

구성체 타당도 검증과 요인분석

Accessing Construct Validity with Factor Analysis

저자 (Authors)	신현중 Hyeon-Joong Shin
출처 (Source)	한국정책과학학회보 18(2) , 2014.06, 217-234(18 pages) korean policy sciences review 18(2) , 2014.06, 217-234(18 pages)
발행처 (Publisher)	한국정책과학학회 Korean Association For Policy Science
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE02450561
APA Style	신현중 (2014). 구성체 타당도 검증과 요인분석. 한국정책과학학회보, 18(2), 217-234
이용정보 (Accessed)	한국외국어대학교 203.253.93.*** 2021/08/19 23:15 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

구성체 타당도 검증과 요인분석*

신현중(한남대학교)

본 연구에서는 구성체 타당도 검증을 위해 기존의 많은 연구들이 사용하고 있는 통계분석 기법들, 특히 주성분분석기법(Principle Component Analysis)과 탐색적 요인분석 기법(Exploratory Factor Analysis)들이 어떤 문제점을 가지고 있는지, 그리고 확인적 요인분석(Confirmatory Factor Analysis)이 왜 구성체 타당도 검증에 가장 적합한 분석방법인지에 대해 방법론적 차원에서 고찰한다. 그 외에도 본 연구에서는 기존 연구들이 구성체 타당도 검증을 위해 요인분석을 사용할 때 간과하는 경향이 있는 적정 사례 수, 측정척도의 문제, 아이겐 값의 한계 및 모형적합도 지수의 적용 등과 관련된 논의도 이루어진다. 본 연구에서 논의하는 내용들이 새로운 방법론의 창출에 대한 것이 아닌 기존 통계 이론의 내용들을 종합 정리한 것이기는 하지만, 기존 연구들이 본 연구에서 다루는 문제점들을 간과하고 있기 때문에 이에 대한 가이드라인으로 논의할 가치가 충분히 있다고 생각된다.

□ 주제어: 구성체 타당도 검증, 요인분석, 확인적 요인분석

I. 서론

최근 행정학 관련 경험연구들은 변수들의 타당도를 높이기 위해 요인분석을 실시하거나 또는 역으로 요인분석을 통해 사용하는 변수들의 타당도가 높음을 보여주려는 검증을 시도하고 있다. 특히, 설문지를 이용한 실증연구에서 이러한 경향이 크다. 비록 타당도와 요인분석이 서로 다른 경로를 통해 그 검증방법들이 각각 발전되어 왔지만 요인분석 자체가 변수의 타당도를 높이는 결과를 보여주는 경우가 많기 때문에 타당도 검증과 요인분석은 서로 밀접하게 연결된 통계기법으로 인식되고 있다.

일반적으로 사회현상의 인과관계는 대부분 특정 개념들 간 구조적 틀 속에서 설명되어지기 때문에 이러한 추상적인 개념들을 얼마나 정확하게 측정하는가가 경험연구의 가장 핵심적인 부분이다. 추상적인 개념의 본질적 특성과는 다른 측정지표를 분석모형에 포함하면 인

* 이 논문은 2014년도 한남대학교 학술연구조성비 지원에 의하여 연구되었음.

과관계 규명 그 자체에 심각한 문제를 야기하며, 이러한 특성을 반영한 측정지표라도 정확도가 떨어진다면(다시 말해서 타당도가 떨어진다면) 인과관계를 정확하게 추정하는데 한계가 있다. 그 이유는 타당도가 떨어진 측정지표의 사용은 통계분석에서 오차가 커지게 되기 때문이다. 사실, 경험연구에서의 오차발생은 사회과학에서는 매우 자연스러운 일인데, 그 이유는 추상적인 개념과 정확하게 일치하는 측정지표를 사용한다는 것이 매우 어렵기 때문이다.

특히, 인간의 특정 인식, 태도 및 성향들을 다루는 연구에서는 이러한 추상적 개념들을 하나의 측정지표로 완벽하게 측정하기는 어렵다. 따라서 복수의 측정지표들로 이러한 추상적 개념들을 측정하는 것은 바람직하다고 볼 수 있다. 복수의 측정지표를 이용해서 개념을 나타내는 데에는 여러 통계기법들이 존재하는데, 이와 같이 복수의 측정지표들이 개념을 제대로 측정하는가를 판단하는 타당도 검증을 위해 많은 연구들이 주성분분석기법(Principle Component Analysis: PCA) 내지 탐색적 요인분석(Exploratory Factor Analysis: EFA)을 사용한다.¹⁾ 특히, 이러한 검증기법은 설문조사 자료를 이용하는 최근의 연구들에서 주로 나타난다.

그러나 PCA 및 EFA를 이용해서 타당도를 검증하는 것은 자칫 그릇된 결과를 나타낸다. 본 연구에서는 기존 연구들이 사용하는 타당도 검증방법의 적실성, 즉 PCA 및 EFA의 타당도 검증의 적정성에 대해 방법론적 차원에서 살펴본다. 더 나아가 이들 분석방법에 대한 적정성 여부와는 다른 차원에서 기존 연구들이 타당도 검증에 사용하는 통계기법의 문제점에 대해서도 심도 있게 고찰한다. 예를 들어, 적정 사례 수, 측정척도의 문제, 아이겐 값의 한계 및 모형적합도 지수의 적용과 관련된 논의가 이루어진다.

본 연구에서 논의하는 분석대상은 설문자료에서 복수의 측정지표를 이용해서 잠재적 요인을 추출하고 이것의 타당도를 검증하는 과정을 거친 경험논문들을 대상으로 한다. 이들 연구들은 잠재적 요인들을 추출한 후 이 요인들을 분산분석이나 회귀분석 또는 경로분석에 사용하고 있다. 따라서 분석모형으로 구조방정식모형을 사용한 연구들은 분석대상에서 제외된다.²⁾ 그 이유는 타당도 검증은 PCA나 EFA가 아닌 확인적 요인분석(Confirmatory Factor Analysis: CFA)을 통해 알 수 있으며, 구조방정식모형은 후자의 분석기법을 이용하기 때문이다.³⁾

한편, 본 연구에서 다루는 내용들은 타당도 검증을 위한 새로운 방법론에 대한 논의가 아니라 기존 방법론에서 제시한 내용들을 종합 정리한 가이드라인이라고 할 수 있다. 그러나 기존 연구들이 본 연구에서 다루는 문제점들을 간과하고 있기 때문에 후속 연구들을 위해서 논의할 충분한 가치가 있다고 판단된다.

1) 대부분의 연구들이 요인분석을 한다고 하지만, 이들 중 상당 수가 주성분분석기법을 이용하고 있다.

2) 그럼에도 불구하고 본 연구에서 논의되는 몇몇 사항들은 구조방정식모형을 사용한 분석결과들도 유념해야 할 필요성이 있다.

3) 주지하다시피 구조방정식모형은 요인분석과 경로분석을 통합한 형태의 분석모형이다.

II. 타당도와 요인분석의 관계

1. 타당도의 종류 및 특성

타당도란 검사(test)할 때 연구자가 주장하는 것 또는 의도했던 것이 어느 정도 제대로 측정하는가를 의미한다(Brown, 1996: 231). 그러나 타당도와 관련된 이슈는 간단하게 해결되는 것은 아니다. 그 이유는 많은 사회현상들을 설명하는 이론들의 인과관계가 추상적 개념들로 이루어져 있고 이러한 추상적 개념을 정확하게 경험의 세계에서 측정한다는 것은 어렵기 때문이다. 타당도에 대한 논의는 Goodwin(1999)에 의하면 1920년대까지 거슬러 올라가지만, 우리가 흔히 타당도를 설명할 때 언급되는 내용(content) 타당도, 기준(criterion-related) 타당도 및 구성체(construct) 타당도는 Cronbach & Meehl(1955)에 의해 처음으로 사용되어졌다.⁴⁾

내용타당도는 측정도구에 포함된 내용, 즉 측정도구를 구성하는 측정지표가 측정하고자 하는 내용을 모두 측정하고 있는지를 보는 것이다. 그러나 내용타당도의 평가는 평가자에 따라 다를 수 있기 때문에 연구자나 전문가의 주관적인 판단에 크게 의존하게 됨으로 해서 본질적인 검증이 어렵다. 기준타당도는 하나의 측정도구를 사용하여 측정한 결과를 다른 기준을 적용하여 측정한 결과와 비교하여 나타난 관련성 정도를 이용하여 타당도를 검증하는 방법이다. 즉, 기준타당도는 측정도구가 어떤 기준 변수의 값을 얼마나 정확하게 예측할 수 있는가를 평가하는 것으로서 좋은 기준을 어떻게 마련하는가가 어려운 문제이다. 구성체타당도는 측정하고자 하는 추상적인 개념이 실제로 측정도구에 의하여 제대로 측정되었는지를 검증하는 방법으로 일반적으로 타당도를 말할 때 이 구성체 타당도를 의미한다. 구성체 타당도는 이론에 의해 정의되며, 많은 검증들의 누적을 통해 구성체 타당도가 높아진다. 예를 들어, 내용분석, 상관관계, 요인분석, 분산분석 및 다속성/다측정방법(multi-trait/multi-method) 등 다양한 통계검증방법을 통해 구성체 타당도의 적실성을 판단할 수 있다.⁵⁾

2. 구성체 타당도와 요인분석

구성체 타당도를 측정하는 방법들 중 요인분석이 요즘 사회과학을 포함한 교육학 및 물리학 분야 등에서 가장 많이 사용되고 있다. 요인분석은 Spearman이 1904년에 처음으로 소개한 통

4) Goodwin(1999)은 타당도에 대한 역사적 논의 전개과정을 4개 시기로 나누어 설명하고 있다. 구체적인 내용은 Goodwin(1999)을 참조하시오.

5) 본 연구에서는 각 타당도와 관련된 내용을 구체적으로 논의하지 않는다. 그 이유는 본 논문의 목적이 각 타당도의 분석방법과 그 장·단점에 대해 고찰하는 것이 아니기 때문이다. 각 타당도에 대한 구체적인 논의는 일반 조사방법론 교재들을 참조하시오.

계분석기법이며, 이 분석기법은 자료행렬(data matrix)에서 잠재적인 구조(underlying structure)를 정의하는 것이 주요 목적인 다변량 통계분석 방법의 하나이다. 이 분석방법은 요인들(factors)이라고 불리는 공통적으로 잠재적인 차원들(underlying dimensions)의 세트를 정의함으로써 많은 변수들(또는 측정척도들) 간 상호관계의 구조를 분석하는 것이다. 요인분석을 통해 연구자는 먼저 구조에서 각 차원을 식별한 후, 각 변수들이 각 차원에 의해 어느 정도 설명되는가를 결정한다(Hair et al., 1995).

구성체 타당도의 검증과 요인분석은 서로 다른 발전과정을 거쳤음에도 불구하고 두 방법은 연관성이 매우 높다. Lovring(1957)는 구성체타당도는 실질적(substantive), 구조적 및 외부적인 세 가지 요소로 이루어졌다고 주장한다. Goodwin(1999:92)는 앞선 두 가지 요소는 타당도의 ‘내부 모형’의 일부분이며 이것들은 각 차원들 간 상호관계에 대한 연구 뿐 아니라 측정의 기본적인 구조(subconstructs) 또는 하위 영역들(subdomains)을 결정하는데 중요한 역할을 수행하는데, 요인분석은 측정의 내부구조를 평가하는데 적절한 통계기법이라고 주장한다.⁶⁾

구성체 타당도에 대한 연구는 측정에 대한 (일반적으로 이론에 바탕을 둔)가설을 세우는 것이며, 가설이 측정의 구조를 다룬다면 CFA가 적절한 분석방법이다. 그 이유는 CFA는 이미 가정된 요인구조의 구체적인 내용을 확증하고 점검하는데 관심이 있어서 이 분석에는 이론을 통해 요인구조가 도출되고 이 구조를 측정지표들을 통해 검증하기 때문이다(심준섭, 2013: 169). 즉, CFA는 측정의 내부구조에 대한 정보를 제공하기 때문에 구성체 타당도를 평가하는데 많은 도움을 준다.

III. 기존 연구들의 잠재적 요인에 대한 타당도 검증의 공통점과 문제점 및 대안

1. 기존 경험연구들의 검증방법의 공통점

잠재적 요인들을 창출한 후 이 요인들에 대한 타당도 검증을 시도하는 경험연구들 중 상당수가 다음과 같은 공통점을 지니고 있다.⁷⁾

6) 타당도는 내부타당도와 외부타당도로 나누어지는데 앞선 두 요소는 내부타당도를, 나머지 하나는 외부타당도를 의미하는 것으로 보인다.

7) 본 연구에서는 잠재적 요인 추출과 타당도 검증을 동시에 시도하지 않은 연구들은 연구대상으로 하지 않는다.

첫째, 이들 연구들의 대부분은 설문자료를 이용하는데, 이는 인식 또는 태도와 관련된 자료는 설문자료와 같은 1차 자료를 통해서만 타당도를 확보할 수 있기 때문이다.

둘째, 이들 자료들은 대부분 5점 리커트 척도로 구성되어 있다(간혹 4점 또는 7점 척도도 있음). 즉 서열척도인 측정척도들로 잠재적 요인을 추출하여 타당도 검증을 시도한다.

셋째, 이들 연구들은 타당도를 측정할 때 구성체 타당도 검증을 시도하지만, 소수의 연구들을 제외하고는 대부분은 잠재적 요인(또는 추상적 개념)을 어떤 이유로 그리고 어떤 근거로 연구자가 사용하는 지표들로 측정하려고 하는지에 대한 이론적 논의가 없거나 매우 부족하다.

넷째, 이론적 논의가 없기 때문에 또는 이론적 논의의 필요성에 대한 인식 없이 PCA나 EFA를 이용하여 잠재적 요인을 추출한다.

다섯째, 잠재적 요인의 타당도는 PCA나 EFA를 통해 나온 각종 통계결과를 이용하는데, 대부분의 연구에서 아이겐 값(Eigen value), 요인부하량(factor loading), 스크리 검사(scree plot test), 총 설명분산 값 및 KMO(Kaiser-Meyer-Olkin)값을 사용한다. 위에 언급한 검증방법들 중 어떤 경험연구들은 하나만 사용하는 경우도 있고 또 다른 연구들은 복수의 검증방법을 이용해서 그들의 요인들이 타당도가 있음을 입증한다. 예를 들어, 아이겐 값은 1.0을, 요인부하량은 $\pm 0.4 \sim 0.5$ 를, 스크리 검사는 각 측정지표들 간 연결선의 급경사가 몇 개인지를 기준으로 잠재적 요인들이 형성되면 타당도가 있다고 간주된다. 때에 따라서는 잠재적 요인을 창출하기 위해 사용된 측정지표들 중 이러한 기준을 만족시키지 못하는 것들은 제외된다.

이와 같은 잠재적 요인 추출방식과 타당도 검증 절차가 방법론적으로 적절한가? 다음 절에서 기존 연구의 검증절차에 대한 방법론적 문제점과 이를 어떻게 해결할 수 있는가에 대해 구체적으로 알아보기로 한다.

2. 기존 경험연구들의 문제점과 대안

1) PCA 및 EFA에 의한 타당도 검증의 문제점 및 해결방안

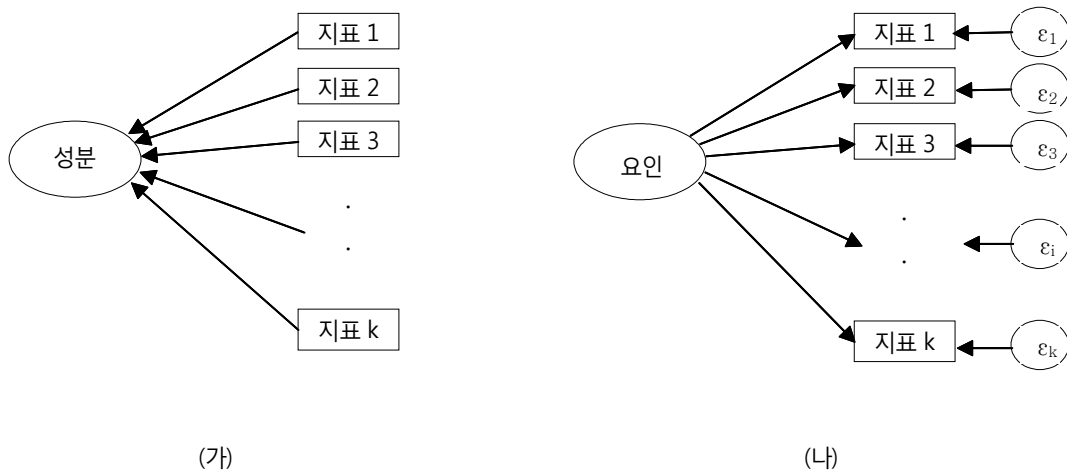
타당도 검증을 시도하는 연구들은 타당도 종류들 중 구성체 타당도 검증을 시도한다고 주장하고 있다. 내용타당도와 기준타당도는 앞서 간단하게 기술했듯이 방법론상의 많은 문제를 내포하고 있기 때문에 구성체 타당도를 통해 변수(또는 측정지표)가 제대로 개념을 측정했는가를 관찰하는 것이 다양한 학문분야에서 일반적으로 받아들여지고 있다. 구성체 타당도는 앞에서 이미 논의했듯이 내적 타당도 여부를 관찰하는 것이다.

일반적으로 측정지표들 간에 상관관계가 높은 것들끼리 하나의 무리(cluster)를 형성하는

경향성이 높으며, 이러한 무리를 통해 잠재적 요인을 추출하는 분석방법으로 PCA나 EFA가 자주 사용된다. 두 분석방법은 잠재적 요인을 추출한다는 데에는 공통점이 있으나 <그림 1>에서 보듯이 두 분석방법은 큰 차이가 있다.⁸⁾

<그림 1>에서 (가)는 PCA의 결과를, (나)는 EFA의 결과를 보여준다. 그림에서 보듯이 두 분석방법은 잠재적 요인과 측정지표들 간의 인과관계에서 상반된 모습을 보이고 있다. 또한 PCA 결과에서는 측정오차가 없는 반면에 EFA 결과에서는 측정오차가 존재한다. 두 분석방법의 차이를 조금 더 구체적으로 설명하면⁹⁾ PCA는 주어진 측정지표(또는 변수)들의 세트로부터 얻을 수 있는 정보를 요약하는데 사용되며 성분(components)의 수를 최소한도로 산출한다(Fabrigar et al., 1999). 따라서 PCA는 전체변량을 이용해서 변량을 극대화하는 성분들을 추출한다. 그러나 EFA는 자료에서 알려지지 않은 잠재적인 공통요인 내지 구조를 추정하는데 사용되며 따라서 변수들 간 변동(variation)과 상호연관성(interrelationships)을 가장 잘 설명하는 잠재적 요인을 탐색함으로써 새로운 이론을 창출하는데 도움이 된다(Henson & Roberts, 2006). 따라서 EFA에서는 전체변량보다는 공통변량으로 잠재적 요인을 추출한다.¹⁰⁾

<그림 1> PCA와 EFA의 차이



8) Blasius & Thiessen(2012)는 때때로 통계관련 문헌은 PCA와 요인분석을 혼동하는 경우가 있으며, 몇몇 통계 프로그램들에서는 PCA가 요인분석모듈에서 디폴트 방법(default method)으로 되어 있다고 지적하고 있다. 이것들 중 가장 대표적이고 빈번하게 사용되는 통계패키지가 SPSS이다.

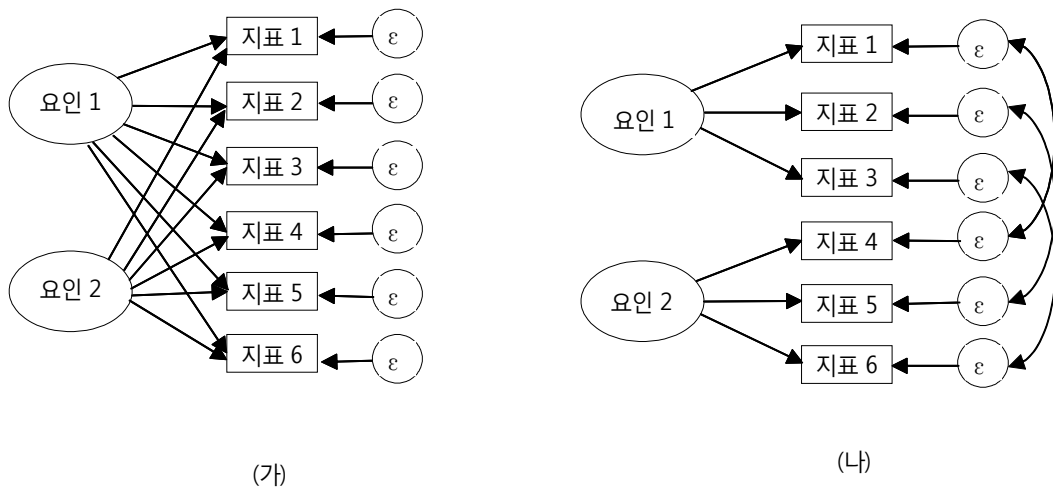
9) 본 연구의 목적이 PCA와 EFA기법 자체에 대해 구체적으로 살펴보는 것이 아니기 때문에 두 분석방법의 차이를 간략하게 기술한다.

10) 전체변량은 공통변량, 특수변량 및 오차변량의 합으로 구성된다. 공통변량은 변수들이 공유하고 있는 변량을 의미하며, 특수변량은 특정 변수만이 가지고 있는 고유의 변량을 말한다. 오차변량은 측정오차 및 표집오차 등을 포함한 변량들을 말한다(양병화, 1998, 275). PCA와 EFA의 차이에 대한 구체적인 논의는 양병화(1998: 275-276)를 참조하시오.

PCA와 EFA는 서로 다른 분석기법이지만 이론적 논의와는 상관없이 자료의 특성에 따라 특정 잠재적 요인을 창출한다는 공통점이 있다. 따라서 이러한 분석기법을 통해 나온 통계결과를 이용해서 구성체 타당도를 검증한다는 것은 방법론상으로 많은 문제를 내포하고 있다. 즉, 이것은 자료에 따라 처리되는(data-driven) 통계분석기법이나 또는 이론을 기준으로 처리되는(theory-driven) 통계분석기법이나의 선택의 문제로 귀결된다. 구성체 타당도 검증은 후자에 속하며, PCA와 EFA는 전자에 해당되기 때문에 구성체 타당도 검증과 PCA 및 EFA는 서로 양립할 수 없다. 이러한 관점에서 CFA가 타당도 검증에 적합하다.

EFA와 CFA의 가장 근본적인 차이는 이론에 기초해서 잠재적 요인과 그 지표들에 대한 가설이 도출되었는가의 여부이다(심준섭, 2013: 169). CFA는 EFA와는 달리 잠재적 요인을 도출해서 새로운 이론을 창출하려는 목적을 지닌 것과는 무관하게 이미 이론적으로 가정된 요인 구조를 확인하고 검증하는데 목적을 두고 있다. <그림 2>는 EFA와 CFA의 차이를 명확하게 보여주는데, (가)는 EFA 모형을 (나)는 CFA모형을 나타내고 있다. 물론 (나)에서 잠재변인과 공통요인 간 상관관계는(오차들 간의 상관관계를 허용) 이론적 근거에 바탕을 두어야 한다.

<그림 2> EFA와 CFA의 차이

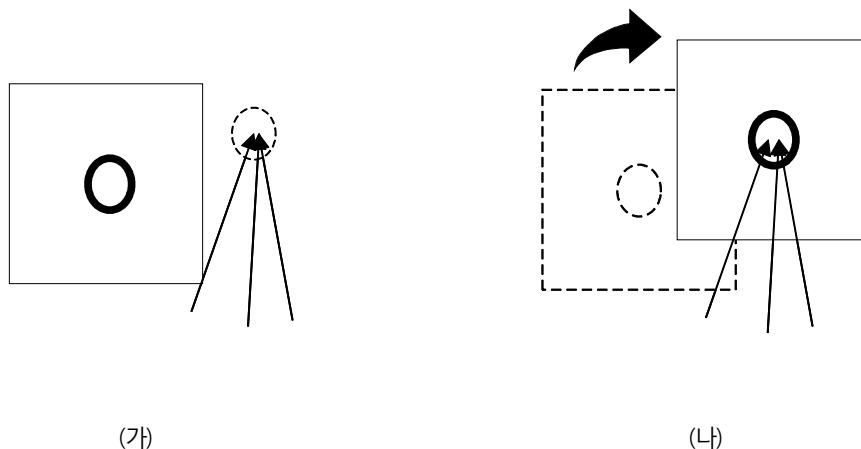


PCA와 EFA의 분석을 통한 타당도 검증 결과와 CFA의 분석을 통한 타당도 검증 결과 간에는 어떤 방법론적 차이가 있으며 전자의 결과가 어떤 방법론적 오류를 범하고 있는가? 이는 <그림 3>을 통해서 간명하게 설명할 수 있다. <그림 3>의 왼쪽 그림인 (가)에서 표적지의 중앙의 굵은 원은 연구자가 측정하고자 하는 개념이며, 화살들이 측정지표들을 의미한다. <그림 3>의 (가)는 화살들을 쫓을 때 이것들이 표적지의 중앙을 제대로 맞추지 못하고 다른 지역

에 무리를 지어 탄착점을 형성했음을 나타내고 있다. 이 그림은 개념과 지표들 간의 이론적 논의가 부족한 상태에서 자료를 중심으로 상관관계가 높은 지표들이 무리를 지은 모습을 보여주는 것이다. <그림 3>의 오른쪽의 그림 (나)는 탄착점이 형성된 곳이 중앙의 표적과 일치시키기 위해 표적지를 이동시킨 모습이다. 이것이 PCA 및 EFA를 통해 나타난 통계결과들(아이젠값, 스크리 검증 등)을 이용해서 타당도 검증을 시도한 결과를 나타낸 것이다. 즉, CFA 결과에서 나타난 다양한 통계결과들이 아닌 PCA 및 EFA에 의한 통계결과들로 타당도 검증을 시도하는 것은 그림 (나)와 같은 의미를 지니기 때문에 방법론적 오류를 범하게 된다.

즉, 특정 개념과 이것을 측정하는 측정지표들에 대한 이론적 논의가 부족한 상황에서 상관관계가 높은 측정지표들을 통해 창출된 잠재적 요인을 추출한 후 이 잠재적 요인을 원래 추정하려던 개념과 일치시키려는 것은 구성체 타당도 검증과 괴리가 있다. 이렇게 창출된 잠재적 요인이 원래 측정하려던 개념과 일치할 수도 있으나 그렇지 않을 수도 있다.

〈그림 3〉 PCA와 EFA를 이용한 타당도 검증의 문제



그렇다면 PCA 또는 EFA의 분석을 통해 잠재적 요인을 추출한 후, 이 결과를 CFA분석을 통해 잠재적 요인에 대한 타당도 검증을 시도하는 것은 방법론적 문제를 해결할 수 있는가? Matsunaga(2010)는 대상 구성체의 요인구조를 확인하기 위해서는 Thompson(2004)이 제안한 세 단계를 거치는 것이 바람직하다고 주장하고 있다: 1) PCA를 통해 측정지표들 세트를 검사와고; 2) 여기에서 걸러진 측정지표들로 EFA분석을 한 후; 3) 추출된 잠재요인들을 CFA로 최종 검증한다.

이러한 방법은 자료에 따라 처리되는 분석기법과 이론을 근거로 처리되는 분석기법을 혼합했기 때문에 일견 바람직한 것처럼 보인다. 그러나 방법론적으로 엄밀한 잣대를 들이대면

이 또한 문제가 생길 수 있는데, 그 이유는 <그림 3>의 (나)와 같이 원래 의도했던 개념 측정이 아닌 새로운 개념 내지 이론을 창출해서 이것의 타당도를 측정하는 것이기 때문이다.

따라서 요인분석을 통해 구성체 타당도 여부를 검증하기 위해서는 먼저 개념(잠재적 요인)과 지표들 간 관계를 충분한 이론적 논의와 기존 연구결과들을 통해 설정한 후, CFA 검증을 통해 구성체 타당도 여부를 결정하는 것이 바람직하다.

2) 표본 크기와 측정적도에 대한 논의

(1) 표본 크기의 적정성

요인분석을 통해 타당도를 검증할 때 표본 수는 어느 정도가 적절한가에 대한 논의는 학자마다 다르다. 사실 통계분석에서 표본 수는 많으면 많을수록 좋다고 하는 것이 일반적인 인식이다. 그러나 Osborne(2013)이 지적하듯이 일정한 통계 검증력(statistical power)을 확보할 정도의 표본 수가 확보된 이후의 사례 추가는 시간과 비용을 더 요구할 뿐이다. CFA를 이용해서 타당도 검증을 할 때 적절한 표본 수가 어느 정도이어야 한다는 논의는 전무하지만 요인분석을 시도할 때 적정 표본 수에 대한 논쟁은 학자들 간 이루어져 왔다. 따라서 CFA에 의한 타당도 검증의 적정 표본 수는 요인분석 시 적정 표본 수에 대한 논의를 통해 유추할 수 있다.

요인분석의 적정 표본 크기에 대한 논쟁은 크게 두 가지 관점에서 이루어져 왔는데, 하나는 단순히 표본 수와 관련된 것이며 다른 하나는 표본 수와 추정모수와의 비율로 적정 표본 수를 추정하는 방식이다. 전자의 경우, Cattell[1978, 양병화(1998: 284) 재인용]은 최소 200개의 표본이 있어야 요인분석의 결과를 해석할 수 있다고 주장한 반면, Gorsuch(1983)는 EFA 분석을 위해서는 표본 수가 적어도 100개 이상은 되어야 한다고 밝히고 있다. 더 나아가 Comrey & Lee(1992)는 표본수가 100개이면 부족하고(poor), 200개는 적절(fair), 300개는 좋고(good), 500개면 매우 좋으며(very good), 1,000개 이상이면 탁월(excellent)라고 구체적으로 표본 수와 요인분석결과와의 관계를 설명하고 있다.

그와 반면에 후자의 경우에는 단순한 표본 수가 아닌 표본 수와 측정지표 수와의 비율로 적정 표본 수를 산정하는데, Everitt(1975)는 이 비율이 적어도 10이어야 한다고 주장한다. 한편, Stone & Sobel(1990)은 몬테칼로 검증방법을 이용해서 10개의 변수와 2개의 잠재변수가 있는 비순차모형에서는 400개 이상의 표본이 필요하다는 결론을 내렸으며, Kline(2005) 및 Tanaka(1987)은 20배 기준으로 그 적정성을 삼았다. 아무튼 Goodwin(1999)의 분석에 의하면 학자들은 그 비율을 3배에서 50배까지 다양하게 권고하고 있다. 이와 같이 표본 수의 적정성 정도는 연구자들마다 상이하지만 양병화(1998)와 심준섭(2013)에 의하면 적어도 200개는 넘

어야 한다는데 학자들이 대체적으로 합의하고 있다고 한다.

그러나 MacCallum et al.(1999)는 이와 같은 표본 수에 대한 논의는 자료 특성에 따라 다르기 때문에 이러한 가이드라인은 잘못될 가능성이 높다고 한다. 이들은 잠재적 요인을 창출하기 위한 측정지표들 간 상관관계가 높다면 표본 수는 100-200개로 작아도 무방하며, 그렇지 않은 경우에는 표본 수가 더 커야한다고 주장한다. 이들의 주장을 받아들인다면 측정지표의 수, 측정지표들 간의 상관관계, 잠재적 요인의 수 등에 따라 적정 표본 수가 결정된다고 유추할 수 있다. 이러한 관점에서 보면 표본 수가 작다고 하더라도(예를 들어 200개보다 작다고 하더라도) 자료구조의 성격에 따라 통계결과의 신뢰성이 높을 수 있다. 따라서 표본 수가 작을 경우에는 통계 검증력의 검사를 통해 적절한 타당도 검증을 위해서 최소한 어느 정도의 표본 수가 필요한지를 관찰하는 것이 바람직하다.¹¹⁾

(2) 서열척도 사용의 적절성 여부

앞에서 요인분석을 통해 타당도 검증을 시도하는 많은 연구들의 공통점 중 하나가 설문조사 자료를 이용해서 잠재적 요인을 추출하고 이 요인에 대한 타당도 검증을 시도한다고 지적하였다. 이때 사용되는 측정지표들은 리커트 척도를 나타내는 서열척도 내지 명목척도이며, 대부분은 전자에 해당된다. 그러나 Stenvens(1946)가 주장하듯이 요인분석은 측정척도로서 적어도 등간척도를 사용해야 한다. Kim & Mueller(1978: 73)는 그 이유가 요인분석을 위해서 입력되는 기본적인 자료가 상관관계나 공분산구조이기 때문이라고 밝히고 있다. 이러한 주장은 통계이론에서 일반적으로 잘 알려진 사실이며, Blasius & Thiessen(2012)은 주성분분석에 대한 논의를 하면서 이 문제를 다시 한 번 상기시켜주고 있다.

그렇다면 요인분석을 포함해서 설문지를 이용해서 등간/비율척도에서 사용할 수 있는 통계기법(예를 들어, 분산분석이나 회귀분석 등)을 사용한 모든 연구들의 통계결과가 왜곡되었는가? 이에 대해서는 논란의 여지가 있다. King(1989: 115)은 7점 리커트 척도를 회귀분석의 종속변수로 사용한 연구들은 각 범주 간 간격이 동일하다는 가정 하에 통계분석을 하지만 이 가정이 적용되지 않으면 통계결과가 왜곡된다고 주장한다.¹²⁾ Berry(1993: 46, 86) 또한 비연속성을 띤 유권자 선호도나 개인의 이데올로기적 속성과 같은 변수들은 회귀분석에 적합하지 않다고 밝히고 있다. 따라서 엄격하게 방법론적 잣대를 들이대면 설문자료로 요인분

11) 통계 검증력은 주지하다시피 잘못된 귀무가설을 올바르게 기각하는 능력을 말한다. 통계 검증력을 관찰하기 위해서 쉽게 구할 수 있는 소프트웨어로는 G*Power가 있으며 이것은 무료로 구할 수 있다. 한편, 심준섭(2013)은 이러한 방법으로 최소한의 표본 수를 정하는 것이 정확한 모수추정에 필요한 표본크기를 정하는 데 만능 약은 아니라면서 이 방법의 무분별한 남용을 경계하고 있다.

12) King(1989)는 최대우도추정방법을 사용하면 이러한 한계를 극복할 수 있다고 주장하고 있다. 이에 대한 구체적인 방법은 King(1989: 115-117)을 참조하시오.

석을 하고 이를 통해 타당도 검증을 하는 것은 통계적 오류를 유발할 가능성이 높다.

그러나 표본 수가 크면 리커트 척도를 사용한 서열척도는 완전선형은 아니지만 단조로운 선형(monotonic linear)의 모습을 갖추기 때문에 이러한 경우에는 리커트척도를 등간척도로 간주해서 사용해도 통계결과가 크게 왜곡되지 않는다는 연구결과들도 있다. 예를 들어, Georgiadis & Manning(2012), Kerr(2011), Alesina & Giuliano(2011) 및 Luttmer & Singhal(2011)은 부의 재분배와 관련된 연구에서 OLS의 분석결과와 서열로짓분석결과가 큰 차이가 없음을 밝히고 있다.

한편, Blasius & Thiessen(2012)은 PCA의 올바른 분석방법에 대한 연구에서 지표가 서열척도인 경우에는 일반적으로 사용하는 PCA가 아닌 범주화 주성분분석(Categorical Principal Component Analysis: CatPCA)을 사용해야 한다고 주장한다. CatPCA는 서열화된 리커트 척도의 특성을 지닌 측정변수들의 잠재변수를 창출할 때 사용하는 통계분석 기법이다.

설문조사에서 응답자들이 리커트 척도의 성격을 띤 지표에 응답할 때 그들이 생각하는 범주 간 거리는 동일하지 않다. 더구나 하나의 잠재적 요인으로 묶인 여러 지표들이 동일한 척도를 지닌 경우에 각 지표의 범주들 간 거리가 동일하지 않을 가능성이 높기 때문에 통계적 오류가 발생할 가능성이 있다.¹³⁾ 특히, 국가 간 비교 시 동일 지표에서의 각 국가의 국민들 간 범주 간 거리가 더 차이가 날 수 있다. 예를 들어, World Value Survey자료에서 특정 인식에 대한 설문문항이 5점 척도로 구성되어 있을 경우, 각 국가의 사회·문화적 차이로 인해 그 측정 척도의 범주 간 간격의 변동성이 더 클 가능성이 농후하다.

이러한 경우에 PCA는 오류를 범할 가능성이 높기 때문에 CatPCA를 사용하는 것이 옳다고 한다(Balasius & Thiessen, 2012). 다만, 이들은 PCA와 CatPCA의 설명되는 분산(explained variance)의 차이가 작을 때는 설문내용의 질의 수준이 높고 이러한 경우에는 PCA를 사용해도 무방하다고 주장한다. 실제 이들은 World Value Survey자료를 이용해서 PCA와 CatPCA의 통계결과를 비교했는데, 그 차이는 크지 않았다.¹⁴⁾ 여기에서 주목해야 할 부분은 이들이 사용한 자료는 적게는 1,000개 내외에서 크게는 2,900여개나 되는 표본 수로 구성되어 있다는 사실이다. 즉, 표본 수가 크면 서열척도를 등간척도로 간주하고 사용해도 큰 무리가 없음을 유추해 볼 수 있다. 이러한 주장은 요인분석에도 동일하게 적용된다고 할 수 있다.

다만, 몇 점의 리커트척도에서, 그리고 적어도 몇 개의 표본 수로 구성되어야 통계결과와 왜곡 현상이 최소화되는지에 대해서는 여러 시뮬레이션들을 통해 알 수 있을 것이다.¹⁵⁾ 따라

13) 이러한 통계적 오류 가능성은 하나의 잠재적인 요인으로 묶인 지표들이 동일한 리커트 척도로 이루어지지 않았을 경우(두 개의 지표가 5점 척도로 구성되어 있고 나머지 하나의 지표가 4점 척도로 구성되어 있을 경우)에 더 커진다.

14) 이들은 명목척도일 경우에는 Multiple Correspondence Analysis(MCA)를 사용해야 한다고 밝히고 있다.

15) 이에 대한 연구가 전무한데, 이를 주제로 한 연구는 그 가치가 충분히 있다고 생각된다.

서 표본 수가 충분치 않은 경우에는 연구자가 서열척도를 등간척도화 해서 요인분석을 한 결과와 서열척도에 적합한 요인분석 결과를 비교해서 더 나은 통계결과를 얻은 분석기법을 사용하는 것이 옳바르다고 할 수 있다.¹⁶⁾

3) 타당도 검증을 위해 사용하는 통계방법에 대한 논의

(1) 이론적 근거에 대한 논쟁이 있을 경우 진짜 요인구조를 추정하는 통계검증 방법

특정 개념을 측정하는 지표들에 대한 이론적 및 기존 경험연구 결과가 확실하다면 CFA를 이용해서 모형 적합도 지수를 구해서 모형 수정을 가할 수 있다. 그러나 학자들 간 논쟁이 있거나 연구결과의 상이성들이 나타나서 진짜 요인구조가 무엇인지에 대한 불확실성이 존재할 경우에는 PCA가 아닌 EFA를 우선적으로 시도할 수 있다.

이때 자료의 진짜 요인구조(true factor structure)를 확인하기 위한 적절한 전략으로 일반적으로 사용하는 기준이 아이겐 값이 1.00 이상인 잠재적 요인을 골라내는 것이다. 즉, 불필요한 요인들을 쏙아내는 기준인데, 불필요한 요인들은 허위적인 잡음(spurious noise)이나 측정 오차를 나타낸다. 이것은 Kaiser-Guttman 기준에 따른 것이며, 아이겐 값이 1.00 이상이면 구성체 타당도가 적합하다고 연구자가 결정한다.¹⁷⁾ 그러나 이 기준은 매우 자의적인데, 예를 들어 아이겐 값이 0.999일 경우 이 잠재적 요인을 배제해야 하는가의 문제가 발생한다(Blasius & Thiessen, 2012: 38). Kaiser-Guttman 기준을 적용하는 문제가 자의적인 수치의 문제만은 아니다. Hayton, Allen & Scarpello(2004)가 지적하듯이 더 큰 문제는 이 기준의 적용이 잠재적 요인의 수를 과다 평가하기 때문이다. 많은 경우 스크리 검정을 통해 이 문제를 보완하려고 하지만, 이 검정 역시 잠재적 요인의 수를 과다 평가하는 경향이 있다. 이들 방법 외에 잠재적 요인의 수를 평가하는 방법에는 최소평균 편상관관계, Bartlett의 χ^2 검정, RMSEA를 기반으로 한 최대우도추정방법 및 PA(Parallel Analysis) 등이 있다.

그러나 불행하게도 이들 검증방법들은 서로 다른 검정결과들을 보여주는 경우가 흔한데, Henson & Roberts(2006) 및 Matsunaga(2010)는 이들 검증방법들 중 PA가 가장 정확하다고 주장한다.¹⁸⁾ PA방법은 요인분석을 통해 병렬 자료(parallel data)를 창출해서 이 자료를 이용하

16) Blasius & Thiessen(2012)은 PCA에 초점을 맞추고 있으나 이는 요인분석에도 적용 가능하다. 이들은 SPSS나 R을 이용해서 이러한 분석들이 가능하다고 하는데, 이 분석들과 관련된 syntax는 www.sage.com.uk/blasius에서 얻을 수 있다.

17) 이 기준은 각 측정지표(변수)들의 설명변량이 1.0이기 때문에(각 측정지표의 자체 상관계수가 1.0) 분석에서 고려된 잠재적 요인의 아이겐 값은 1.0보다 커야 한다는 데 기인한다.

18) Matsunaga(2010: 102)에 의하면 스크리검정은 Kaiser-Guttman 기준과 유사하게 잠재적 요인을 과다 창출할 뿐 아니라 연구자의 주관적 결정이 개입하게 된다고 한다. 또한 그는 Bartlett의 χ^2 검정은 표본 크기에 민감하게 반응하는 카이제곱의 유의성 검정을 기반으로 하고 있으며, RMSEA를 기반으로 한 최대우도추정

여 500-1000번 정도의 요인분석을 한 후 각 아이젠 값들의 산술평균을 구한 다음, 이 산술평균을 기준으로 원자료로부터 구한 아이젠 값이 더 큰 것을 잠재적 요인으로 간주한다.¹⁹⁾

이와 같이 개념을 어떤 측정지표(또는 변수)가 제대로 측정하는가에 대한 이론적 논쟁이 있는 경우에 EFA 방법을 시도한다고 하더라도 마지막 단계에서는 CFA를 통해 타당도 검증을 시도해야 한다.²⁰⁾

(2) 모형적합도 지수를 이용한 타당도 검증의 필요성

설문자료를 이용해서 잠재적 요인에 대한 구성체 타당도 검증을 시도하는 대부분의 경험 연구들은 모형적합도 지수에 대한 논의를 간과한다. 그러나 CFA를 이용해서 구성체 타당도 검증을 하기 위해서는 모형적합도 지수를 관찰해서 타당도 검증을 하는 것이 바람직하다. 대부분의 경험연구들이 모형적합도 지수를 이용하지 않는 이유는 이들 연구들이 CFA가 아닌 PCA나 EFA를 이용해서 타당도 검증을 시도하고 이를 위해 SPSS 프로그램을 사용하기 때문이다. 주지하다시피 SPSS 프로그램은 모형적합도 지수를 쉽게 산출할 수 없다.

모형적합도 지수를 통해서 타당도 검증을 하기 위해 주로 사용하는 통계 프로그램들에는 AMOS, LISREL 및 EQS 등이 있다.²¹⁾ 모형적합도 지수들은 각기 다른 기준에 의해 수많은 지수가 개발되었다.²²⁾ 모형의 적합도를 완벽하게 설명하는 지수는 아직까지 존재하지 않으며, 따라서 어떤 지수가 적절한지에 대한 학자들 간의 일치된 견해가 없다. 따라서 전문가들은 서로 다른 지수들을 이용해서 모형 적합도를 평가할 것을 권장하고 있다(Bollen & Long, 1993; Tanaka, 1993; Gerbing & Anderson, 1993; 신현중, 2000).

이들 지수들 각각에 대한 특성 및 적합도 여부에 대한 기준과 관련된 내용은 구조방정식과 관련된 여러 문헌에서 논의하고 있기 때문에 본 연구에서는 이에 대해서는 구체적으로 다루지 않기로 한다. 다만 여기에서는 어떤 특성을 지닌 지수들이 타당도 검증(또는 모형평가)을 할 때 바람직한가에 대해 간략하게 알아보기로 한다. 신현중(2000: 158-160)은 Bollen &

방법은 다른 방법들에 비해 상대적으로 정확하기는 하지만 최대우도추정추정량이 사용될 때에만 검증이 가능하다고 그 한계를 지적하고 있다.

19) PA검증방법에 대한 구체적인 내용은 Matsunaga(2010)을 참조하시오. 또한, SPSS에서 PA검증방법을 사용할 수 있는 소프트웨어 프로그램은 <http://www.softpedia.com/get/Others/Home-Education/Monte-Carlo-PCA-for-Parallel-Analysis.shtml>에서 무료로 얻을 수 있다.

20) EFA방법을 시도할 경우 요인들 간 상관관계를 허용하는 사각회전(oblique rotation) 방법을 사용해야 한다. SPSS는 “varimax”를 디폴트로 설정하고 있는데, 이는 요인들 간 상관관계를 허용하지 않는 직각회전(orthogonal rotation) 방법이다.

21) LISREL 프로그램의 경우 <http://www.ssicentral.com/lisrel/student.html>에서 학생용 9.0버전을 무료로 얻을 수 있는데, 구조방정식모형(Structural Equation Model)을 위해서는 16개의 변수까지 사용할 수 있다.

22) 이순목(1990)에 의하면 모형적합도를 나타내는 지수는 수십 개에 달한다고 하지만 일반적으로 사용되는 지수는 십여 개 안팎이다.

Long(1993), Tanaka(1993) 및 Gerbing & Anderson(1993)의 연구들에서 바람직하다고 주장하는 각각의 지수의 선택 기준들 중 공통된 내용을 바탕으로 다음과 같은 지수 선택의 가이드라인을 제시하고 있다: 첫째, 표본 크기에 상관없이 지수의 표본 평균값이 일정한 지수를 선택할 것; 둘째, 모형의 간명화 추구를 고려한 지수를 선택할 것; 셋째, 모형의 적합도를 비교할 수 있는 상대지수(relative index)를 사용할 것; 넷째, 모형의 추정방법에 따른 지수의 민감성이 낮은 지수를 사용할 것(이것은 다양한 모형의 추정방법과 상관없이 동일한 자료를 통계 분석하면 동일한 결과가 나와야 한다는 것을 의미함).²³⁾

IV. 맺음말

본 연구에서는 기존 연구들이 잠재적 요인들에 대한 구성체 타당도 여부를 검증할 때 방법론적으로 어떤 오류를 범하고 있는지에 대해 살펴보았다. 대부분의 기존 연구들은 잠재적 요인(또는 개념)에 대한 구성체 타당도 여부를 PCA나 EFA를 통해 검증하고 있는데, 구성체 타당도 여부에 대한 검증은 CFA를 기반으로 한 분석결과를 이용해야 한다.

구성체 타당도 여부를 검증하기 위해서는 CFA 방법을 사용해야 한다는 것은 많은 요인분석 내지 구조방정식모형 관련 교재 및 논문들에서 밝히고 있어서 새로운 사실은 아니다. 따라서 본 연구에서 다루는 내용들도 새로운 방법론적 관점에서 이 문제를 다루는 것이 아니라 기존에 이미 알려진 방법론적 내용을 정리한 수준에 불과하다고 혹자는 폄하할 수도 있다. 그러나 구성체 타당도 검증을 위해 PCA 및 EFA방법이 너무 무분별하게 사용되기 때문에 이에 대해 논의할 필요성이 있는 것도 사실이다. 경험연구의 질적 발전을 위해 통계분석방법이 상당한 공헌을 한 것은 사실이지만 특정 통계분석방법이 요구하는 자료의 특성을 파악해서 올바른 적용을 해야 통계 및 해석의 오류를 최소화시킬 수 있다.

따라서 구성체 타당도 여부를 요인분석을 통해 경험적 검증을 시도하기 위해서는 지표들과 개념(또는 잠재적 요인)들 간 이론적 논의가 우선적으로 이루어져야 하고, 이를 기반으로 CFA를 통해 타당도 검증 및 수정을 시도해야 한다. 또한, 이론적으로 논쟁이 되는 부분에 대해서는 PCA가 아닌 EFA를 통해 요인구조를 확인하고 이를 CFA로 타당도 검증을 시도해야 한다. 더구나 본 연구에서는 CFA로 타당도 검증을 시도할 때 표본 크기 및 측정척도에 대해 방법론적 논의를 했는데 분석 시 항상 유념해야 할 사안들이다.

특정 사회현상의 인과관계를 경험연구를 통해 적절한 방법으로 검증하기 위해서는 올바른 가설의 설정 및 적정 통계분석방법의 사용 등과 같은 분석모형 틀이 제대로 구축되어야

23) 각 조건을 만족시키는 적합도 지수들에 대한 구체적인 내용은 신현중(2000)을 참조하시오.

한다. 이와 더불어 사회적 인과관계를 추적하기 위해 특정 개념을 어떤 측정지표들(또는 변수들)로, 그리고 어떻게 사용할 것인가는 모든 경험연구를 시도하는 연구자들이 항상 주의를 기울이고 있음에도 불구하고 그 중요성을 강조해도 지나침이 없다.

【참고문헌】

- 신현중. (2000). “순차경로모형(Recursive Path Model)을 사용할 경우의 유의사항에 대한 논의.” 「정책분석평가학회보」. 10(1): 145-164.
- 심준섭. (2013). 「구조방정식모형: 이해와 활용」. 파주시: 법문사.
- 양병화. (1998). 「다변량 자료분석의 이해와 활용」. 서울: 학지사.
- 이순목. (1990). 「공변량구조분석」. 서울: 성원사.
- Alesina, A., and Giuliano, P. (2011). “Preferences for Redistribution.” In *Handbook of Social Economics*. Edited by A. Bissin, M. O. and J. Benhabib. pp. 93-131. San Diego, CA: North-Holland.
- Berry, W., D. (1993). *Understanding Regression Assumptions*. London: SAGE Publications Ltd.
- Blusius, J. and Thiessen, V. (2012). *Assessing the Quality of Survey Data*. London: SAGE Publications Ltd.
- Bollen, K. A., and Long, J. S. (1993). “Introduction.” In K. A. Bollen and J. S. Long. Eds. *Testing Structural Equation Models*. pp. 1-9. Newbury Park: SAGE Publications, Inc.
- Brown, J. D. (1996). *Testing in Language Programs*. Upper Saddle River, NJ: Prentice Hall Regents.
- Cattell, R. B. (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Science*. New York: Plenum.
- Comrey, A. L., and Lee, H. B. (1992). *A First Course in Factor Analysis*. Hillsdale, NJ.: LEA. Do’s, Don’ts, and How-To’s of Factor Analysis 26.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*. 52: 281-302.
- Everitt, B. S. (1975). “Multivariate Analysis: The Need for Data and Other Problems.” *British Journal of Psychiatry*. 126: 237-240.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., and Strahan, E. J.(1999). “Evaluating the Use of Sample Size, Estimation Methods, and Model Specification on Structural Equation Modeling Fit Indexes.” *Structural Equation Modeling*. 6: 56-83.
- Georgiadis, A., and Manning, A. (2012). “Spend It like Beckham? Inequality and Redistribution in the UK, 1983-2004.” *Public Choice*. 151: 537-563.
- Gerging, D. W., and Anderson, J. C. (1993). “Monte Carlo Evaluations of Goodness-of-fit Indices for

- Structural Equation Models.” In K. A. Bollen and J. S. Long. Eds. *Testing Structural Equation Models*. pp. 40-65. Newbury Park: SAGE Publications, Inc.
- Goodwin, Laura, D. (1999). “The Role of Factor Analysis in the Estimation of Construct Validity.” *Measurement in Physical Education and Exercise Science*. 3(2): 85-100.
- Gorsuch, R. L. (1983). *Factor Analysis*. 2nd ed. Hillsdale, NJ.: LEA.
- Hair, J. F. Jr., Anderson, R. E., Tatham, R. L., and Black, W. C. (1995). *Multivariate Data Analysis*. 4th ed. Englewood Cliffs, NJ: Prentice Hall.
- Hayton, J. C., Allen, D. G., and Scarpello, V. (2004). “Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis.” *Organizational Research Methods*. 7: 191-205.
- Henson, R. K., and Roberts, J. K. (2006). “Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice.” *Education and Psychological Measurement*. 66: 393-416.
- Kerr, W. (2011). “Income Inequality and Social Preferences for Redistribution and Compensation Differentials.” *NBER Working Paper No. 17701*.
- Kim, Jae-On, and Mueller, C., W. (1978). *Factor Analysis: Statistical Methods and Practical Issues*. London: SAGE Publications Ltd.
- King, G. (1989). *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Cambridge: Cambridge University Press.
- Kline, R. b. (2005). *Principles and Practice of Structural Equation Modeling*. 2nd ed. New York: Guilford.
- Loevinger, J. (1957). “Objective Tests as Instruments of Psychological Theory.” *Psychological Reports*. 3: 635-694.
- Luttmer, E. F. P., and Singhal, M. (2011). “Culture, Context and the Taste for Redistribution.” *American Economic Journal: Economic Policy*. 16(3): 367-410.
- MacCallum, R. C., Widaman, K. F., Zhang, S., and Hong, S. (1999). “Sample Size in Factor Analysis.” *Psychological Methods*. 4: 84-99.
- Matsunaga, Masaki (2010). “How to Factor-Analyze Your Data Right: Do’s, Don’ts, and How-to’s.” *International Journal of Psychological Research*. 3(1): 97-110.
- Osborne, J. W. (2013). *Best Practices in Data Cleaning*. London: SAGE Publications, Ltd.
- Stevens, S. S. (1946). “On the Theory of Scales of Measurement.” *Science*. 103: 677-680.
- Stone, C. A., and Sobel, M. E. (1990). “The Robustness of Estimates of Total Indirect Effects in Covariance Structure Models Estimated by Maximum Likelihood.” *Psychometrika*. 55: 337-352.
- Tanaka, J. S. (1987). “How Big Is Big Enough Sample Size and Goodness of Fit in Structural Equation Models with Latent Variables.” *Child Development*. 58: 134-146.

- Tanaka, J. S. (1993). "Multifaceted Conceptions of Fit in Structural Equation Models." In K. A. Bollen and J. S. Long. Eds. *Testing Structural Equation Models*. pp. 10-39. Newbury Park: SAGE Publications, Inc.
- Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis*. Washington, DC: American Psychological Association. 38: 1-10.

* 申 鉉 仲: 미국 Florida State University에서 정치학박사학위를 취득하고(논문: The Growth Patterns of Public Pension Expenditures of Sixteen Advanced Industrial Democracies in the Period of 1961-1984: Pooled Cross-sectional and Time Series Analysis, 1995), 현재 한남대학교 행정학과 교수로 재직 중이다. 주요 관심분야는 공공정책, 계량행정 및 비교정치경제학이다. 최근의 주요 논문으로는 “18개 OECD국가의 아동빈곤을 결정요인에 대한 연구: 거시적 차원에서의 복합적 인과관계 분석을 중심으로”(2012), “정부 이데올로기 측정지표들의 한계와 극복방안: 대안지표 개발을 중심으로”(2013) 및 “국민의 복지선호도와 정부 대응성의 순환적 인과관계에 대한 고찰”(2013) 등이 있다(phjshin@gmail.com).

ABSTRACT

Accessing Construct Validity with Factor Analysis

Hyeon-Joong Shin

The objective of this research is to discuss why both principle component analysis and exploratory factor analysis are not available to test construct validity even though many researchers are used to evaluate it with these methods. The study shows why confirmatory factor analysis is proper to estimate construct validity. The study, in addition, treats several issues which researchers in general tend to ignore when they use factor analysis: number of cases, measurement problem, the limit of eigenvalue, and application of goodness of fit statistics. The problems dealt in this paper are not for a new methodological review. But there is enough valuable to deal with them because researchers do not consider a possibility of serious statistical errors which may be occurred by misunderstanding them.

Key Words: Construct Validity, Factor Analysis, Confirmatory Factor Analysis