

NEO-PI-3 변안검사의 구성개념 타당도

최 상 민¹⁾ 정 재 칭²⁾ 서 동 기^{1)†}

¹⁾한림대학교 심리학과, ²⁾PSI 컨설팅

본 연구는 성격 이론 중 가장 널리 사용되고 있는 성격 5요인 이론에 기반한 NEO PI-R의 개정판인 NEO-PI-3 검사의 한국판 타당화를 목적으로 하였다. 이를 위해 검사를 번역하고 수정하여 검사의 구조가 이론적 구조와 같은지 검증하는 절차를 거쳤다. 연구 과정에서는 원검사를 번역하고 예비검사, 수정검사, 최종검사의 세 단계를 거쳐 검사의 이론적 구조를 검증하였다. 예비검사는 직장인 및 대학생 1,934명을 대상으로 실시하였으며, 탐색적 요인분석을 통해 요인구조를 탐색하고 고전검사이론 및 문항반응이론에 기반하여 하위 요인의 문항들이 적절한지 문항분석을 통해 문항을 수정 보완하였다. 수정검사에서는 1,280명의 응답을 이용하여 확인적 요인분석을 실시한 후 추가적인 문항 수정을 진행하였다. 최종검사에서는 345명의 응답을 이용하여 탐색적 구조방정식을 적용하고 목표회전 요인분석을 실시하여 검사 구조가 이론적인 성격 5요인 구조로 구성되어 있음을 확인하였다. 더 나아가 성별에 따른 측정 단위 동일성을 검증하였으며, 성별에 따른 각 점수의 차이가 존재하는지 확인하였다. 최종검사의 검사-재검사를 통해 상위 5개 요인은 0.9이상, 하위 30개 요인은 “가치”요인을 제외하고 0.7이상의 적절한 신뢰도 지수를 확인하였다. 마지막으로 본 연구의 의의와 제한점을 논의 부분에 자세히 제시하였다.

주요어 : 성격 5요인, NEO-PI-3, 변안 타당화, 요인분석

† 교신저자: 한림대학교 심리학과, 부교수 / 한림응용심리연구소, 강원도 춘천시 한림대학길 1

Tel: 033-248-1727, E-mail: wmtmote@hallym.ac.kr

 © 2025, Korean Social and Personality Psychological Association.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://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.

성격은 사람을 이해하는 아주 중요한 특질 지표이다. 서로의 성격을 이해하지 못하면 부모와 자녀, 교사와 학생, 관리자와 직원은 서로 다른으로 인하여 갈등이 발생할 수 있다. 반대로 자기 혹은 타인의 특성을 잘 이해하면 집단에 소속된 사람들 각자가 편안하고 안정되어 서로에 대한 존중감과 소속감을 가질 수 있을 뿐 아니라 효과적이고 효율적인 업무 수행도 가능하다(민병모, 이경임, 정재창, 2002). 따라서 사람의 성격을 측정하는 것은 우리 사회에서 여러 도움을 제공하기 때문에 많은 성격검사들이 개발되고 있다.

성격 혹은 인성은 인간이 환경에 적응하고 살아가는 과정에서 시간과 상황에 걸쳐 지속적이고 일관되게 나타나는 개인을 구별하는 특징적인 사고, 정서, 행동 등의 독특한 행동 유형이며(Phares, 1984), 한 사람의 행동에 잠재적인 영향을 미치는 심리적 특성으로, 다른 사람과 구별되게 하는 독특한 성향이다(Ozer, Benet-Martinez, 2006; Randolph & Dahling, 2013). 또한 성격은 환경에 대한 사람의 고유한 적응을 결정하는 개인 내의 정신적 신체적 체계의 역동적 조직화라고 불리며(Allport, 1937), 개인의 타고난 기본적인 성향과 환경에 대한 적응이 개인에게 발현되어 적응적인 측면과 함께 부적응적인 측면도 같이 내포할 수 있다(Judge & LePine, 2007; Smith, Hill, Wallace, Recendes, & Judge, 2018). 따라서 성격을 이해하고 이를 연구하기 위해서는 개인이 타고난 기본적인 성향과 더불어 환경과 상호작용하여 형성된 성격의 적응을 이해해야 한다(Costa & McCrae, 2017; Costa & McCrae, 1992a; 최정우, 한여울, 박두진, 최기홍, 서동기, 2021).

성격 5요인

성격 특질을 밝혀내고 이해하기 위해서 지금 까지 많은 연구가들은 이론적 접근, 어휘적 접근, 심리측정학적 접근이라는 세 가지 접근 방

법을 조합해 왔다(Miserandino, 2011). Allport는 인간 행동의 내적, 심리적인 근저에는 일관되고 안정적이며, 개인차를 잘 나타내는 특성인 심리적 구조가 있다고 가정하였다. 그리고 성격 특성 구조에 대한 구성 요소를 밝히고자 어휘적 접근을 이용해 영어 대사전에서 성격을 기술하는 용어의 목록을 작성하였다(Allport & Odbert, 1936). 이후 Allport의 연구는 심리측정학적 접근으로 확장되어 요인 분석을 실시하여 16개의 요인으로 묶이게 되었다(Cattell, 1943). 이러한 연구를 이어받아 Tupes와 Christal(1961)은 많은 사람들에게 일관적이고 반복되게 나타나는 5개의 성격 요인이 존재한다는 것을 밝혀냈다. 성격 5요인설은 이후 연구에서도 재확인이 되어(Norman, 1963), 최종적으로 5요인(Big Five; Goldberg, 1981) 혹은 5요인 모형(Five Factor Model; Costa & McCrae, 1992a)이라 명명하였다. 성격 5요인 이론은 이후에 많은 연구자들이 개인차를 설명하는 성격구조로 인정하게 되었다 (Costa & McCrae & Dye 1991; Digman, 1996). 이러한 5가지 요인은 일반적으로 심리적 민감성(N: Neuroticism)¹⁾, 외향성(E: Extraversion), 개방성(O: Openness to Experience, Culture, Intellect), 우호성(A: Agreeableness), 성실성(C: Conscientiousness, Will to Achieve)을 일컫고 있으며, 여러 문화권에서 (Benet-Martinez & John, 1998) 공통으로 발견되고 있다.

NEO PI의 역사

성격/인성검사는 대개 두 가지 방식으로 개발되고 있다. 첫째는 특정 성격 이론을 기초로, 그 중심적인 구성개념을 측정하는 것이며, 두 번째는 직업적인 성공이나 정신병리와 같은 변인을 예언하기 위하여 개발되는 방식이다(민병모 외,

1) 다수 연구에서는 신경증적 경향성으로 불리지만, 채용, 선발 등 여러 장면에서 사용 할 때에 해당 요인의 이름으로 인해 부정적인 선입견을 제공할 수 있기에, 심리적 민감성으로 표현하였다.

2002). 성격 5요인을 측정하기 위해 제작된 몇 가지의 검사 중 NEO PI(NEO Personality Inventory; McCrae & Costa, 1985)는 이중 가장 대표적인 초기 성격 5요인 검사이다.

NEO PI-R은 개발 과정에서 여러 성격 검사들을 종단적으로 연구 후 결합 요인분석(joint factor analysis)하여 개발되었으며, 기존 NEO PI를 개정한 검사다(Costa & McCrae, 1992b). NEO PI-R은 5 가지 요인이 각각 6가지의 하위 요인으로 구성되어 최종적으로 30개의 하위요인으로 세분화되었다. 이러한 특징으로 인해 NEO PI-R은 개인의 인성을 기술해 주는 주요 측면들을 모두 포함하는 다차원적인 인성검사로 개발되었다. NEO PI-R의 상위 5개 요인은 개인의 중요한 성격 특징을 빠르게 파악하는 데에 도움을 줄 수 있으며, 하위 요인은 중요한 특성을 상세하게 분석할 수 있게 도와준다. 즉 성격에 대한 포괄적인 이해와 동시에 특별한 상황에서 독특한 개인의 행동을 예측하는 데에 도움을 준다(McAdams, 1992; 김동일, 홍성두, 인효연, 2006).

NEO PI가 1989년 NEO PI-R로 개정된 이후 2010년에 두 번째 개정(NEO-PI-3)이 실시되었다. 이 개정을 위해 McCrae와 동료들은 1단계로 대학생 1,959명을 대상으로 기존의 NEO PI-R 검사에 응답하는데 ‘이해가 잘 안되는 문항이 있다면 응답을 하지말고 비워두라고 하여 전체 응답자의 2% 이상이 ‘이해가 잘 안 된다’고 응답한 30개 문항을 폐하였다. 청소년과 문해력이 부족한 사람들이 이해하기 어려운 형용사들이 문제의 원인으로 파악되었다. 이 30개 문항 외에 성인과 고등학생 표본집단 간의 상관이 낮은 24개 문항을 합한 54개 문항 중 중복되는 6개 문항을 제외한 48개 문항이 수정 후보 문항으로 선정되었다(McCrae, Costa, et al., 2005). 이 문항들을 대체할 수정 문항을 개발 후 청소년 500명을 대상으로 검사를 실시한 결과 48개 문항 중 37개 문항의 심리측정 속성이 현저히 개선되었다. NEO PI-R검사와 이 수정된 37개 문항을 적용한 검사를 성인들을 대상으로 동시에 진단하

였더니 응답자의 2% 이상이 ‘이해하기 어렵다’고 응답한 문항이 수정한 검사에서는 하나도 없었으며 종전의 37개 문항에 비해 새로 개발 된 37개 문항 총점 상관이 .36에서 .45로 증가하였다. 수정한 37개 문항으로 대체한 검사를 NEO-PI-3이라고 명명하고 12세-13세의 중학생들 449명을 대상으로 NEO-PI-3 검사를 실시하여 분석한 결과 내적 일관성, 요인 구조, 수렴 및 판별 타당성 측면에서 본 검사로 활용할 수 있음을 검증하였다. 이처럼 개정판인 NEO-PI-3은 McCrae와 Costa(2010)가 기존 NEO PI-R의 240 문항 중 37개 문항에 대해 문항 자체를 바꾼 것이 아니라 기존 검사에서는 만 17세 이상만 평가할 수 있었던 데에 반해 12세 이상의 청소년과 교육수준이나 문해력이 낮은 성인들이 이해하기 어렵거나 구별하기 어려운 형용사를 이해하기 쉽도록 개정하여 심리측정학적인 특성이 개선되도록 바꾼 것이다. NEO-PI-3은 NEO PI-R의 요인구조와 해석을 기존 그대로 유지하고 있지만 더 넓은 범위의 응답자에게 적용할 수 있게 문항이 구성 되었다(McCrae & Costa, 2010).

NEO PI 국내 타당화 연구

NEO PI-R의 다양한 활용 가능성으로 인하여 1992년 이후부터 NEO PI 검사의 국내 이용을 위한 타당화 작업이 지속되면서 국내 여러 연구자들(안창규, 채준호, 1997; 유태용, 김명언, 이도형, 1997; 유태용, 이도형, 1997; 이경임, 1994; 이경임, 1995a; 이경임, 1995b; 이경임, 1997; 이경임, 안창규, 1996)이 타당화 논문을 발표하였다. 원저자(Robert R. McCrae, Paul T. Costa, Jr)의 승인과 저작권자인 PAR(Psychological Assessment Resources)의 허가를 받아 민병모와 정재창이 NEO PI-R을 1998년에 번역한 후 저작권자에 의한 역번역 검증을 거쳐 타당화를 실시하였고 이러한 국내 연구를 반영하여 NEO PI-R 검사 매뉴얼이 작성되었다(민병모 외, 2002). 이후 2010년에 NEO-PI-3 검사가 개정되자 다양한 연구 분

야와 채용 현장에서 활용 가능성이 확대되었다. 이러한 상황에서 국내에서는 원저자의 승인을 받은 공인된 도구가 아닌 NEO 검사를 표방한 성격 5요인 기반의 성격진단 도구가 사용되고 있다. 한국판 NEO-PI-3 검사가 성인 뿐 아니라 대학생, 고등학생 및 중학생들을 대상으로 사용되기 위해서는 통합적인 타당성 연구나 각 대상별 타당화 과정이 필요하다. 본 연구는 1차적으로 성인을 대상으로 NEO-PI-3의 원검사를 국내 표본을 통해 타당성 검증을 실시하여, 원저자가 측정하고자 한 개념을 보다 정확하게 반영하는 도구로 변안하는 것을 목표로 하였다.

연구의 목적

새로 만들어진 NEO-PI-3을 국내에서 사용하기 위해서는 한국인을 대상으로 타당화하는 과정이 필요하다. 특히 단순히 문항을 번역(translation)하는 것이 아닌 번안(adaptation)하는 과정이 필요하다. 교육 및 심리검사의 적절한 개발과 사용을 위해 구성된 국제 검사 위원회(International Test Commission: ITC)에서는 2017년에 검사의 번역과 번안에 대한 지침을 작성하였다. ITC 4번 지침에서는 해외 검사를 국내에서 이용하기 위한 번안 타당화를 위해서는 단순히 언어적인 의미만이 아니라 영어권과 한국어권의 심리적/문화적인 차이를 고려해야한다고 기술하였으며, ITC 지침 8에서는 국내에서 번안검사를 사용하기 위해 단순히 해외에서 제작한 검사를 번역만 하여 사용하지 말고 국내 문화와 표본을 반영한 검사가 될 수 있도록 타당화 연구를 통해 번안검사의 제작을 제안하였다(ITC, 2017).

따라서 본 연구는 이러한 ITC 국제 지침의 제안 사항을 반영하여 대한민국 표본과 한국문화에 맞는 NEO-PI-3 한국판 검사의 타당화를 실시하였고 구체적인 연구 가설을 요약하자면 다음과 같다.

1. 한국표본에서 NEO-PI-3을 정확하게 사용하기 위해 적절한 번역 설계와 타당화 절차를 사

용하였는가?

2. 문항분석, 신뢰도 평가를 통해 문항들이 적절하게 국내 표본에 반응을 하는가?
3. 국내 표본 집단에 비추어 NEO-PI-3 검사가 측정하는 구성개념의 정의와 내용이 문항 내용에 충분히 반영되었는가?
4. NEO-PI-3 검사가 상위 5개 요인과 하위 30개 요인으로 이론적으로 구조화되었는가?
5. NEO-PI-3 검사가 성별 간에 구성개념 동등성, 측정 동등성 및 문항 동등성이 충족되었는가?

방법

연구 대상

본 연구는 NEO-PI-3 한국판 타당화 연구를 위하여 수집한 자료를 바탕으로 실시하였으며, 이 자료는 기존 자료를 활용하기 위해 H 대학교 생명윤리위원회에서 심사를 받아 승인된 후 진행되었다(HIRB-2023-074). 자료 수집은 NEO-PI-3의 타당화 연구를 위해 전국 각지의 국내 기업 및 공공기관에 근무하는 직장인 및 대학생을 대상으로 한 응답으로 이루어졌다. 자료 수집을 위해 국내 소재 기업과 공공기관을 구분하여 성별, 연령, 직급, 직장 소재지를 나누어 표본을 추출하였다. 본 연구에서는 세 차례의 표집과정을 거쳤는데, NEO-PI-3 원검사를 최초 번역한 첫 검사를 **예비검사**, 예비검사 결과를 바탕으로 문항을 수정한 검사를 **수정검사**, 마지막으로 수정검사의 문항을 분석하여 문항을 재수정한 검사를 **최종검사**로 명명하였다. 예비검사에서는 총 1,934명의 직장인 표본을 수집하였으며, 수정검사 과정에서는 1,280명의 응답을, 최종 검사에서는 직장인 345명의 응답을 수집하여 분석하였다. 예비, 수정, 최종검사 모두 응시자가 온라인에서 진행하였다. 표본을 수집하는 과정에서, 개인 응답자로부터 정보를 수집하기

위해 URL에 접속하여 개인정보 활용에 ‘동의’를 한 응답자에게만 이메일이나 전화번호를 입력하도록 하였다. 기업이나 단체로부터 응답을 수집할 때는 해당 단체에 요청하여 개인정보의 재위탁 제한, 개인정보보호법 준수, 그리고 정보 파기에 대한 확인을 한 뒤 진행하였다. 이를 통해 참여자들의 정보를 단체 내부로 송부하였으며, 외부로의 정보 유출을 방지하기 위한 조치를 취하였다. 예비검사와 수정검사 및 최종검사의 표본 정보는 부록 1에 첨부하였다.

측정 도구

국내에서 널리 이용되고 있는 NEO PI-R검사가 2010년에 원저자에 의해 NEO-PI-3 검사도구 (McCrae & Costa, 2010)로 개정됨에 따라 본 연구는 개정된 NEO-PI-3을 번안 타당화 연구에 사용하였다. NEO-PI-3의 문항 수는 NEO PI-R과 동일한 240개이며 사회적 바람직성과 비전형 척도 16개를 추가하여 총 256개 문항으로 구성되어 있다. 1997년 번역한 후 미세 수정과 매년 규준 데이터의 업데이트를 거쳐 사용해 오던 NEO PI-R 문항을 근거로 NEO-PI-3에서 수정된 37개 문항을 내용 전문가가 번역한 후 저작권자의 역 번역 검증을 거쳐 최종 확정된 문항을 사용하였다. 문항 구성 개념은 NEO PI-R과 같다. 검사는 각 하위요인에 따라 8개의 문항으로 구성되어 있고 각 문항의 척도는 5점 척도로 제작되었다.

연구 절차

NEO-PI-3 발표 후 2016년 원저자의 허락을 받아 국내 적용의 타당성을 확인하기 위한 연구에 착수하여 국내 97개 기업, 21개 공공기관에 근무하는 성인 직장인들을 대상으로 예비검사용 1,934명 정보를 수집하여 분석을 실시하였다. 예비검사의 분석 절차는 아래와 같다. 첫째, 원검사의 분석 절차와 마찬가지로 전체 응답을 각 요인의 하위요인들의 문항총점을 평균점수로 문

항 묶음 처리를 한 이후, 이를 바탕으로 탐색적 요인분석을 실시하였다. 탐색적 요인분석을 진행하기 전에 문항들이 공통요인으로 묶이고 요인분석이 가능한지 알기 위해 KMO와 Bartlett 검정을 실시하였다. 그리고 스크리 도표 및 평행분석을 이용하여 적절한 요인의 개수를 탐색하였다. 결정 된 요인의 개수로 각 요인에 문항들이 적절하게 묶이는지, 요인 부하량이 .3이 상이 되는지 등을 확인하였다. 탐색적 요인분석을 위한 추정방법으로 최대우도법(Maximum likelihood)을 이용하였고, 요인 회전 방법은 Geomin 사각 회전을 이용하였다. 둘째, 상위요인에서 요인부하량이 낮은 문항 묶음 내의 각 문항을 분석하기 위해 고전검사이론과 문항반응이론을 바탕으로 문항을 하나씩 탐색하였다. 고전검사이론 기반의 문항 변별도인 문항-총점 상관의 경우 .30 이상이면 적절하다고 판단하였다 (Gable & Wolf, 1993). 문항반응이론의 경우 단일 차원성 가정으로 인하여 상위 성격 요인에 따라 독립적으로 진행하였으며, 사용한 모형은 부분점수모형(partial credit model; PCM, Masters, 1982)을 이용하여 문항의 난이도 모수를 점검하였다. 또 각 문항의 적합도 지수값인 infit과 outfit의 값이 평정적도의 기준인 0.6과 1.4 사이에 위치하는지 확인하였다(Wright & Linacre, 1994).

문항 분석을 거친 이후 문항이 내용적으로, 통계적으로 적절하지 않은 문항은 내용전문가가 수정하였다. 문항의 수정 과정은 심리측정 전문가와 성격심리 전문가의 논의를 통해 통계적으로 적절하지 않는 문항을 중심으로 국내 문화에 적합한 문항으로 수정하였다. 이후 수정검사는 성인 1,280명의 응답을 바탕으로 교차타당도 검정을 위해 확인적 요인분석을 실시하였다. 모형의 적합도는 카이자승 검증과, 표본 크기에 대한 민감성이 덜한 CFI(Comparative Fit Index), RMSEA(Root Mean Square Error of Approximation)를 이용하였다. 이때 각 수치의 적절성은 Browne과 Cudeck(1992) 및 Hu와 Bentler(1999)가 설정한 기준을 바탕으로 CFI와 TLI는 0.9 이상, RMSEA

와 SRMR은 0.08 이하인 경우에 양호한 모형으로 판단하였다. 이후 다시 측정학적으로 문제가 있는 문항을 선별하여 성격심리학자가 내용적으로 문항을 검토하였고 수정이 필요한 문항은 다시 수정하여 최종검사를 구성하였다. 마지막으로 최종검사를 직장인 345명에게 실시하여 이론적인 5요인 구조를 다시 확인적으로 검증하였다. 최종 요인분석에서 탐색적 요인분석과 확인적 요인분석의 한계점을 보완하는 방법으로 탐색적 구조방정식 모형(Exploratory Structural Equation Model; 이하 ESEM) 방법을 이용하였다. ESEM 방식은 탐색적 요인분석 관점에서 척도 내에서 문항 간에 나타날 수 있는 방법효과를 통제할 수 있으면서, 확인적 요인분석 차원에서 각 문항들이 복수의 요인을 측정할 수 있는 분석 방식이다(남궁준재, 이순묵, 김효선, 2013). 이러한 ESEM의 장점을 이용하여 유사한 개념, 혹은 방법효과를 가지고 있다고 판단되는 문항묶음에 잔차간 상관을 허용하여 분석하였다.

회전 부분에는 부분제약 목표회전(임경민, 김수영, 2019; Browne, 1972)을 실시하였다. 부분제약 목표회전은 연구자가 일차로 실시한 탐색적 결과에 근거하여 요인형태계수 행렬 내의 작은 값을 0으로 제약하거나, 0에 가깝도록 추정하는 회전 방법이다(이순묵, 윤창영, 이민형, 정선호, 2016). 즉 교차부하를 허용하면서 동시에 이론적 구조를 가정하여 탐색적인 절차와 확인적인 절차를 동시에 시도하는 분석방식이다.

마지막으로, 번안된 NEO-PI-3 검사가 남녀 간에 성격의 요인을 동일하게 개념화하고 있는지, 즉 남녀에 따라 동일한 성격 구조를 가지고 있는지 등의 검토를 위해 최종검사 자료를 바탕으로 측정 동일성 검증을 실시하였다. 모든 통계적 분석은 R 4.3.1 프로그램을 이용하였다.

결 과

예비검사의 탐색적 요인분석 결과

한국판 NEO-PI-3 예비검사에서 성격 5요인 구조가 나타나는지 탐색적 요인분석을 통해 검증하였다. 이를 위해 수집된 자료가 탐색적 요인분석을 실시하기에 적절한 자료인지 확인하고자 KMO와 Bartlett 검정을 실시하였는데, KMO 값이 0.93이고, Bartlett 검정에서 영가설을 기각하여 자료가 탐색적 요인분석에 적절함을 확인하였다.

이후 예비검사 자료를 바탕으로 스크리 도표와 평행 분석을 이용하여 요인의 개수를 탐색하였다. 평행 분석 결과 적절한 요인의 개수는 성격 5요인과 마찬가지로 5개의 요인으로 나타났다. 예비검사를 이용하여 요인의 수를 결정하기 위해 진행한 평행분석 결과는 그림 1과 같다.

5개 요인을 바탕으로 탐색적 요인분석을 실한 결과는 부록 2에 첨부하였다. 탐색적 요인분석 결과 대다수의 하위요인의 문항묶음이 각각 예상되는 상위 성격 5요인에 할당이 되었으나, 개방성 요인의 '가치' 하위요인과(.202) 우호성 요인의 '겸손' 하위요인(.283)은 0.3 미만의 요인 부하량을 가지고 있었으며, 이외의 '충동', '긍정', '사교', '온정', '활동', '감정', '성취' 하위요인의 경우 다른 상위요인과 중복되는 변수복합도가 존재하였다. 특히 외향성 요인의 '온정' 하위요인의 경우 우호성 요인이 설명하는 정도 (.680)가 외향성이 설명하는 정도(.456)보다 더 높았다. 예비검사의 탐색적 요인분석 결과를 보았을 때, 한국 문화에서는 '온정'이라는 개념이 외향성보다는 우호성에 가깝다고 볼 수 있다. 이는 내용전문가가 문항을 보고 내용적으로 판단했을 때에도 온정이 우호성에 가까움을 확인하였다.

예비검사의 문항분석 결과

한국판 NEO-PI-3 예비검사의 각 문항을 고전검사이론에 기반하여 문항평균값이 중심경향을

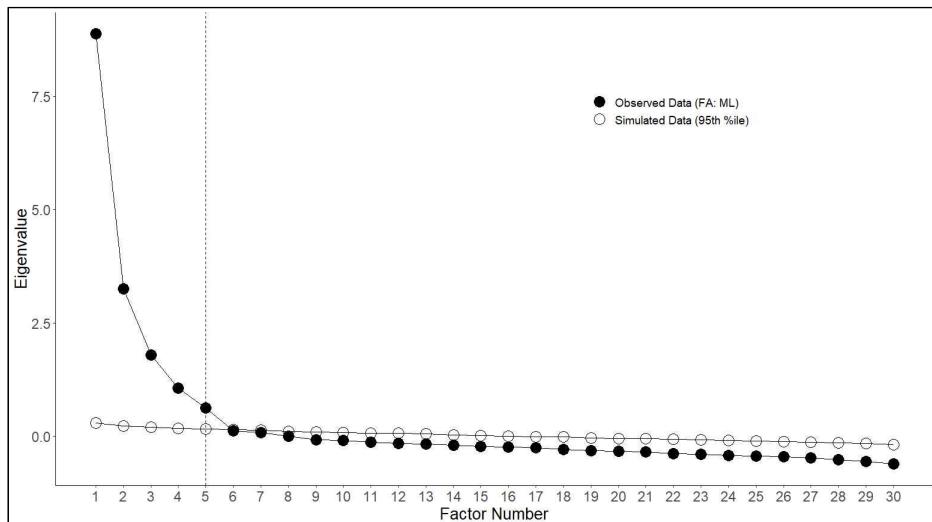


그림 1. 예비검사의 평행분석 결과

가지고 문항-총점 상관이 .3 이상의 값을 보인 문항을 양호한 문항이라 판단하였다(총점의 경우 각 상위 요인의 총점). 문항반응이론 분석에서는 다분문항반응이론 중 하나인 부분점수모형을 이용하여 각 문항의 심각도(severity)를 계산하기 위해 경계모수(boundary parameter)를 추정하였다. 또한 문항의 적합도 지수인 infit 지수와 outfit 지수를 산출하였는데, 평정척도에서 주로 사용하는 기준인 0.6 이상 1.4미만의 적합도를 가지는 문항을 양호한 문항이라 판단하였다. 이러한 통계적 기준을 바탕으로 적합도가 낮은 46개 문항을 선별하였으며, 46개 문항 모두 낮은 문항-총점 상관을 보였고, 46개 문항 중 문항적 합도 infit과 outfit 지수가 양호하지 않은 8개 문항이 포함되어 있었다.

문항분석 결과 탐색적 요인분석 결과와 마찬가지로 개방성의 ‘가치’ 하위요인과 우호성의 ‘겸손’ 하위요인은 문항 단위에서도 양호하지 않은 값이 나타남을 알 수 있었다. 특히 개방성의 ‘가치’ 하위요인의 경우 8개 문항 모두 양호하지 않은 문항임을 알 수 있다. 이러한 통계적인 분석 결과와 내용 전문가들의 판단을 모두 고려하여 총 46개 문항을 수정하였는데, 가령

‘가치’ 하위요인의 문항 중 ‘25살이 되도록 소신이 확고하지 못하다면 문제가 있는 사람이다.’ 혹은 ‘전통적 가치는 의심의 여지없이 존중되어야 한다.’ 등의 문항은 ‘도덕적 문제에 대한 결정을 내리기 위해서는 하나님의 기준을 따라야 한다고 생각한다.’, ‘전통적 가치는 존중되어야 한다.’와 같이 NEO-PI-3의 개발 취지에 맞게 보다 이해하기 쉽고 직접적인 질문으로 변경하여 수정검사를 구성하였다.

수정검사 확인적 요인분석 결과

예비검사의 문항을 수정한 수정검사를 바탕으로 1,280명의 표본을 재수집하여 이를 통해 확인적 요인분석을 실시하였다. 이때 검사의 구조는 예비검사의 탐색적 요인분석과 기존 해외의 원검사가 설정해 두었던 원검사의 검사 구조를 바탕으로 실시하였다. 분석하는 자료는 탐색적 요인분석과 마찬가지로 원 자료에서 하위요인을 문항묶음 처리하여 분석을 실시하였다. 수정검사의 확인적 요인분석 결과는 부록 3에 첨부하였다. 분석 결과 모형 적합도에서 CFI는 .702, TLI는 .672, RMSEA는 .113, SRMR은 .115를

보여 본 연구에서 설정한 기준인 CFI와 TLI는 0.9 이상, RMSEA와 SRMR은 0.08 이하 기준에 미치지 못하는 값을 보였다. 그러나, 각 모형의 요인 부하량은 대부분 예비 검사의 탐색적 요인분석과 유사하게, 각 상위요인이 하위요인을 적절하게 설명하는 것으로 나타났다. 또한 예비검사에서 낮은 요인 부하량으로 설명이 되었던 개방성 요인의 '가치' 하위요인은 유의한 수준의 요인 부하량으로 나타나 적절하게 문항이 수정되었다고 판단하였다. 그러나 '겸손'에 해당하는 요인 부하량이 예비검사 결과와 마찬가지로 낮은 수준의 부하량(.017)으로 나타났다. 따라서 '겸손'에 해당하는 부분을 별도로 고전검사이론 기반의 문항분석을 실시하였는데, 겸손 부분의 5번째 문항(-.12)과 6번째 문항(-0.9)이 낮은 수준을 넘어 역상관 관계의 문항-총점 상관을 보였고, 8번째에 해당하는 문항은 문항-총점 상관이 0에 가까웠다. '겸손' 하위개념을 구성하는 3개의 문항을 이론 및 내용상 적절하다고 판단되도록 다시 수정하였는데, 가령 '나의 재능은 부족함이 없다.'와 같은 문항을 '나는 남들에 비해서 탁월한 점이 많다고 생각한다.'와 같이 내용을 직설적으로 변경하여 최종검사를 구성하였다. 최종적으로 수정한 46개의 문항은 부록4에 자세히 제시하였고 통계적 판단뿐만 아니라 내용전문가의 내용적 판단과 요인구조의 확인을 통해 문항을 번역하고 수정하였다.

최종검사 분석 결과

수정검사의 문항을 수정한 최종검사를 바탕으로 345명 대상에게 다시 응답을 수집하여 자료를 구성하고, 예비검사 및 수정검사와 마찬가지로 각 하위요인을 문항묶음 처리하였다. 이후 부분제약 목표회전 요인분석 기반의 ESEM 분석을 실시하였으며, 수정 지수와 해당 측정오차 간 상관이 설명 가능한지에 대한 내용적 판단을 검토하여 측정 오차 간의 상관을 부여하였다. 그 결과는 표 1과 같으며, 모형 적합도는 CFI는

.928, TLI는 .891, RMSEA는 .068, SRMR은 .030으로 나타나 본 연구에서 설정한 기준인 CFI와 TLI 0.9 이상, RMSEA와 SRMR은 0.08 이하 기준에서 만족할 정도의 적합도를 보였다. TLI는 0.891로 기준치에 약간 미치지 못하였으나, 다른 적합도 수치가 기준을 충족하였고, TLI의 기준이 미비한 차이임을 고려하여, 해당 모형이 연구에서 제시한 기준에 적합한 모형이라 판단하였다. 각 상위요인이 설명하는 요인 부하량은 0.30 이상으로 유의한 수준이었다. 따라서 최종 검사의 요인구조를 통해 NEO-PI-3 성격검사가 이론적으로 측정하고자 하는 성격 5요인의 구성 개념을 타당하게 측정하고 있음을 검증하였다.

측정오차 간 상관을 고려했을 때, 심리적 민감성의 '분노'는 '이타'와 약한 음의 상관(-.256)을 가지는 것을 볼 수 있다. 이는 분노적인 성격 특성을 가지는 사람들이 낮은 이타적 특성을 가질 수 있기에 해당 측정오차 간 상관이 타당하다고 내용전문가가 판단하였다. 반면에 '분노'는 '솔직'과 약한 양의 상관관계(.217)이 나타났다. 이는 분노 표현을 하는 사람의 경우 솔직하게 표현을 하는 사람일 수 있기에 해당 오차 간 상관이 타당하다고 내용전문가가 판단하였다.

우호성의 '신뢰' 하위요인의 경우 같은 우호성의 '솔직'(.259)과 '온유'(.246) 하위요인과 각각 낮은 양의 측정오차 간 상관을 보였다. 이는 타인을 신뢰하는 경우 솔직한 행동을 할 수 있고, 타인을 도울 수 있기 때문에 해당 오차간 상관이 타당하다고 내용전문가가 판단하였다. 또한 해당 문항의 잔차 상관을 통해 우호성의 하위요인들이 설명하지 못하는 또다른 변인이 있을 수 있음을 시사한다.

성실성의 '유능감' 하위요인의 경우 우호성의 '겸손' 하위요인과 낮은 음의 상관(-.291)을 보였다. 이는 겸손한 성격 특성을 가진 사람들이 자신의 유능감을 다소 낮게 평가하는 경향이 있음을 시사하며, 이러한 내용전문가의 판단은 해당 측정오차 간 상관이 타당함을 뒷받침하였다. '이타'는 '유능감'과 약한 양의 상관(0.235)을 나타

표 1. NEO-PI-3 최종검사의 ESEM 분석 결과

RMSEA								
χ^2	df	p-value	RMSEA	C.I.	CFI	TLI	SRMR	
754.973	289	.00	.068	.062	.074	.928	.891	.030
요인 이름								
상위 요인	문항 뮤음	심리적 민감성	외향성	개방성	우호성	성실성		
심리적 민감성	N(불안)	.886	-.102	.002	.112	.054		
	N(분노)	.786	.105	-.024	-.348	.032		
	N(심약)	.698	-.076	-.082	.068	-.277		
	N(열등)	.678	-.304	.010	.027	-.090		
	N(위축)	.858	-.029	.023	.029	-.047		
	N(충동)	.502	.191	.081	-.273	-.302		
외향성	E(긍정)	-.107	.527	.260	.199	-.024		
	E(사교)	-.128	.753	-.034	.173	.014		
	E(자극)	.113	.510	.165	-.215	-.079		
	E(온정)	-.036	.676	.067	.323	.139		
	E(주장)	-.162	.444	.210	-.311	.246		
	E(활동)	-.056	.754	-.039	-.065	.120		
개방성	O(가치)	-.276	-.006	.330	.176	-.097		
	O(감정)	.403	.373	.393	.211	.077		
	O(독창)	-.102	-.265	.752	-.137	.188		
	O(상상)	.046	-.049	.798	-.011	-.158		
	O(시도)	-.292	.342	.380	-.159	-.198		
	O(심미)	.177	-.061	.630	.150	.083		
우호성	A(이타)	-.028	.200	.160	.684	.045		
	A(겸손)	-.086	-.286	-.097	.557	-.058		
	A(솔직)	-.282	-.025	.048	.464	.095		
	A(순응)	-.010	-.096	-.081	.771	.033		
	A(신뢰)	-.239	.270	.024	.559	.018		
	A(온유)	.204	.186	.081	.560	.068		
성실성	C(성취)	-.019	.275	.095	-.118	.700		
	C(신중)	-.007	-.226	.048	.079	.692		
	C(유능감)	-.202	.050	.075	-.052	.686		
	C(절제)	-.169	.073	-.043	.000	.758		
	C(체계)	-.014	-.052	-.134	-.025	.676		
	C(책임감)	.070	.077	-.009	.161	.793		
측정오차 간 상관	A(겸손) -	A(솔직) -	N(분노) -	A(이타) -	A(신뢰) -	N(분노) -		
	C(유능감)	A(신뢰)	A(이타)	C(유능감)	A(온유)	A(솔직)		
	-.291***	.259***	-.256***	.235***	.246***	.217***		
심리적 민감성								
요인 간 상관								
외향성	외향성	-.213***						
	개방성	-.067	.494***					
	우호성	-.147**	.061	.024				
	성실성	-.484***	.230***	.178**	.224***			

* $p < .05$, ** $p < .01$, *** $p < .001$.

냈다. 이는 이타적인 성향을 가진 개인이 자신의 유능감을 높게 평가할 수 있으며, 해당 상관이 타당하다고 내용전문가가 판단하였다.

최종요인을 확인한 후 국내 표본을 이용한 최종검사의 모형 구조와 미국 원본 NEO-PI-3 검사(Costa & McCrae, 2017)의 모형 구조를 비교하였다. 심리적 민감성의 하위요인을 살펴보면, '분노'와 '충동' 요인은 한국과 미국 표본 모두에서 우호성과 높은 변수복합도를 보였다. 반면, '심약' 요인은 미국 원본에서는 성실성에서 높은 음의 부하량을 보였으나, 한국 표본에서는 유의미한 요인 부하량을 보이지 않았다.

외향성과 관련해서는, '온정' 요인이 두 표본 모두에서 우호성과 높은 양의 부하량을 보였고, '주장' 요인은 우호성과 높은 음의 부하량으로 변수복합도가 높았다. 그러나 '활동' 요인은 미국 원본에서 성실성과 양의 부하량을, '자극' 요인은 우호성과 음의 부하량을 보였으나, 한국 표본에서는 이러한 결과가 나타나지 않았다.

외향성의 '감정'과 '시도' 하위요인은 미국과 한국 모두에서 각각 외향성과 심리적 민감성과 높은 양의 부하량을 보였다. 특히, 한국 표본에서는 '감정' 요인이 심리적 민감성과, '시도' 요인이 외향성과 양의 부하량을 보였다.

우호성 요인에서는 미국 원본의 '이타' 요인이 외향성과 정적 관계를 보였으나, 한국 최종 검사에서는 우호성의 각 하위요인이 정확하게 우호성 요인으로 묶였다. 성실성 요인에서도 미국 원본의 '유능감' 요인이 심리적 민감성과 음의 부하량을 보였으나, 한국 최종검사에서는 성실성 하위요인이 모두 정확하게 성실성 요인으로 묶였다.

최종검사의 요인구조를 확인한 후 검사의 신뢰도를 확인하기 위해 검사-재검사 신뢰도를 계산하였다. 최종검사를 실시한 표본중 64명을 대상으로 1개월의 시차를 두고 재검사를 실시하여, 해당 정보를 바탕으로 각 상위요인과 하위요인에 따라 시점에 따른 점수 간의 상관을 산출하였다. 해당 검사-재검사 신뢰도 지수는 표 2와 같다.

상위요인과 하위요인의 신뢰도를 보았을 때, '가치' 하위요인을 제외한 모든 하위요인은 .70 이상의 값으로, 각 성격 지수가 신뢰로운 검사임을 알 수 있었으나, '가치' 하위요인의 경우 유의한 상관을 보이긴 하였으나 낮은 검사-재검사 신뢰도 지수(.43)를 보였다. 따라서 현장에서 검사를 사용할 때 '가치' 하위요인 점수를 해석할 경우 낮은 신뢰도를 고려할 필요가 있다.

표 2. NEO-PI-3의 검사-재검사 신뢰도 지수

5요인		하위요인					
심리적 민감성(N)	(.921 ***)	불안	(.910 ***)	분노	(.864 ***)	심약	(.848 ***)
외향성(E)	(.927 ***)	긍정	(.835 ***)	사교	(.900 ***)	자극	(.862 ***)
개방성(O)	(.930 ***)	가치	(.431 ***)	감정	(.849 ***)	독창	(.920 ***)
우호성(A)	(.907 ***)	이타	(.792 ***)	겸손	(.787 ***)	상상	(.893 ***)
성실성(C)	(.943 ***)	성취	(.856 ***)	신중	(.878 ***)	솔직	(.854 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)
						순응	(.856 ***)
						신뢰	(.895 ***)
						온유	(.705 ***)
						시도	(.932 ***)
						활동	(.898 ***)
						자극	(.839 ***)
						감정	(.895 ***)
						독창	(.856 ***)
						상상	(.893 ***)

이후 성격 5요인 구조와 하위요인 30개의 검사 구조가 하위 집단에서도 동일하게 적용될 수 있는지 탐색하고자 성별로 구분하여 측정 동일성 분석을 실시하였다. 측정단위 동일성 분석을 위해서 문항들간의 교차부하가 존재하지 않고, 하나의 요인이 하나의 문항묶음만을 설명하는 모형으로 실시하였다. 분석한 결과는 표 3과 같다. 표 3을 보았을 때, 측정단위 동일 모형의 경우 측정틀 동일 모형과 통계적으로 유의한 차이가 존재하지 않았다($\Delta CFI=.000$, $\Delta RMSEA=-.002$). 그러나 측정원점 동일 모형은 측정단위 동일 모형과 통계적으로 유의한 차이가 존재하여 측정원점 동일성과 요인평균 동일성은 존재하지

않는 것으로 판단하였다. 따라서 최종적으로 NEO-PI-3 최종검사의 구조는 성별에 따른 약한 측정 동일성이 존재한다고 볼 수 있다. 측정원점과 요인평균이 남여 집단에 동질하지 않게 나타나서 어떤 상위요인과 하위요인에서 성별차이가 있는지 구체적으로 확인하기 위해 Welch T 검정을 실시하였다. 그 결과를 표 4에 제시하였고, 상위요인 점수 중 심리적 민감성 요인에서 성별에 따른 유의미한 차이가 발견되었으며, 여성이 남성보다 높은 점수를 나타내는 것으로 확인되었다($p < .001$).

하위요인에서 살펴보았을 때, 심리적 민감성에 속하는 모든 6개 하위요인에서도 통계적으로

표 3. NEO PI-3 최종검사의 측정 동일성 분석 결과

제약모형	$\Delta\chi^2$	Δdf	p	ΔCFI	$\Delta RMSEA$
측정틀 동일	-	-	-	-	-
측정단위 동일	24.617	25	.484	.000	-.002
측정원점 동일	97.885	25	<.000	-.011	.000
요인평균 동일	36.537	5	<.000	-.005	.001

표 4. 최종검사의 각 요인의 성별에 따른 평균(M) 및 표준편차(SD)와 t검정 결과

상위요인 및 하위요인	전체 표본		남성		여성		t검정(성별)	
	M	SD	M	SD	M	SD	t	p
심리적 민감성	120.68	25.12	114.77	23.35	126.99	25.48	-4.65***	.000
외향성	161.11	20.99	162.65	20.28	159.48	21.66	1.40	.162
개방성	161.30	18.82	160.85	17.79	161.77	19.90	-0.45	.653
우호성	166.60	18.27	165.31	17.73	167.98	18.78	-1.36	.176
성실성	178.25	21.09	179.78	19.50	176.62	22.61	1.39	.165
N(불안)	23.49	5.69	22.34	5.42	24.71	5.73	-3.95***	.000
N(분노)	18.21	4.79	17.29	4.63	19.19	4.77	-3.76***	.000
N(심약)	19.43	4.70	18.37	4.30	20.56	4.85	-4.44***	.000
N(열등)	19.79	4.58	19.01	4.47	20.63	4.55	-3.33**	.001
N(위축)	19.17	5.30	17.90	4.87	20.53	5.43	-4.73***	.000
N(충동)	20.59	4.78	19.85	4.49	21.37	4.96	-2.98**	.003

표 4. 최종검사의 각 요인의 성별에 따른 평균(M) 및 표준편차(SD)와 t검정 결과 (계속)

상위요인 및 하위요인	전체 표본		남성		여성		t검정(성별)	
	M	SD	M	SD	M	SD	t	P
E(긍정)	27.62	4.61	27.35	4.70	27.90	4.51	-1.10	.274
E(사교)	26.74	5.49	27.20	5.31	26.26	5.66	1.60	.110
E(자극)	22.95	4.31	23.31	3.91	22.56	4.68	1.62	.106
E(온정)	30.41	4.25	30.63	4.19	30.17	4.31	1.00	.320
E(주장)	27.54	4.35	27.72	4.06	27.36	4.65	0.77	.444
E(활동)	25.85	4.65	26.43	4.49	25.23	4.74	2.41*	.016
O(가치)	27.07	2.81	27.16	2.76	26.96	2.87	0.66	.512
O(감정)	29.31	4.23	28.67	4.21	29.98	4.15	-2.90**	.004
O(독창)	27.75	5.53	28.39	5.28	27.07	5.72	2.24*	.026
O(상상)	25.66	5.46	26.09	4.61	25.21	6.23	1.50	.135
O(시도)	23.94	4.82	24.22	4.72	23.65	4.91	1.11	.266
O(심미)	27.56	5.28	26.31	5.27	28.90	4.96	-4.69***	.000
A(이타)	31.21	3.80	31.12	3.83	31.30	3.77	-0.44	.658
A(겸손)	25.59	4.37	25.71	4.67	25.46	4.03	0.52	.601
A(솔직)	27.28	4.35	27.23	4.30	27.34	4.42	-0.22	.823
A(순응)	26.86	4.40	26.40	4.20	27.35	4.57	-2.00*	.047
A(신뢰)	27.55	4.47	27.29	4.35	27.83	4.59	-1.11	.268
A(온유)	28.11	4.58	27.56	4.97	28.71	4.07	-2.34*	.020
C(성취)	30.43	4.31	30.90	4.11	29.93	4.48	2.10*	.037
C(신중)	28.54	4.47	28.83	4.38	28.23	4.56	1.23	.219
C(유능감)	29.32	3.74	29.59	3.77	29.02	3.71	1.40	.161
C(절제)	29.74	4.15	29.90	4.04	29.57	4.28	0.75	.454
C(체계)	28.58	6.18	29.03	5.61	28.09	6.72	1.42	.156
C(책임감)	31.65	3.74	31.53	3.62	31.78	3.87	-0.62	.535

* $p < .05$, ** $p < .01$, *** $p < .001$.

유의한 차이가 나타났으며, 여성의 점수가 남성보다 일관되게 높았다. 외향성의 ‘활동’ 하위요인에서는 남성이 여성보다 더 높은 점수를 보였으며, 이 차이는 통계적으로 유의했다($p = .016$). 개방성에 속하는 하위요인들 중 ‘감정’($p = .004$,

‘독창’($p = .026$), ‘심미’($p < .001$)에서 성별에 따른 유의미한 차이가 관찰되었다. ‘감정’과 ‘심미’는 여성이 더 높은 점수가 나타난 반면, ‘독창’에서는 남성이 더 높은 점수가 나타났다. 우호성에 속하는 ‘순응’($p = .047$), ‘온유’($p < .001$)

하위요인에서도 여성이 더 높은 점수를 보였다. 성실성의 하위요인인 ‘성취’에서도 성별에 따른 유의미한 차이가 나타났으며($p = .037$), 남성이 더 높은 ‘성취’ 점수를 보였다. 이러한 남녀 차이는 원본 검사(Costa & McCrae, 2017)에서도 비슷하게 나타났다.

논 의

본 연구는 성격 5요인 이론을 바탕으로 사람의 성격을 측정하는 NEO-PI-3 검사의 변안 타당도를 검증한 연구이다. 이를 위해 내용전문가가 번역한 예비검사를 바탕으로 심리측정전문가가 성격심리전문가가 이론적인 면과 측정학적 면을 동시에 고려하여 수정과정을 거쳐 수정검사를 제작하였으며, 수정한 검사를 이용하여 다시 표본을 수집하여 타당화한 최종검사를 결정하였다. 본 연구 결과에 대한 의미는 다음과 같다.

첫째, 본 연구에서 탐색적 요인분석, 확인적 요인분석, 그리고 탐색적 구조방정식 모형(ESEM)을 실시하여 국내 표본을 이용한 NEO-PI-3의 요인구조를 확인할 수 있었다. 외국 검사를 국내 검사로 변안할 경우에도 새로운 검사를 개발하는 과정과 마찬가지로 검사 구조의 타당성을 증명하기 위해서 국내 표본으로 다시 실시해서 요인구조와 문항내용의 적절성을 확인해야 한다. ITC(International Test Commission, 2017) 국제 지침에서는 원검사 집단과 변안 집단 간에 구성개념의 동등성, 측정 방법 동등성 및 문항 동등성에 대한 통계적 수치를 제공해야 한다고 기술되어 있다(ITC, 2017). 원검사와 변안검사의 구성개념 동등성을 확인하기 위한 대표적인 분석방법이 탐색적 요인분석과 확인적 요인분석이다. 탐색적 요인분석은 요인의 개수와 구조를 탐색하는 용도로 사용되고 확인적 요인분석은 탐색적 요인분석에서 나타난 요인의 구조를 다른 표본을 이용해서 확인하고 검정하는 절차이다. 많은 연구자들이 동일한 표본에서 탐색적 요인분-

석과 확인적 요인분석을 같이 실시하는 경향이 있는데 동일한 표본에서 실시한 탐색적 요인분석과 확인적 요인분석의 결과로 검사의 요인구조를 일반화하여 해석하는데 한계가 있다(Kline, 2019). 따라서 본 연구는 변안타당화 과정에서도 일반적인 검사개발과정과 마찬가지로 두 차례 이상의 표집과정을 통해 탐색과 확인적 분석을 독립적으로 진행하였다. 연구별 연구참가자가 완전히 동일하지 않지만 연령과 성별을 비율적으로 고르게 표집하였기 때문에 연구의 결과를 어느정도 일반화하였다고 볼 수 있다. 특히, 최종검사에서 ESEM을 적용하였고, 목표회전을 실시하였는데, 확인적 요인분석을 실시한 이후 여러 차례 문항 수정 후 확인적인 정신을 위배하지 않고 탐색적인 목적까지 달성할 수 있는 부분제약 목표회전이 효과적인 방법임을 보였다(Browne, 2001; 이순록 외, 2016).

둘째, 하위요인에 있는 문항들을 고전적인 문항분석과 함께 문항반응이론에 기반한 문항분석을 실시하여 번역 및 의미오류(semantic error), 문해력이나 문화 차이가 있을 수 있는 문항들을 정교하게 탐색하였다. 총점과 상관이 낮은 문항, 상관이 존재하지 않는 문항, 역상관이 되는 문항 및 문항반응모형에 적합하지 않는 문항들을 검토할 수 있었다. ITC(2017) 국제 지침서에서도 변안을 실시할 때, 예비검사에서 문항분석을 실시할 것을 제안하였고 실질적으로 현장에서 변안검사를 사용하기 전에 변안된 검사의 측정학적 증거들을 확인하는 것이 필요하다고 하였다. 본 연구에서는 예비검사에서는 전체 문항을 분석하였고, 수정검사에서 낮은 요인부하량을 보이는 하위요인에 해당하는 문항을 분석하여 내용전문가가 오류를 바로 잡고 더 쉽고 직접적인 질문으로 수정하는데 도움을 제공할 수 있었다.

셋째, 측정전문가와 내용전문가의 상호 논의를 통해 문항의 수정을 반복적으로 시도하여 검사의 구조를 이론적인 모형과 경험적으로 확인할 수 있었다. 예비검사에서 측정학자가 분석한 결과를 바탕으로 내용전문가의 문항 수정과정을

거쳐 수정검사를 제작하였고, 이러한 분석과 수정과정을 반복하여 최종검사를 결정하여 NEO-PI-3 검사의 이론적 구조가 국내에서도 원본검사와 동일하게 나타남을 검증하였다. 예비검사에서 적절하지 않은 하위요인이었던 ‘가치’ 하위요인은 수정을 통해 확인적 요인분석에서는 양호한 수준의 요인부하량(.314)으로 나타났다. 예비검사와 수정검사에서 지속적으로 적절하지 않은 요인부하량이 나타난 ‘겸손’ 하위요인 또한 문항 수정을 통해 최종적으로 수행한 확인적 요인분석에서 양호한 수준의 요인부하량(.565)을 보였다. ‘겸손’이란 성격요인은 모든 문화권에서 내재하는 일반적인 특질이지만 겸손을 정의하고 보여주는 방식은 소속한 문화권에 따라 차이를 보일 수 있다. 예를 들어, ‘겸손’을 정의할 때 원본검사에서는 자기를 낮추어서 자기를 표현하되 (humble and self-effacing) 자신감과 자존감은 부족하지 않는 것으로 정의하였다(not lacking in self-confidence and self-esteem). 그러나 한국은 ‘겸손’을 사실과 다르게 자신을 낮추는 것만을 말한다(최상진, 김시업, 김은미, 김기법, 2000). 이와 같이 겸손을 바라보고 정의하는데 있어 문화 차이가 있을 수 있음에도 불구하고 번안(adaptation) 검사를 번역(translation)검사로 오인해서 진행하는 경우가 많았다. 검사를 수정하는 과정에서는 원검사의 문항을 그대로 번역하는 것이 아니라 응답자의 의미오류를 예방할 수 있고 문해력의 차이로 인한 문항이해도의 차이가 발생하지 않도록 직접적으로 서술하거나 문화적인 차이가 있을 수 있는 문항을 같은 질문의도나 취지를 살리되 다르게 표현하는 절차도 포함되어 있다. 즉, 검사의 번안 과정은 원검사가 제시한 이론적 구조를 반영하여 번안자가 국내 표본에 맞추어 새로이 검사를 개발하는 과정이라고 할 수 있다.

넷째, 본 연구에서는 성별에 따라 NEO-PI-3가 동일하게 측정이 될 수 있는지 알아보고자, 성별에 따른 측정 동일성을 살펴 보았다. 본 연구는 이러한 목적을 위하여 다집단 분석을 시도하

였으며, 그 결과 NEO-PI-3 검사가 측정구조와 측정 단위가 성별에서 동일한 약한 동일성을 보임을 알 수 있었다. 동일성 검증의 주된 목적 중 하나는 하나의 문화 속 많은 하위 집단 간의 일반화를 검토하고자 할 때 요구될 수 있다(이순묵, 김한조, 2011). 동일성 검증을 통해 척도의 일반화 가능성에 대한 양적인 접근을 시도할 수 있으며, 본 연구에서 적용한 결과 NEO-PI-3가 성별에 따라 동일한 구조와 요인 부하량을 가지고 있다고 판단하였다. 반면에 측정원점과 요인 평균 동질성을 확보하지 못하였는데 이는 성별에 따른 특정 상위요인 점수와 하위요인점수의 차이에서 확인할 수 있었다. 성별에 따른 상위요인 및 하위요인에 따른 점수 차이를 확인했을 때, 상위요인에서는 심리적 민감성에서 성별에 따른 유의한 차이가 나타났고, 하위요인에서는 심리적 민감성의 모든 하위요인, ‘활동’, ‘감정’, ‘독창’, ‘심미’, ‘순응’ ‘온유’, ‘성취’에서 성별에 따른 유의한 차이가 나타났다. 이러한 결과들은 성별에 따라 NEO-PI-3의 특정 상위요인 점수에서 유의미한 차이가 존재함을 보였다. 이는 기존 성격연구들(Feingold, 1994; Hyde & Linn, 1988)에서 알려진 것처럼 성별에 따라 일부 성격 점수의 차이가 나타남을 다시 한번 확인할 수 있었다. 이를 반영하여 남녀의 규준을 별도로 구분하여 분석할 필요가 있다.

본 연구는 NEO-PI-3를 한국판으로 번안하고 타당화한 첫 연구로서 본 타당화 연구결과 상위 5개 요인 및 하위 30개 각 하위요인의 구조가 원본과 동일하게 나타났고 본 검사가 신뢰로운 검사로 검증이 되었지만, 하위요인들을 살펴보았을 때, ‘가치’ 하위요인에서는 낮은 신뢰도가 나타나 검사 결과를 해석하는 과정에서 내용적 판단을 통하여 이를 고려할 필요가 있다. 또한 하위요인의 검사 구조에 있어서 하위 요인 구조가 단순히 하나의 공통요인으로 묶이는지, 이중 요인 구조인지(second-order), 또는 이원요인 구조인지(bi-factor)를 다양한 요인모형을 적용해 볼 필요가 있다.

둘째, 최종적으로 수정한 NEO-PI-3 검사가 실제적으로 직무성과나 성격장애를 예측하는지에 대한 예측 타당도를 검증할 필요가 있다. 이러한 예측 타당도를 확인하기 위해 NEO-PI-3 검사를 실시한 응시자들을 이후에 다시 업무성과 척도나 성격장애 척도를 가지고 측정할 수 있어야 한다. 이러한 준거관련 타당도는 응시자 표본을 가지고 NEO-PI-3 검사와 준거척도를 동시에 실시하여 두 검사들의 상관을 보는 동시타당도 (concurrent validity) 검정을 통해서도 가능한데, 본 연구에서는 NEO-PI-3 문항 자체가 길어서 준거 척도와 함께 통합하여 검사를 진행하지 못하였다. 추후 연구에서는 한국판 NEO-PI-3가 다른 준거 척도와 관련이 있는지 또는 NEO-PI-3 검사가 예측하고자 하는 행동 특성을 정확하게 예측하는지를 확인할 수 있다면, 실제 현장에서 보다 신뢰롭고 타당한 검사로 사용될 수 있을 것이다.

셋째, 예비문항 표집부터 최종문항 까지 각 연구의 표본 특성이 현실적인 상황으로 인해 고르게 표집되지 않았다. 가령, 1차로 수정한 문항에서는 최종 표본에 비해 남성이 더 많았으며, 예비 문항과 1차 문항에는 대학생 표본이 포함되기도 하였다. 그러나, 본 연구는 주 대상이 성인 대상 NEO-PI-3 검사의 변안 타당화 연구라는 점, 예비검사, 수정검사, 최종검사의 표본의 특성이 달라도 해당 표본에서 유사한 검사 구조를 확인했다는 점에서 NEO-PI-3 검사가 성격 5요인의 구조를 보이고 있으며, 성격 5요인을 측정할 수 있는 검사의 문항을 타당하게 변안하였다고 볼 수 있다. 그러나, NEO-PI-3 검사의 개발 취지에 맞게 앞으로 대학생, 고등학생, 중학생이 범용적으로 사용할 수 있기 위해서는 추가적인 타당화 연구가 필요하다. 또한, 지속적으로 신뢰할 수 있는 검사가 되기 위해서는 정기적으로 문항 분석을 진행해야하며, 각 연구 과정 동안 제한적이었던 표집의 한계를 보완하고자 실제 이용자를 대상으로 표본을 꾸준히 모아서 규준(norm)을 개선할 필요도 있다. 특히 본 연구에서 낮은

신뢰도 지수를 보인 ‘가치’ 하위요인의 문항 보완과 내용 타당도를 높이기 위한 지속적인 노력이 필요하다. 하나의 검사를 계속해서 활용하기 위해서는 검사의 신뢰성과 타당성을 지속적으로 확인하는 과정이 필요한데 현 시대는 기술적, 사회적으로 급격히 변화하고 있을 뿐만 아니라 세대간의 라이프스타일, 가치 등의 변화 주기가 빨라졌기 때문에 심리측정학적 평가를 통해 신뢰도와 타당도를 주기적으로 입증하는 과정이 중요하다(Gehrig, 2019). 따라서 주기적으로 표본을 수집하여 지속적으로 분석하고 검사의 규준을 수정/보완하는 작업이 필요하다. 지속적으로 문항의 양호도와 검사 구조를 확인하고 수정/보완을 할 때 좀 더 NEO-PI-3 검사가 정확하게 인간의 성격을 측정하고 예측할 수 있는 범용적 검사 도구가 될 것이다.

저자소개

최상민은 한림대학교에서 심리학 석사학위를 취득하였다. 주요 연구 관심사는 심리측정, 문항반응이론, 혼합모형, 성장모형이다.

정재창은 중앙대학교에서 인적자원개발 박사학위를 받고 HR 컨설팅사인 PSI컨설팅의 대표로서 성격이 조직행동과 수행에 미치는 영향을 파악하여 인재선발, 배치, 육성을 위한 HR 빅데이터를 제공하며 조직의 효과성을 높이기 위한 솔루션을 개발하는 데 관심을 갖고 있다.

서동기는 University of Minnesota Twin-cites에서 심리학 박사학위를 받았으며, 현재 한림대학교 심리학과 교수로 재직 중이고 AI 평가 플랫폼 개발 회사인 더캣코리아 대표이다. 주요 연구 관심사는 심리측정과 문항반응이론(item response theory)을 적용한 컴퓨터 기반 개별 맞춤형 검사(computerized adaptive testing)이다.

참고문헌

- 김동일, 홍성두, 인효연 (2006). Big Five 성격 요인에 기초한 청소년 성격검사의 개발 및 타당화. *열린교육연구*, 14(1), 289-305.
- 남궁준재, 이순묵, 김효선 (2013). 상황판단검사에서 시나리오 효과를 통제한 탐색적 요인분석. *한국심리학회지: 산업 및 조직*, 26(4), 599-624
- 민병모, 이경임, 정재창 (2002). NEO인성검사 (NEO PI-R), PSI 컨설팅.
- 안정원, 이순묵 (2015). 조직몰입 3요소 모형의 내적구조 검토: 탐색적 구조방정식 모형 (ESEM)의 적용. *한국심리학회지: 산업 및 조직*, 28(4), 795-827.
- 안창규, 채준호 (1997). NEO-PI-R의 한국 표준화를 위한 연구. *한국심리학회지: 상담과 심리치료*, 9(1), 443- 473.
- 유태용, 김명언, 이도형 (1997). 5요인 성격검사의 개발 및 타당화 연구. *한국심리학회지: 산업 및 조직*, 10(1), 85-102.
- 유태용, 이도형 (1997). 다양한 직군에서의 성격의 5요인과 직무수행간의 관계. *한국심리학회 연차 학술대회 학술논문집*, 687-701.
- 이경임, 안창규 (1996). NEO인성검사의 임상적 활용. *한국심리학회지: 상담과 심리치료*, 8(1), 65-79
- 이경임 (1994). 성격에 있어서 "Big Five" 구조의 발달. *부산대학교 학생생활연구소 연구보 제29집*, 39-68.
- 이경임 (1995). 한국인의 NEO인성검사 요인 구조와 부적응 집단 프로파일. *부산대학교 대학원 박사학위논문*.
- 이경임 (1995). 임상적 증후들에 대한 성격 문제: NEO인성검사. *효정 장학표 박사 회갑 기념 논문집*.
- 이경임 (1997). MMPI 타당도 척도에 따른 NEO 인성검사 프로파일 분석: V형과 역V형을 중심으로. *부산교육심리학연구*, 1(1), 95-108.
- 이순묵, 김한조 (2011). 구조방정식 모형의 일반화 또는 집단차 연구를 위한 다집단 분석의 관행과 문제점. *사회과학*, 43(1), 63-112.
- 이순묵, 윤창영, 이민형, 정선호 (2016). 탐색적 요인분석: 어떻게 달라지나?. *한국심리학회지: 일반*, 35(1), 217-255.
- 임경민, 김수영 (2019). 요인분석에서 목표회전의 적용 가능성 탐색. *한국심리학회지: 일반*, 38(3), 377-400.
- 최상진, 김시업, 김은미, 김기범 (2000). 한국 여성들의 사회 규범적 행위 양식에 대한 조망. *한국심리학회지: 여성*, 5(2), 103-115.
- 최정우, 한여울, 박두진, 최기홍, 서동기 (2021). 한국형 다차원 성격검사 단축형(Bright and Dark Personality Inventory-Short Form: BDPI-SF) 개발. *한국심리학회지: 사회및성격*, 35(1), 43-67.
- Allport, G. W. (1937). *Personality: A Psychological Interpretation*. New York, NY: Holt, Rinehart & Winston.
- Allport, G. W., & Odber, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, 471, Whole No. 211.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1985). *Standards for educational and psychological testing*. American Psychological Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. American Educational Research Association.
- American Psychiatric Association (1987). *Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R)*. Washington, D. C., American Psychiatric Association.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders(DSM-*

- IV). Washington, D. C., American Psychiatric Association.
- Benet-Martínez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, 75(3), 729-750.
<https://doi.org/10.1037/0022-3514.75.3.729>
- Browne, M. W. (1972). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, 25(1), 207-212.
<https://doi.org/10.1111/j.2044-8317.1972.tb00482.x>
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36(1), 111-150.
- Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, 21(2), 230-258.
<https://doi.org/10.1177/0049124192021002005>
- Brooner, R.K., Costa, P.T., Jr., Feltch, L.J., Rousar, E.E., Bigelow, G.E., & Schmidt, C.W. (1991). The personality dimensions of male and female drug abusers with and without antisocial personality disorder. In L.S. Harris (Ed.), *Problems of drug dependence: Proceedings of the 53rd Annual Scientific Meeting*, Committee on Problems of Drug Dependence. Rockville, MD: National Institute on Drug Abuse.
- Butcher, J. N., Atlis, M. M., & Hahn, J. (2004). The Minnesota Multiphasic Personality Inventory-2 (MMPI-2). In M. J. Hilsenroth & D. L. Segal (Eds.), Comprehensive handbook of psychological assessment, Vol. 2. Personality assessment (pp. 30-38). John Wiley & Sons, Inc..
- Cattell, R. B. (1943). The description of personality: basic traits resolved into clusters. *The Journal of Abnormal and Social Psychology*, 38(4), 476-506.
<https://doi.org/10.1037/h0054116>
- Costa, P. T., Jr., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, 13(6), 653-665.
- Costa, P. T., Jr., & McCrae, R. R. (1992). Revised NEO Personality Inventory and NEO Five-Factor Inventory Professional Manual. Florid: *Psychological Assessment Resources*
- Costa, P. T., Jr., & McCrae, R. R. (2017). The NEO Inventories as instruments of psychological theory. In T. A. Widiger (Ed.), *The Oxford handbook of the Five Factor Model* (pp. 11-37). Oxford University Press.
- Costa, P. T., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences*, 12(9), 887-898.
[https://doi.org/10.1016/0191-8869\(91\)90177-D](https://doi.org/10.1016/0191-8869(91)90177-D)
- Costa, P. T., Martin, T. A., & McCrae, R. R. (2005). Age Trends and age norms for the NEO Personality Inventory-3 in adolescents and adults. *Assessment*, 12, (pp. 363-373)
- Digman, J. M. (1996). The curious history of the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 1-20). Guilford Press.
- Feingold, A. (1994). A Gender differences in personality: A meta-analysis. *Psychological Bulletin*, 116(3), 429-456.
doi: 10.1037/0033-2909.116.3.429.
- Gable, R. K., & Wolf, M. B. (1993). *Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings* (2nd ed.). Boston: Kluwer Academic Publishers.
- Gehrig, E. (2019). *American Psychological Association Guidelines on Psychometric Assessment Validity and*

- Reliability Requirements.* Target Training International.
- Hyde, JS., & Linn, MC. (1988). Gender differences in verbal ability A meta-analysis. *Psychological Bulletin, 104*, 53-69.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.
<https://doi.org/10.1080/10705519909540118>
- International Test Commission. (2017). *The ITC Guidelines for Translating and Adapting Tests.*
- Kline, Rex B. (2019). *구조방정식모형: 원리와 적용* (이현숙, 장승민, 신혜숙, 김수진, 전경희, 공역). 학지사. (2019).
- Lehne, G. K. (2002). The NEO Personality Inventory and the Millon Clinical Multiaxial Inventory in the forensic evaluation of sex offenders. In P. T. Costa, Jr. & T. A. Widiger (Eds.), *Personality disorders and the five-factor model of personality* (pp. 269-282). American Psychological Association.
<https://doi.org/10.1037/10423-016>
- Masters, G.N. (1982). A rasch model for partial credit scoring. *Psychometrika, 47*, 149-174 .
<https://doi.org/10.1007/BF02296272>
- McAdams, D. P. (1992). The five-factor model in personality: A critical appraisal. *Journal of Personality, 60*(2), 329-361.
<https://doi.org/10.1111/j.1467-6494.1992.tb00976.x>
- McCrae, R. R., & Costa, P. T. (1985). Updating Norman's "adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology, 49*(3), 710-721.
<https://doi.org/10.1037/0022-3514.49.3.710>
- McCrae, R. R., & Costa, P. T. (1986). Personality, coping, and coping effectiveness in an adult sample. *Journal of Personality, 54*(2), 385-405.
<https://doi.org/10.1111/j.1467-6494.1986.tb00401.x>
- McCrae, R. R & Costa, P. T. (2010). *NEO Inventories Professional Manual.* PAR
- Marianne Miserandino. (2011). *성격심리학(정영숙, 조옥귀, 조현주, 장문선, 공역)*. 시그마프레스. (2013).
- Mutén E. (1991). Self-reports, spouse ratings, and psychophysiological assessment in a behavioral medicine program: an application of the five-factor model. *Journal of personality assessment, 57*(3), 449-464.
https://doi.org/10.1207/s15327752jpa5703_5
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology, 66*(6), 574-583.
<https://doi.org/10.1037/h0040291>
- Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annual review of psychology, 57*, 401-421.
<https://doi.org/10.1146/annurev.psych.57.102904.190127>
- Randolph, K. L., & Dahling, J. J. (2013). Interactive effects of proactive personality and display rules on emotional labor in organizations. *Journal of Applied Social Psychology, 43*(12), 2350-2359.
<https://doi.org/10.1111/jasp.12184>
- Smith, M. B., Hill, A. D., Wallace, J. C., Recendes, T., & Judge, T. A. (2018). Upsides to dark and downsides to bright personality: A multidomain review and future research agenda. *Journal of Management, 44*(1), 191-217.
<https://doi.org/10.1177/0149206317733511>
- Tupes, E.C. and Christal, R.E. (1961) Recurrent Personality Factors based on Trait Ratings. *USAF ASD Tech. Rep. No. 61-97*, Lackland Airforce Base, TX: US Air Force.
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement*

최상민 · 정재창 · 서동기 / NEO-PI-3 변안검사의 구성개념 타당도

Transactions, 8, 370-371.

1 차원고접수 : 2024. 03. 05.

수정원고접수 : 2024. 07. 09.

최종제재결정 : 2024. 11. 14.

Korean Version NEO-PI-3 Factor Structure Validity

SangMin Choi¹⁾ JaeChang Jeong²⁾ Dong Gi Seo¹⁾

¹⁾Department of Psychology, Hallym University

²⁾PSI Consulting

In this study, we developed and modified the Korean version of the NEO-PI-3 test, a revision of a test measuring personality based on one of the most widely used personality theories, the Five-factor personality traits, to validate its structure. The process involved three stages: a preliminary test, a modified test, and a final test, with analyses conducted at each stage. Initially, with permission from the original copyright holders, content experts translated the original test to create the preliminary test. Data were collected from 1,934 employees and university students for the preliminary test. Based on this data, exploratory factor analysis, classical test theory, and item response theory-based item analysis were conducted to identify and modify items in subfactors with inadequate factor structures, leading to the creation of the modified test. The modified test, which was responded to by 1,280 participants, underwent confirmatory factor analysis to further modify items in subfactors that did not demonstrate satisfactory factor loadings, to construct the final test. The final test was then subjected to target rotation factor analysis to confirm that the test structure adhered to the Five-factor personality structure. Measurement invariance across genders was established, confirming weak invariance at the level of metric invariance. In addition, test-retest reliabilities of all factors were relatively high except for only sub-factor 'values'. Finally, limitations and implications of this study were described in discussion part.

Key words : Five-factor model, NEO-PI-3, Adaptation Validation, Factor analysis

최상민 · 정재창 · 서동기 / NEO-PI-3 변인검사의 구성개념 타당도

부록 1. 예비검사, 본검사, 최종검사 표본 정보

		구분	빈도	%	구분	빈도	%	
예비검사	성별	남자	990	51.19%	연령	10대	7	0.36%
		여자	942	48.71%		20대	616	31.85%
		무응답	2	0.1%		30대	621	32.11%
	연령	합계	1,934	100%		40대	546	28.23%
						50대	127	6.57%
						60대	13	0.67%
						무응답	4	0.21%
						합계	1,934	100%
		구분	빈도	%	구분	빈도	%	
수정검사	성별	남자	944	73.75%	연령	10대	3	0.23%
		여자	336	26.25%		20대	364	28.44%
		무응답	0	0%		30대	386	30.16%
	연령	합계	1,280	100%		40대	322	25.16%
						50대	194	15.16%
						60대	11	0.86%
						무응답	0	0%
						합계	1,280	100%
		구분	빈도	%	구분	빈도	%	
최종검사	성별	남자	178	51.59%	연령	10대	7	2.03%
		여자	167	48.41%		20대	99	28.7%
		무응답	0	0%		30대	102	29.57%
	연령	합계	345	100%		40대	101	29.28%
						50대	36	10.43%
						60대	0	0%
						무응답	0	0%
						합계	345	100%

부록 2. NEO-PI-3 예비검사의 탐색적 요인분석 결과

상위 요인	문항 뜻	요인이름				
		심리적 민감성	외향성	개방성	우호성	성실성
심리적 민감성	N(불안)	.902	-.076	.009	.026	.191
	N(분노)	.640	.307	-.056	-.488	-.008
	N(심약)	.732	-.009	-.080	-.014	-.244
	N(열등)	.784	-.123	.006	-.035	.014
	N(위축)	.843	.034	.036	-.091	-.011
외향성	N(충동)	.510	.420	.016	-.255	-.281
	E(긍정)	-.103	.476	.178	.373	-.057
	E(사교)	-.098	.584	-.209	.441	.019
	E(자극)	.098	.585	.192	.003	-.182
	E(온정)	.012	.480	-.009	.621	.099
개방성	E(주장)	-.248	.593	.002	-.155	.232
	E(활동)	-.017	.694	-.161	.016	.370
	O(가치)	-.078	-.016	.199	.257	-.001
	O(감정)	.308	.408	.425	.078	.009
	O(독창)	-.230	-.009	.653	-.094	.120
우호성	O(상상)	-.032	.133	.646	-.018	-.248
	O(시도)	-.262	.218	.343	.086	-.136
	O(심미)	.034	.101	.592	.140	.064
	A(이타)	.020	-.001	.168	.737	.161
	A(겸손)	.107	-.503	-.027	.329	-.102
성실성	A(솔직)	-.280	-.380	.110	.346	.016
	A(순응)	-.064	-.388	.010	.773	-.028
	A(신뢰)	-.201	.068	-.006	.635	-.013
	A(온유)	.014	.026	.218	.502	.057
	C(성취)	.018	.397	.024	-.128	.661
	C(신중)	-.105	-.321	.064	.018	.705
	C(유능감)	-.344	.016	.096	-.024	.611
	C(절제)	-.237	.032	-.050	.086	.650
	C(체계)	-.024	-.092	-.063	-.019	.655
	C(책임감)	-.021	-.023	.000	.075	.697
		심리적 민감성	외향성	개방성	우호성	성실성
심리적 민감성						
외향성		-.279*				
개방성		.056	.372*			
우호성		-.309*	.158*	.116		
성실성		-.521*	.213*	.201	.466*	

* $p < .05$, ** $p < .01$, *** $p < .001$.

부록 3. NEO-PI-3 수정검사의 확인적 요인분석 결과

RMSEA								
χ^2	df	p-value	RMSEA	C.I.	CFI	TLI	SRMR	
6824.757	395	.00	.113	.110 .115	.702	.672	.115	
상위 요인	문항 묶음		요인 이름					
		심리적 민감성	외향성	개방성	우호성	성실성		
심리적 민감성	N(불안)	.770						
	N(분노)	.680						
	N(심약)	.849						
	N(열등)	.794						
	N(위축)	.887						
	N(충동)	.605						
외향성	E(긍정)		.702					
	E(사교)		.749					
	E(자극)		.405					
	E(온정)		.849					
	E(주장)		.679					
	E(활동)		.707					
개방성	O(가치)			.314				
	O(감정)			.691				
	O(독창)			.601				
	O(상상)			.689				
	O(시도)			.526				
	O(심미)			.711				
우호성	A(이타)				.780			
	A(겸손)				.017			
	A(솔직)				.436			
	A(순응)				.431			
	A(신뢰)				.691			
	A(온유)				.592			
성실성	C(성취)					.751		
	C(신중)					.667		
	C(유능감)					.834		
	C(절제)					.836		
	C(체계)					.559		
	C(책임감)					.749		
심리적 민감성 외향성 개방성 우호성 성실성								
요인 간 상관	심리적 민감성							
	외향성	- .533***						
	개방성		-.103***	.517***				
	우호성			.583***	.407***			
	성실성				.219 ***	.436 ***		

* $p < .05$, ** $p < .01$, *** $p < .001$.

부록 4. NEO-PI-3 최종 수정 문항들

요인	하위요인	영어 원문	최종 문항
N	열등	I seldom feel self-conscious when I'm around people.	나는 스스로가 무가치한 인간이라고 느끼지 않는다.
	충동	When I am having my favorite foods, I tend to eat too much.	물건을 너무 쉽게 사고 나중에 후회하곤 한다.
O	온정	I really like most people I meet.	나는 내가 만나는 사람들을 진정으로 좋아하는 편이다.
E	주장	I am dominant, forceful, and assertive.	반대가 있더라도 주장할 것은 주장한다.
	활력	I have a laid-back style in work and play.	몸을 움직이는 것보다 정적인 일을 할 때 마음이 편하다.
O	주장	Sometimes I don't stand up for my rights like I should.	내 주장을 제대로 하지 않아 손해볼 때가 있다.
	온정	I'm known as a warm and friendly person.	사람들은 나를 다정한 사람이라고 생각한다.
E	활력	My work is likely to be slow but steady.	한 자리에 오래 앉아 있어도 답답함을 느끼지 않는다.
	온정	Many people think of me as somewhat cold and distant.	다른 사람들과 쉽게 친해지기가 힘든 편이다.
O	온정	I find it easy to smile and be outgoing with strangers.	나는 친하지 않은 사람과도 허물없이 이야기를 나눈다.
	온정	I have strong emotional attachments to my friends.	나는 친구들을 다정하게 대한다.
O	주장	I don't find it easy to take charge of a situation.	책임이 두려워서 내 주장을 못할 때가 있다.
	감정	Without strong emotions, life would be uninteresting to me.	인생에 있어서 풍부한 감정이 없다면 재미가 없을 것이다.
O	활동	I'm pretty set in my ways.	새로운 변화나 시도를 하기 보다는 실수하지 않는 것을 더 중요시한다.
	가치	I believe letting students hear controversial speakers can only confuse and mislead them.	학생들에게 논란의 여지가 있는 강사의 이야기를 듣게 하는 것은 그들을 혼란과 잘못된 길로 인도하게 한다.
O	상상	I try to keep all my thoughts directed along realistic lines and avoid flights of fancy.	나는 막연한 환상보다 객관적인 사실에 근거해서 생각하려고 노력한다.
	감정	I rarely experience strong emotions.	최근 몇 년간 드라마나 슬픈 영화를 보고 눈물을 흘려본 적이 거의 없다.
O	가치	I believe that laws and social policies should change to reflect the needs of a changing world.	나하고 다른 가치관을 가진 사람과 만나는 게 불편하지 않다.
	활동	I like the old-fashioned methods I'm used to.	새로운 방법보다 나에게 익숙한 방식을 더 좋아한다.

부록 4. NEO-PI-3 최종 수정 문항들

(계속 1)

요인	하위요인	영어 원문	최종 문항
	가치	I believe we should look to our religious authorities for decisions on moral issues.	나는 나와 다른 가치관을 갖고 있는 사람들을 이해하려고 노력한다.
	가치	Our ideas of right and wrong may not be right for everyone in the world.	우리가 가지고 있는 옳고 그름에 대한 기준이 세상의 모든 사람들에게 똑같이 적용되지 않는다.
	가치	I believe that it's better to stick to your own principles than to be open-minded.	상황에 맞게 기준을 적용하는 것보다 일관되게 원칙을 고수하는 것이 더 낫다고 생각한다.
	상상	If I feel my mind starting to drift off into daydreams, I usually get busy and start concentrating on some work or activity instead.	공상을 하느니 차라리 낮잠을 자는 것이 낫다고 생각한다.
	감정	I seldom notice the moods or feelings that different environments produce.	내가 어떤 감정을 느끼는지는 나에게 중요하지 않다.
O	가치	I consider myself broad-minded and tolerant of other people's lifestyles.	사람들이 자신만의 생활 스타일을 가지고 있는 것을 존중한다.
	활동	On a vacation, I prefer going back to a tried-and-true spot.	아무리 신제품이 좋다고 해도 검증되지 않은 상품은 사지 않는다.
	가치	I think that if people don't know what they believe in by the time they're 25, there's something wrong with them.	도덕적 문제에 대한 결정을 내리기 위해서는 하나의 기준을 따라야 한다고 생각한다.
	심미	I enjoy reading poetry that emphasizes feelings and images more than story lines.	나는 사실 중심의 전기나 다큐멘터리 보다 비유가 많은 문학작품을 더 좋아한다.
	활동	I follow the same route when I go someplace.	어딘가를 갈 때, 전에 간 적이 있는 길로 가는 것이 마음 편하다.
	가치	People should honor traditional values, not question them.	전통적 가치는 존중되어야 한다.
	겸손	I don't mind bragging about my talents and accomplishments.	내 재능과 성취에 대해 남들이 더 알아주기를 원한다.
	온유	When making laws and social policies, we need to think about who might be hurt.	어려운 상황에 처한 사람의 일이 마치 내 일처럼 느껴질 때가 있다.
A	온유	I don't worry much about the homeless.	불쌍한 사람들에게 동정심을 갖는 것은 그들에게 도움이 되지 않는다.
	솔직	I couldn't deceive anyone even if I wanted to.	불리함을 당하더라도 거짓말은 하지 않는다.
	겸손	I'm better than most people, and I know it.	나는 자기를 좀 내세우는 편이다.
	솔직	I would hate to be thought of as a hypocrite.	속마음을 솔직히 잘 드러내는 편이다.
	겸손	I have a very high opinion of myself.	내 분야에서 나보다 일을 잘하는 사람은 별로 없다.

부록 4. NEO-PI-3 최종 수정 문항들

(계속 2)

요인	하위요인	영어 원문	최종 문항
	겸손	I feel that I am no better than others, no matter what their condition.	나는 진심으로 자신을 낮추려고 노력한다.
	솔직	I'm pretty slick when it comes to dealing with people.	원하는 것을 얻기 위해 마음에도 없는 소리를 하기도 한다.
A	순옹	I sometimes get into arguments.	나는 생각이 다른 사람들과 논쟁을 자주 하는 편이다.
	겸손	I'm a superior person.	나는 남들에 비해서 탁월한 점이 많다고 생각 한다.
	온유	I would rather be known as "merciful" than as "just."	냉정한 논리보다 인간적인 측면을 고려하여 의사결정한다.
	성취	I'm not very ambitious.	내 분야에서 최고가 되겠다는 생각을 해 본 적이 별로 없다.
C	소명	I ignore a lot of silly little rules.	덜 중요한 일은 대충대충 처리 할 때가 있다.
	성취	I don't feel like I'm driven to get ahead.	남보다 잘 하려고 노력해봐야 소용없다.
	소명	I try to do jobs carefully, so they won't have to be done again.	일을 할 때 실수없이 하려고 노력한다.