어 전반에 걸쳐 공통적으로 사용할 수 있는 지표로는 스크리 도표, 평행분석, 그리고 모형 합치도지수(예: χ^2)를 들 수 있으며, 최대우도법을 사용하여 요인을

추출한 경우에 대해 *Mplus*, *jamovi*, 그리고 R 패키지는 대안적 합치도지수(예: CFI, RMSEA, SRMR)를 제시하고 있어 모형의 평가가 좀 더 수월해졌다.

탐색적 요인분석을 이용한 요인개수의 결정 단계에서 주의해야 할 사항은 다음과 같다. 첫째, 연구자는 카이저 룰(Kaiser-Guttman rule)에만 기반한 요인개수의 결정을 지양해야 한다. 이 지표는 1 이상의 고유치를 갖는 요인의 개수만큼 요인을 추출하는 방식으로, 판단이 직관적이라는 점에서 널리 사용된 적이 있다. 그러나 요인의 개수를 과대 혹은 과소 추정하는 경우가 많아 안정적인 결과를 산출하지 않는다는 점에서 현재는 추천하지 않는 준거로 볼 수 있다(Brown, 2015; Fabrigar, Wegener, MacCallum, & Strahan, 1999). 둘째, 각 평가지표는 서로 다른 결과를 가리키고 있을 수 있으므로 연구자는 다양한 평가지표에 복합적으로 근거하여 요인개수를 결정할 필요가 있다(Brown 2015; O'Connor 2000). 또한 단순히 통계적 결과에만 의존할 것이 아니라 이론에 기반한 해석가능성을 필수적으로 고려하여 요인의 개수를 결정할 필요가 있다.

마지막으로 살펴볼 탐색적 요인분석의 특징으로는 요인구조의 회전을 들 수 있다. 탐색적 요인분석 모형은 유일해(unique solution)가 아닌 다수의 해를 갖는 모형으로, 자료와 추출법이 같다면 동일한 모형 합치도지수를 갖는 여러 개의 요인구조를 산출할 수 있다(Brown, 2015). 이때 요인구조의 회전은 다요인 모형에 대해 특정 요인과 문항 간의 관계를 부각하여 해석이 수월하게 하는 과정으로, 특정 요인과 관계된 요인부하량을 작게 하거나 크게 하여 요인-문항 간 해석이 더욱 용이한 구조를 찾을 수 있도록 한다(Brown, 2015; Flora

& Flake, 2017). 연구자는 요인 간 상관에 대해 가설을 세우고 회전법을 선택할 수 있으며, 요인 간 상관이 없다고 가정하는 경우와 있다고 가정하는 경우에 대해 각각 직각회전(요인상관=0)과 사각회전(요인상관≠0)에 해당하는 회전법을 선택하고 요인구조를 산출한다. 일반적으로 사회과학 연구에서는 요인 간 상관이 있음을 가정하므로, 특별한 이론적 근거가 없다면 사각회전법을 사용하는 것이 타당하다.

회전 이후에는 요인과 문항의 적합성을 고려하여 모형을 수정하거나 문항을 삭제하여 최종 요인구조를 결정한다. 먼저 추출한 요인이 적합하지 않다고 판단하는 경우, 연구자는 더 적은 수의 요인을 추출하는 전략을 취할 수 있다. (a)요인에 부하하는 문항의 수가 적어서 해당 요인에 대한 측정모형을 타당화할 수 없을 때, (b)문항 *i*에 대한 요인 *j*의 요인부하량 λ_{ij} 이 작게 추정되고 공통분산($h_i^2 = \sum \lambda_{ij}^2$)의 크기가 작은 것으로 나타낼 때, 그리고 (c)요인 간 변별이 명확하지 않은 경우가 이에 해당한다. 반대로 검사 문항을 삭제하는 전략이 필요한 때도 있다. (d)다수 요인에 대한 특정 문항의 요인부하량의 크기가 비슷하여 요인-문항 관계를 명확히 할 수 없을 때, 또는 (e)문항의 모든 요인부하량이 1.3이나 .4보다 작은 경우에는 해당 문항을 삭제하기도 한다. 참고로 각 상황에 대한 구체적인 준거는 통계적 타당화에 근거했다기보다는 응용 연구자들의 경험에 기반하여 마련된 것으로 보인다(Brown, 2015).

확인적 요인분석은 요인구조에 대한 명확한

1) 요인부하량이 0.3, 0.4인 경우, 공통분산은 각각 $0.09 = (0.3)^2$ 과 $0.16 = (0.4)^2$ 으로, 요인이 문항의 분산을 9%, 16% 설명한다고 볼 수 있다.

가설에 기반하므로, 탐색적 요인분석과는 달리 요인의 추출 및 요인개수의 결정에 대한 고민을 덜 수 있다는 장점이 있다. 각 문항은 한 개의 요인에만 회귀하는 단순구조에 대한 모수를 추정하므로 회전 과정이 불필요하며, 참조변수에 해당하는 요인부하량을 1로 고정하여 요인에 척도를 부여함으로써 모형을 식별한다. 확인적 요인분석 모형에 대한 평가는 모형 합치도지수를 이용하며, 가설 모형에 대한 합치도지수가 준거 합치도지수에 근접할 때 요인구조가 적합하다고 평가한다. 일반적으로 다수의 합치도지수를 종합적으로 고려하여 모형의 적합성을 판단하며 .95 이상의 상대적 합치도 지수(incremental fit index; 예: CFI,

TLI), 0.08 미만의 절대적 합치도 지수(absolute fit index; 예: RMSEA, SRMR)을 나타내는 모형을 적합하다고 판단한다(Hu & Bentler, 1999; Kline, 2013).

확인적 요인분석 또한 탐색적 요인분석과 마찬가지로 요인과 문항의 적합성을 고려하여 모형을 수정하고 최종 모형을 결정하며(표 1 참고), 추가적으로 수정지수(modification index)를 활용하여 모형을 수정하기도 한다. 수정지수는 특정 모수에 대한 제약 조건이 얼마나 부적합한지를 나타내는 지수로 측정모형 내에서 0으로 고정된 가설이 해제될 때 감소할 수 있는 χ^2 의 정도로 해석할 수 있다. χ^2 의 감소량이 클수록 모형의 합치 정도가 높아지는

표 1. 요인분석 모형의 설정, 평가 및 수정

| | 탐색적 요인분석 | 확인적 요인분석 |
|-----------|---|---|
| 모형설정 | 요인추출 - 추출: 최대우도법, 주축분해법 | |
| | 요인개수 결정 - 준거: 스크리 도표, 평행검사, 모형합치도 | - 이론에 기반하여 요인개수와 요인-문항 간의 관계를 설정 |
| | 요인구조의 회전 - 사각회전(예: Varimax), 직각회전(예: Promax) | |
| 모형평가 및 수정 | 모형 평가 준거 - 모형합치도 - 요인부하량 및 공통분산 크기 | 모형 평가 준거 - 모형합치도 - 요인부하량 및 공통분산 크기 |
| | 모형 수정 요인개수의 축소 - 공통요인에 해당하는 문항이 적은 경우 - 공통요인에서 비롯된 요인부하량 및 공통분산이 작은 경우 - 요인 간 변별이 어려운 경우 문항 삭제 - 요인부하량이 다수 요인에 대하여 비슷한 경우 - 요인부하량 < 0.3 또는 0.4 | 모형 수정 문항 삭제 - 한 문항의 요인부하량이 다수 요인에 대하여 비슷한 경우 - 요인부하량 < 0.3 또는 0.4 수정지수 활용 |

것을 의미하기 때문에 큰 값을 갖는 수정지수를 고려하여 요인분석 모형을 수정한다. 그러나 수정지수의 크기에만 전적으로 의존하는 자료 중심의 분석(data-driven analysis)이 되지 않도록 이론적으로 설명이 가능한 수정지수를 선택하여 모형을 수정하는 것이 필요하다(Whittaker, 2012).

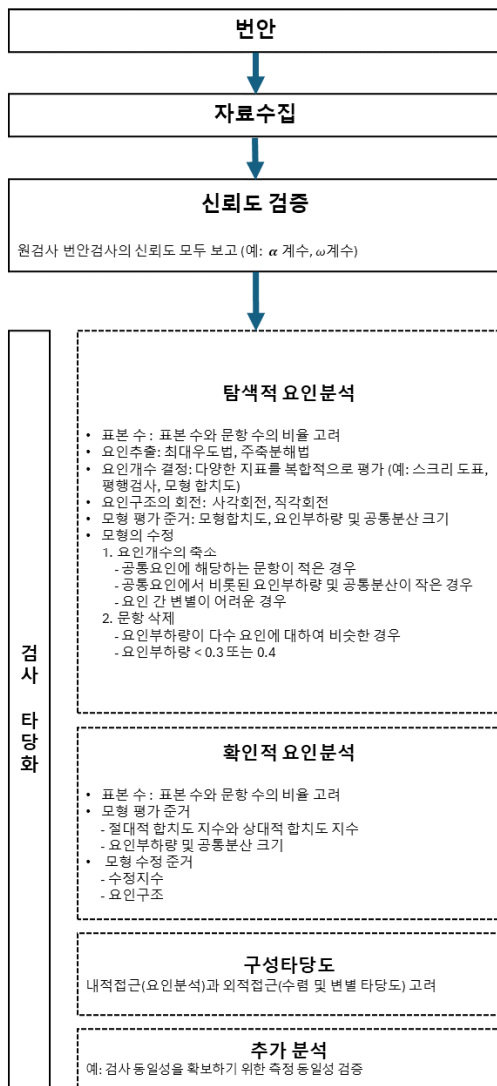


그림 3. 변안검사 타당화 과정

연구자는 요인분석 과정 및 결과와 관련된 주요 사항을 보고할 의무가 있다. 모형의 설정과 분석 과정을 설명하고 증거에 근거한 모형 평가 결과를 보고하며, 참고한 증거에 대해서도 명시하도록 한다. 모형을 수정하는 경우 또한 그 증거와 과정을 자세히 기술하여 수정 모형의 타당성을 뒷받침할 수 있어야 하며, 이렇게 선택된 최종 모형에 대해서는 분석 결과, 즉 요인구조를 보고함으로써 구성 타당도에 대한 이해를 용이하게 할 필요가 있다. 앞서 설명한 탐색적 요인분석 모형과 확인적 요인분석 모형의 설정 및 평가 과정은 표 1과 같이 요약할 수 있다. 표 1은 연구자가 각 과정에서 고려해야 하는 대표적인 증거 또한 포함하고 있다. 또한 변안검사 타당화 과정을 구조화하여 그림 3으로 나타내었다.

요인분석과 표본

연구모형의 안정적인 수렴과 추정을 위해서는 충분히 큰 표본에 기반하여 분석할 필요가 있다. 그러나 연구 조건과 분석 모형에 따라 충분한 표본크기가 상이하기 때문에 적정 표본크기에 대한 확일적 기준을 마련하는 것은 어려운 일이다. 일반적으로 추정해야 하는 모수가 많을수록, 자료의 분포가 비정규성을 가질수록 근사 정규성에 기반하여 모수를 추정하기 위해 더 큰 표본이 필요하며, 비정규성을 통제하기 위한 추정법(예: 가중 최소 제곱법; weighted least square method)을 사용하는 경우에도 최대우도법을 사용한 경우보다 더 큰 표본이 필요하다(Kline, 2011). 같은 맥락으로, 요인분석도 표본의 크기가 클수록 작은 표본 오차와 정확한 요인부하량을 산출할 수 있으므로(Browne, 1968) 충분한 크기의 표본에 기반

하여 검사를 타당화하는 것이 중요하다.

탐색적 요인분석의 경우 연구자는 $N:p$ 비율, 즉 문항의 수(p)에 대비한 최소 표본크기(N)를 고려하여 적정 표본크기를 설계할 수 있다. 예를 들어 Everitt(1975)는 $N:p$ 비율이 최소 10 이상의 값을 가질 것을 제안하였고, 이는 1개의 문항에 대해 표본크기가 10배 이상이어야 함을 의미한다. $N:p$ 비율에 대한 확실적인 기준은 없지만, 이순목(1994)에 따르면 해당 비율이 5 미만인 경우에 대해서 여러 선행 연구가 우려를 나타냈다고 하므로 연구자들은 이를 참고하길 바란다. 그밖에 요인분석에 필요한 절대적인 표본크기에 대한 연구도 있다. MacCallum, Widaman, Zhang 그리고 Hong (1999)은 추정된 요인구조의 복잡도에 따라 필요한 표본크기가 다름을 확인하였으며, 요인 간 변별이 수월하여 추출된 요인의 수가 적고, 각 요인 당 부하하는 문항이 3-4개인 경우에는 최소 300 이상의 표본이 필요하고, 그렇지 않은 경우에는 500 이상의 표본에 기반해야 안정적인 추정이 가능하다고 하였다. 그러나 실제 분석 이전에 연구자가 요인구조에 대해 판단하기는 어려우므로 500 이상의 표본을 확보하는 것이 타당한 것으로 보인다. 최근 연구도 이와 같은 맥락을 공유한다. Auerswald와 Moshagen(2019)은 공통분산과 요인부하량이 낮을 때, 요인 간 상관성이 높을 때, 그리고 요인 당 부하하는 문항의 수가 적을 때 요인개수를 결정하는데 어려움이 더 커지는 경향이 있음을 밝혔으며, 이에 따라 요인개수 결정 준거가 서로 불일치한 결과를 지지할 경우에는 500 이상의 표본에 기반하여 더욱 안정적인 요인 추출을 계획할 것을 제시하였다.

확인적 요인분석을 실시할 경우에는 $N:q$ 비율을 참고할 수도 있다. 추정하는 모수의 개

수(q)에 대비하여 최소 표본의 크기(N)를 고려하는 방식이며, Jackson(2003)은 이상적인 $N:q$ 비율을 20:1, 즉 한 개 모수를 추정할 때 필요한 표본은 20이라고 제안하였다. Kline(2011)은 최소 비율로 10:1을 언급하였으며, 한 모수 당 최소 10배 이상의 표본이 갖추어지지 않으면 표본크기가 작을수록 분석 결과에 대한 명확성이 줄어들 것임을 이야기하였다. 추가적으로, 구조방정식모형의 분석을 위해서는 200 이상의 표본이 필요하다는 선행 연구가 있으며(Barrett, 2007; Cattell, 1978), 다요인 모형에 대해서는 200, 500, 그리고 1,000의 표본크기를 작은 크기, 중간 크기, 그리고 큰 크기의 표본으로 고려할 수 있으니(Li, 2016), 연구자들은 이를 참고하여 자료수집과 분석을 계획하는 것이 필요하다.

실증 자료를 이용한 분석은 표집 오차(sampling error)를 마주하기 마련이며, 요인분석 또한 표집 오차의 영향을 받을 수 있다. 즉, 현재의 결과는 다른 표본을 이용하여 재분석한 결과와는 동일하지 않을 수 있기 때문에 반복 검증을 통한 교차 타당도의 확보가 중요하다(Flora & Flake, 2017). 만약 같은 표본을 이용하여 반복적으로 검증하는 경우에는 동일한 결과를 도출할 가능성이 높아지므로 교차 검증의 의미가 퇴색할 수 있다. 많은 연구자들이 표집의 어려움 때문에 동일하게 표집한 표본을 무작위로 나누어 요인분석 결과를 반복적으로 확인하기도 한다. 그러나 이와 같은 자료 이용법에 대한 명확한 이론적 근거는 찾아볼 수 없으며 연구자들의 경험에 근거한 것으로 보인다. 동일 검사에 대한 요인분석일지라도 다른 표본에 기반한 재분석은 언제나 필요한 과정이며(Cham, Hughes, West, & Im, 2014), 번안검사의 타당화가 중요한 이유가 여

기에 있다.

결과보고

앞서서 요인분석 모형의 설정, 평가, 그리고 수정과 관련된 구체적인 정보를 보고해야함을 명시한 바 있다. 마찬가지로 최종 요인구조를 도출한 후에는, 그 결과를 상세히 보고하고 해석하여 후속 연구자들이 검사를 사용하고 검사 타당도를 재검토할 수 있도록 해야 한다. 이를 위해 연구자는 (1) 최종 요인구조, (2) 모형합치도, (3) 그리고 최종 검사 문항을 보고하도록 한다. 최종 요인구조에는 요인부하량과 요인 간 상관관이 포함되며 문항-요인 간의 관계와 요인-요인 간의 관계를 살펴볼 수 있도록 하기 위함이다. 또한 최종 모형과 관계된 합치도 지수를 보고함으로써 모형 적합성에 대한 근거를 제공하고, 수정지수를 이용한 경우에는 해당 지수와 관계된 통계적/이론적 근거 또한 보고하도록 한다. 마지막으로, 최종 검사를 이루는 문항을 보고하여 다른 연구자들도 타당한 검사를 이용할 수 있도록 돕고, 후속 연구가 타당화 작업을 지속할 수 있도록 독려할 필요가 있다.

변안 타당화에서의 요인분석

변안 타당화의 맥락에서도 요인분석 모형은 언어와 문화에 따른 검사 구조의 동일성을 확인하려는 방법으로 활용된다. 교육 및 심리검사의 표준(AERA, APA, & NCME, 2014)은 원검사의 축소판 검사를 개발하거나 원검사를 변안, 혹은 수정하는 경우에 요인구조를 재평가할 것을 강조하였는데, 이는 모집단이 변경되거나 원검사에 변경이 이루어지는 모든 검사

에 대하여 요인구조를 재검증하는 것이 필요함을 시사하는 것으로 볼 수 있다. ITC 지침의 경우에도 변안 검사의 동일성 확인을 위하여 구체적으로 탐색적 요인분석, 확인적 요인분석 등을 이용할 것을 제안한 바 있다. 최근에는 검사 동일성을 확보하기 위한 측정 동일성(measurement invariance)의 성립이 중요해지고 있으며(Chen et al. 2020; Flora et al., 2017), 이 또한 요인분석을 통해 검증할 수 있는 심리측정적 속성이다. 요인분석 모형에 기반한 측정 동일성에 대해서는 김미림(2023), 주영신과 장승민(2023)의 연구를 참고하길 바란다.

앞서 설명한 이론적 배경을 바탕으로, 다음 장에서는 한국심리학회 산하 학술지에서 출간된 변안검사 타당화 연구에 대해서 살펴보도록 한다. 본 연구는 첫째, 출간된 변안검사 연구 결과를 바탕으로 변안검사 연구자가 검사의 심리측정적 속성(신뢰도와 타당도)을 어떻게 다루고 있는지 그 현황을 알아보았고 둘째, 타당화 작업에 쓰인 요인분석의 시행과 보고를 살펴봄으로써 심리검사 변안 연구에서의 요인분석 관행을 알아보았다.

방 법

연구 대상 및 자료수집

한국심리학회(<https://www.koreanpsychology.or.kr>)에 출판되는 연구 논문은 심리학회 산하의 17가지 하위분야에 따라 각기 다른 세부 주제를 다루고 있다. 본 연구는 전반적인 한국 심리검사 변안연구의 관행을 살펴보고자 하였으므로, 모든 하위분야의 연구를 대상으로 체계적 검토를 실시하였다. 구체적으로, 2017년부터

2023년까지 한국 심리학회지에 게재된 연구 논문 중, 척도 변안 및 타당화를 목적으로 하는 연구 논문을 검토의 대상으로 포함하였다. 검토 대상 연구의 색출을 위하여 해당하는 연도의 심리학회 산하 학회지 문헌들을 모두 검토하였다. 구체적으로, 각 권호에 실린 문헌의 제목과 키워드에 ‘척도’, ‘변안’, ‘번역’, ‘타당화’ 들어간 논문들을 우선적으로 선별했고, 추가로 이에 해당하는 연구의 초록을 검토하여 저자가 척도를 개발한 경우와 직접 변안을 하지 않는 경우를 제외하고, 척도 변안 및 타당화를 연구 주제로 한 논문들만 색출하였다. 그 결과 표 2에 보고한 것과 같이 한국 심리학회지의 하위 13개의 학술지로부터 총 107편의 논문을 체계적 검토의 대상 연구로 선정하였다.

표 2. 논문 분류

| 논문 하위 영역 | N(Proportion) |
|---------------------------------------|---------------|
| 건강 | 21(23%) |
| 문화 및 사회문제 | 12(11%) |
| 발달 | 5(5%) |
| 법 | 1(1%) |
| 사회 및 성격 | 4(4%) |
| 산업 및 조직 | 9(8%) |
| 상담 및 심리치료 | 18(17%) |
| 여성 | 4(4%) |
| 일반 | 3(3%) |
| 임상심리 연구와 실제 | 11(10%) |
| 중독 | 1(1%) |
| 학교 | 5(5%) |
| Korean Journal of Clinical Psychology | 13(12%) |
| 합계 | 107(100%) |

자료 코딩과 분석 방법

본 연구는 선정된 107개의 논문을 검사 정보, 검사의 심리측정적 속성, 그리고 연구에서 실시한 요인분석을 검토하고 분석하였다. 검사 정보로는 변안검사와 원검사와의 차이 여부와 최종 변안검사 문항의 보고 여부를 확인하였고, 심리측정적 속성으로는 각 연구의 신뢰도와 타당도 정보를 정리함으로써 107개 변안검사의 신뢰도 분포와 연구가 근거한 타당도의 종류를 살펴보았다.

또한 요인분석 모형이 변안검사 연구에서 빈번하게 사용되고 있음을 확인하고, 탐색적 요인분석과 확인적 요인분석으로 나누어 변안 타당화를 위한 요인분석의 시행과 결과에 포함된 정보를 분석하였다. 먼저 각 연구에서의 표본크기와 문항 대비 표본크기의 비율($N:p$)을 알아봄으로써 요인분석에 사용한 표본이 충분하였는지 살펴보고, 모형 적절성 판단과 최종 모형의 선택 및 수정을 위해 근거한 준거를 검토하였다. 또한 최종 모형에 기반하여 요인 부하량과 요인 간 상관 정보를 고지하였는지를 포함하여 변안검사 타당도에 대한 충분한 정보를 제공하였는지도 확인하였다. 체계적 검토를 통해 산출한 모든 통계량은 전체 107개 연구에 기반하였으나, 각 연구에서 다수의 값을 보고했을 경우(예: 하위 요인에 대한 신뢰도 분석을 추가로 실시하여 보고), 해당 연구 내의 평균값을 사용하여 전체 통계량 분석에 활용하였다. 이와 같은 체계적 검토 전략에 기반하여 분석한 결과는 다음 장에 보고하였다.

결 과

검사 정보

107개의 변안검사 중 총 75편의 연구가 원 검사와 최종 변안검사의 요인구조나 문항이 다르다고 보고하였으며, 이 중 14편의 연구는 변안의 유연성을 위해(예: 국내 정서에는 적합하지 않은 문항: $N = 5$) 또는 통계적 기준을 충족하기 위해(예: 문항이 정규성 기준을 충족하지 못함: $N = 11$) 요인분석 이전에 삭제한 문항이 있다고 밝혔다. 또한 총 95편의 연구가 최종 변안검사의 문항을 보고했으나, 12편의 연구는 보고하지 않은 것으로 나타났다.

검사도구의 신뢰도 및 타당도

표 3은 검토 대상 연구가 다룬 심리측정적 속성을 나타낸다. 107개의 변안 검사 중 총 105편의 연구가 검사 문항의 내적 일치도를 나타내는 Cronbach's α 를 신뢰도로 보고하였으며, 105편의 연구에 기반한 변안검사의 신뢰도 평균은 .87이었다(최대값=.98, 최소값=.55). 신뢰도의 계산은 각 연구가 보고한 전체 검사의 α 에 기반하였으며, 각 하위 척도의 α 만을 보고 한 경우에는 하위 척도의 평균값을 전체 검사의 α 로 고려하였다. 더불어, 총 28편의 연구가 검사-재검사 신뢰도를, 6편의 연구는 반분 신뢰도를 확인하였다. 이 외에도 6편의 연구에서 ω 계수를 보고하였고 해당 신뢰도 평균은 .88이었다(최대값=.97, 최소값=.77).

변안 검사는 구성개념에 대한 측정을 주목적으로 하므로 구성 타당도의 확인이 필수적이다. 107편의 모든 연구가 구성 타당도를 확인하였다고 보고하였으며, 대다수의 연구가 요인분석을 기반으로 문항과 요인 간 관계를 살펴봄으로써 내적 접근에 기반하여 타당도를

표 3. 검사도구의 신뢰도 및 타당도 보고 현황

| 심리측정 속성 | N |
|----------------------|-----|
| 신뢰도 | 107 |
| Chronbach's α | 105 |
| 검사-재검사 신뢰도 | 28 |
| 반분 신뢰도 | 6 |
| ω | 6 |
| 구성 타당도 | 107 |
| 수렴 타당도 | 83 |
| 변별 타당도 | 63 |
| 준거 타당도 | 63 |
| 공존 타당도 | 28 |
| 예언 타당도 | 6 |
| 충분 타당도 | 18 |
| 내용 타당도(문화적 타당도) | 7 |

평가한 것으로 보인다. 107편의 연구 중 추가로 외적 접근에 기반한 구성 타당도를 확인한 연구도 있었으며, 수렴 타당도와 변별 타당도를 확인한 연구는 각각 83편과 63편이었다. 그 밖에, 준거 타당도를 살펴본 연구가 63편이었으며, 특별히 준거 타당도의 일종인 공존 타당도와 예측 타당도를 조사했다고 명시한 연구는 각각 28편과 6편이었다. 또한 충분 타당도와 내용 타당도를 확인한 연구는 각각 18편과 7편으로 나타났다. 이와 같은 결과는 연구자의 주관적인 분류에 기반한 것이 아니라 각 변안연구에서 언급하고 정의한 타당도에 근거하여 검토하고 분석한 결과임을 밝혀둔다.

요인분석의 시행과 표본크기

대부분의 변안 및 타당화 연구가 구성 타당

도의 확인을 위해 탐색적 요인분석과 확인적 요인분석을 사용하였다고 밝혔다. 탐색적 요인분석과 확인적 요인분석을 모두 실시한 연구의 경우 70편이며 반면 한 가지 요인분석만 실시한 연구는 34편(탐색적 요인분석 9편, 확인적 요인분석 25편)이었다. 각 요인분석 모형에 따라 살펴보면 총 79편의 연구가 탐색적 요인분석을, 95편의 연구가 확인적 요인분석을 통해 타당도를 검증하였다고 볼 수 있다. 탐색적 및 확인적 요인분석을 모두 실시하지 않은 연구는 3편이었으며, 상관분석을 통해 구성 타당도를 확인했다고 보고하였다.

각 요인분석에 사용된 표본크기를 표 4에 요약 및 정리하였다. 분석 결과, 평균적으로 탐색적 요인분석과 확인적 요인분석에 사용된 표본크기는 약 283와 407로 확인적 요인분석에 사용된 표본크기가 큰 것으로 나타났다. 표본크기의 최솟값은 탐색적 요인분석의 경우 100, 그리고 확인적 요인분석의 경우 124로 그 크기가 비슷하였으나 최댓값은 각각 846과 2,478로 큰 차이를 보이는 경향이 있었다. 추가로, 본 연구는 문항 개수 대비 표본크기, 즉 $N:p$ 를 확인함으로써 모형의 복잡도 대비 표본크기를 검토하였으며, 평균적으로 탐색적 요

인분석(16.72)보다 확인적 요인분석(30.56)의 $N:p$ 비율이 높았다.

탐색적 요인분석 세부 사항

탐색적 요인분석을 실시한 79편의 연구 결과에 기반하여 모형의 설정, 평가, 그리고 수정과 관계된 사항을 표 5에 정리하였다. 요인 추출법으로는 최대우도법($n = 38$)과 주축분해법($n = 29$)이 가장 많았으며, 주성분 분석과 가장 최소 제곱법을 사용하였다고 보고한 연구가 각각 5편, 3편, 그리고 추출법을 보고하지 않은 연구가 4편 있었다. 회전 방법을 보고한 연구는 79편 중 74편으로 사각회전($n = 65$)을 적용한 연구가 직각회전($n = 7$)을 적용한 연구보다 많았다. 사각회전 중에서는 직접 오블리민($n = 38$), 프로맥스($n = 16$), Geomin($n = 7$) 순서로 빈번히 사용되었으나, 구체적으로 어떤 사각회전을 택하였는지 명시하지 않은 연구도 4편 있었다. 직각회전의 경우 4편의 연구가 베리맥스를 적용하였다고 하였으며, 2편의 연구의 경우 회전법을 중복으로 보고하는 등 사용한 회전법이 명확하지 않기에 기타로 분류하였다. 나머지 5편의 연구는 회전법

표 4. 요인분석에 사용된 표본크기

| | Mean | SD | Minimum | Maximum |
|-----------------------|--------|--------|---------|---------|
| 탐색적 요인분석 (79편) | | | | |
| 표본크기 | 282.88 | 126.92 | 100 | 846 |
| $N:p$ | 16.72 | 15.47 | 1.06 | 112.60 |
| 확인적 요인분석 (95편) | | | | |
| 표본크기 | 406.64 | 349.28 | 124 | 2,478 |
| $N:p$ | 30.56 | 52.22 | 4.63 | 495.60 |

참조. $N:p$ 는 표본크기:문항개수의 비율을 나타냄.

표 5. 탐색적 요인분석 모형의 설정과 평가(N=79)

| 보고 사항 | N |
|----------------------------|----------|
| 요인 추출법 | 총 75편 보고 |
| 최대우도법 | 38 |
| 주축분해법 | 29 |
| 주성분분석 | 5 |
| 가중 최소 제곱법 | 3 |
| 회전 방법 | 총 74편 보고 |
| 사각회전 | 65 |
| 직접 오블리민(Direct Oblimin) | 38 |
| 프로맥스(Promax) | 16 |
| Geomin | 7 |
| 직각회전 | 7 |
| 베리맥스(Varimax) | 5 |
| 기타 | 2 |
| 모형의 평가 및 요인개수 결정 준거 | |
| 고유치 | 61 |
| 스크리 도표 | 57 |
| 해석 가능성 | 31 |
| 설명량 | 23 |
| 평행검사 | 21 |
| 모형 합치도 | 13 |
| 요인부하량 | 5 |
| 요인 내 최소 문항 수 | 2 |
| 모형의 수정 및 문항 선택 준거 | |
| 요인부하량 | 62 |
| 요인부하량 $\geq .30$ | 18 |
| 요인부하량 $\geq .40$ | 32 |
| 요인부하량 $\geq .50$ | 6 |
| 요인부하량 $\geq .60$ | 1 |
| 기타 (구체적인 준거 미보고) | 3 |
| 교차부하 | 38 |
| 공통분 | 13 |
| 이론적 근거 | 9 |
| 문항 변별도(총점과 문항 간의 상관) | 7 |
| 신뢰도 증감 | 5 |
| 모형합치도 | 2 |

을 보고하지 않았다.

각 번안검사 연구는 다양한 준거를 이용하여 모형을 평가하고 요인개수를 결정하였으며, 평균적으로 약 3개(평균 = 2.56)의 준거를 사용하는 것으로 나타났다. 표 5를 살펴보면 고유치($n = 61$), 스크리 도표($n = 57$), 해석 가능성($n = 31$), 설명량($n = 23$), 평행검사($n = 21$), 모형합치도($n = 13$), 요인부하량($n = 5$), 그리고 요인 내 최소 문항 수($n = 2$) 순으로 빈번하게 고려하여 요인개수를 결정하였음을 알 수 있다. 하나의 준거만을 살펴보고 모형을 결정한 연구는 11편이었다.

모형의 수정 및 문항 선택을 위한 준거로는 요인부하량($n = 62$)을 가장 많이 참고한 것으로 나타났으며, 부하량 0.40 이상을 보이는 문항을 타당한 문항으로 판단한 연구가 가장 많았다($n = 32$). 이외에도 문항의 교차부하 여부($n = 38$), 공통분산의 크기($n = 13$), 이론적 근거($n = 9$), 그리고 문항 변별도($n = 7$), 문항의 포함에 따른 신뢰도의 증감 여부($n = 5$), 모형합치도($n = 2$)에 근거하여 적합하지 않은 문항을 삭제하고 최종 모형을 추정하였음을 알 수 있었다.

확인적 요인분석 모형의 설정과 평가

확인적 요인분석을 실시한 연구는 총 95편이었다. 각 연구는 구조방정식모형의 모형합치도에 기반하여 모형을 평가하였으며, 95개 번안검사의 최종 요인구조에 대한 모형합치도지수의 정보를 요약하면 표 6과 같다. 선행 연구에서 제시한 준거(RMSEA, SRMR < 0.08 ; CFI, TLI $> .95$)에 적합하지 않은 모형합치도 값을 보이는 모형도 있었는데, 다수의 합치도 지수를 모두를 고려한 연구도 있는 반면 그중

표 6. 확인적 요인분석 세부 사항(N=95)

| | N | Mean | SD | Minimum | Maximum |
|--------------------------|----|------|-----|---------|---------|
| 모형의 평가 | | | | | |
| RMSEA | 94 | .07 | .02 | .01 | .19 |
| CFI | 95 | .92 | .05 | .56 | .99 |
| TLI | 86 | .91 | .06 | .54 | .99 |
| SRMR | 41 | .06 | .01 | .03 | .08 |
| 모형의 수정 및 문항 선택 준거 | | | | | |
| 모형 간 모형합치도 비교 | | | 44 | | |
| 수정지수 검토 | | | 19 | | |
| 요인부하량 | | | 15 | | |
| 이론적 배경 | | | 13 | | |
| ω 계수 | | | 3 | | |
| 다중상관자승 | | | 2 | | |
| 표준잔차행렬 | | | 2 | | |
| 요인 간 상관계수 | | | 1 | | |

일부에만 기반하여 모형의 적절성을 판단한 경우도 있었기 때문이다.

탐색적 요인분석과 마찬가지로 확인적 요인 분석도 모형의 수정과 문항 선택을 위한 판단이 필요하다. 검토 결과, 44편의 연구가 모형 간 모형합치도 비교를 통해 모형의 수정 여부를 선택하였고, 수정지수를 이용한 연구가 19편이었다. 또한 최종 문항의 선택과 관련해서는 이론적 배경과 요인부하량을 고려한 연구가 각각 15, 13편, ω 계수, 즉 신뢰도를 고려한 연구가 3편이었다. 이 외에도 다중상관자승과 표준잔차행렬을 토대로 문항을 선택했다고 보고한 연구는 각각 2편씩 있었는데, 요인부하량의 크기에 따라 증감하는 공통분산과 고유분산을 고려한 것으로 요인부하량을 살펴보는 것과 같은 맥락이라고 볼 수 있다. 그밖에

모형 수정을 위해 요인 간 상관을 고려한 연구는 1편이었다.

최종 요인구조의 보고 요인부하량과 요인 간 상관계수 보고

연구자는 최종 요인분석 모형에 기반하여 측정모형의 정보를 보고할 필요가 있으며, 문항-요인 간의 관계를 나타내는 요인부하량과 요인 간 상관계수(다요인 모형의 경우)를 보고한다. 본 연구는 검토 대상 연구가 보고한 최종 요인분석 모형의 정보를 살펴보고 그 결과를 표 7에 정리하였다.

먼저 요인부하량 정보를 살펴보면, 탐색적 요인분석을 사용한 사례($n=79$) 중 78편의 연구가 요인부하량을 표나 그림을 이용하여 보

표 7. 측정모형의 보고

| 요인분석 | N |
|---|----|
| 탐색적 요인분석(79편) | |
| 요인부하량 보고 | 78 |
| 요인 간 상관 보고 (1요인:11편, 다요인:68편) | |
| 상관계수와 유의성 | 19 |
| 상관계수만 보고 | 2 |
| 모두 보고하지 않은 논문 | 47 |
| 확인적 요인분석(95편) | |
| 요인부하량 보고 | 75 |
| 요인 간 상관 정보 (1요인:9편, 다요인:82편 고차요인:4편) | |
| 상관계수와 유의성 | 50 |
| 상관계수만 보고 | 17 |
| 모두 보고하지 않은 논문 | 15 |
| 기타 | 7 |

고하였고, 1편의 연구는 관련 정보를 보고하지 않았다. 확인적 요인분석($n=95$)의 경우에는 75편의 연구가 요인부하량을 보고하였으나 20편의 연구는 보고하지 않았다. 요인 간 상관의 보고 여부는 단일요인 사례(탐색적 요인분석: 11건, 확인적 요인분석: 9건)를 제외하고 다요인을 분석한 사례에 대해서만 그 현황을 살펴보았다. 68개의 탐색적 요인분석 모형에 대하여, 요인 간 상관계수와 그 유의성을 모두 보고한 연구는 19편, 상관계수만 보고한 연구는 2편이었으며, 모두 보고하지 않은 연구는 47편으로 가장 많이 나타났다. 82개의 다요인 확인적 요인분석 모형에 대해서는 요인 상관계수와 그 유의성을 모두 보고한 연구가 50편으로 가장 많았고, 상관계수만 보고한 연구는 17편, 그리고 모두 보고하지 않은 연

구는 15편이었다. 요인 간 상관계수를 보고했으나, 어떤 요인분석에서 기반한 결과인지 명시하지 않은 경우는 기타로 분류하였고 7편의 논문이 이에 해당하였다.

검토 대상 중, 김광태 등(2023)의 연구는 앞서 살펴본 사항들을 전반적으로 다루고 다양한 근거에 기반하여 번안검사를 타당화한 연구로 볼 수 있다. 번안, 신뢰도 및 타당도 검증을 위한 절차를 자세히 기술하고 있으며 의사 결정을 위해 근거한 증거와 연구 결과에 대해서도 구체적으로 보고하고 있다. 특별히 증거를 복합적으로 검토하여 의사를 결정하고, 이에 대해 구체적으로 보고하고 있다는 면을 주목할 필요가 있다.

논의 및 제언

본 연구는 국내 심리검사 번안의 검수와 보고에 대해 제언하는 것을 목표로 2017년부터 2023년까지의 한국 심리검사 번안 타당화 연구를 대상으로 체계적 검토를 실시하고 타당화 현황을 살펴보았다. 먼저 교육 및 심리검사의 표준과 ITC 검수 지침에 기반하여, 번안검사의 심리측정 속성에 대한 정보가 충분히 보고되고 있는지 또한 각 연구가 타당도 확보를 위해 적합한 분석을 실시했는지 조사하였다. 특별히 번안검사 타당화 검수 작업에서 사용된 요인분석의 시행과 결과보고에 대해 면밀히 살펴보았다.

요약하자면, 검토 대상 중 두 편을 제외한 모든 번안검사 타당화 연구가 신뢰도를 보고하였고 측정의 타당성을 근거하기 위해 구성 타당도를 검증하였음을 알 수 있었다. 외부검사와의 관계성에 기반한 구성 타당도, 즉 수

럼 타당도와 변별 타당도를 추가로 검증한 연구는 각각 83, 63편이었다. 준거변수와의 관계에 기반하여 준거 타당도를 고려한 연구는 63편으로, 이 중 28편과 6편의 연구가 구체적으로 각각 공존 타당도와 예언 타당도를 살펴본 있음을 언급했다. 검사 번안에 대한 비교 문화 심리학 분야의 관심이 높아지고 있는데 (Gregoire, 2018), 이는 검사를 문화적 배경에 더욱 적합한 검사로 각색하고 활용해야 하기 때문이다. 이에 따라 7편의 연구가 문항내용의 타당성과 문화적 적합성에 근거하여 내용 타당도를 참작하였음을 직접 보고하였다. 좀 더 세부적인 검토 결과에 기반한 논의와 제언은 신뢰도, 타당도, 그리고 요인분석 순으로 아래에 언급하고자 한다.

대다수의 연구가 Cronbach's α 를 신뢰도로 보고한 것과 관련하여, 본 연구는 α 계수에 대한 지나친 의존에서 벗어날 필요가 있다는 점을 언급하고자 한다. α 계수는 다수 문항이 한 검사를 구성할 때 산출하는 지수로, 전체 검사점수의 분산에 대비한 문항 간 공분산 평균²⁾의 크기에 기반하여 산출한다(Geldhof, Preacher, & Zyphur, 2014). 이는 문항 간 내적 일관도를 나타내는 지수로 전체 분산에서 공통요인에 기반한 문항 간 공분산이 차지하는 비율로 해석할 수 있으므로 신뢰도를 의미한다. 그러나 엄밀히 말한다면 α 계수는 타우동형(tau-equivalent) 검사 조건을 만족하지 않는 검사에 대해 제약이 있다. 타우동형 검사 조건을 충족하는 검사는 모든 문항의 표준화 요인부하량이 거의 같은 크기여야 하지만 이를

만족하는 검사는 현실에서 찾아보기 어렵기 때문이다. 또한 타우동형 검사 조건을 만족하지 않는 경우에 α 계수는 과소 추정되기 때문에(McNeish, 2018; Raykov & Marcoulides, 2011) 선행연구는 α 계수의 대안으로 McDonald (1999)의 ω 계수를 언급한다. 본 연구가 검토한 번안연구 중에도 과소추정 문제로 ω 계수를 보고한 연구를 살펴볼 수 있었다(정서영, 박희웅, 손우영, 2023). ω 계열의 신뢰도는 타우동형 검사보다 제약이 적은 동일구조(congeneric) 검사에 기반하고 있으며 모든 문항의 요인부하량이 동일하지 않아도 산출할 수 있다(이순목, 조은경, 1998; McNeish, 1995). 그러나 아직까지 α 계수의 보고가 주를 이루고 있는 것으로 보이며, 한국심리학회 산하 학술지에서도 ω 계수를 보고한 연구는 적은 편인 것으로 나타났다. 과거에는 소프트웨어의 접근성을 이유로 α 계수의 보고가 주를 이룬 것으로 보이거나 현재는 R, jamovi와 같은 소프트웨어에서 쉽게 산출할 수 있으므로 연구자들은 ω 계수의 보고를 활발히 할 필요가 있다. 본 연구는 ω 계수를 소개하는 것에서 글을 마치지만, ω 계열의 신뢰도를 자세히 알고 싶은 연구자는 신재은과 이태현(2017), 그리고 McNeish(1995)의 연구를 참고하기 바란다.

본 연구는 앞서 교육 및 심리검사의 표준과 미국심리학회의 정의를 따라 구성 타당도와 준거 타당도를 중점적으로 설명하였다. 그러나 증분 타당도를 검증하고 보고한 연구의 수(18편)가 적지 않은 것을 확인하고 후속 연구자들을 위해 설명을 보충하고자 한다. 증분 타당도는 예측력을 논의하는 평가 현장에서 활용되고 있으며(Clark & Watson, 2019; Hunsley & Meyer, 2003), 인사 혹은 임상 현장에서 근거하는 타당도의 종류로 볼 수 있다. Hunsley

2) $\alpha = (p^2 \overline{\sigma_{ij}}) / \sigma_X^2$, p = 검사 문항 수; $\overline{\sigma_{ij}}$ = 문항 i 와 j 간의 공분산 평균; σ_X = 척도 점수의 분산

와 Meyer(2003)는 증분 타당도가 구성 타당도의 추가 근거로 활용되고 있음을 언급한 바 있다. 예를 들어, 기존 풀 배터리 검사(full battery test)에 새로운 소검사나 문항을 추가하는 경우, 개정된 검사가 기존 검사보다 더 많은 정보를 제공해야 증분된 타당도를 갖췄다고 볼 수 있다.

다음으로 요인분석 모형의 선택에 관하여 논의하고자 한다. 번안검사는 선행 타당화 작업을 거친 원검사를 바탕으로 하지만, 번안에 따른 검사의 개정이 이루어지므로 별도의 타당화 작업이 필요하다(AERA, APA, NCME, 2014; Gregoire, 2018). 원검사와 번안검사의 심리측정적 속성은 다를 가능성이 높아서 번안 검사에 대한 추가 검증이 필수적이기 때문이다. 연구자는 언어적/문화적 차이에 의해 번안 검사의 요인구조가 원검사와 다를 수도 있다는 점에 유의하고, 이를 반영한 검수 방법을 택해야 한다. 실제로 본 연구가 검토했던 107편의 번안검사 타당화 연구 중 75편의 연구가 원검사와 번안검사의 요인구조 간에 차이가 있음을 보고하였다. 이러한 맥락에서 볼 때, 연구자는 번안검사가 원검사와 다른 요인구조를 가질 수 있음을 가정하고 타당화 작업을 시행하는 것이 적합한 것으로 보인다. 총 107편의 번안연구 중 탐색적 요인분석을 활용하여 번안검사의 요인구조를 새롭게 탐색하고 확인적 요인분석을 통해 이를 교차 타당화한 연구는 70편이었다. 34편의 연구는 두 가지 요인분석 중 한 가지만 이용하여 요인구조 모형을 결정하였으나, 원검사와 다른 요인구조를 예상하는 경우에는 타당도에 대한 더욱 충분한 근거를 확보하기 위하여 탐색 및 확인적 요인분석을 모두 실시하는 것이 적절할 수도 있다.

체계적 검토를 통하여 요인분석이 번안검사 타당화 연구에서 거의 필수적으로 사용하고 있는 방법임을 다시 한번 확인할 수 있었다. 그러나 국내 번안 타당화 연구 중 일부는 요인분석과 관계된 사항을 충분히 보고하지 않고 있었기에, 응용 연구자들에게 도움이 되고자 몇 가지 제언을 하고자 한다. 첫째, 검사 타당화에 요인분석을 사용하는 연구자는 안정적인 추정을 위해 충분한 크기의 표본을 확보할 필요가 있다. 검토 결과를 살펴보면, 탐색적 요인분석의 경우 $N:p$ 의 최소값은 1.06로 선행연구에서 제시한 문항 수 대비 5배의 표본 크기에 한참 미치지 못하였으며, 최소 표본크기는 100으로 선행연구가 제시한 최소 표본크기(300)보다 매우 작았다. 확인적 요인분석은 좀 더 충분한 표본크기에 기반했으나 역시 주의할 필요가 있는 것으로 보인다. $N:p$ 의 최소값이 4.63로 이는 Kline이 제시한 10보다 작은 비율이었으며, 최소 표본크기는 124로 일반적으로 구조방정식 모형에서 고려하는 작은 표본크기(200)보다 작았다. 연구자들은 연구설계 단계에서 문항/모수의 수를 고려하여 충분한 표본크기를 계획해야 하며, 이를 확보하기 위해 노력해야 함을 다시 한 번 강조한다.

둘째, 요인분석 이후에는 최종 요인구조를 보고함으로써 모형의 타당성을 뒷받침하고 다른 연구자들의 이해를 도울 필요가 있으나, 이에 대한 보고가 충분하지 않은 경우가 있음을 확인할 수 있었다. 요인부하량은 문항에 대한 요인의 공통분산을 확인할 수 있는 정보이며 요인 간 상관계수는 추출된 요인이 이론에 적합한지의 여부, 또는 요인의 추출이 과다하지는 않은지의 여부(예: 요인 간 상관이 너무 큰 경우)를 판단할 수 있는 정보다. 그러나 모든 연구가 이에 해당하는 결과를 보고하

지는 않았으며 특히, 요인 간 상관의 경우 탐색적 요인분석은 19편, 그리고 확인적 요인분석의 경우 50편의 연구만 상수계수와 그 유의성을 보고하였다. 변안검사의 타당성을 주장하고 후속 연구자들에게 검사 사용을 권장하기 위해서는 필요한 정보를 충분히 보고하는 연구자들의 노력이 필요하다.

셋째, 탐색적 요인분석의 요인구조의 회전과 요인개수의 결정 과정에 대해서 추가적으로 제안하고자 한다. 본 연구에서 검토한 결과 탐색적 요인분석을 실시한 79건의 사례 중 6건이 직각회전을 실시하였으나 그 이유에 대한 설명이 충분하지 않은 경우가 대부분이었다. 그러나 사회과학 연구에서는 요인 간 상관의 0이라는 가설이 적합하지 않은 경우가 많으므로(이순목, 1994), 요인 간 상관을 0으로 가정하고 직각회전을 실시한 것에 대해 설명이 필요하다. 추가로, 요인개수에 대한 결정을 단일 준거에만 기반하는 것은 지양할 필요가 있음을 강조하고자 한다. 본 연구의 검토 결과, 단일 준거에 기반한 연구가 총 11편 있었으며 그 중 카이제 룰에만 기반하여 요인모형을 결정한 사례가 2편임을 확인하였다. 공통적으로 선행연구들은 여러 개의 준거를 함께 고려한 요인개수 결정을 추천하고 있으며, 특히 카이제 룰은 단일 준거로의 사용이 적합하지 않다는 것이 반복적으로 확인되었으므로(Brown, 2015) 연구자들의 주의가 필요하다.

넷째, 준거값에 가깝지 않은 합치도지수에 기반하여 확인적 요인분석 모형을 평가한 사례들이 있어 이에 대해 언급하고자 한다. 최종 모형의 CFI와 TLI의 최솟값이 각각 .56과 .54로 굉장히 낮은 연구 사례와 RMSEA의 최댓값이 .19로 비교적 높은 사례가 있었다. 모형의 적합성은 다수의 모형 합치도지수에 기

반하여 판단하므로, 연구자는 일부 합치도지수가 적합하지 않아도 적합한 나머지에 기반하여 전반적인 평가를 실시했을 수도 있다. 그러나 CFI와 TLI는 대상 모형이 가장 제약이 많은 모형과 비교했을 때 상대적으로 얼마나 나은지를 나타내는 상대적 합치도 지수이며, RMSEA와 SRMR은 자료와 모형 간에 합치하는 정도를 살피는 절대적 합치도 지수로서 모형의 서로 다른 면모를 평가한다. 따라서 연구자는 전반적 모형 평가를 위해서 절대적 합치도 지수와 상대적 합치도지수 모두 적정 준거 점수를 만족하고 있는지 살펴볼 필요가 있음을 강조하고자 한다. 다양한 모형 합치도 지수를 종합적으로 판단하는 방법은 김수영(2017), 홍세희(2000), 그리고 Kline(2013) 연구를 통해 살펴볼 수 있다.

본 연구는 변안검사 연구가 주로 요인분석을 사용하고 있는 것을 확인함과 동시에 타당도 확보를 위한 통계모형이 전통적인 요인분석에만 머물러 있다는 점을 알 수 있었다. ITC 지침 또한 이미 집단 간 동일한 변안검사의 타당화에 대해서 강조하고 측정 동일성의 확보를 위한 검증법을 자세히 기술한 바 있으나 국내 107편의 연구 중에 측정 동일성을 검증한 연구는 단 12편이었다. 더욱 다양한 각도에서의 타당도 검증을 실시하고 이에 기반하여 풍부한 근거를 확보하는 국내 연구자들의 노력이 필요한 것으로 보인다.

마지막으로 심리학 연구에서 사용하는 학술 용어의 통합을 제언한다. 한국심리학회에서는 영문 용어에 해당하는 한글 용어를 제공함으로써 연구자들의 공통된 용어 사용을 지향하고 있는 것으로 보이지만, 검토 결과 같은 개념에 대해 서로 다른 용어를 사용하는 경우가 빈번한 것을 확인하였다. 예를 들어, 심리학

용어사전(<https://www.koreanpsychology.or.kr/psychology/term.html>)에는 구성 타당도로 등재된 용어를 구인 타당도로, 공존 타당도로 등재된 용어를 공인 타당도로 보고하는 사례들을 들 수 있다. 또한 사전에 등록되지 않았으나 이미 널리 쓰이고 있는 개념의 경우에는 연구자끼리 서로 다른 용어를 빈번하게 사용했다. 그 예로는 orthogonal rotation에 대한 직각회전과 직교회전, content validity에 대한 내용 타당도와 문화적 타당도를 들 수 있다. 이에 따라 본 연구에서 체계적 검토를 실시하는 과정에서도 용어에 해당하는 원개념을 추적하기 위해 추가로 시간을 소비해야만 했는데, 다른 연구자들도 동일한 어려움을 겪을 수 있을 것으로 예상된다. 따라서 이와 같은 혼동을 피하기 위해서라도 한국심리학회 산하의 학술지에서 출판하는 연구논문은 이미 등재된 용어를 사용할 필요가 있으며, 새로운 용어는 계속해서 사전에 등재하는 등의 노력이 필요해 보인다.

마무리하며, 본 연구는 검사 번역/번안을 위한 ITC 지침에 기반하여 국내 번안 심리검사 타당화 연구의 현황을 검토하고 번안검사 타당화 연구가 고려해야 할 사항에 대해 제언함으로써 연구자들에게 도움이 되고자 하였다. 그러나 새로운 검사의 타당화 과정 또한 번안이라는 과정을 제외하고는 번안검사 타당화 연구와 마찬가지로 심리측정적 속성의 검증은 중요시하므로, 본 연구의 검토 결과 및 제언은 번안검사 뿐만 아니라 새로운 검사의 개발 및 타당화를 목표로 하는 연구자들에게도 도움이 될 것으로 기대한다. 체계적 검토 대상에 포함된 우수한 연구들³⁾ 또한 참고문헌에

표기하였으므로, 실제 사례를 살피는데 도움이 될 것으로 본다.

참고문헌

- *강규림, 현명호 (2022). 한국판 실존불안 질문지 타당화 및 관련 변인 탐색. *한국심리학회지: 임상심리 연구와 실제*, 8(4), 683-711.
<https://doi.org/10.15842/CPKJOURNAL.PUB.8.4.683>
- *강병은, 신현숙 (2017). 청소년 자의식 척도의 타당화. *한국심리학회지: 학교*, 14(1), 105-128.
<https://doi.org/10.16983/kjisp.2017.14.1.105>
- *강수경, 김해미, 정미라 (2017). 한국판 태아애착 척도(MAAS)의 타당화 연구. *한국심리학회지: 여성*, 22(2), 89-112.
<https://doi.org/10.18205/kpa.2017.22.2.001>
- 강태훈, 김명연 (2023). 이해하기 쉬운 교육평가. 서울: 박영story.
- *강지훈, 김재웅, 석정호, 구본훈, 류진선, 신현경, 윤석호, 홍혜정, 최현정 (2023). Validation of the Korean version of the Borderline Symptom List Short Version (K-BSL-23). *Korean Journal of Clinical Psychology*, 42(2), 23-34.
<http://doi.org/10.15842/KJCP.PUB.42.2.23>
- *강효신, 김빛나 (2021). 미래시간조망 척도의 탐색적 및 확인적 요인분석과 연령 집단 비교. *한국심리학회지: 발달*, 34(1), 63-80.
<https://doi.org/10.35574/KJDP.2021.3.34.1.63>
- *권두리, 신나나 (2021). 유아기 일상의 실행기 필요한 연구자는 해당 문헌을 확인하기 바람.

3) 참고문헌 목록에 표기된 *표기는 본 연구가 검토한 번안 타당화 연구이므로, 구체적인 예시가

- 능 척도(REEF)의 타당화 연구. *한국심리학회지: 발달*, 34(4), 67-90.
<https://doi.org/10.35574/KJDP.2021.12.34.4.67>
- *권은정, 김울리, 김미리혜, 광경화, 양재원 (2023). 한국판 ICD-11 성격장애 심각도 평가 (PDS-ICD-11)의 타당화. *한국심리학회지: 건강*, 28(1), 205-226.
<https://doi.org/10.17315/kjhp.2023.28.1.0>
- *권혁진, 권석만 (2017). 한국판 자해기능 평가지(The Functional Assessment of Self-Mutilation)의 타당화 연구: 대학생을 중심으로. *한국심리학회지: 임상심리 연구와 실제*, 3(1), 187-205.
- *김경아, 장혜인 (2022). 모호한 상실의 측정: 북한이탈주민 경계보호성 척도(BAS-NK)의 예비 타당화 연구. *한국심리학회지: 일반*, 41(4), 349-386.
<https://doi.org/10.22257/kjp.2022.9.41.4.349>
- *김경훈, 전영민, 이길전 (2018). 한국어판 단도박변화추진활동척도(K-POCS-G)의 타당화 연구. *한국심리학회지: 중독*, 3(1), 1-18.
<https://doi.org/10.23147/ADDICTPSY.PUB.3.1.1>
- *김광태, 이혜원, 손영우 (2023). 다차원적 조용한 사직 척도(MQQS) 타당화 연구. *한국심리학회지: 산업 및 조직*, 36(4), 557-583.
<https://doi.org/10.24230/kjiop.v36i4.557-583>
- *김다혜, 안정광 (2021). 한국판 사회불안 안전 행동 질문지(SBQ) 타당화 연구. *Korean Journal of Clinical Psychology*, 40(3), 280-298.
<https://doi.org/10.15842/kjcp.2021.40.3.009>
- *김도희, 김희정, 정주리 (2022). 한국판 청소년용 플로리시 척도 타당화 연구: EPOCH 모델을 중심으로. *한국심리학회지: 학교*, 19(3), 187-213.
<https://doi.org/10.16983/kjisp.2022.19.3.187>
- 김미림 (2023). 영과잉 자료의 측정동일성 검증: 다집단 확인적 요인분석과 2-부분 요인분석 모형의 적용과 비교. *교육평가연구*, 36(3), 445-472.
<http://dx.doi.org/10.31158/JEEV.2023.36.3.445>
- 김수영 (2016). 구조방정식 모형의 기본과 확장: Mplus 예제와 함께. 학지사.
- *김승민, 박 경 (2017). 한국판 양육 수용행동 척도(Korean Parental Acceptance Questionnaire: K-4-PAQ) 타당화 연구. *한국심리학회지: 건강*, 22(3), 531-549.
<https://doi.org/10.17315/kjhp.2017.22.3.004>
- *김유나, 안정광 (2023). 한국판 겸손 반응 척도(K-MRS) 타당화 연구. *한국심리학회지: 사회 및 성격*, 37(2), 215-235.
<https://doi.org/10.21193/kjspp.2023.37.2.005>
- *김은하, 김도연, 박한솔, 김수용, 김지수 (2017). 한국어판 정당한 세상에 대한 믿음 척도(Belief in a Just World Scale: K-BJWS)의 타당화. *한국심리학회지: 상담 및 심리치료*, 29(3), 689-710.
<https://doi.org/10.23844/KJCP.2017.08.29.3.689>
- *김은하, 김현지 (2020). 한국판 남성에 대한 양가적 태도 척도 타당화 연구. *한국심리학회지: 문화 및 사회문제*, 26(4), 525-549.
<https://doi.org/10.20406/kjcs.2020.11.26.4.525>
- *김준현, 유금란 (2023). 한국판 남성 동성애자의 이성애불편감(Heterophobia) 척도(K-HGM)의 타당화. *한국심리학회지: 문화 및 사회문제*, 29(1), 25-51.
<https://doi.org/10.20406/kjcs.2023.2.29.1.25>
- *김지은, 안현의 (2021). 한국판 조직배반 척도의 타당화: 조직 내 성폭력 피해를 중심으로

- 으로. 한국심리학회지: 여성, 26(2), 99-121.
<https://doi.org/10.18205/KPA.2021.26.1.005>
- *김홍주, 김은영 (2018). 한국판 청소년 성찰기 능척도 타당화. 한국심리학회지 : 상담 및 심리치료, 30(2), 297-316.
<https://doi.org/10.23844/kjcp.2018.05.30.2.297>
- *남기은, 이선희 (2022). 다차원 일중독 척도 (Multidimensional Workaholism Scale, MWS) 타당화 연구. 한국심리학회지: 산업 및 조직, 35(1), 65-87.
<https://doi.org/10.24230/kjiop.v35i1.65-87>
- *남지선, 박형인. (2021). 분리된 관심의 척도 타당화 및 직무열의와의 관계 연구. 한국 심리학회지: 산업 및 조직, 34(4), 629-662.
<https://doi.org/10.24230/kjiop.v34i4.629-662>
- *류지영, 신희천, 김은하 (2020). 한국판 기본 심리적 욕구에 기반한 대인관계 행동 척도 (IBQ)의 타당화. 한국심리학회지: 상담 및 심리치료, 32(3), 1203-1224.
<https://doi.org/10.23844/KJCP.2020.08.32.3.1203>
- *문선현, 엄진섭, 최원일, AllysonBrothers, 노수림 (2023). 한국판 연령 관련 변화에 대한 인식 척도 타당화: 중년 및 노인을 대상으로. 한국심리학회지: 건강, 28(3), 789-915.
<https://doi.org/10.17315/kjhp.2023.28.3.010>
- *문영국, 이종현 (2021). 직무배태성 개념의 확장: 가정배태성 척도 타당화 연구. 한국심리학회지: 산업 및 조직, 34(4), 723-750.
<https://doi.org/10.24230/KJIOP.V34I4.723-750>
- *문찬기, 이세라 (2023). Validating Korean Narcissistic Admiration and Rivalry Questionnaire (NARQ): Its Relations with Big Five, Self-esteem, NPI and Benign and Malicious Envy. Korean Journal of Clinical Psychology, 42(3), 45-57.
<https://doi.org/10.15842/kjcp.2023.42.3.001>
- *박가현, 김시형, 이동훈 (2020). 한국판 정서에 대한 신념 척도의 타당화. 한국심리학회지: 건강, 25(1), 97-114.
<https://doi.org/10.17315/kjhp.2020.25.1.006>
- *박경우, 장혜인 (2021). 한국어판 강박적 성행동 장애 척도(K-CSBD-19)의 타당화 연구. 한국심리학회지: 건강, 26(5), 859-879.
<http://dx.doi.org/10.17315/kjhp.2021.26.5.002>
- *박경우, 유현종, 장혜인, 이상규, 이은지 (2022). 한국어판 음란물 사용동기 척도 (K-PUMS)의 타당화 연구. 한국심리학회지: 건강, 27(5), 763-788.
<https://doi.org/10.17315/kjhp.2022.27.5.003>
- *박기라, 김소은, 양은주 (2018). 한국판 대학생 진로타협 경향성 척도의 타당화 연구. 한국심리학회지: 학교, 15(2), 155-175.
<https://doi.org/10.16983/kjsp.2018.15.2.155>
- *박도담, 유성경 (2019). 한국판 성적 지향 마이크로어그레션 척도 (Korean version of the Sexual Orientation Microaggressions Scale; K-SOMS) 타당화. 한국심리학회지: 상담 및 심리치료, 31(3), 899-927.
<https://doi.org/10.23844/kjcp.2019.08.31.3.899>
- *박선영, 이종은, 이정애, 오강섭 (2023). 한국판 노인불안척도(K-GAS)의 신뢰도 및 타당도 연구. 한국심리학회지: 임상심리 연구와 실제, 9(2), 235-251.
<https://doi.org/10.15842/CPKJOURNAL.PUB.9.2.235>
- *박세란 (2022). 따뜻함과 안전 초기기억 척도 타당화. 한국심리학회지 : 상담 및 심리치료, 34(2), 413-431.

- <https://doi.org/10.23844/kjcp.2022.05.34.2.413>
*박세란 (2022). 자비적 참여행동 척도 타당화 연구. 한국심리학회지: 건강, 27(3), 563-587.
<https://doi.org/10.17315/kjhp.2022.27.3.007>
- *박용욱, 설정훈, 최진수, 이혜주, 손영우 (2022). 다차원적 일 의미감 척도(CMWS) 타당화 연구. 한국심리학회지: 산업 및 조직, 35(2), 213-245.
<https://doi.org/10.24230/kjiop.v35i2.213-245>
- *박정민, 안현의 (2022). 한국판 과거 연애 관계 사고 척도(Positive and Negative Ex-Relationship Thoughts Scale) 타당화. 한국심리학회지: 문화 및 사회문제, 28(4), 627-659.
<https://doi.org/10.20406/kjcs.2022.11.28.4.627>
- *박주하, 양수진 (2022). 미취학 자녀 양육모를 대상으로 하는 한국판 코로나19 스트레스 척도 타당화. 한국심리학회지: 발달, 35(3), 1-19.
<https://doi.org/10.35574/KJDP.2022.9.35.3.1>
- *박홍석, 이정미 (2018). 한국판 상황적 자기인식척도(K-SSAS)의 타당화 연구. 한국심리학회지: 학교, 15(2), 177-196.
<https://doi.org/10.16983/kjsp.2018.15.2.177>
- *배라영, 최지영 (2018). 한국판 관계적 공격성 척도의 타당화 연구: 대학생을 대상으로. 한국심리학회지: 상담 및 심리치료, 30(1), 55-79.
<https://doi.org/10.23844/kjcp.2018.02.30.1.55>
- *서원진, 이수민, 김미리혜, 김율리, 김경희, Chad Ebesutani, 김다미, 황보인, 도현정, 박유진 (2018). 한국판 체중 걱정 척도 (Weight Concern Scale)의 타당화. 한국심리학회지: 건강, 23(4), 925-938.
<https://doi.org/10.17315/kjhp.2018.23.4.006>
- *서자경, 이기학 (2017). 한국판 다면적 성실성 척도(K-CCS)의 타당화 연구. 한국심리학회지: 사회 및 성격, 31(4), 51-78.
<https://doi.org/10.21193/kjspp.2017.31.4.003>
- *서장원 (2021). 개정판 테이트 폭력 질문지의 타당화 연구. 한국심리학회지: 건강, 26(1), 109-123.
<https://doi.org/10.17315/kjhp.2021.26.1.007>
- *서중환 (2022). 면담기반 사이코패시 성격 종합평가-기관평가척도(CAPP-IRS) 도구 타당화 연구. 한국심리학회지: 사회 및 성격, 36(1), 25-46.
<https://doi.org/10.21193/kjspp.2022.36.1.002>
- *손옥선, 김진숙 (2021). 한국판 매우 민감한 사람 척도(K-HSPS-18)의 재타당화. 한국심리학회지: 상담 및 심리치료, 33(3), 1049-1075.
<https://doi.org/10.23844/kjcp.2021.08.33.3.1049>
- *손은솔, 송현주 (2022). 한국형 부모의 포괄적 식사양육 실행척도(K-CFPQ)의 타당화 연구. 한국심리학회지: 발달, 35(4), 71-92.
<https://doi.org/10.35574/KDJP.2022.12.35.4.71>
- *송연주, 하문선. (2020). 한국 단축형 사랑중독 척도(Korean-Love Addiction Questionnaire-Short Form: K-LAQ-SF) 타당화 연구. 한국심리학회지: 문화 및 사회문제, 26(4), 501-524.
<https://doi.org/10.20406/kjcs.2020.11.26.4.501>
- *송원영 (2021). 한국판 성적 이미지 기반 학대 통념 수용 척도(K-SIAMA)의 타당화. 한국심리학회지: 여성, 26(3), 185-200.
<https://doi.org/10.18205/kpa.2021.26.1.009>
- *신문혜, 이지연 (2020). 한국판 긍정도식척도 (YPSQ)의 타당화. 한국심리학회지: 상담

- 및 심리치료, 32(3), 1125-1151.
<https://doi.org/10.23844/kjcp.2020.08.32.3.1125>
- *신성만, 윤지혜, 조요한, 고은정, 박명준 (2018). 예일음식중독척도 2.0(Yale Food Addiction Scale 2.0) 국내 타당화 연구. 한국심리학회지: 여성, 23(1), 25-49.
<https://doi.org/10.18205/KPA.2018.23.1.002>
- 신재은, 이태현 (2017). 쌍요인(Bifactor) 모형을 이용한 심리척도의 측정적 속성 연구방법 개관. 한국심리학회지: 일반, 36(4), 477-504.
<https://doi.org/10.18205/kpa.2018.23.1.002>
- 신진아, 시기자, 성태제 (2021). 검사제작과 분석. 서울: 학지사.
- *안명희, 정유선 (2023). 한국판 정신화된 정서성 척도(K-MAS)의 타당화 연구. 한국심리학회지: 상담 및 심리치료, 35(2), 413-440.
<https://doi.org/10.23844/kjcp.2023.05.35.2.413>
- *안선미, 현영권 (2023). 대학생을 대상으로 한 비판적 의식 척도 타당화. 한국심리학회지: 문화 및 사회문제, 29(4), 595-616.
<https://doi.org/10.20406/kjcs.2023.11.29.4.595>
- *안선영, 황순택 (2019). 한국판 DSM-5 성격질문지 정보제공자보고형(K-PID-5-IRF)의 신뢰도와 타당도. Korean Journal of Clinical Psychology, 38(1), 82-101.
<https://doi.org/10.15842/kjcp.2019.38.1.007>
- *안재경, 최이문 (2020). 한국판 자유의지와 결정론 척도(Free will and Determinism Plus: FAD+)의 타당화 연구. 한국심리학회지: 법, 11(2), 191-210.
<https://doi.org/10.53302/KJFP.2020.07.11.2.191>
- *양나연, 이수정 (2018). 한글판 의미 만들기 척도(K-MMS)의 타당화. 한국심리학회지: 상담 및 심리치료, 30(1), 81-100.
<https://doi.org/10.23844/kjcp.2018.02.30.1.81>
- *양진원, 권석만 (2021). 한국판 보상적 섭식 욕구 척도의 타당화. 한국심리학회지: 건강, 26(6), 985-1003.
<https://doi.org/10.17315/kjhp.2021.26.6.002>
- *양진원, 권석만 (2022). 한국판 음식기대척도(AEFS)의 타당화. 한국심리학회지: 임상심리 연구와 실제, 8(2), 249-271.
<https://doi.org/10.15842/CPKJOURNAL.PUB.8.2.249>
- *오주용, 안도연. (2023). 한국어판 죽음태도 척도 개정판(DAP-R)의 타당화. 한국심리학회지: 임상심리 연구와 실제, 9(2), 345-371.
<https://doi.org/10.15842/CPKJOURNAL.PUB.9.2.345>
- *유지혜, 설경옥 (2018). 한국판 물질주의척도의 타당화 연구. 한국심리학회지: 문화 및 사회문제, 24(3), 385-410.
<https://doi.org/10.20406/kjcs.2018.8.24.3.385>
- *윤정미, 김진영 (2019). 한국판 청소년용 지각된 스트레스 척도의 타당화 연구. 한국심리학회지: 건강, 24(3), 569-586.
<https://doi.org/10.17315/kjhp.2019.24.3.003>
- *이덕희, 남슬기, 이동훈 (2021). 초월적 시간관 척도(TFTPS) 타당화. 한국심리학회지: 건강, 26(6), 961-983.
<https://doi.org/10.17315/kjhp.2021.26.6.001>
- *이덕희, 김성현, 정다송, 이동훈 (2023). 자살 사고 속성 척도(Suicidal Ideation Attributes Scale; SIDAS) 타당화 연구. 한국심리학회지: 문화 및 사회문제, 29(1), 1-23.
<https://doi.org/10.20406/kjcs.2023.2.29.1.1>
- *이동현, 김향숙 (2021). 고통 과잉 감내 척도의 타당화 연구: 한국 대학생 집단을 대

- 상으로. *Korean Journal of Clinical Psychology*, 40(2), 143-155.
<https://doi.org/10.15842/kjcp.2021.40.2.003>
- *이동훈, 엄희준, 이덕희 (2022). 트라우마와 사별 경험에 대한 개인의 의미통합 척도 (K-ISLES): 한국판 타당화를 위한 탐색적 연구. *한국심리학회지: 상담 및 심리치료*, 34(3), 719-744.
<https://doi.org/10.23844/kjcp.2022.08.34.3.719>
- *이동훈, 엄희준, 이덕희 (2022). 한국판 트라우마 분노반응척도-5(DAR-5-K)의 중단 타당화 연구*. *한국심리학회지: 일반*, 41(2), 133-161.
<https://doi.org/10.22257/kjp.2022.6.41.2.133>
- *이보라, Eunjoo Kim (2019). 한국판 일유인가 척도 타당화 연구. *한국심리학회지: 상담 및 심리치료*, 31(1), 283-302.
<https://doi.org/10.23844/kjcp.2019.02.31.1.283>
- *이선영, 안현의 (2020). 한국판 가족돌봄의무 척도(Filial Responsibility Scale-Adult)의 타당화. *한국심리학회지: 문화 및 사회문제*, 26(3), 259-282.
<https://doi.org/10.20406/kjcs.2020.8.26.3.259>
- *이세라, 신현균 (2018). 한국판 모멸감 척도 (K-HI)의 타당화 연구. *Korean Journal of Clinical Psychology*, 37(1), 119-129.
<https://doi.org/10.15842/kjcp.2018.37.1.010>
- *이소정, 김은하 (2022). 한국판 동성애자/양성애자 다차원 정체성 척도 타당화. *한국심리학회지: 문화 및 사회문제*, 28(2), 133-161.
<https://doi.org/10.20406/kjcs.2022.5.28.2.133>
- 이순목 (1994). 요인분석의 관행과 문제점. *한국심리학회지: 산업 및 조직*, 7(1), 1-27.
- 이순목 (1995). SPSS를 사용한 공통요인분석의 문제점. *교육평가연구*, 8(1), 5-33.
- 이순목, 조은경 (1998). 고전검사이론의 가정 검증: 불필요한가 또는 불가능한가. *교육평가연구*, 11(2), 98-109.
- *이순행, 이희연, 정미라 (2018). 한국판 성인 놀이성 척도(K-APTS) 타당화 연구. *한국심리학회지: 건강*, 23(2), 397-425.
<https://doi.org/10.17315/kjhp.2018.23.2.006>
- *이승민, 이혜진 (2021). 한국판 신체자비 척도 (BCS)의 타당화. *한국심리학회지: 건강*, 26(5), 835-858.
<https://doi.org/10.17315/kjhp.2021.26.5.001>
- *이은진, 최보윤, 한주옥 (2022). 한국판 다차원적 문화적 겸손성 척도 타당화 연구. *한국심리학회지: 상담 및 심리치료*, 34(2), 461-491.
<https://doi.org/10.23844/kjcp.2022.05.34.2.461>
- *이재은, 정보영 (2021). 한국판 칼레이도스코프 경력태도 측정도구 타당화 연구. *한국심리학회지: 산업 및 조직*, 34(1), 51-79.
<https://doi.org/10.24230/kjiop.v34i1.51-79>
- *이종환, 임종민, 장문선 (2018). 주체통각검사를 활용한 사회인지와 대상관계 척도 타당화 연구. *Korean Journal of Clinical Psychology*, 37(4), 540-557.
<https://doi.org/10.15842/kjcp.2018.37.4.007>
- *이주원, 유정아, 송원영 (2022). 한국판 단축형 어둠의 성격 4요소(SD4-K) 척도의 타당화 연구. *한국심리학회지: 건강*, 27(6), 999-1023.
<https://doi.org/10.17315/kjhp.2022.27.6.010>
- *이현지 & 김가림, & 권유리, & 신윤정 (2021). 한국형 트랜스젠더에 대한 태도 및 믿음 척도(K-TABS) 타당화 연구. *한국심리학회지: 상담 및 심리치료*, 33(3), 1077-1108,

- <https://doi.org/10.23844/kjcp.2021.08.33.3.1077>
*임선영 (2023). 외상 후 성장 척도 확장판 (PTGI-X)의 심리측정적 속성: 한국판 척도의 요인구조와 유용성 재검토. 한국심리학회지: 임상심리 연구와 실제, 9(1), 161-181.
<https://doi.org/10.15842/CPKJOURNAL.PUB.9.1.161>
- *임소희, 황순택, 권혜수, 김지혜, 박은영, 박중규, 이수정, 이은호, 홍상환 (2018). PAI-A(Personality Assessment Inventory for Adolescent) 재표준화 연구. 한국심리학회지: 임상심리 연구와 실제, 4(3), 435-454.
<https://doi.org/10.15842/CPKJOURNAL.PUB.4.3.435>
- *임형민, 이슬아, 권석만 (2021). 한국판 미래심상과제의 타당화. Korean Journal of Clinical Psychology, 40(1), 103-113.
<https://doi.org/10.15842/kjcp.2021.40.1.008>
- *임혜선, 김정윤, 홍혜영 (2023). 한국판 공격자 동일시 척도(K-IAS)의 타당화. 한국심리학회지: 상담 및 심리치료, 35(3), 997-1022.
<https://doi.org/10.23844/kjcp.2023.08.35.3.997>
- *장은영 (2021). 한국판 도덕손상 사건 척도 및 도덕손상 경험 척도 개발 연구. 한국심리학회지: 일반, 40(3), 301-327.
<https://doi.org/10.22257/kjp.2021.9.40.3.301>
- *장지윤, 안현의 (2018). 한국판 동성애자/양성애자 긍정적 정체성 척도 타당화. 한국심리학회지: 상담 및 심리치료, 30(2), 273-295.
<https://doi.org/10.23844/kjcp.2018.05.30.2.273>
- *전영민, 이정임, 박은경 (2018). 치료효과 평가도구로서의 한국판 도박회복추진척도 (GFS-K) 타당화. Korean Journal of Clinical Psychology, 37(2), 225-235.
<https://doi.org/10.15842/kjcp.2018.37.2.009>
- *정경미, 김수연 (2017). 한국형 아동 섭식행동질문지(K-CFQ)의 타당화 연구. 한국심리학회지: 건강, 22(2), 317-338.
<https://doi.org/10.15842/10.17315/kjhp.2017.22.2.006>
- *정경미, 양윤정, 정승민, 이경숙, 박진아 (2019). 한국판 부모 양육스트레스 검사 4판 단축형 (K-PSI-4-SF)의 표준화 연구. 한국심리학회지: 건강, 24(4), 785-807.
<https://doi.org/10.17315/kjhp.2019.24.4.001>
- *정다송 & 이보라, & 이덕희, & 이동훈 (2023). 사별 경험 이후 사회적 의미 만들기 척도 타당화 연구. 한국심리학회지: 상담 및 심리치료, 35(4), 1361-1397.
<https://doi.org/10.23844/kjcp.2023.11.35.4.1361>
- *정서영, 박희웅, 손우영 (2023). 한국판 세대친화적 조직문화척도(K-WICS) 타당화 연구. 한국심리학회지: 문화 및 사회문제, 29(4), 429-453.
<https://doi.org/10.20406/kjcs.2023.11.29.4.429>
- *정수인, 안현의 (2019). 한국판 성인용 놀이성 척도의 타당화. 한국심리학회지: 문화 및 사회문제, 25(4), 353-375.
<https://doi.org/10.20406/kjcs.2019.11.25.4.353>
- *정주리. (2021). 한국판 제로섬 신념 척도 타당화 연구. 한국심리학회지: 문화 및 사회문제, 27(3), 285-303.
<https://doi.org/10.20406/kjcs.2021.8.27.3.285>
- *조성훈, 권정혜 (2017). 한국판 인터넷 게임장에 척도의 타당화. Korean Journal of Clinical Psychology, 36(1), 104-117.
<https://doi.org/10.15842/KJCP.2017.36.1.010>

- *조수현 (2020). 한국판 직무소진평가척도 (K-BAT) 타당화를 위한 예비연구. *한국심리학회지: 산업 및 조직*, 33(4), 461-499.
<https://doi.org/10.24230/KJIOP.V33I4.461-499>
- *조은호, 최지은, 한지수, 허윤성, 김현경 (2023). 한국형 미디어 리터러시 척도 (K-NMLS)와 한국형 미디어 리터러시 단축형 척도(K-NMLS-SF)의 타당화. *한국심리학회지: 발달*, 36(4), 45-64.
<https://doi.org/10.35574/KJDP.2023.12.36.4.45>
- *조혜현, 현명호 (2023). 한국판 뒤셀도르프 건강음식집착 척도(K-DOS)의 타당화. *한국심리학회지: 건강*, 28(2), 561-579.
<https://doi.org/10.17315/kjhp.2023.28.2.015>
- *조호진, 이풍가, 마 흥, 옥지수 (2022). 한국어판 단축형 어둠의 3요소 척도(K-Dirty Dozen)의 타당도와 요인구조 연구. *한국심리학회지: 산업 및 조직*, 35(2), 299-326,
<https://doi.org/10.24230/kjiop.v35i2.299-326>
- 주영신, 장승민 (2023). 정렬법을 이용한 범주형 자료의 근사 측정동일성 분석. *한국심리학회지: 일반*, 42(2), 119-140.
<http://dx.doi.org/10.22257/kjp.2023.6.42.2.119>
- *주현영, 김나래 (2023). 한국판 실수 반추 척도(K-MRS)의 타당화 연구: 대학생을 대상으로. *한국심리학회지: 상담 및 심리치료*, 35(3), 971-994.
<https://doi.org/10.23844/kjcp.2023.08.35.3.971>
- *지은혜, 조용래, 김선영 (2022). 코로나 19 스트레스 척도(COVID Stress Scales, CSS)의 타당화: 한국 성인표본을 대상으로. *한국심리학회지: 임상심리 연구와 실제*, 8(4), 659-681,
<https://doi.org/10.15842/CPKJOURNAL.PUB.8.4.659>
- *차윤지, 이은호, 황순택, 홍상황, 김지혜. (2020). 한국어판 개정된 아동발현불안척도 2판의 심리측정적 속성에 대한 연구. *Korean Journal of Clinical Psychology*, 39(3), 203-214.
<https://doi.org/10.15842/KJCP.2020.39.3.001>
- *최영환, 이은호, 황순택, 홍상황, 김지혜. (2020). 한국어판 백자살사고척도의 신뢰도 및 타당도 연구: 일반 성인 집단을 대상으로. *Korean Journal of Clinical Psychology*, 39(2), 111-123.
<https://doi.org/10.15842/KJCP.PUB.39.2.111>
- *최재광, 한지현, 김민범, 송원영 (2023). 한국판 노모포비아 척도 타당화. *한국심리학회지: 건강*, 28(2), 581-600, <https://doi.org/10.17315/kjhp.2023.28.2.016>
- *최정금, 김명소 (2018). 한국판 직무 번영감 척도의 타당화 연구. *한국심리학회지: 산업 및 조직*, 31(3), 715-739.
<https://doi.org/10.24230/KSIOP.31.3.201808.715>
- *최지선, 이정미 (2018). 한국판 균형적 시간조망척도(K-BTPS)의 타당화 연구. *한국심리학회지: 건강*, 23(3), 701-720.
<https://doi.org/10.17315/kjhp.2018.23.3.006>
- *최환규, 이정미 (2017). 한국판 일의 의미 척도(K-WAMI)의 타당화 연구. *한국심리학회지: 사회 및 성격*, 31(4), 1-25.
<https://doi.org/10.21193/kjspp.2017.31.4.001>
- *표소휘, 양은주 (2020). 한국판 학생 진로구성 척도(Korean Student Career Construction Inventory)의 타당화 연구: 초기 성인기를 중심으로. *한국심리학회지: 학교*, 17(2), 145-164.
<https://doi.org/10.16983/kjsp.2020.17.2.145>
- *표지은, 안정광 (2021). 한국판 사후처리 특성 및 상태 척도(PEPI) 타당화. *한국심리학회*

- 지: *임상심리 연구와 실제*, 7(3), 283-308.
<https://doi.org/10.15842/CPKJOURNAL.PUB.7.3.283>
- *허민희, 황순택, 박광배, 유성은, 김지혜, 박중규, 이은호, 홍상황. (2021). 한국판 밀론 다축 임상성격검사 4판의 신뢰도와 타당도. *Korean Journal of Clinical Psychology*, 40(1), 91-102.
<https://doi.org/10.15842/kjcp.2021.40.1.007>
- *허주연, 황성훈 (2023). 한국판 Perth 정서 반응성 척도의 타당화. *한국심리학회지: 임상심리 연구와 실제*, 9(3), 425-453.
<https://doi.org/10.15842/CPKJOURNAL.PUB.9.3.425>
- *현혜민, 박기환 (2018). 한국판 저장 척도 개정판의 타당화 연구. *한국심리학회지: 건강*, 23(3), 721-738.
<https://doi.org/10.17315/kjhp.2018.23.3.007>
- 홍세희 (2000). 특별기고: 구조 방정식 모형의 적합도 지수 선정기준과 그 근거. *한국심리학회지: 임상*, 19(1), 161-177.
- *홍혜정, 박중규 (2019). 야식증후군 진단질문지(NEDQ)의 타당화 연구. *한국심리학회지: 임상심리 연구와 실제*, 5(1), 65-90.
<https://doi.org/10.15842/CPKJOURNAL.PUB.5.1.65>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, 42(5), 815-824.
<https://doi.org/10.1016/j.paid.2006.09.018>
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186-3191.
<https://doi.org/10.1097/00007632-200012150-00014>
- Brislin, R. W. (1970). Back-translation for crosscultural research. *Journal of Cross-Cultural Psychology* 1(3), 185-216.
<https://doi.org/10.1177/135910457000100301>
- Browne, M. W. (1968). A comparison of factor analytic techniques. *Psychometrika*, 33, 267-334. <https://doi.org/10.1007/BF02289327>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*(2nd ed.). The Guilford Press.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research*, 1(2), 245-276.
https://doi.org/10.1207/s15327906mbr0102_10
- Cattell, R. B. (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. New York: Plenum.
- Chen, M. Y., Liu, Y., & Zumbo, B. D. (2020). A Propensity Score Method for Investigating Differential Item Functioning in Performance Assessment. *Educ Psychol Meas*, 80(3), 476-498.
<https://doi.org/10.1177/0013164419878861>
- Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychol Assess*, 31(12), 1412-1427.
<https://doi.org/10.1037/pas0000626>
- De Winter, J. C., & Dodou, D. (2012). Factor

- recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of applied statistics*, 39(4), 695-710. <https://doi.org/10.1080/02664763.2011.610445>
- Everitt, B. S. (1975). Multivariate analysis: The need for data, and other problems. *British Journal of Psychiatry*, 126(3), 237-240. <https://doi.org/10.1192/bjp.126.3.237>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods*, 4(3), 272. <https://doi.org/10.1037/1082-989X.4.3.272>
- Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. *Canadian Journal of Behavioural Science-Revue Canadienne Des Sciences Du Comportement*, 49(2), 78-88. <https://doi.org/10.1037/cbs0000069>
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19, 72-91. <https://doi.org/10.1037/a0032138>
- Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate behavioral research*, 25(1), 33-39. https://doi.org/10.1207/s15327906mbr2501_3
- Gregoire, J. (2018). 검사 번역/변안을 위한 국제 지침: 한국어판 2017년 2판 [ITC guidelines for translating and adapting tests (2nd ed.)] (서동기와 이순목 역). *International Journal of Testing*, 18(2), 101-134. <https://doi.org/10.1080/15305058.2017.1398166>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185. <https://doi.org/10.1007/BF02289447>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N: q hypothesis. *Structural equation modeling*, 10(1), 128-141. https://doi.org/10.1207/S15328007SEM1001_6
- Jenkins, M., & Griffith, R. (2004). Using Personality Constructs to Predict Performance: Narrow or Broad Bandwidth. *Journal of Business and Psychology*, 19(20), 255-269. <https://doi.org/10.1007/s10869-004-0551-9>
- Kline, R. B. (2013). Exploratory and confirmatory factor analysis. In Y. Petscher & C. Schatsschneider (Eds.), *Applied quantitative analysis in the social sciences* (pp. 171-207). New York, NY: Routledge.
- Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936-949. <https://doi.org/10.3758/s13428-015-0619->
- Loevinger, J. (1957). Objective tests as instruments

- of psychological theory. *Psychological reports*, 3(3), 635-694.
<https://doi.org/10.2466/PRO.3.7.635-694>
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84-99.
<https://doi.org/10.1037/1082-989X.4.1.84>
- McDonald, R. P. (1999). Test theory: A unified approach. Mahwah, NJ: Erlbaum.
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychol Methods*, 23(3), 412-433. <https://doi.org/10.1037/met0000144>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
<https://doi.org/10.1002/j.2333-8504.1994.tb01618.x>
- *Moon, C., & Lee, S. (2023). Validating Korean Narcissistic Admiration and Rivalry Questionnaire (NARQ): Its Relations with Big Five, Self-esteem, NPI and Benign and Malicious Envy. *Korean Journal of Clinical Psychology*, 42(3), 45-57.
<https://doi.org/10.15842/kjcp.2023.42.3.00>
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior research methods, instruments, & computers*, 32(3), 396-402.
<https://doi.org/10.3758/BF03200807>
- Raykov, T., & Marcoulides, G. A. (2011). Introduction to psychometric theory. Routledge.
- Thurstone, L. L. (1947). Multiple factor analysis. Chicago: University of Chicago Pres.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327.
<https://doi.org/10.1007/BF02293557>
- Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in Health*, 8(2), 94-104.
<https://doi.org/10.1111/j.1524-4733.2005.04054.x>
- Whittaker, T. (2012). Using the Modification Index and Standardized Expected Parameter Change for Model Modification. *Journal of Experimental Education*, 80(1), 26-44.
<https://doi.org/10.1080/00220973.2010.531299>

1차원고접수 : 2024. 05. 27

최종게재결정 : 2024. 07. 30

A Systematic Review of Translating and Adapting Psychological Test: Practices and Recommendations

Mirim Kim

Yeji Im

Korea University

This current study conducted a systematic review of translating and adapting psychological tests based on the Standards for Educational and Psychological Testing and the ITC Guidelines. We reviewed a total of 107 KAPA articles published from 2017 to 2023 and examined validation and reporting practices. The current study examined whether sufficient information about the psychometric properties of the tests was reported and whether each study conducted adequate analyses to ensure validity. Specifically, we reviewed the implementation and reporting of factor analysis for validation. In short, we found that most studies reported reliability and used factor analysis for testing construct validity; therefore, we suggested recommendations in this regard. We also indicated that academic terminologies in psychology research should be unified as the studies often used different terms for the same concept. Although the study results are based on the translating and adapting tests, the findings and recommendations will be also useful for development and validation new tests.

Key words : translating and adapting tests, confirmation guidelines, systematic review, psychometric properties, factor analysis