탐색적 요인분석 : 어떻게 달라지나?

Exploratory Factor Analysis : How has it Changed?

저자 (Authors)	이순묵, 윤창영, 이민형, 정선호 Soonmook Lee, Chang-Young Youn, Minhyung (Mina) Lee, Sunho Jung

출처 (Source)	한국심리학회지: 일반 35(1), 2016.3, 217-255(39 pages) THE KOREAN JOURNAL OF PSYCHOLOGY : GENERAL 35(1), 2016.3, 217-255(39 pages)
발행처 (Publisher)	한국심리학회 The Korean Psychological Association
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06648074
APA Style	이순묵, 윤창영, 이민형, 정선호 (2016). 탐색적 요인분석 : 어떻게 달라지나?. 한국심리학회지: 일반, 35(1), 217-255

한국심리학회지 : 일반
Korean Journal of Psychology: General
2016, Vol. 35, No. 1, 217-255

탐색적 요인분석: 어떻게 달라지나?*

이 순 묵† 윤 창 영 이 민 형 정 선 호

성균관대학교 대구대학교 성균관대학교 경희대학교

본 연구에서는 21세기 전후하여 새롭게 제안된 탐색적 요인분석(EFA)에 대한 지침들을 정리하고 실제 자료의 분석 예를 제시하였다. 대략적인 요인수효를 결정하기 위한 발견법(heuristics)의 내용가운데 평행성 분석에 대한 평가가 정리되었고 랜덤자료에서의 고유치로서 기존의 Horn(1965) 방식이나 주축요인방식이 아닌, 최소계수요인 방식(MRFA: minimum rank factor analysis)에서의 고유치가 권고된다. 요인수효 결정을 위한 추론적 접근에서 합치도의 참조는 카이제곱 검증뿐만 아니라 표본 크기에 영향을 덜 받는 다양한 판단적 합치도(예: CFI, RMSEA 등)를 함께 참조할 수 있고 이로 인해 요인수효 결정에서 "다양한 정보의 종합적 사용"이 가능해졌다. 여기에 서열자료 분석에 사용될 수 있는 추정법이 개발되면서, 문항점수들을 연속변수에 준하는 것으로 보고 피어슨상관을 구하여 고전적 요인분석을 하는 관행을 벗어나 문항의 범주별 반응형태를 반영하는 문항요인분석이 현실화되었다. 요인구조의 회전에 있어서는 사각구조의 추정이 용이해졌고, 임의적인 파라메터의 설정 없이 복잡도 함수만을 최소화함으로써 단순구조를 추구할 수 있게 되었다. 또한 탐색과정에서 연구자의 내용적 판단을 반영하는 목표행렬을 주고 그 방향을 따르도록 회전하는 부분제약 목표회전의 사용이 가능해져 이전의 기계적인 회전을 벗어나게 되었다. 요인구조의 해석 가능성에서 가장 큰 변화로 볼 수 있는 것은 측정오차 간 상관을 허용하는 탐색적 구조방정식 모형(ESEM: Exploratory Structural Equation Modeling)이 개발되어, EFA를 할 때 측정오차 간 상관이 없다는 종래의 강한 가정을 완화시키면서 현실적이고 해석 가능한 구조를 산출하게 되었다. 실제 자료의 분석 예시에서는 탐색적 요인분석에서 새로운 지침들이 어떻게 활용되고 있는지를 상세히 설명하고 있다.

주요어 : 탐색적 요인분석, 공통요인분석, 탐색적 회전, 목표회전, 탐색적 구조방정식 모형

---

* 이 글의 초고에 많은 조언을 주신 익명의 심사위원님들, 그리고 조영일, 이태헌, 서동기 교수님들께 감사드린다.

† 교신저자: 이순묵, 성균관대학교 명예교수, 서울시 종로구 창경궁로 265, 101-1409 호
Tel: 010-7144-1580, E-mail: smlyhl@chol.com

이 글은 최근 21세기를 전후하여 탐색적 요인분석(EFA: Exlploratory Factor Analysis)에서 발생한 현저한 변화를 소개하고, 그로 인하여 앞으로의 경험연구에서 EFA의 관행이 어떻게 달라져야 할 것인지를 실제 자료에 대한 요인분석을 통해 제시하는 것을 목적으로 한다. 일반지능에 대한 Spearman(1904)의 논문으로 인하여 지능이론은 물론 요인분석에서도 중요한 출발점이 되었다. 이 때부터 행동과학에서의 요인분석은 많은 측정치들(당시에 학교에서 보는 각 과목 시험점수) 간의 상관을 초래하는 잠재변수인 공통요인을 탐색하는 '탐색적 공통요인분석(EFA: Exploratory Common Factor Analysis)'이었다. 요인구조가 탐색이 된 후에는 새로운 자료에 확인적 요인분석(CFA: Confirmatory Factor Analysis)을 하게 되는데 CFA에서의 요인 역시 공통요인을 가리킨다. 행동과학 및 사회과학에서 요인(factor)이라고 하면 인간의 여러 행동에 효과를 미치는 공통요인을 의미한다는 자부심으로 인하여 "공통"이란 수식어를 생략하는 경우가 많다. 이 글에서도 그러한 전통을 따른다.

요인분석의 발달과정에서 단일요인이 아닌 다요인을 추정함에 따라, 모수가 많아지면서 모형식별(model identification: 각 모수가 유일한 값으로 추정)을 위해 임의적인 제약이 필요하게 되었다. 그런 임의성 중의 하나가 요인구조의 회전에서 발생한다. 회전을 통해 기초구조와 수학적으로 동등한 많은 최종구조를 산출할 수가 있지만 그 중에서 최적의 해석을 제공하는 해를 구하는 것이 EFA의 과제이다. 이러한 과제를 추구하는 가운데 다음의 세 가지 문제는 오랫동안 경험 연구자들을 괴롭혀 왔다. 즉, 회전하기 전에 기초구조를 구하기 위해서 요인의 수효를 결정하는 문제, 회전 시 제약을 표시하는 많은 분석적 기준들 가운데 어느 것을 선택할 것인가의 문제, 끝으로 Spearman 이래로 가정되어 온 "측정오차 간 상관 0"이 현실에서 위반되는 문제가 꾸준히 척도개발 및 모형 검증 시에 연구자와 사용자들을 괴롭혀 왔다.

이제 이 세 가지 문제들이 최근에 어떻게 보다 발전된 수준으로 해결되고 있는지를 살펴보고, 실제 자료에 EFA를 적용하는 분석에서 그러한 해결안들을 어떻게 활용하는지 알아볼 필요가 있다. 그래서 이 글에서는 우선 탐색적 요인분석의 개념을 간단히 제시한 후 세 가지 문제들이 최근 20여년 간에 어떻게 발전적으로 해소되고 있는지를 살펴보았다. 끝으로, 발전된 해결대안들을 실제 자료에 적용하여 분석하는 예를 제시하였다.

탐색적 요인분석의 개념

탐색적 요인분석은 행동과학에서 이론적 개념을 탐색하는 데 가장 널리 쓰이는 통계 방법이다. 넓은 의미의 요인분석에는 주성분분석과 공통요인분석이 포함되지만 이 글에서는 특별히 주성분이라고 언급하지 않는 한 "요인"은 공통요인을 의미한다. 주성분 모형과 공통요인 모형의 구분을 위해서는 이순묵(1995, 2000)이 적절한 참조가 된다. 행동과학/사회과학에서는 하나의 구성개념(요인, 이론변수, 가설적 개념, 잠재변수)을 측정하고자 복수의 측정치들을 사용하게 되면서 측정치들 간에 공유되고 있는 공통요인을 추출하는 모형이 요인분석의 모형이다. 요인 즉, 이론적 개념은 그것의 존재를 나타내는 지표(indicator)들인 측정변수(측정치) 간의 관계(예: 상관)로부터 추

정된다.

각 측정치의 행동이 수량화된 값인 분산(variance)에는 다수의 측정치에 효과를 미치는 공통요인이 설명하는 부분과 그 측정치 특유의 고유요인이 설명하는 부분으로 나뉜다. 고유요인은 측정방법에 특유한 부분과 무선오차의 부분이 포함되는데 분석과정에서는 이 둘을 묶어서 이론변수 측정의 오차 또는 줄여서 측정오차로 간주한다. 측정치는 인간의 내적/외적 행동에 일관성 있게 숫자를 매긴 결과인데, 연속변수로서의 측정치와 소수의 이론변수(요인) 간 1차식의 관계(EFA 모형에 대한 가정)를 고전적인 요인분석(Classical factor analysis, Knol & Berger, 1991; Common linear factor model, Wirth & Edwards, 2007)의 수리적 모형으로 나타내면 아래와 같다.

행동에 대한 설명모형:

$$Z_i = a_{i1}F_1 + a_{i2}F_2 + u_i \quad \cdots\cdots (1)$$

Z_i: i번째 측정변수(평균 0, 분산 1),

F: 공통요인(평균 0, 분산 1)으로서 서로 상관 가능.

u_i: 고유요인(또는 측정오차)으로서 F 또는 다른 u와 상관 "0".

a_{i1}: 요인 1이 측정치 Z_i에 대해서 가지는 요인계수.

요인분석에서 측정치 Z_i의 신뢰도는 식(1)에서의 R^2, 즉 측정치의 분산 가운데 공통요인들이 설명하는 부분(공통분, communality)이다. 만일에 p개의 측정치가 있고 m개의 요인이 있다면, 측정치 간 상관행렬 R과 요인분석에서 추정되는 모수들 간의 관계는 아래와 같다.

기초(요인)구조 추출모형:

$$R = AA' + \Psi \quad \cdots\cdots (2)$$

기초구조의 회전을 위한 변환행렬 $\boldsymbol{T}$ 도입:

$$AT = \Lambda \quad \cdots\cdots (3)$$

최종(요인)구조 산출모형:

$$R = \Lambda\Phi\Lambda' + \Psi = S\Lambda' + \Psi \quad \cdots\cdots (4)$$

R은 Z_i들 간 p×p의 상관행렬, A는 기초구조에서의 p×m 요인계수(형태계수=구조계수) 행렬, A'와 Λ'는 A와 Λ에 대한 전치행렬, T는 기초구조를 최종구조로 회전시키는 변환행렬, Λ는 최종구조에서 p×m의 요인형태계수(factor pattern loading, 요인과 측정치 간 "표준회귀계수") 행렬, Φ는 최종구조에서 m×m의 요인 간 상관행렬, Ψ는 u_i의 분산으로 구성되는 p×p의 대각행렬(회전의 영향을 받지 않음). $\Lambda\Phi = S$는 최종구조에서 p×m의 요인구조계수(factor structure loading, 요인과 측정치간 "상관") 행렬.

식 (1)은 Z라는 행동을 F_1과 F_2라는 이론적 변수(공통요인)에 의해서 설명하고 나머지는 잔차(고유요인, 측정오차)로 간주한다는 행동과학의 모형이다. 여기서 공통요인은 여러 측정치들(Z) 간에 상관을 맺어주는 이론적 변수이므로, 그 상관을 분석하여 기초구조를 추출하는 모형이 식 (2)가 된다. 즉, 상관의 대각선에서 u의 분산을 뺀 나머지인 축소상관행렬 $(R-\Psi)$에서 기초구조를 나타내는 행렬인 $\boldsymbol{A}$를 얻는 것이 기초구조의 추출이다(Preacher & MacCallum, 2003). 수학적 능률에 따라 추정된 기초구조를 해석에 편리한 최종구조로 바꾸는 변환행렬의 도입은 식 (3)에 제시되었다. 회전

은 기하적 개념이지만, 대수(algebra)적으로는 기초요인계수 행렬 A에 변환행렬 T를 곱하여 최종구조에서의 요인형태계수 행렬인 Λ를 얻는 것이고, T에 기반하여 요인간 상관 Φ를 구하는 것이다(식 (4)). 여기서 최종구조는 Λ, Φ, 그리고 이 두 행렬의 곱인 요인구조계수 행렬 S로 구성된다. 이와 달리 주성분모형에서는 식 (1)과 같은 행동과학적 모형이 없이, 측정치들을 1차식으로 결합하여 주성분을 정의하므로 측정오차의 개념이 없다. 따라서 식 (1)의 u가 없고 식(2)에서는 Ψ가 없으므로 축소상관행렬이 아닌 원상관행렬을 분석한다. 비록 특정한 조건에서는 공통요인 모형과 주성분 모형이 비슷해지긴 하지만 일반적으로 기대되는 사항은 아니다(이순묵, 2000; Bentler & Kano, 1990).

식 (2)와 (4)에서 등호의 우측은 모두 상관자료에 기초하여 추정되는 모수(parameter)들인데, 이들을 추정하는 방법으로서 크게 ML (Maximum Likelihood)법과 LS(Least Square)법이 있다. ML법은 관찰된 자료를 산출할 수 있는 가능도(likelihood)를 최대화하는 것을 기준으로 하여 모수를 추정한다. 여기서의 가능도는 통계적 개념으로서 관찰자료가 발생할 확률밀도 함수가 되는데, 관찰된 모든 변수들이 정규분포를 따른다는 가정하에 그 함수들의 정확한 형태가 주어지고 모수가 추정된다. 그러나 LS법에서는 식(2)에서 산출된 "모수" 추정치들을 다시 그 식의 "모수" 위치에 넣음으로써 연구자의 모형을 대변하는 상관행렬을 추정한다(이것을 재생산 상관행렬 또는 모형상관행렬이라 함). 그런 다음 표본상관행렬과 모형상관행렬의 차이제곱의 합을 최소화하는 것을 기준으로 하여 모수를 추정한다. LS법에서는 상관 추정치에 대한 가중치를 적용하여 요인구조의 모수를 추정하는 WLS(Weighted Least Square) 계열과 비가중의 ULS(Unweighted Least Square) 계열이 있다.

이러한 공통요인의 탐색 또는 탐색적 요인분석에서 앞서 언급된 세 가지 문제에 대한 해결을 위해 개발된 현저한 변화의 내용을 제시하고 경험자료에 활용함으로써 보다 정교한 분석을 강조하는 것이 이 글의 목적이다. 이 글의 성격상 20세기 마지막 5년부터 지난 20여년 간의 변화를 중심으로 하였기에, 전통적으로 안정된 부분에 대해서는 언급을 회피하였으며, 그런 부분에 대해서는 다른 문헌들(예: 이순묵, 1995, 2000; 장승민, 2015, Harman, 1976)을 참조할 수 있다. 또한 최근의 요인분석 전문서인 Mulaik(2010)에서 요인수효 결정 그리고 기초해 산출후 요인구조 회전의 논리에 대한 많은 부분을 잘 포괄하고 있으나 그러한 논리를 구현하는 실용적 방식에 대한 소개는 누락되어 있어 그 부분을 채워주는 것은 물론 포괄하지 못한 부분을 제시하는 것 역시 이 글에서 목적의 일부이다.

요인수효 결정

요인의 수효는 곧 요인의 모형(예: 2요인 모형, 3요인 모형 등)을 의미하며 그에 따라 기초구조를 추정하고, 다음 순서(요인구조 회전, 해석)로 진행한다. 요인수효의 결정을 위한 방식은 크게 세 가지로 볼 수 있다. 첫째로 특별한 통계적 원리보다는 직관적 논리에 의해서 제안된 발견법(heuristics), 둘째로 통계적 논리에 기반하여 개발된 추론적 접근, 끝으로 다요인이론의 초기 개척자인 Thurstone 이래로 참조되어 온 해석가능성이 있다. 요인분석 시

가장 중요한 결정이 요인수효 결정인 만큼, 가급적 많은 방법을 사용하여 종합적으로 판단할 것이 권고된다(Preacher & MacCallum, 2003). 그 많은 방법 가운데 해석가능성은 그 논의가 서술적 해설을 넘어 탐색적 구조방정식 모형(ESEM: Exploratory Structural Equation Modeling)이라고 하는 통계방법의 적용을 통해 기술적 비약이 이루어졌기에 뒷부분에 “해석가능성 논의의 확대”에서 독립시켜 논의하기로 하고 여기서는 발견법과 추론적 접근에서의 최근 변화를 소개하였다.

발견법

고유치 검사(scree test, Cattell, 1966)와 평행성 분석(PA: parallel analysis, Horn, 1965; Montanelli & Humhreys, 1976)을 가장 최근의 요인분석 전문서인 Mulaik(2010)에서 발견법으로 부르고 있는데 거기에 누적분산비율(Gorsuch, 1983; Harman, 1976)도 함께 포함해서 발견법으로 볼 수가 있다. 모두가 통계적 논리가 아닌 직관적 논리를 바탕으로 적절한 요인의 수효를 판단하는 방식이기 때문이다. 고유치 검사는 축소상관행렬의 고유치를 검토하여, 부스러기(scree) 값이라고 할 만한 것은 버리고 해석(양의 고유치는 공통분산) 대상이 될 만큼 큰 값을 가지는 고유치의 수효를 요인수효로 하는 것이다. 누적분산비율은 고유치에 의해 설명되는 공통분산의 비율이 1번 고유치부터 누적되면서 75～85%가 되기까지의 고유치 수효를 요인수효로 하자는 것이다.

그러나 고유치 검사와 누적분산비율 모두 주관성이 높다는 문제가 있어, 보다 객관적인 방법을 찾자는 데서 평행성 분석이 제안되었다. 즉, 주어진 측정치 수효와 표본크기에 의해서 얻어진 “경험자료에서의 고유치”는 특별히 내용이 들어있지 않은 “무선자료의 고유치”보다 높은 값에서 출발하는데, 별 의미가 없는 고유치에 이르러서는 두 고유치 간에 차이가 없거나 비교가 역전될 수도 있다(경험자료 고유치≦무선자료 고유치). 따라서 경험자료 고유치가 무선자료 고유치보다 큰 값이 유지되는 범위에 있는 고유치 수효를 요인수효로 하자는 것이 평행성 분석의 논리이다.

이러한 세 가지의 발견법에서 최근의 큰 변화라고 하면, 평행성 분석이 많이 연구되었고 이제는 그 기능에 대해서 충분히 알려졌다는 것이다. 발견법 가운데 거의 사용되지 않았던 이 방법이 그 동안에 사용되어 온 고유치 검사와 누적분산비율의 주관성을 보완하는데 유용하고 정확도가 높음이 보고되었다(Timmerman & Lorenzo-Seva, 2011). 평행성 분석을 Horn(1965)이 처음 제안할 때는 연속변수 자료에 주성분 분석을 하는 것을 염두에 두었기 때문에 원상관행렬의 고유치와 무선자료에서의 고유치를 비교하는 것이었다. 이것을 ‘Horn의 평행성 분석’이라고 하자. 이 Horn의 평행성 분석은 그 후 공통요인의 수효를 판단하는 데도 사용되었다(예: Finch & Monahan, 2008). 즉 Horn의 방식에서는 추출할 만한 주성분의 수효를 제공하는데, 그 수효를 그대로 공통요인의 수효로 사용하는 관행도 존재하는 것이다. 현재 상용 소프트웨어인 Mplus 7판(Muthen & Muthen, 1998-2012)의 EFA 부분에서 도입된 평행성 분석(EFA에서 PARALLEL 명령문, 연속변수 대상)도 Horn의 방식이다.

그런데 Horn의 방식을 공통요인 분석에 적용하는 것에는 이론적 근거가 약하다. 공통요인 분석은 원상관행렬이 아닌 축소상관행렬을 분석하기 때문이다. 그래서 Humphreys와 동료

들(Humphreys & Ilgen, 1969; Montantanelli & Humphrey, 1976)은 축소상관행렬에서의 고유치를 구하여 무선자료에서의 고유치와 비교하는 공통요인 분석용 평행성 분석을 개발하였다. 축소상관행렬을 산출하기 위해서, 원상관행렬의 대각선에 각 변수의 공통분 추정치로 다중상관제곱치(각 측정치를 다른 모든 측정치에 회귀시키는 회귀분석에서의 R^2)를 사용하였다. 이것은 공통요인 분석에서 모수 추정법으로서 ULS 계열인 주축요인 방식(PAFA: principal axis factor analysis)에서 사용되는 평행성 분석(PA)이다. 이것을 '주축요인 평행성 분석(PA-PAFA)'이라고 하자. 주축요인을 산출하는 방식으로 경험자료의 고유치를 구하기 때문이다. O'Connor(2000)는 연속변수를 대상으로 Horn의 방식은 물론 주축요인 방식(Montanelli & Humphreys, 1976에 본질적으로 동등한 방식)으로 평행성 분석을 하는 것을 제시하고 있다.

그 동안 공통요인 분석을 위한 평행성 분석에서 Horn의 방식이 가지고 있는 약한 이론적 근거에도 불구하고 PA-PAFA와 경쟁을 해왔으나 어느 것도 완벽하지는 않다. Horn의 방식은 주성분 모형을 염두에 둔 것이기에 공통요인 분석에 적용하는 이론적 근거가 약하고, PA-PAFA는 공통요인 모형을 염두에 둔 것이긴 하나 요인수효를 과다 추정하는 문제가 있어왔다. 이 문제에 대한 이유중 하나는 주축요인 방식으로 기초구조를 구할 때 축소상관행렬의 고유치 가운데 음수가 나오는 것이다. 경험자료 분석에서 음수인 고유치가 있으면 양수인 고유치는 상대적으로 큰 값들이 많이 나오게 되고 그에 따른 PA-PAFA에서 요인수효가 큰 값으로 나온다.

그리고 지금까지의 평행성 분석에 대한 연구는 측정변수가 연속변수일 때를 대상으로 한 것인데 행동/사회과학에서 요인분석 시에 측정치는 연속변수라고 할 수 있는 검사(척도) 점수를 사용하는 경우가 없지는 않으나, 오히려 평정척도의 1개 문항(예: 2점척도, 3점척도, 4점척도 등으로 된 1개 문항)으로 구성되는 경우가 매우 많고 이것은 연속변수가 아니다. 그에 따라 연속변수가 아닌 서열변수 자료를 대상으로 하는 평행성 분석도 필요하다. 최근에 Timmerman과 Lorenzo-Seva(2011)는 경험자료의 고유치 산출시에 양수로 제약되는 최소계수 요인분석(MRFA: minimun rank factor analysis)[1] 방식에서 산출된 고유치를 사용하여 서열자료에 대한 평행성 분석을 연구하였다. 그런데 경험자료 고유치의 설명분산과 무선자료 고유치의 설명분산을 비교할 때, 무선자료 고유치는 하나의 값이 아니라 확률분포를 가질 수가 있고 통상은 평균값 또는 95백분위수를 구하여 경험자료 고유치의 설명분산과 비교한다.

서열척도 자료에 대한 Timmerman과 Lorenzo-Seva(2011)의 연구에서 평행성 분석을 위한 Horn의 방식, PA-PAFA, 그리고 PA-MRFA간 비교를 목적으로 모의실험한 결과 PA-MRFA가 가장 좋은 수행을 보였는데, 서열변수 자료에서 특수상관(2, 4, 5점 척도였고 다분상관)을 구하여 평행성 분석을 하는 것이 단순히 연속

1) 통계학에서는 행렬의 rank(독립적인 벡터의 수효)를 계수(階數)로 번역한다. coefficient에 대한 번역도 계수(係數)지만 한자가 다르다. 최소계수요인분석(ten Berge & Kiers, 1991)은, 나중에 표1에서 제시되는 기초해 산출방식인 ML(최대가능도)과 LS(최소제곱)계열의 추정법과 다른 제3의 추정법인데 아직 널리 쓰이지는 않지만, ML이나 LS 방식에서 경험자료의 고유치가 음수가 나올 수 있는 것에 대한 대안으로서 모든 고유치가 양수가 되도록 한 기초해 추정방식임.

변수처럼 간주하고 피어슨 상관을 구하여 평행성 분석을 하는 것보다 좋은 수행을 보였다. 단, 특수상관(포괄적으로 “다분상관”으로 부를 수 있음)은 계산과정에서 수렴이 안 되는 경우가 많아서(수렴이 전체에서 37%) 어쩔 수 없이 피어슨 상관을 구해서 평행성 분석을 해야할 경우도 많이 있다. 따라서 경험자료 고유치와 비교할 때 다분상관에서 무선고유치 분포의 “95백분위수”를 사용하거나 그것이 수렴을 안 하면 피어슨 상관에서 무선고유치의 “평균값”을 사용하는 것이 추천되었다. PA-MRFA를 포함하여 세 가지 평행성 분석(연속변수 및 서열변수 대상)은 무료 소프트웨어인 Factor 10.3(Lorenzo-Seva & Ferrando, 2015)에서 제공되고 있다.

추론적 접근: 모형의 합치도 참조

요인수효 결정에서 발견법은 요인수효의 대략적인 범위를 파악하는 데 유용하다. 그러나 일단 범위가 파악된 후에는 추론적 접근으로써 모형을 설정하고, 검증용 합치도(χ^2)와 판단적 합치도(RMSEA, CFI, TLI, 연속변수에 SRMR, 범주변수에 WRMR)를 참조하여 요인의 수효결정을 할 수가 있다. EFA에서 제공되는 합치도는, EFA보다 더 일반적인 확인적 요인분석(CFA: Confirmatory Factor Analysis)을 위해서 개발된 것들을 사용하고 있어서 EFA 맥락에서 어떻게 기능할지 좀더 많은 시험을 거쳐야 하지만, 기초구조를 구하기 전 요인수효 결정에서 참조할 것이 권고된다(Fabrigar, Wegener, MacCallum, & Strahan, 1999; Lee, 2010). 따라서 Hu와 Bentler(1999)의 엄격한 기준(연속변수 자료에 대한 연구였음, RMSEA는 .06이하, CFI와 TLI는 .95 이상, SRMR은 .08 이하)도 있으나, Vandenberg와 Lance(2000)의 전통적 기준들도 (RMSEA는 .08 이하, CFI와 TLI는 .90 이상, SRMR은 .10 이하) 참조함이 바람직하다.

추론적 접근에는 발견법에 더하여 이렇게 다양한 합치도가 있어서 요인수효 결정에 사용될 수 있음은 21세기 요인분석의 새로운 면모가 될 것이다. 또한 Preacher와 MacCallum (2003)이 요인수효 결정 과정에 대해서 강조한 ‘다양한 기준의 종합적 사용’을 실천할 수 있는 여건이 되었음을 의미한다. 일단 요인의 수효가 결정되면 곧바로 기초구조를 산출하면서 모수가 추정된다(식(2)에서 A의 원소들이 추정됨). 그런데 이 추정에도 역시 지난 20년간에 현저한 발전이 있어 고전적으로 실시된 EFA를 초월하는 원동력이 되고 있기에, ‘요인의 수효결정’에 연결된 내용이긴 하나 별도의 큰 제목으로 다룬다.

모수 추정법의 다양화로 고전적 EFA에서 문항점수 EFA로 발전

요인분석의 과정을 크게 둘로 나누어 기초해 산출(factor extraction, factoring)과 요인구조의 회전(factor rotation)이라고 하면 기초해 산출(모수 추정 포함)이 수학적인 기본 절차이고 요인구조의 회전은 이미 추정된 기초해를 보다 해석이 용이한, 그러나 수학적으로는 기초해와 동등한 성질의 동치해(equivalent solution)로 유도하는 과정이다. 따라서 요인분석의 이론서에서는 좁은 의미의 요인분석(factor analysis)이라고 하면 요인추출(factoring)을 의미하고 회전은 별도로 서술하는 경우도 있다(예: Harman, 1976; Mulaik, 2010).

표 1을 보면 EFA는 크게 고전적 EFA와 문

표 1. 기초해 산출 방식의 분류 및 모수 추정법

분류	방식(approach, method)	원자료 (측정치)	EFA용 분석자료	모수 추정법[a]	추정시 가정[주의]
고전적 EFA (검사점수 요인분석) (선형 요인분석)	연속변수 방식	검사(척도) 점수	피어슨상관	**ML**[b], ULS,	ML: 검사점수 간에 다변량 정규분포. [표본이 작거나, 분포가 크게 편포(skewed), 또는 부요인(weak factor)의 존재가능시 ULS 사용]
	연속변수 근사방식	연속변수에 근사한 문항점수	피어슨상관	**ML**[b], ULS,	ML: 문항점수들의 분포가 다변량정규분포에 근사. [문항의 척도에서 눈금수효가 충분히 많고, 문항점수 분포에서 봉우리가 하나이며 대칭에서 크게 이탈하지 않을 것이 필요조건. 이 조건이 안되거나 표본이 작으면 ULS 사용]
문항점수 EFA (문항 요인분석) (비선형 요인분석)	제한정보 가중방식	문항 점수	다분상관 및 가중치행렬	**WLSMV***, WLSM*, WLS*	①문항점수는 반응과정에 잠재하는 연속변수인 잠재반응변수에서의 값을 경계값들(thresholds)에 따라서 범주화(예: 2점척도, 3점척도 등)한 것이다. ②잠재반응변수 쌍들이 이변량 정규분포라는 가정하에 다분상관 추정 [그 상관의 추정이 수렴 안할 수 있음.][표본이 크거나, 편포가 크지 않을 것]
	제한정보 비가중방식	문항점수	다분상관	**ULSMV***, ULS	①문항점수는 반응과정에 잠재하는 연속변수인 잠재반응변수에서의 값을 경계값들(thresholds)에 따라서 범주화(예: 2점척도, 3점척도 등)한 것이다. ②잠재반응변수 쌍들이 이변량 정규분포라는 가정하에 다분상관 추정 [그 상관의 추정이 수렴 안할 수도 있음.][표본이 너무 작지않고 모형오류가 크지 않을 것]

주:

[a] 분석용 소프트웨어로서 Mplus를 사용하는 경우의 추정법들을 제시하였음.
Mplus의 경우 ANALYSIS 블록에서 "TYPE=EFA; ESTIMATOR=추정법;"을 표시
각 방식에 대한 추정법들 가운데 맨 앞에 굵은 글씨로 쓴 방식이 선호되며, 추정법의 우측에 * 표시는 상관자료 아닌 원자료를 입력할 때만 가능한 방법.
[b] 고전적 EFA에서는 ML 과 ULS 만이 사용 가능.

항점수 EFA로 나뉜다. 고전적 EFA라면, 다문항으로 구성된 검사에서의 점수를 기반으로 하여 표준화된 연속변수를 가정하고 앞서의 식 (1)~(4)에서 제시된 전통적인 선형적(linear, 종속변수와 독립변수간 1차식 관계) 요인분석이다. 추정법이 발달하지 않았을 때는 연속변수에 근사한 문항점수에 대해서도 연속변수로 간주하고 고전적 EFA를 적용했으며 그 전통은

지금도 강하게 남아있다. 그러나 문항점수를 기반으로 하는 비선형적 문항요인분석이 1970년대부터 제시되었고 이제는 널리 사용되는 정착단계로 볼 수 있다.

표 1에서 모수 추정법의 세로줄을 보면, ML 계열의 추정법으로는 ML법만이 제시되어 있으나 LS 계열의 추정법으로는 ULS 계열과 WLS 계열이 제시되어 있다. 일반적으로는 이상에서 소개한 ML법과 LS법 사이에 ML법에 가까운(정규분포성이 약하게 가정) GLS (Generalized Least Square)법이 있으나 현실적으로 ML이 강내성(robust)있고 추정이 더 정확하므로 GLS를 사용할 자료라면 ML을 권하게 되며(예: Olsson, Foss, Troye, & Howell, 2000), Mplus에서 “ANALYSIS” 블록에서 탐색적 요인분석을 위해 “Type=EFA” 를 하면 GLS는 사용되지 않는다. 최근 20여년 간에 현저하게 발전된 EFA의 분석을 충분히 수용하는 소프트웨어로는 Mplus가 거의 유일하므로 이 글에서도 Mplus에서 사용하는 추정법 용어들을 중심으로 설명하였다.

고전적 EFA

고전적 EFA에는 측정치가 연속변수인 연속변수 방식과, 연속은 아니지만 연속변수에 근사하여 편의상 연속변수로 간주되는 연속변수 근사방식이 있다. 연속변수 방식에서의 원자료는 다문항척도(검사)점수가 되며 식 (2)에서의 상관은 전형적인 피어슨(적률)상관이다. 모수를 추정하기 위한 추정법으로서 최대가능도법(ML: Maximum Likelihood Method)과 비가중최소제곱법(ULS)이 사용된다.

연속변수 근사방식에서 원자료는 문항점수가 되며, 단일문항의 점수를 식 (1)에서의 종속변수(y)로 사용하기 위해서는 눈금의 수효가 충분히 많아야 하고 분포에서 봉우리가 하나이며 대칭에서 크게 이탈하지 않을 것이 필요하다. 이 때 서열 변수의 눈금이 적어도 몇 개 이상이어야 하는가에 대해 오랜 연구가 있었다. 예로서, 4점 척도 이상이면 무난하다는 견해가 있다(예: Bentler & Chou, 1987). 그러나 최근의 문헌 고찰에 의하면 5점척도 이상일 때를 권하고 있다(Bovaird & Koziol, 2012). 한편 5점척도라 해도 실제 응답 시에 사용이 안 되는 눈금이 있으므로 좀 더 보수적인 견해를 취하는 경우도 있다(예: Millsap & Kim, in press 에서는 7점척도 부터를 연속변수에 근사하다고 본다). 따라서 서열변수 자료에서 피어슨 상관을 구하고자 할 때 연구자는 자료의 성질을 충분히 알고서 결정해야 할 것이다.

자료가 충분하고 모형이 정확한 경우 ML법과 LS계열간 추정결과가 동일하진 않아도 매우 유사하게 된다. 한편 ML법은 사용에 대한 조건을 충족시켜야 하지만 LS 계열은 그 조건들이 덜 맞는 경우에도 무난하게 사용될 수 있다. 예로서 연속변수 EFA에서 표본이 작거나, 측정치가 크게 편포되어 있거나, 주요인(major factors)뿐 아니라 부요인(weak factors, 요인계수가 0.4~0.5 정도, Briggs & MacCallum, 2003에서 측정치가 연속변수인 경우에 대하여 연구)까지 있는 경우 ULS가 더 좋은 수행을 보인다. ULS를 연구자에 따라서는 OLS(ordinary least square)라고도 부르며(예: Briggs & MacCallum, 2003; MacCallum, 2003), 그 논리에 대한 가장 오래된 계산적 접근은 반복주축요인법(IPAF: iterated principal axis factoring) (Thomson, 1934)이다.[2] 그러나 이 방법의 문제

2) ULS와 IPAF를 각기 다른 방식인 것으로 언급하는 경우가 있는데(예: Knol & Berger, 1991) 실제

는 반복과정에서 공통분을 최대값인 1.0 이상으로 과대추정하는 헤이우드사례가 발생한다는(ML법에서도 발생) 것이고, 그걸 피하고자 ULS법의 또 다른 계산적 접근으로 최소잔차법(MINRES: MINimizing RESiduals)(Harman & Jones, 1966)이 개발되었다. 그러나 IPAF와 MINRES는 거의 같은 결과를 가져오므로(Knol & Berger, 1991) 실제에서는 계산과정이 상대적으로 단순한 반복주축요인법이 더 널리 사용된다(주축요인을 1회 추출로 제한하는 단일주축요인법은 앞서 요인수효 결정에 대한 "발견법"의 절에서 언급된 축소상관행렬의 고유치 산출, 누적분산비율 계산, 및 PA-PAFA 실시에서 주로 사용). 끝으로 연속변수 EFA에서 추정법으로서 ML과 ULS는 요약자료(피어슨상관)만으로도 가능하다. ML은 통계적 가정이 있어서 추정치에 대한 표준오차 및 모형에 대한 많은 합치도를 제공하지만 ULS는 그렇지 못하다.

문항점수 EFA

표 1에서의 문항점수 EFA에는 제한정보 방식이 제시되어 있는데 이것은 현실적으로 잘 쓰이지 않는 완전정보 방식에 대비된 명칭이다. 완전정보 방식에서는 모든 문항들에 대한 반응정보(문항수 p 개면 p차원의 빈도표)를 반영하여, 상관행렬을 추정하지 않고(요인분석의 모형이 다름) ML법으로 추정한다(Bock & Aitkin, 1981; Bock, Gibbons, & Muraki, 1988). 반면에 제한정보 방식에서는 문항쌍별(pairwise)로 정의되는 2원 빈도표에서의 반응정보를 사용하여 다분상관을 추정하고 식 (2)~(4)의 모형에 따라 LS법으로 모수를 추정하는 것을 가리킨다(가중방식에서는 다분상관+가중치행렬이 분석됨). 완전정보 방식은 다차원 문항반응이론에 기초하여 모수가 추정되며 이 글의 주된 내용인 요인분석의 전통이 아닌 문항반응이론의 전통에서 유래되었고 그 두 전통의 이론을 함께 다루기에는 이 글의 범위를 넘기에 표 1에서 제외하였다.

제한정보 방식에서는 식 (1)의 Z_i 가 연속변수인 잠재반응 변수이며 이 변수들 간의 상관인 특수상관을 구하여 거기에 식 (2)~(4)의 모형을 적용시키면 모수가 추정된다. 이 글에서는 특수상관으로서 사분상관(tetrachoric correlation), 다분상관(polychoric correlation), 이연상관(biserial correlation), 및 다연상관(polyserial correlation)을 포함하여 다분상관으로 부르기로 한다.

측정치가 평정척도(2점척도 이상)로 된 단일문항이라면 식 (1)의 Z_i는 반응과정에서 응답자의 마음속에 존재하는 잠재반응변수로서의 연속변수이고 실제 관찰변수는 아래와 같은 문항점수 x이다. x는 1, 2, 3, 4의 값이 매겨진 4점척도라고 하자.

$$\begin{cases} x=1 \Leftrightarrow Z<\tau_1 \\ x=2 \Leftrightarrow \tau_1 \le Z<\tau_2 \\ x=3 \Leftrightarrow \tau_2 \le Z<\tau_3 \\ x=4 \Leftrightarrow \tau_3 \le Z \end{cases} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (5)$$

여기서 τ는 경계값이고 관찰된 문항점수 x는 개인의 반응과정에서 잠재반응변수의 값이

로 Knol과 Berger(1991)에서도 두 방식은 동일한 결과를 가져왔고, 통계 소프트웨어 SAS에서의 Proc Factor에서도 M=PRINIT(IPAF를 실시)와 M=ULS를 주고 돌렸을 때 동일한 결과를 얻는다. 즉, 소프트웨어에서 요인추출 방식으로 ULS라고 할 때 실제로 계산용 알고리즘으로는 IPAF 또는 그에 아주 유사한 것이 들어있는 것으로 보인다.

어떤 경계값의 아래에 또는 위에 있는가에 따라서 매겨진 서열변수이다. 따라서 고전적 EFA가 식 (1)~(4)로 모형화된다면, 문항점수 EFA는 거기에 추가로 식 (5)가 있어야 완결이 된다. 즉, 문항 x에서의 반응자료(두 문항일 경우 두 문항간 2원 빈도표)에 기초하여 반응과정에 존재함직한 잠재반응변수 및 그 연속선상에 있는 경계값들을 추정하고 나면 다분상관이(또는 상관추정치의 가중치행렬까지) 추정되고 거기에 식 (1)~(4)의 전통적 EFA 모형을 적용시키는 것이 문항점수 EFA이다. 식 (1)~(4)만 보면 선형모형이지만 식 (5)가 추가되면서 측정치 x와 요인 F 사이에는 비선형적 관계가 되므로 문항요인분석은 비선형적 요인분석이 된다.

다분상관을 추정하는 과정에서 문항쌍별로 이변량 정규분포가 가정되는데, 이것은 ML법으로 추정하는데 필요한 다변량 정규분포보다 완화된 조건이므로 문항점수 EFA에서의 분석자료에 추정법으로서 ML법이 사용될 수 없고 LS법이 사용된다.

다분상관에 LS법을 적용할 때 다분상관행렬의 원소들에 대한 가중치행렬을 적용하면 가중방식이 되어 WLS 계열의 추정이 되고 적용하지 않으면 비가중방식으로서 ULS 계열의 추정이 된다. 그런데 WLS에 적용되는 가중치행렬의 크기는 문항수의 증가에 대해 폭발적으로 커지는 문제가 있다. 문항수효가 p이면 상관행렬내 원소의 수효는 $u=p(p+1)/2$가 되고 이 원소들에 대한 가중치행렬의 크기는 $u \times u$가 된다. 문항이 10개라면 가중치행렬은 55x55가 되겠으나 20개 문항이면 210x210의 크기가 되고 계산해야 되는 가중치의 수효는 22,155개가 되어 매우 큰 표본이 아니면 정확한 추정에 큰 어려움이 있다. 따라서 표 1에서 추정법으로서의 WLS를 실제로 사용하기는 어려워진다 (Wirth & Edwards, 2007).

이러한 문제에 대한 현실적 대안으로 가중치행렬의 전체 원소가 아닌 대각선만 사용하는 방식(Christofferson, 1975; Joreskog & Sorbom, 1989에서 “DWLS”로 명명; Muthen, 1993에서 “robust”한 방법이라고 소개; Muthen., du Toit, & Spisic, 1997에서 “robust WLS”라고 명명)이 사용되었다. 그런데 가중치행렬의 대각선만 사용하는 것은 최적이 아니므로 그에 따른 χ^2 통계량도 부정확하게 되어 평균을 수정하게 되었다(표 1에서 WLSM의 M은 평균수정의 의미를 명시한 것). 나아가 분산까지 수정을 하면 WLSMV가 된다(자세한 논리는 Muthen et al., 1997 참조).

Muthen등(1997)의 시뮬레이션 연구에서 WLSM을 사용한 χ^2 검증에서는 참모형을 너무 자주 기각하는 문제가 있어서 권하기 어려운 반면에, WLSMV는 모수 추정과 χ^2 검증에서 수행이 좋았다. 요인분석에서 WLSMV 사용시 자료의 분포가 대칭이면 n=200의 작은 표본에서도 모수 추정과 χ^2 검증의 수행이 무난하였다. 또한 표본이 크면(예: n=400을 초과) 어느 정도의 편포에도 모수추정과 χ^2 검증의 수행이 무난하였다. 그러나 자료의 분포에 크게 편포가 있으면서 표본크기가 작은 경우에 WLSMV를 사용하는 것은 피해야 할 것이다.

문항점수 EFA에서 원자료의 분포에 크게 편포가 있거나 표본크기가 충분하지 않으면 다분상관 추정치에 대한 가중치를 구하기가 어려우므로, 다분상관만을 분석자료로 하는 ULS계열의 추정법을 사용해야 할 것이다. 다분상관에 ULS계열의 분석방법을 적용하는 연구는 아주 최근에(Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Lee, Zhang, & Edwards, 2012; Lorenzo-Seva & Ferrando, 2015) 시작되고 있어서 참고할 문헌이 많지 않은 상태이다. WLS계열의 추정법이 상대적으로 더 정교하고 그에 대한 연구가 집중적으로 이루어졌으나, 심리학의 문항점수 자료는 표본크기가 작은 경우도 매우 많고 분포가 대칭에서 크게 이탈하는 경우도 많아서 항상 WLS계열의 추정법을 사용할 수 있는 것은 아니다. 이 때는 ULS 계열의 추정이 유일한 대안이 된다.

Lee, Zhang, 및 Edwards(2012)에서는 대칭인 분포(5점척도의 서열변수)에서 표본크기를 100, 228, 400, 1000으로 하고 모집단 모형이 5요인 모형인 경우에 대한 시뮬레이션 연구를 하였다. 그 연구에서 가중치행렬이 없는 다분상관 자료에 ULS 추정법(ULSMV는 사용되지 않았음)을 적용하고 추정(모수추정치, 표준오차 추정치, 신뢰구간)의 정확도를 검토하였다. 연구의 잠정적 결론으로서 요인계수에 대한 추정은 표본크기가 228 이상일 때 편향(bias)이 수용할만한 정도로 작고 추정치들의 95% 가량이 95% 신뢰구간에 포함되었다. 요인 간 상관은 참값보다 약간 작게 나오는 경향이 있었으나 표본크기가 228 이상일 때 대체로 수용할 만한 추정의 정확도를 보였다. 이 시뮬레이션 연구에서 시사하는 점은 표본크기가 아주 작은 경우(예: n=100)가 아니면 ULS 추정은 좋은 선택이라고 할 수 있다. 물론 편포가 없는 대칭분포의 자료였으므로 편포의 정도가 어느 정도가 되면 다분상관에 대한 ULS 추정이 부정확한 결과를 가져올 것인지 아직은 의문이다. Forero 등(2009)에서는 편포도가 1.5이상인 경우 추정이 부적절할 수 있음을 시사하고 있다. 또한 ULSMV(원자료 입력시 사용가능)로 추정할 때 제공되는 χ^2 검증의 정확도에 대한 것도 연구되어 있지 않다. 그러나 ULS계열을 사용하려면 적어도 다분상관 추정치의 신뢰도(안정성)를 보장할만한 정도의 표본크기는 되어야 할 것이다. 예로서, 표본크기 200이 안 되면 다분상관의 안정성은 많이 저하된다(Chen & Choi, 2009). 이 때 다분상관의 계산에서 수렴이 안 되는 경우가 자주 있고 (Lorenzo-Seva & Ferrando, 2015), 그에 기초한 요인분석 역시 부정확해 질 것이다. 따라서 표본이 작거나 편포가 큰 자료에 ULS계열의 추정법을 사용하려면, 표본이 너무 작지는 않아야 하고 정보의 손실을 최소화하기 위해 응답에서 결측치가 최소화되어야 하며, 끝으로 모형에 대한 오류(예: 요인수효, 요인간 상관유무, 측정오차간 상관 "0"등에서의 오류)가 크지 않아야 할 것이다.

탐색적 요인분석에서 모형의 명세(specification)라고 하면 요인의 수효, 요인간 상관 유무, 그리고 측정오차간 상관 "0"이라고 할 수 있다. 이 중에서 요인의 수효는 다양한 방법들을 적용하고 종합해서 결정해야 할 것이며(Preacher & MacCallum, 2003), 요인 간 상관은 사각회전을 실시함으로써 자료내 상관의 정도를 반영하는 추정을 한다면 모형오류를 방지할 수 있다. 그러나 EFA의 중요 가정인 측정오차 간 상관 "0"은 척도 내 문항 간에 동일방법을 공유할 경우(예: 조사가 동일, 핵심단어가 유사, 동일맥락 속에서 인접된 위치 등) 성립되지 않고 모형오류가 된다. 그럼에도 불구하고 이것은 모형의 식별(모든 모수가 유일한 값으로 추정)을 위해서 반드시 가정해야 하는 부분이므로 연구자는 문항들에서 동일방법을 공유하는 정도가 적은 것들만으로 척도를 구성하거나, 피할 수 없다면 이 글의 "해석가능성의 확대"에서 제시하는 탐색적 구조방정식 모형(ESEM)을 적용하여 측정오차간 상관을 감안하

면서 EFA를 할 수 있는 방안을 강구해야 할 것이다.

이상과 같이 21세기에 와서 잘 준비된 문항점수 EFA의 기여는 모든 반응자료에 대하여 본질적으로는 연속변수임을 가정하는 고전적 EFA의 범위를 초월하게 하는 것이다. 따라서 심리학에서 문항을 측정변수로 하여 수집된 많은 자료에 대한 EFA를 할 때 고전적 EFA가 지니는 제약을 벗어나는 대안으로서 문항점수 EFA가 적절하다. 이제 그런 자료에 대한 요인분석을 통해서 다른 이론변수들과의 관계를 보고자 하는 경우, 문항점수 EFA의 일상적 사용을 촉구하는 시점이 되었다. 이것은 많은 심리학 자료에 대한 EFA가 고전적 요인분석에서 문항요인분석으로 일대 전환을 해야 함을 의미한다.

요인구조의 회전

사각회전의 일반적 사용을 위한 여건 성숙

기초요인구조를 회전할 때 요인 간 직각(상관 0)을 유지하면서 진행하는 직각회전과, 사각(상관이 0이 아님)을 허용하면서 진행하는 사각회전 가운데 후자가 더 현실에 적절하다는 이야기는 아주 오래되었다. 현실에서는 대부분의 요인들은 상관이 되어있기 때문이다(Browne, 2001; Cattell & Dickman, 1962; Cattell & Gorsuch, 1963; Guilford, 1981; Thurstone, 1947). 그러나 몇 가지 이유로 그 동안 직각회전이 더 선호되어 왔다(Fabrigar, Wegener, MacCallum, & Strahan, 1999). 그리고 그 동안의 사각회전 방식들에서는, 사각회전 시의 파라메터(SPSS에서 direct oblimin을 위한 δ, SAS에서 orthoblique 회전을 위한 HKP) 값을 연구자가 모르는 가운데 어떤 값이든 결정해서 주어야 하는 어려움이 있었다. 그래서 직각회전 후, 요인 간 상관을 간편방식으로 추정해서 보고하는 편법을 사용하는 것이 제시되기도 하였다(예: 이순묵, 2001).

그러나 이제는 사각회전의 이론적 강점과 실용상 편리성(사각회전이 용이해짐) 때문에 사각회전의 사용을 더 이상 회피할 수 없고 일반적으로 사용해야 하는 시점에 왔다. 사각회전의 이론적 강점을 보면, 직각회전에 비해 사각회전이 현실에서의 요인 간 상관을 반영한다는 중요한 차이점 외에도, 사각회전은 제약이 적어서(요인상관을 허용) 요인구조에 대한 복잡(도)함수(complexity function)의 값을 저하시키고 직각회전보다 더 잘 단순구조를 산출한다. 여기서의 단순구조는 Thurstone(1935, 1947)에서 제시한, 해석을 위해 용이한 구조를 말한다. Thurstone이 처음에(1935년) 제시한 단순구조의 3조건은 다음과 같다. 즉, 요인계수(factor loading) 행렬에서 각 행에 적어도 하나 이상의 "0"이 있고, m개의 공통요인이 있을 때 요인계수 행렬의 각 열에 적어도 m개의 0이 있고, 요인계수 행렬에서 어느 두 개의 열을 비교해 볼 때 몇 개의 변수가 하나의 열에서 0이 아닌 계수를 가지면 다른 열에선 0의 계수를 가져야 한다. 이상과 같은 조건들을 만족시키면 결과적으로 요인계수가 큰 값은 아주 커서 돋보이고 작은 것은 아주 작아서 "0"에 가까이 가도록 된다. 그런데 현실에서 요인 간 상관이 존재하는데 분석시에 그것을 0으로 제약하면(직각회전), 그 상관은 추가적 요인을 생성하게 되어 요인구조는 간명함을 상실하게 된다(Brogden, 1969). 하나의 요인이 추가됨은 단순구조의 추구와 반대되는 방향이

다. 따라서 단순구조의 추구에서는 직각구조보다 사각구조를 선호한다.

그러면 사각구조이면서 현실적인 단순구조는 어떤 것인가? 물론 해의 복잡도가 낮아야 할 것이다. 해의 복잡도는 행(요인계수 행렬의 행)복잡도 또는 변수 복잡도를 기초로 정의되며, 그것을 하나의 값으로 나타내기 위한 함수를 복잡(도) 함수(complexity function)라고 한다(Browne, 2001). 요인수효를 m이라 할 때 하나의 측정치가 몇 개의 요인에 대한 지표가 되는지를 나타내는 변수복잡도(variable complexity, Browne, 2001; Rummel, 1970)는 최소값이 1이고 최대값이 m-1로서, 1을 넘는 것이 현실에서 가능한 일인데, 이러한 구조를 복잡군집해라고 한다. 그 반대는 독립군집해(independent cluster solution, McDonald, 1985; 또는 완벽군집해 perfect cluster solution, Borwne, 2001)라고 한다. 독립군집해는 단순구조의 특별한 경우(McDonald, 1985)이지만, 경험자료에서는 불가능하고(“practically impossible with empirical data”, Harman, 1976, p.99), Thurstone도 복잡군집해가 현실에 맞는 것이라고 하였다(“complex patterns originally regarded by Thurstone as representative of reality”, Browne, 2001, p.113). 즉 현실적인 단순구조는 복잡군집해가 된다는 것이다.

복잡도 함수는 회전기준(예: VARIMAX, direct oblimin)에 따라서 달리 정의되며 그 함수값을 가급적 최소화함으로써 최대한 단순구조가 되게 하려는 것이 회전의 목적이다. 그런데 현재 요인분석이 포함된 범용 소프트웨어인 SAS나 SPSS에서 널리 쓰이는 회전기준들은 복잡군집해를 효과적으로 다룰 수가 없다. 즉, varimax 및 Direct Oblimin(with parameter delta=0) 기준은 독립군집해를 다루기에만 적절하다(Browne, 2001). 사각회전 시 복잡군집해를 다루기에는 Geomin(요인계수 행렬에서 행계수 제곱들의 기하평균을 최소화) 기준이 가장 적절하다(Browne, 2001; McDonald, 2005). 이 Geomin 기준이 포함된 소프트웨어로는 CEFA와 Mplus가 있어서 사각회전을 위한 실용적 여건이 갖추어져 있는 것이다.

현재의 사각회전 기준인 direct oblimin(SPSS)이나 Orthoblique(SAS)에서는 파라메터의 값을 연구자가 여러가지로 주어서 회전하는 가운데 적절한 해를 선택하게 하는 것인데, 연구자로서는 어떤 값을 주어야 할지 알 수가 없어서 대체로 0을 주는 경향이 있다. 이것은 자료 내 구조에서 요인 간 각도를 최대한 좁혀서 요인 간 상관을 크게 하는 결과를 가져올 수가 있다. 그러나 복잡도 함수를 정의하고 그 값이 최소화되는 기준에서 최종해를 산출하는 방식이면 파라메터값을 연구자가 임의로 줄 필요가 없는 진일보한 사각회전이 되는데, 이것이 종래의 사각회전 프로그램에서 갖추지 못한 최근 프로그램들(CEFA, Mplus)의 실용성이다.

탐색적 회전에 부분제약 목표회전의 추가

요인구조의 회전에서 직각구조 또는 사각구조를 추구할 때 회전방식에는 탐색적 회전과 목표회전이 있다. 탐색적 회전(exploratory rotation)은 해의 복잡도 함수가 최소화되는 기준에서 기초구조를 회전하여, 최종해에 있는 “모든 모수의 추정”을 한다는 의미에서 기계적인 과정이다. 예로서, 최종해가 p×m의 행렬이면 모수로서의 요인형태계수는 pm개가 추정되어야 한다. 그에 반해 1차적으로 실시된 탐색적 회전 결과를 참조하여 해석가능한 크

기의 요인계수들에 대해서는 자유모수(미지수)로 하여 추정하고 나머지에 대해서는 값을 고정 또는 제약하므로서 간명한 목표구조를 주고서 그에 맞게 회전하는 것이 목표회전이다. 목표회전에는 탐색적 회전에서 구한 요인형태계수 행렬의 모든 원소를 전반적으로 고정하는 '전반적 고정 목표회전'과, 요인계수의 일부에 대해서만 연구자의 견해에 따라 값을 제약(예: 값이 작은 경우에 0으로 제약)하는 '부분제약 목표회전'이 있다.

과거에 많이 쓰였던 전반적 고정 목표회전인 Procrustes 회전은 목표구조에 과대하게 충실하여 자료 내 구조를 왜곡시킬 수가 있고, Promax 회전은 컴퓨터가 직각회전 결과에 기초하여 내부적으로 목표구조를 생산하므로 연구자의 견해가 반영되기 어려웠다. 사각구조를 추구할 때, 이러한 두 극단을 피할 수 있는 목표회전이 Browne(1972)의 부분제약 목표회전이다. 부분제약이란 연구자가 일차로 탐색적 회전(예: 사각 Geomin)을 하고나서 그 결과에 근거하여 요인형태계수 행렬 내 작은 값을 0으로 제약(0에 가깝게 추정하도록 하는 제약임)하고 나머지 값들만을 자유모수로 하여 추정하도록 회전하는 방식이다. 이 때 동일한 자료에서의 분석이라는 점에서 탐색의 연속이긴 하지만 연구자의 잠정적 가설을 직접적으로 확인해 준다는 의미에서 확인적 측면이 있다. 물론 일단 사용된 목표구조를 바꾸어서 다시 부분제약된 목표회전을 시킬 수도 있으므로 이 때는 확인적이라기 보다는 "인간의 판단에 기초하여 실시되는 비기계적인 탐색과정"이 된다(Browne, 2001, p. 125).

부분제약 목표회전에서는 연구자가 표시한 목표행렬 내 제약 모수의 지정이 잘못될 경우(모형오류), 그 제약모수에 대하여 실제 가능한 추정치를 제공함으로써 잘못 제약되었음에 대한 직접적인 신호를 주는데, 이것은 CFA에는 없는 효율적 기능이다. 또한 자료가 한 세트밖에 없는 가운데 연구자의 내용적 전문성을 기반으로 여러 차례의 탐색을 통해 자료를 잘 개략하는 해를 산출할 수 있다는 장점이 있다. 따라서 CFA를 실시한 후 모형 수정을 위해 여러 차례의 수정을 함으로써 확인적 정신을 위반하는 것보다는 '탐색적 회전과 부분제약 목표회전이 가능한 EFA를 하는 것이 더 바람직하다'(Browne, 2001, p.113 참조).

해석가능성 논의의 확대

앞서의 요인수효 결정에서 가장 중요한 것은, Thurstone 이래로 해석가능성으로 언급되어 왔다. Thurstone은 .3~.4 이상의 요인계수가 해석가능한 크기이며 m요인 구조에서 하나의 측정치는 두 개 이상의(최대 m-1개) 요인에 대한 지표로 볼 수 있음을 밝힌 바 있으며, Gorsuch(1983)는 수학적 결과로 산출된 요인들을 "모두 다 해석할 필요는 없다"(p.206)고 하였다. 또한 하나의 요인에 해석가능한 지표가 3개 이상일 것이 바람직하지만 구조가 분명할 경우 2개 밖에 없는 경우도 해석 가능하다(예: Browne, 2001; Preacher & MacCallum, 2003).

그런데 기초구조를 회전해서 얻은 최종구조의 해석에서 과거의 관행과 달라지는 점이 몇 가지가 있다. 첫째는 요인계수를 해석할 때 계수의 크기뿐만 아니라 추정의 정확도(precision)까지 보아야 한다. 어떤 기준값 이상이면 일괄적으로 해석하고 그 미만은 해석하지 않는 기계적 규칙보다는 표본크기, 모형크기, 추정법, 회전방식 등이 반영되는 표준오

차(standard error)를 고려하는 것이 중요하다(Cudeck & O'Dell, 1994). 아울러 모수추정치와 표준오차의 비율을 참조하는 것도 실용적이다.

둘째, 사각회전 시 요인구조의 해석에서 참조해야 하는 두 가지 요인계수 행렬 중 한 가지 일변도로 가고 있는 것의 문제성에 대한 환기가 필요하다. 최종구조의 요인계수 행렬에는 식 (4)에서와 같이 형태계수와 구조계수 행렬의 두가지가 있다. 수학적으로 보면 전자는 측정치를 종속변수로 하고 요인들을 예측변수로 하는 표준회귀분석에서의 회귀계수이고 후자는 측정치와 요인 간의 영차상관(다른 요인들의 효과가 통제되지 않은 온전한 상관)이다. 표준회귀계수는 측정과 요인 간 영차상관에서 그 요인이 다른 요인들과 중복되는 부분을 통제한 결과이다. 직각구조이면 요인 간 상관이 0이라서 서로 간에 통제할 것이 없으므로 표준회귀계수(형태계수)와 영차상관(구조계수)이 같고 요인계수행렬은 한 가지 명칭(통상 "형태계수")으로 주어진다. 그러나 일반적으로 사용하게 되는 사각회전에서는 요인 간에 서로 상관되는 부분이 통제되면서 형태계수가 산출되므로 이 계수는 구조계수보다 크기가 작은 값이 된다.

그 동안 구조계수를 중심으로 해석하는 것을 지지하는 연구들이 있고(예: Brogden, 1969; Gorsuch, 1983), 형태계수를 중심으로 해석하는 것을 지지하는 연구들이(예: Harman, 1976; Rummel, 1970) 있어왔다. 그러나 최근에 압도적으로 많이 사용되는 CFA모형을 포함한 구조방정식 모형에서 해석되는 요인계수가 모두 형태계수가 되면서 EFA에서도 형태계수를 해석하는 경향이 강해졌다. 그런데 구조방정식 모형에서는 연구자가 사전에 요인의 의미를 부여하므로 이미 그 의미가 정해져 있으니 특별히 의미해석을 위해 (형태계수와 구조계수 중) 어느 것을 참조해야 할지의 탐색적 논의가 필요하지 않다. 단지 유의한 계수가 많을수록 요인 추정의 신뢰도가 높아질 뿐이다. 반면에 EFA에서는 요인의 의미부여를 위해서 분석을 하는 것이므로 각 측정치가 요인과 가지는 관계를 온전히 파악하는 것이 중요하다. 그 관계의 전체는 측정치와 요인 간 상관인 구조계수에 포함되어 있고, 관계 가운데 해당 요인이 측정치에 가지는 직접효과(해당 요인이 다른 요인들과 가지는 상관을 통제한 상태에서 측정치에 가지는 효과)가 표준회귀계수인 형태계수로 나타난다. 따라서 비록 형태계수를 보고 해석하는 것이 요인 간에 잘 변별되는 해석을 제공하는 면이 있기는 하지만, 의미의 탐색을 위해서는 거기에 요인 간 상관, 그리고 구조계수까지 항상 참조하는 것이 필요하다(Lee, 2010).

끝으로 탐색적 구조방정식 모형(Exloratory Structural Equation Modeling)의 출현(Asparouhov & Muthen, 2009)으로 종래의 탐색적 요인분석에 큰 변화가 생겼다. 이 모형은 종래의 EFA와 구조방정식 모형(SEM: Structural Equation Modeling)을 복합하여 EFA-SEM의 의미를 가진 것으로 EFA 모형과도 다르고 SEM 과도 다르며 줄여서 ESEM으로 부른다. 이 모형은 SEM의 폭넓은 기능을 보여주고 있으나 EFA 연구자들에게는 종래의 EFA에서 고유요인(측정오차) 간 상관을 모두 0으로 고정하는 강한 가정을 완화한 상태에서 EFA를 구현하는 길을 열어주었다. 또한 구조가 잘 확인된 일부에 대해서는 CFA로 모형 표시를 하고, 나머지 부분에 대해서만 EFA를 실시할 수도 있게 되었다. 따라서 측정치에 따라서는 탐색적 요인과 확인적 요인에 동시적으로 지표가 될 수 있다.

EFA의 최근 소프트웨어(CEFA, Mplus)에서는 결과의 출력에서, 측정오차 간 상관이 0으로 고정된 곳에 대해서 자유모수로 할 경우 감소되는 χ^2인 MI(modification index)값을 제공한다(원자료 입력시에만 가능). 물론 EFA 모형에서는 측정오차 간 상관을 자유모수로 표시를 할 수가 없으므로 MI값이 주는 정보를 활용할 수가 없다. 그러나 ESEM 모형에서는 모형표시에 측정오차 간 상관을 자유모수로 하여 추정케 하고, 동시에 EFA를 실시함으로써 측정오차 간 상관이 탐색적 요인구조의 추정에 미칠 수 있는 영향을 통제할 수가 있다. 그 결과로 EFA에서는 약간 부풀려졌던 요인 간 상관이나 요인계수들이 약간 감소되면서 요인 간 변별이 더 잘 되는 모습이 될 수 있다. 요인 간 상관이나 요인계수의 추정과정에 혼입되었던 측정오차 간 상관이 제 위치를 찾아 들어갔기 때문이다. 결과로 해가 더 간명해 질 수도 있다.

ESEM 사용의 가장 중요한 효과는, EFA에 의해서 개발된 선언척도가 경험자료에서 어떤 실제척도의 모습을 가지는지 알려주는 것이다. 선언척도는 연구자가 제시한 구조로서 대부분의 경우 측정오차간 상관이 모형에 반영되어 있지 않다. 이 선언척도는 경험자료에 적용될 때 측정오차 간 상관을 통제하는 ESEM 분석을 통해서 요인 간 상관이 감소되고 요인의 수효가 감소될 수도 있는, 보다 간명한 실제척도의 모습으로 드러날 수 있다. 따라서 척도개발 시 ESEM의 사용은 종래에 해당 척도 중심으로 EFA를 적용한 폐쇄적인 방식보다 더 개방적인 문항분석을 가능하게 하고, 그 결과로 보다 정확한 요인구조를 해석가능하게 한다(김종규, 이순묵, 윤창영 2015; 남궁준재, 이순묵, 김효선, 2013; 안정원, 이순묵, 2015; Asparouhov & Muthen, 2009).

경험적 자료에 대한 탐색적 요인분석

EFA에서 발생하고 있는 변화를 반영하여, 경험자료에 그에 일관된 절차를 따라 EFA를 실시하였다. 4점척도의 18개 문항으로 구성된 L척도(낮은 수준의 도박행동을 측정하는 척도; 김교헌, 권선중, 김세진, 이순묵, 2011)의 척도개발용 표본인 368명의 자료에 대해 원저자들은 도박행동이 본질적으로 연속변수라고 판단하고 피어슨 상관을 구하여 고전적인 EFA(연속변수 근사방식)를 실시하였다. 그러나 본 연구에서는 이제껏 논의된 새로운 방식들의 사용을 보이기 위한 문항요인분석을 실시(제한정보 가중방식)하였다(표 1 참조). 응답 형식은 4점 평정척도(0, 1, 2, 3)이고, 문항 내용은 표 11(V5 문항은 분석과정에서 제외)에 있다. 불성실한 응답자와 극단적 응답자를 제거하고 나서 남은 368명 자료에 대하여 결측치 처리(SAS에서 다중대체의 기본방식인 MCMC 방식으로 10회 실시 후 평균값으로 대체)를 한 결과의 요약자료(김교헌 등, 2011)가 제시되었는데(표 2 참조), 맨 끝의 SMC(다중상관제곱치)만 본 연구에서 추가한 것이다. 여기서의 SMC는 우선 다분상관을 구하여 저장한 후[3] 그 상관행렬에 기초하여 각 측정치를 다른 측정치 전체에 회귀하여 얻은 R^2이다. 분석용 소프트웨어로는 Mplus 7판을 사용하였다.

EFA의 분석절차는 4단계로 나뉜다: ㈀ 요인분석 가능성 검토 및 요인수효 결정, ㈁ 탐색적 회전, ㈂ 부분제약 목표회전, ㈃ 측정오차

---

3) Mplus에서 DATA: File=filename.dat; Format= ;
VARIABLE: Usevariable= ; Categorical= ;
ANALYSIS: Type=basic;
SAVEDATA: Sample=poly.dat; Type=correlation;
을 입력.

표 2. 문항의 요약자료

문항	평균	표준편차	다분 상관(Polychoric Correlation)																		SMC
			V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	
V1	1.77	0.95	1.00																		.488
V2	1.88	1.05	.65	1.00																	.651
V3	1.37	0.98	-.42	-.43	1.00																.497
V4	0.63	0.80	-.30	-.29	.48	1.00															.408
V5	1.27	1.05	-.31	-.35	.34	.30	1.00														.278
V6	1.70	1.06	.46	.58	-.38	-.24	-.32	1.00													.520
V7	0.62	0.89	-.40	-.47	.34	.32	.37	-.43	1.00												.579
V8	0.88	0.93	-.24	-.15	.38	.38	.24	-.19	.32	1.00											.439
V9	1.59	1.13	.53	.64	-.49	-.37	-.38	.64	-.52	-.26	1.00										.642
V10	0.94	0.91	-.32	-.34	.58	.50	.34	-.41	.47	.54	-.46	1.00									.594
V11	0.84	0.93	-.25	-.29	.33	.42	.32	-.24	.43	.46	-.27	.49	1.00								.480
V12	1.61	1.09	.53	.65	-.42	-.33	-.40	.62	-.50	-.35	.67	-.46	-.36	1.00							.646
V13	1.02	0.92	-.34	-.33	.39	.42	.32	-.28	.48	.34	-.38	.37	.42	-.36	1.00						.411
V14	1.13	1.00	-.39	-.49	.40	.26	.39	-.47	.60	.26	-.59	.39	.42	-.58	.47	1.00					.593
V15	1.17	1.09	-.24	-.36	.21	.18	.31	-.29	.43	.22	-.36	.24	.27	-.30	.30	.45	1.00				.298
V16	0.49	0.83	-.34	-.49	.34	.41	.30	-.39	.61	.43	-.48	.55	.55	-.50	.45	.47	.32	1.00			.769
V17	0.69	1.09	-.31	-.40	.34	.28	.21	-.37	.37	.36	-.41	.48	.33	-.37	.30	.39	.18	.72	1.00		.612
V18	0.56	0.81	-.30	-.40	.30	.30	.36	-.34	.47	.35	-.39	.32	.38	-.44	.45	.50	.36	.50	.32	1.00	.429

간 상관 추정. 통상은 요인수효 결정의 마지막 단계인 해석가능성 검토를 위해 탐색적 회전을 실시하게 된다. 탐색적 회전의 결과(요인형태계수 행렬)에서 큰 값을 중심으로 목표행렬을 지정하여 부분제약 목표회전을 한다. 목표회전(엄밀하게는 사각목표회전인데 “사각”이 기본이므로 생략) 결과에서 측정오차 간 상관에 대한 수정지수(MI)가 나오는데 이를 참조하여 ESEM에서 측정오차 간 상관을 일부 자유모수로 하여 추정하고, 동시에 모든 변수에 대하여 탐색적 요인분석(목표회전 포함)을 한다. 4점척도 자료를 사용한 본 연구에서는 WLS계열의 추정법으로 계산의 편리를 위해 가중치 행렬에서 대각선만을 사용하고, 그에 따라 χ^2 통계량의 정확도를 위해 평균과 분산을 조정한 방식인 WLSMV를 사용하였다.

요인분석 가능성 검토

변수 간 중복가능성 검토를 위해서 표 2의 상관계수를 보면 지나치게 큰 값은 없다. 먼저 요인분석을 할 만한 자료인지에 대한 판단을 위해서 무료 소프트웨어인 Factor 10.3 (Lorenzo-Seva & Ferrando, 2015)에서 제공하는 Bartlett 검증(H_0: 상관행렬에서 대각선을 제외한 나머지 값이 0이다)과 표집적절성 지수(KMO, Cerny & Kaiser, 1977)를 참조한 결과, Bartlett 검증은 기각되었고 KMO지수는 .912로 요인분석에 적절한(0.6 이상일 것) 자료로 판단되었다.

다음은 Mplus에서 회귀분석(V_i를 모든 다른 측정치들의 집합에 회귀시킴)을 하여 SMC, 즉 R^2를 계산하였다(부록 2의 Syntax 0). 이 SMC

는 측정치의 분산에 들어있는 공통분의 추정치이기도 하지만 다른 변수와의 중복을 나타내는 다중공선성의 지표이기도 하다. 따라서 SMC가 너무 작으면 공통분이 부족한 것으로 보고 문항을 버려야 하지만 너무 큰 변수는 다른 변수들과 중복이 크므로 삭제하거나 의미가 유사한 다른 변수와 통합이 가능하다. 표 2를 보면 각 측정치의 공통분 추정치인 SMC가 대부분 .3~.7 사이로 너무 작거나 너무 큰 경우가 없어서 조정없이 모든 변수를 분석에 사용하기로 하였다.

요인수효 결정 및 기초해 산출

사회과학에서의 구성개념을 측정하는 과정에서 개별 측정치에는 측정오차가 무시할 수 없는 정도임을 고려하여, 고유요인을 제거하고 공통요인을 추출하는 공통요인분석(common factor analysis)을 하였다. 즉, 원상관행렬이 아닌 축소상관행렬(대각선을 1이 아닌 공통분 추정치로 대체)을 분석하였다.

대략적인 요인수효를 결정하기 위한 발견법을 위해서는 FACTOR 10.3(Lorenzo-Seva & Ferrando, 2015)을(부록 1 참조), 추론적 접근에서는 합치도(검증적/판단적 합치도) 참조를 위해서 Mplus 7(Muthen & Muthen, 2012)을 사용하였다. 끝으로 해석 가능성을 검토하기 위해서 Mplus 7판에서의 탐색적 요인분석(EFA)과 탐색적 구조방정식 모형(ESEM) 분석 결과를 해석 및 비교하였다.

발견법으로 대략적 수효 결정

소프트웨어 FACTOR 10.3에서 구한 다분상관(표 2 참조) 자료의 일반적인(주축요인 추출 방식에서의) 고유치(음수값이 가능)를 구한 값을 표 3에 제시하였다.

먼저 고유치 검사(scree test)(Cattell, 1966)를 위해 표 3에서 축소상관행렬의 고유치 차이를 보면 2번과 4번에서 다음으로 가면서 고유치가 현저히 떨어지는데(고유치 차이: 0.55, 0.31) 5번 고유치부터는 크게 감소되지 않고 평준화되는 것을 볼 수 있다. 그렇다면 평준화되기 직전인 4개 요인이 적절할 수 있다. 다음으로 누적분산 비율(Gorsuch, 1983)을 검토하면, 통상 75~85% 정도면 무난하다고 할 때 3개부터 75%를 상회하여 5개면 85%를 상회한다. 따라서 3~5개의 요인을 생각할 수 있다. 끝으로 평행성 분석(Parallel Analysis)을 하였다.

Timmerman과 Lorenzo-Seva(2011)에 따라 PA-MRFA로 구해진 무선자료 고유치 분포에서의 95 백분위 값(다분상관일 때)을 경험자료 고유치(Factor 10.3에서는 분산%로 제시)에 비교하여 평행성 분석을 하였다(표 4). 결과로 이미

표 3. 다분상관자료의 고유치 및 누적분산비율

번호	고유치 (eigenvalue)	고유치 차이	분산 비율	누적분산 비율
1	7.50	6.10	0.58	0.58
2	**1.40**	**0.55**	0.11	0.69
3	0.85	0.03	0.07	**0.76**
4	**0.82**	**0.31**	0.06	0.82
5	0.51	0.06	0.04	**0.86**
6	0.44	0.13	0.03	0.89
7	0.31	0.07	0.02	0.92
8	0.25	0.02	0.02	0.94
9	0.22	0.02	0.02	0.95
10	0.20	0.08	0.02	0.97
⋮	⋮	⋮	⋮	⋮

2번 요인에서 무선자료 고유치가 경험자료 고유치를 앞지르므로 1개의 요인을 생각할 수 있다. 종합적으로 발견법을 통한 요인수효의 범위는 1-5개로 판단되었고 다음의 추론적 접근에서 다시 검토하였다.

표 4. 다분상관자료에 대한 PA-MRFA 결과

번호	경험자료고유치[a] 분산 %	무선자료고유치[b] 분산 %
1	46.1	12.5
2	**9.1**	**11.2**
3	6.9	10.4
4	5.5	9.6
5	4.3	8.9
6	4	8.2
7	3.4	7.6
8	3.3	7
9	3.1	6.4
10	3	5.8
⋮	⋮	⋮

주: a: 축소상관행렬의 모든 고유치가 음수가 되지 않는 MRFA 방식으로 계산하므로 각 고유치가 설명하는 분산%가 표 3(음수가 가능한 일반적 고유치)과 다름.
b: 무선자료는 500세트를 계산하여 각 번호의 고유치에 대한 분포에서 95 백분위 값을 취한 것의 분산%

추론적 접근

Mplus에서 1~5요인 모형의 해를 구하면서 (부록 2의 Syntax 1) 검증적 합치도 및 판단적 합치도를 참조하였다(표 5 참조). 요인이 1개씩 증가할 때마다 χ^2는 유의하게 낮아지고 있다. 여기서 사용된 추정법 WLSMV의 카이제곱차이 값은 이론적인 카이제곱분포를 따르지 않아 Muthen 등(1997)의 방식으로 수정된 값이다. 한편 1요인 모형은 합치도가 나쁘고, 2요인 모형부터 좋은 편인데 3요인 모형에서 현저하게 좋아졌고, 4요인부터는 합치도가 매우 높다(χ^2값이 유의하지 않고 판단적 합치도가 완벽에 가까움). 그래서 5요인 모형은 제외하고 3-4요인 모형에 대해 기초요인구조를 회전하고 최종해에서 해석가능성을 검토하였다.

모형선택, 회전 및 해석가능성 검토

여기서의 해석에서 참조되는 표 6, 7, 8, 10에 제시된 문항들은 실제 반응 변수가 아니라

표 5. 탐색적 1~5요인 모형의 합치도

모형	$\chi^2(df)$	$\Delta\chi^2(\Delta df)$	CFI	TLI	RMSEA	SRMR
1요인 모형	598.628(135)**		.911	.900	.097	.081
2요인 모형	335.527(118)**	202.541(17)**	.958	.946	.071	.055
3요인 모형	232.597(102)**	90.440(16)**	.975	.963	.059	.043
4요인 모형	120.302(87)	87.878(15)**	.989	.986	.032	.026
5요인 모형	69.744(73)	47.708(14)**	1.000	1.001	.000	.020

주: 카이제곱차이 검증은 Mplus의 EFA에서 기본으로 제공하는 방식. ** p<.001.

표 6. 3요인/4요인 모형 탐색적 회전 결과의 형태계수

문항	3요인 모형			4요인 모형			
	F1	F2	F3	F1	F2	F3	F4
V1	.761	-.026	.119	.700	-.038	.002	.006
V2	.857	.124	-.004	.784	.122	-.077	-.072
V3	-.504	.530	-.183	-.505	.537	-.050	-.088
V4	-.226	.546	.008	-.203	.531	.092	.006
V5	-.341	.124	.152	-.176	.156	.380	-.095
V6	.748	.060	-.003	.680	.039	-.086	-.026
V7	-.326	.010	.494	-.103	.009	.622	.146
V8	.010	.535	.243	.005	.503	.181	.167
V9	.788	.008	-.062	.679	-.015	-.193	-.014
V10	-.232	.609	.145	-.314	.585	.003	.204
V11	.001	.370	.432	.104	.340	.429	.204
V12	.735	.003	-.116	.620	-.009	-.227	-.059
V13	-.200	.219	.338	-.003	.233	.525	.017
V14	-.449	-.075	.429	-.177	-.047	.696	-.019
V15	-.252	-.077	.345	-.012	-.050	.591	-.056
V16	,020	.116	.894	-.068	-.036	.322	.861
V17	-.024	.142	.616	-.284	.096	-.015	.562
V18	-.250	.032	.440	-.008	.039	.628	.065
요인 간 상관	1.000			1.000			
	-.349	1.000		-.214	1.000		
	-.589	.401	1.000	-.628	.343	1.000	
				-.288	.311	.361	1.000

반응과정에 내재하는 것으로 가정된 연속변수인 잠재반응변수임에 유의해야 할 것이다. 반응 자료에 근거하여 다분상관 및 가중치가 구해지면 잠재반응변수를 종속변수로 하여 식 (1)~(4)의 모형이 적용되기 때문이다.

우선 3요인 모형과 4요인 모형에 대하여 기초해를 구한 후, 탐색적 회전은 Browne(2001)에 따라 가장 수행이 우수했던 사각 GEOMIN을 선택하였다(부록 2의 Syntax 2). 탐색적 회전을 할 때는 단순구조의 가능성을 높이고 해석을 용이하게 하기 위해 기초해의 행(row) 을 표준화 시킨다. 여기서는 표본이 적절한 크기일

때 사용되는 Kaiser(1958) 방식을 사용하였다.

모형 선택

표 5에서의 3～4 요인모형에 실시된 탐색적 회전결과를 표 6에서 보면 4요인 모형에서 네 번째 요인은 지표가 되는 문항이 두 개밖에 없고 문항의 내용도 어떤 방향성을 나타낸다고 보기 어려워 해석하기 곤란하다. 따라서 3요인 모형을 선택하여 문항 분석을 하고 그 결과에 따라 다시 분석한 결과를 가지고 해석을 하였다.

3요인 모형 1차 (문항)분석

표 6에서 3요인 모형을 보면, 요인 2와 3이 양의 상관인데, 요인 1은 이 두 요인과 음의 상관이다. 문항들(표 11 참조)에 대한 요인계수들을 볼 때, 요인 2와 3은 대체로 도박을 하는 행동을 나타내고 있다. 요인 2는 사교용으로 하는 내용이고, 요인 3은 약간의 위험성, 즉 저위험의 도박행동을 나타낸다. 그에 반해 요인 1은 도박에 관심이 없는 내용을 나타내고 있다. 그러한 1차적 해석을 바탕으로 각 문항들이 몇 개의 요인에 대한 지표(indicator)가 되는지를 나타내는 변수 복잡도를 보면 요인계수의 크기(절대값으로 0.3 이상)로 보아 대체로 1이지만 2인 경우도 있다. 3요인 모형이므로 단순구조가 되려면 변수 복잡도는 1이나 2가 바람직하고 0이나 3이 되는 문항은 버려야 한다.

우선 변수복잡도가 1인 문항들은 V1, V2, V4, V5, V6, V8, V9, V10, V12, V13, V15, V16, V17, V18 로 14개인데 이 중에서 V5(도박에서 기술과 능력이 중요)는 약간의 위험성을 나타내는 요인 3의 지표라야 하는데, 요인 1에서 음의 값을 보이는 지표로 나타났다. 역으로 해석하면 ‘도박에 관심이 없지 않음’을 나타내는 것으로 볼 수 있다. 그런데 요인 2는 사교상 관심이고, 요인 3은 약간의 사행성 관심이므로 V5는 두 요인중 하나에도 적절한 크기의 값을 보여야 할 것인데 그렇지 않다. 그렇다면 내용상 일관성이 없는 문항이 되므로 제거하였다.

변수복잡도가 1인 문항들을 제외한 나머지 4개 문항들(V3, V7, V14, V11)은 모두 변수복잡도가 2인 문항들이다. 그 중 V3(여가 목적의 도박)은 사교도박에 대한 지표인데 요인 1에도 큰 값(-.504)을 보인다. 즉 여가 목적의 도박을 하므로 도박 무관심(요인 1)과는 반대라서 요인 1에 큰 음의 계수를 보이는 것은 이론적으로 문제가 없다. 요인간 상관에서도 요인 1과 요인 2는 음의 상관을 보였으로 V3이 두 요인의 지표로서 상반된 부호를 보이는 것도 이해가 된다. V7(큰 돈 도박에 충동적 참여 경험)은 약간의 위험성이 있으니 요인3의 지표라야 하는데 요인1에도 해석을 하자면 가능한 크기(-.326)를 보인다. ‘도박에 관심이 없을수록 큰 돈 도박에 충동참여 낮음’으로 해석되므로 지표가 될 수 있다. V14(돈을 따기 위해 도박한 경험)는 의미상 요인 3에 지표가 되어야 하는데 요인 1에도 비슷한 정도로 큰 음의 계수(-.449)를 보이므로 방향을 바꾸어 도박에의 관심, 즉 약간의 위험성을 나타내는 것에 부합되는 것으로 보고 요인 1에서도 지표로 볼 수 있다. 마지막으로 V11(도박을 통해 그날의 운을 시험)은 원래 사교성 도박의 지표로 생각하고 제작한 것인데, 요인 3(저위험 도박)에 더 큰 계수를 보였다. 제작자들은 전문가들인데 응답자들은 일반인으로서 이 정도면 위험하다고 보았기에 변수복잡도가 2인 문항이 된 것으로 판단되었다. 이상의 문항분석

결과로 V5문항을 제외하고 다시 3요인 모형의 2차 분석을 하였다(부록 2의 Syntax 3 참조).

3요인 모형 2차 분석/3차 분석

표 7에는 V5문항이 삭제된 3요인 모형에 대한 탐색적 회전 결과(2차분석)와 그에 따라 지정된 목표행렬, 그리고 부분제약 목표회전 결과(3차 분석)가 제시되었다. 표 7의 요인계수는 모두가 형태계수이다. 그 중 '탐색적 회전'에 대한 요인계수들은 표 8에 그것의 표준

표 7. 3요인 모형 2차분석(탐색적 회전) 및 3차분석(목표회전) 결과의 형태계수

문항	탐색적 회전			목표행렬			목표회전		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
V1	.758	-.029	.114		0	0	.659	-.128	.044
V2	.852	.118	-.010		0	0	.777	.033	-.117
V3	-.507	.526	-.179	0		0	-.300	.687	-.221
V4	-.227	.543	.010	0		0	-.041	.675	-.060
V6	.745	.054	-.007		0	0	.665	-.030	-.091
V7	-.316	.014	.503	0	0		-.245	.061	.542
V8	.006	.538	.240	0		0	.173	.641	.153
V9	.784	.003	-.067		0	0	.681	-.096	-.148
V10	-.239	.616	.136	0		0	-.024	.764	.058
V11	.005	.373	.433	0			.135	.447	.378
V12	.729	-.002	-.120		0	0	.629	-.096	-.196
V13	-.192	.219	.347	0	0		-.087	.287	.336
V14	-.435	-.075	.443	0	0		-.378	-.030	.508
V15	-.238	-.080	.357	0	0		-.212	-.061	.401
V16	.016	.141	.878	0	0		.100	.174	.869
V17	-.031	.163	.602	0		0	.052	.203	.589
V18	-.235	.031	.452	0	0		-.172	.071	.479
요인 간 상관	1.000						1.000		
	-.338	1.000					-.490	1.000	
	-.587	.390	1.000				-.509	.590	1.000

주: 합치도는 $\chi^2_{(88)} = 218.109$, p-value=.000, RESEA=.063(90% CI, .053～.074), CFI=.975, TLI=.961, WRMR=.759. WRMR의 해석(대체로 1.0 이하일 것)은 Yu(2012)를 참조.
탐색적 회전과 목표회전 결과에서 음영은 해석대상이 되는 계수, 목표행렬에서 음영은 회전후 검토요점 1~5.

표 8. 3요인 모형 2차 분석(탐색적 회전)의 표준오차(형태계수/표준오차 비율) 및 구조계수

문항	표준오차(형태계수/표준오차 비율)			구조계수		
	F1	F2	F3	F1	F2	F3
V1	.06(12.45)	.08(-.38)	.08(1.47)	.70	-.24	-.34
V2	.04(19.66)	.06(1.89)	.04(-.24)	.82	-.17	-.46
V3	.11(-4.75)	.08(6.38)	.12(-1.56)	-.58	.63	.32
V4	.10(-2.25)	.07(8.01)	.01(.66)	-.42	.62	.36
V6	.05(16.12)	.06(.84)	.05(-.13)	.73	-.20	-.42
V7	.09(-3.36)	.03(.41)	.08(5.99)	-.62	.32	.69
V8	.03(0.22)	.08(6.59)	.11(2.16)	-.32	.63	.45
V9	.05(15.25)	.04(.08)	.07(-.90)	.82	-.29	-.53
V10	.09(-2.70)	.07(8.98)	.11(1.26)	-.53	.75	.52
V11	.03(.16)	.11(3.31)	.09(4.68)	-.38	.54	.58
V12	.07(11.28)	.03(-.09)	.09(-1.34)	.80	-.30	-.55
V13	.08(-2.56)	.08(2.67)	.10(3.58)	-.47	.42	.55
V14	.09(-4.70)	.09(-.84)	.09(4.95)	-.67	.25	.67
V15	.09(-2.69)	.11(-.76)	.09(3.87)	-.42	.14	.47
V16	.09(.18)	.20(.72)	.11(8.09)	-.55	.48	.92
V17	.10(-.31)	.15(1.06)	.10(6.15)	-.44	.41	.68
V18	.11(-2.23)	.12(.26)	.11(4.32)	-.51	.29	.60

오차가 제시되었고 괄호 안에 추정치/표준오차 비율이 제시되었다. 이 값의 분포는 대략 "표준정규분포인 *Z*값을 따른다"(Bentler, 1980, p.428). 따라서 절댓값으로 본 크기가 대략 2.0 이상이면 유의하다고 해석할 수 있다. 또한 그 탐색적 회전결과에서의 구조계수를 표 8의 우측에 제시하였다.

2차분석인 탐색적 회전 결과를 해석할 때는 표 7의 형태계수, 표 8의 표준오차와 추정치/표준오차 비율, 구조계수, 그리고 가장 중요한 것으로서 문항의 의미를 염두에 두고 진행하였다. 그 결과 각 요인에서 해석가능한 크기의 요인계수(대략 0.3을 초과하면서, 추정치/표준오차 비율의 절대값이 2 이상이고, 계수 값이 큰 집단에 속함)에는 음영을 하였다. 표 7의 탐색적 회전 결과에서, V1과 V2는 내용상 도박 무관심을 나타내는데 F1로 부터의 형태계수가 큰 값이고 계수/오차 비율이 충분히 큰 편이며, 구조계수는 .70, .82으로 큰 값이다. 즉 F1의 좋은 지표가 된다. V3과 V4는 내용상 사교도박을 의미하는데 F2와의 형태계수가 큰 값이고 계수/오차 비율이 충분히 큰 편이고, 구조계수는 .63, .62로 큰 값으로 좋은 지표가 된다. 그 중에서 V3(여가 목적으로 도박한다)

은 F1에 대하여 형태계수가 -.507로 '도박무관심이 낮을수록 여가 도박을 함' 의 해석이 가능하지만 원래 제작시에 정방향은 F2에 일관된 것이었고, 현재 계수(.526)도 무난한 크기이다. 따라서 간명한 해를 구하기 위한 목표회전에서는 F1에 대한 요인계수를 "0"으로 제약하였고 이 가설의 적절성 여부에 대하여는 그 결과를 보고 검토할 예정이다(**검토요점** 1).

V6(도박생각을 해본 적 없음) 요인계수의 크기 및 계수/오차 비율이 충분히 크고 문항의 의미상 F1의 좋은 지표가 된다. V7(큰 돈 도박에 충동참여 경험)은 내용상 저위험성의 도박인데 F3에 무난한 크기의 형태계수이고, 계수/오차 비율이 크며, 구조계수는 .69를 보이므로 F3에 대해서 좋은 지표이다. 그런데 이 문항은 F1과도 -.316의 형태계수(음의 방향으로 해석가능)를 보이고, 계수/오차 비율이 유의하며 구조계수도 -.60이라는 높은 크기를 보이지만 F1의 세로줄에 있는 해석가능 계수들의 집단에서 매우 작은 크기이다. 따라서 F3만의 지표로 보고 목표행렬에서는 F1과의 요인계수를 0으로 제약하였다. 이 가설은 목표회전 후 검토될 것이다(**검토요점** 2).

V8은 내용상 사교도박인데 F2에 무난한 크기의 형태계수, 계수/오차 비율이 충분히 크고 F2와 .63의 구조계수를 보이므로 F2의 좋은 지표이다. 유사한 논리에서 V9와 V10은 각각 F1과 F2에 좋은 지표이다.

V11은 도박을 통해 그 날의 운을 시험해 보는 것인데 그 정도는 저위험성 도박의 수준이 아니라고 생각하고 제작한 문항이었는데 사교도박의 문항들이 지표로 표시되는 F2에서보다는, 저위험성 도박행동 문항들이 지표가 되고 있는 F3에서 더 높은 형태계수를 보였고 계수/오차 비율도 유의하다. 그런데 F2 및 F3과의 구조계수를 보면 .54와 .58로 F3에서 더 높다. 문항 제작자들은 전문가들이었는데, 응답자인 일반인으로서는 그 정도도 위험하다고 볼 수 있는 것 같다. 이 척도는 도박을 많이 하는 집단이 아닌 적게 하는 집단용이었으므로 적용 대상이 될 응답자들의 생각도 존중하기로 하고 변수 복잡도가 2인 문항으로 유지하기로 하였다. 따라서 목표행렬에서는 F2와 F3 모두에서 자유모수로 하고 F1에서만 0으로 제약하였다(**검토요점** 3).

유사한 방식으로 문항의 의미와 표 7 그리고 표 8을 검토하여 V12는 F1의 지표로, V13/V15/V16/V18은 모두 F3의 지표로 판단하고, 지표로서의 해석을 제공하기에 미약한 계수를 목표행렬에서 모두 0으로 제약하였다.

V14(돈을 따기 위해 도박)는 저위험성 도박인 F3의 지표가 되는 것은 당연하고, F1에 대해서도 역방향의 해석, 즉 '도박에 무관심할수록 돈따는 도박행동 낮음'을 나타내는 지표로 기능하고 있는데 계수/오차 비율이 유의하고 구조계수가 무시할 수 없는 정도(-.67)이다. 그러나 저자로서는 내용이 우선이라고 보고 정방향인 F3의 지표로만 간주하여 목표행렬에서는 F3에 대한 요인계수만을 자유모수로 하였고 나머지는 모두 0으로 제약하였다. 이 가설 역시 결과를 보고서 검토할 것이다(**검토요점** 4). V17은 의미상 사교도박인데 F3에만 지표로 나타났다. 그러나 내용상 사교도박 문항들이 지표로 나타나는 F2의 지표라야 하므로 목표행렬에서는 F2에만 자유모수로 하고, F1과 F3 모두에서 0으로 제약하였다(**검토요점** 5).

결론적으로 3요인 모형의 2차분석 결과는 1차분석의 대략적 해석을 구체화하여 F1은 도박 무관심 요인, F2는 사교도박 요인, F3은 저위험도박 요인으로 해석하였으며 F1은 나머지

두 요인과 음의 상관, F2와 F3 간에는 양의 상관이 산출되었다(표 7의 탐색적 회전 맨 아래).

목표행렬에서의 제약에 따라 3요인 모형의 3차 분석은 목표회전이 되었다(부록 2의 Syntax 4 참조). 회전을 한 결과의 요인계수 행렬을 보면(표 7의 '목표행렬', '목표회전' 참조) **검토요점 1과** 2에 대해서는 저자가 가설화한 대로 0으로 제약하는 데 문제가 없었다. 즉 V3과 F1간, V7과 F1간 요인계수 모두 해석가능한 크기가 되지 않으므로 저자의 판단이 지지되고 있다. **검토요점** 3인 V11은 목표행렬과 같이 F2 및 F3에 대한 요인계수가 모두 해석가능한 크기로 추정되어 변수 복잡도가 2라고 하는 저자의 판단이 지지되었다. 그러나 **검토요점** 4인 V14에는 F1과의 요인계수에 0으로 제약을 주었는데도 해석가능한 크기의 요인계수(-.378)가 산출되어서 좀더 검토가 요망된다. **검토요점** 5인 V17 역시 위험도박이 아닌데도 계속 F3의 지표로만 확인되고 있어서 좀 더 검토가 필요하다. 그런데 아직 해소되지 않은 **검토요점 4와** 5의 문제가 ESEM을 적용하는 과정에서 해소되었다(다음 절에 상세 서술). 표 7의 탐색적 회전과 목표회전은 기초구조가 동일하므로 모형의 합치도 역시 동일하며 표 아래에 제시된 바와 같이 무난한 편이다.

목표회전 결과에서 10 이상의 수정지수(MI 값)는 V16과 V17의 측정오차 간 상관에 대해 80.063, V1과 V2 간 측정오차 상관에 대해 25.975, V14와 V16 간 측정오차 상관에 대해 18.822, V7과 V17 간의 측정오차 상관에 대해 11.074로 나타났다. V16과 V17은 "도박을"이라는 동일 단어가 방법효과의 원인이 될 수 있어 높은 수정지수 값이 이해될 수 있다. 따라서 우선적으로 이 수정지수를 고려하여 ESEM 분석을 하였다.

측정오차 간 상관을 고려한 ESEM

여기서는 표 7의 목표회전에 추가로 측정오차 간 상관을 1개 추정하도록 하는 ESEM(수정모형 1)을 분석하였다. ESEM 결과에서 요인형태계수들이 추정되는 부분은 곧 탐색적 요인분석의 결과이므로 해석을 할 때 형태계수, 형태계수/표준오차 비율, 구조계수(탐색적 회전에서는 기본적으로 제시되지만 목표회전이나 ESEM에서는 OUTPUT에서 TECH4를 요청해서 요인과 측정치 간 상관 참조), 요인 간 상관, 문항의 의미를 참조해야 할 것이다. 표 9와 표 10에서 수정모형 1에 대한 부분이 ESEM 분석(부록 2의 Syntax 5 참조) 결과이다. 표 10의 음영은 해석 대상의 요인계수를 나타낸 것이다. V16과 V17 간의 측정오차 상관은 .618로 큰 값으로 추정되었다. 이 오차 간 상관을 추정하고 나니 V14와 F1간 요인계수가 해석 대상이 안 되는 정도로 감소되어 바람직한 상태가 되었다(**검토요점** 4의 해소). 또한 V17의 요인계수가 요인 2에 대해 커진 것을 확인할 수 있다. 즉, 문항의 의미에 맞게 요인 2에 대한 지표가 되어 해석에 보다 적절하게 되었다(**검토요점** 5의 해소).

수정모형 1의 분석 결과에서 수정지수 중 10 이상의 값은 V1과 V2 간 측정오차상관에 대해 23.107로 나타났다. V1과 V2는 문항 내용이 거의 같고 척도 내에서 위치가 근접한 것이 동일 방법효과를 가져올 수 있는 점이 있어, 높은 수정지수 값이 이해될 수 있다. 이에 대해 추가적으로 수정을 하여 분석하였다(수정모형 2, 부록2의 Syntax 6 참조). 결과로 V16과 V17 간의 오차 상관은 .622, V1과 V2 간의 오차 상관은 .348로 추정되었다. 그런데

표 9. 다분상관자료에 대한 목표회전과 ESEM(수정모형) 결과의 합치도

모형	$\chi^2(df)$	$\triangle\chi^2(\triangle df)$	CFI	TLI	RMSEA	WRMR
3요인 모형 (표 7의 목표회전 결과)	218.109(88)**		.975	.961	.063	.759
수정모형 1	134.378(87)**	30.103(1)**	.991	.986	.038	.573
수정모형 2	111.288(86)**	18.217(1)**	.995	.992	.028	.510

주: 카이제곱차이 검증을 위해 Mplus에서 DIFFTEST 사용. 즉 수정모형 1과 2를 분석 후, Syntax 4와 5의 Analysis 블록에 각기 “DIFFTEST=ESEM1.dat”(Syntax 5의 SAVEDATA 블록에서 저장시킨 합치도 파일임. 이것을 Syntax 4에서 불러서 차이분석), “DIFFTEST=ESEM2.dat” (Syntax 6의 SAVEDATA 블록에서 저장된 것을 Syntax 5에서 불러서 차이분석)를 추가하고 다시 돌림. WRMR의 해석(대체로 1.0 이하일 것)은 Yu(2012)를 참조.

표 10. ESEM 추정 결과의 형태계수 및 요인 간 상관

문항	수정모형 1 (V16 with V17)			수정모형 2 (V16 with V17, V1 with V2)		
	F1	F2	F3	F1	F2	F3
V1	.679	-.126	.063	.531	-.159	-.017
V2	.803	.022	-.062	.692	.026	-.147
V3	-.324	.622	-.163	-.330	.621	-.163
V4	-.037	.619	.013	-.028	.612	.029
V6	.683	-.038	-.053	.738	-.047	.004
V7	-.100	.002	.705	-.108	-.002	.701
V8	.200	.614	.212	.201	.610	.217
V9	.675	-.081	-.150	.724	-.088	-.103
V10	-.066	.790	.011	-.082	.789	.000
V11	.237	.402	.506	.256	.392	.534
V12	.609	-.081	-.210	.640	-.085	-.182
V13	.031	.208	.504	.039	.200	.519
V14	-.241	-.125	.689	-.266	-.122	.663
V15	-.081	-.145	.571	-.086	-.149	.569
V16	-.010	.248	.589	-.008	.239	.599
V17	-.196	.352	.116	-.203	.351	.112
V18	-.029	-.005	.653	-.032	-.012	.657
요인 간 상관	1.000			1.000		
	-.475	1.000		-.471	1.000	
	-.626	.628	1.000	-.630	.627	1.000

수정모형 2에서 모수(요인계수, 요인상관) 추정치들은 수정모형 1과 유사하고(표 10) 판단적 합치도에는 큰 차이가 없다. 따라서 보다 간명한 수정모형 1을 최종 해석하였다.

표 10의 수정모형 1의 요인계수를 보면 V11은 변수복잡도가 2인 상태로 유지되고 있다. V3(여가를 위한 도박)의 경우 F1과의 요인계수가 -.324로서 내용상 필요하다면 해석이 가능한 경우이다. 그러나 이미 표 7에서 목표행렬 설정시에 F2의 지표로만 추정하기로 한 것이 목표회전 결과에서도 지지되었고 표 10에서도 F1의 세로줄에서 보면(수정모형 1) 해석대상인 나머지 요인계수들의 크기에 비해 현저하게 작으므로 F1(도박 무관심)에서는 해석하지 않고 F2(사교도박)만의 지표로 보는 것이 적절하다. 최종적으로 해석을 위한 요인구조는 표 10의 수정모형 1에 따라서 표 11에 제시되었다.

표 11. 다분상관자료에 대한 요인구조

도박에 대한 무관심(F1): 도박에 대해 관심이 없는 상태	
V1	나는 다른 사람들이 도박을 해도 거기에 참여하지 않는다.
V2	나는 도박과는 거리가 멀다.
V6	나는 도박할 생각을 해본 적이 거의 없다.
V9	나는 도박을 해본 적이 거의 없다.
V12	나는 도박을 해본 지가 언제인지 모르겠다 .
사교 도박(F2): 사교나 여가를 위해 도박을 즐기는 상태	
V3	나는 여가를 즐기기 위한 목적으로 도박을 하기도 한다.
V4	나는 도박을 통해 가족 및 친척들과 더 많은 시간을 함께 보낸다.
V8	나는 단지 에너지를 재충전하기 위해 도박을 한다.
V10	나는 도박을 통해 여가를 즐긴다.
V11	나는 도박을 통해 그날의 운을 시험해 본다.
V17	나는 도를 넘는 수준으로 도박을 한 적은 없다.
위험 도박(F3): 위험 수준으로 도박에 빠진 경험이 있는 상태	
V7	나는 내 형편을 넘어서는 큰 돈을 거는 도박에 충동적으로 참여해본 적이 있다.
V11	나는 도박을 통해 그 날의 운을 시험해 본다.
V13	나는 도박을 하자고 권하는 친구나 동료 등의 청을 거절하지 못해서 도박을 한 적이 있다.
V14	나는 돈을 따기 위해 도박을 한 적이 있다.
V15	나는 잃어도 전혀 상관없는 정도를 넘는 금액을 도박에 걸어본 적이 있다.
V16	나는 불안하거나 울적한 기분을 잊기 위해 도박을 할 때도 있다.
V18	나는 적정한 선을 넘기도 한다.

논 의

요약

심리학 자료에서 측정치를 본질적으로 연속변수로 가정한 고전적 요인분석은 오래전에 정착되었다. 한편 연속변수에 근사할 수 없는 문항점수로 된 측정치들의 자료에 대해서는 서열자료에 대한 문항요인분석으로의 일대 전환이 필요하다. 물론 이 글의 내용은 고전적 요인분석과 문항요인분석 모두에 적용된다. 우선 요인수효 결정에 추가된 내용들은 보다 합리적이고 풍부한 정보에 근거하여 기초구조를 산출할 수 있게 해줄 것이다. 즉 발견법에서 PA-MRFA의 실용성, 추론적 접근을 위해 참조가능한 합치도들은 모두 종합적으로 사용될 때 요인수효를 정확하게 결정하는 효과성을 높일 것이다. 또한 주어진 자료에 기반하여 탐색적인 사각회전이 끝난 후 모형을 최대한 탐색하고 보다 간명한 모형을 산출하는 부분제약 목표회전(Browne, 1972)은 전반적 고정 목표회전이 가지고 있는 특징인 기계적 회전의 한계를 넘어, 연구자의 전문적 판단을 입력할 수 있는 진일보한 탐색 기회를 제공하고 있다.

ESEM의 강점

21세기에 탐색적 요인분석에서 발생한 가장 현저한 변화는 전통적인 EFA에서 "측정오차 상관 0"이라는 강한 가정이 주는 제약을 벗어나게 되었다는 것이다. Asparauhov와 Muthen (2009)의 ESEM은 현실적으로 척도 내 문항들 간에 다양한 동일방법으로 인하여 발생하는 측정오차 간 상관을 표시하고 추정 및 검증할 수 있는 기능을 제공하므로서 EFA의 한계를 극복하는 결과를 가져왔다. EFA에서 지난 100여년 동안 가정한 "측정오차 간 상관 0"은 단지 모형식별을 위한 임의적 제약 중 하나였으나, 척도 개발자들에게는 척도 내 구조를 탐색할 때 반드시 가정하므로서 척도의 실제 모습을 왜곡시킬 수 있는 문제를 가지고 있었다.

그러한 왜곡의 예가 표 7의 목표회전 후 요인계수 행렬에서 두 가지로 나타났다. 첫째는 V14와 F1 간에 요인계수가 해석가능한 크기의 음수 값(-.378)인 것이고(**검토요점** 4), 둘째는 V17과 F2 간에는 요인계수가 해석할만한 크기가 안 되고 엉뚱한 F3과의 관계가 .589로 해석 대상이 되는 모습이었다(**검토요점** 5). 그런데 목표회전 결과의 MI값이 V16과 V17 간에 80.063이라는 값은 모형의 χ^2값이 218.109인 것에 비해 매우 큰 값이었기에 그것을 반영하여 측정오차 간 상관을 자유모수로 한다면 모형 내 왜곡된 부분에 대한 큰 개선효과가 기대되는 상황이었다. 그렇게 수정하려면 고전적 EFA로는 불가능하고 SEM의 기능인 측정오차 간 상관 추정의 기능을 EFA에 추가함으로써만 가능한 것이다. ESEM을 적용하여 V16과 V17 간 측정오차 간 상관이라는 하나의 모수를 고정모수에서 자유모수로 하여 추정하고 나니 모형의 합치도 향상은 물론 예기치 않게도 **검토요점 4와** 5의 문제가 해소되었다. 즉 V14와 F1 간 요인계수가 음의 큰 값이었는데 해석 대상이 안 될 정도로 바뀌었고, V17이 F3의 지표로만 나타나던 이제껏의 EFA 분석 결과와 달리 올바르게 F2의 지표만으로 나타나는 큰 변화가 있었다. 이러한 변화가 모형의 내용적 해석에 부합하는 방향이기에 ESEM은 EFA의 관행에 현저한 변화를 가져올 잠재력을 가지고 있다고 판단된다.

ESEM 사용상 주의

그러나 ESEM을 잘못 사용할 수 있는 소지도 있으므로 주의가 필요한데 그 중 두 가지만 언급하기로 한다. 첫째로, ESEM은 CFA를 대체하는 것이 아니다. CFA가 적절할 정도로 구조가 안정된 자료에서 굳이 탐색적으로 ESEM을 사용할 필요는 없다. 그 때는 MTMM (다특질 다방법)이나 전통적인 CFA처럼 처음부터 확인적 접근을 하는 것이 바람직하다 (Asparouhov & Muthen., 2009). ESEM의 본래 명칭인 EFA-SEM에서 보듯이, 측정모형에 대한 탐색을 해야 하는 경우에 보다 유연하게 하기 위해서 SEM의 기능을 도입하는 것이 ESEM이다. 둘째로, 측정오차 간 상관을 자유모수로 하여 추정하고자 할 때 자료 주도(예: 수정지수 MI 값 중심)의 분석 보다는 충분히 이론적으로 의미가 있는 경우에 한해서 추정하여야 할 것이다. 단순히 합치도만 높이고자 측정오차 간 상관을 다수 추정한다면 반복가능한 일반성 있는 결과가 되지 않을 것이다. 특히 다수의 척도가 포함된 검사집(battery)을 사용해서 얻은 자료를 분석할 때 하나의 척도에서 측정오차 간 상관을 자료주도로 많이 추정한다면 모형을 검증하는 자유도의 대부분이 나머지 척도들의 분석에서 발생하여 모형 검증이 왜곡될 수 있다(익명의 심사위원의 제언).

이상의 주의점을 염두에 두고 ESEM을 활용한다면 탐색적 연구에 새로운 지평을 열어줄 것이다. 앞으로의 요인구조 탐색에서 요인수효 결정 방식의 개선, 요인구조의 탐색적 회전과 부분제약 목표회전, 그리고 해석가능성의 확대는 21세기에 탐색적 요인분석에 변혁을 가져올 것으로 기대할 수가 있고 아울러 많은 구성개념의 탐색에 보다 정교한 도구가 되어 심리학의 발전에 크게 기여할 것으로 기대된다.

너무 작은 표본에는 요인분석을 할 수 없는가?

표 1에서 보면 표본이 너무 작으면(예: n=100 전후 또는 그 이하) ULS 계열의 추정법도 EFA의 모수 추정에 권하기 어려움을 알 수 있다. 표 1의 추정법들은 모집단 크기가 큰 경우를 중심으로 개발되었기 때문이다. 그러나 모집단의 원래 크기가 매우 작은 경우에도 집단원들의 행동에 일정수의 잠재변수가 있을 것으로 보는 행동과학적 모형을 적용할 수가 있어야 할 것이다. 예로서, 희귀병 환자들은 모집단이 10명도 안 될 수가 있고 특정 범죄자들의 모집단이라면 50명도 많을 수 있지만 그들의 행동적 특징을 충분히 수집하여 그에 대한 요인구조를 추정하고자 할 수가 있다. 이 때 Jung과 Lee(2011)의 방식이 사용가능하다.

이 새로운 접근법은 정형화된 탐색적 요인분석(REFA: Regularized Exploratory Factor Analysis)이라고 한다. 정형화(regularization)는 표본의 크기가 충분하지 않은 상황에서 모수추정의 안정성과 정확성을 향상시키기 위해 제안된 통계 방법으로서(예: 이미 회귀분석에 "ridge regression" 으로 적용) 탐색적 요인분석에 적용한 것이 REFA이다. 연구자들은 표본 크기가 작은 문제에 직면하면 보통 이를 해결하기 위해 표본을 더 수집하는 방향으로 나아가지만, 원래 모집단 크기가 작아서 표본 크기를 늘릴 수 없다면 추정해야 할 모수의 수를 줄이는 것이 대안이다. 이 대안이 EFA의 축소상관행렬($R-\Psi$)을 추정하는 단계에서 도입될 수 있다. 통상은 이 단계에서 측정변수 수효(p) 만

큼의 모수(공통분 또는 반대로 고유분산)가 추정되는 것인데, 측정변수 대비 표본 크기가 작은 경우에는 안정적으로 추정되기 어렵다. REFA는 이러한 문제를 해결하고자 공통분 추정 단계에서 아무리 p 가 커도 추정되어야 할 모수의 수는 단 1개로 가능하게 하였다. REFA의 방법론적 원리는 수학적으로 도출된 $\Psi = \lambda L$ (Jung & Takane, 2008)에 토대를 두고 있다. 이 식에서 추정해야 할 모수는 λ(정형화 모수: $0 < \lambda \leq 1$)이다. 그리고 L 은 고유분산의 초기값인데 역-이미지(anti-image) 행렬을 주로 선택한다(Jung & Lee, 2011). λ는 이 글의 표 1에 있는 EFA의 ML 또는 ULS 추정방법과 동일한 합치함수(fitting function)를 사용해서 추정된다.

참고문헌

김교헌, 권선중, 김세진, 이순묵 (2011). 저수준 도박행동 연구를 위한 개념화 및 척도개발. 한국심리학회지: 일반, 30(2), 599-628.

김종규, 이순묵, 윤창영 (2015). 핵심자기평가의 내적 구조 검토: 탐색적 구조방정식 모형(ESEM)의 적용을 통한 선언척도와 실제 척도의 차이 검토. 한국심리학회지: 산업 및 조직. 28(3), 355-384.

남궁준재, 이순묵, 김효선 (2013). 상황판단검사에서 시나리오 효과를 통제한 탐색적 요인분석. 한국심리학회지: 산업 및 조직. 26(4), 599-624.

안정원, 이순묵 (2015). 조직몰입 3요소 모형의 내적구조 검토: 탐색적 구조방정식 모형(ESEM)의 적용. 한국심리학회지: 산업 및 조직. 28(4), 795-827.

이순묵 (1995). 요인분석 I. 서울: 학지사

이순묵 (2000). 요인분석의 기초. 서울: 교육과학사.

장승민 (2015). 리커트 척도개발을 위한 탐색적 요인분석의 활용, 한국심리학회지: 임상, 34(4), 1079-1100.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 397-438.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology, 31*, 419-456.

Bentler, P. M. & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*(1), 78-117.

Bentler, P. M. & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research, 25*(1), 67-74.

Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443-459.

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-Information item factor analysis. *Applied Psychological measurement, 12*(3), 261-280.

Bovaird, J. A. & Koziol, N. A. (2012). Measurement models for ordered-categorical indicators. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp.495-511). NY: Guilford Press.

Briggs, N. E. & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research,* *38*(1), 25-56.

Brogden, H. E. (1969). Pattern, structure, and the interpretation of factors. *Psychological Bulletin, 72*, 375-378.

Browne, M. W. (1972). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology, 25*, 207-212.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*(1), 111-150.

Cattell, R. B (1966). The Scree test for the Number of Factors. *Multivariate Behavioral Research*, *1*, 245-276.

Cattell, R. B. & Dickman, K. (1962). A Dynamic Model of hysical influences demonstrating the necessity of oblique simple structure. *Psychological Bulletin, 59*, 389-400.

Cattell, R. B. & Gorsuch, R. L. (1963). "The Uniqueness and Significance of Simple Structure Demonstrated by Contrasting Organic 'Natural Structure' and 'Random Structure' Data." *Psychometrika, 28*, 55-67.

Cerny, B. A. & Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *The Journal of Multivariate Behavioral Research, 12*, 43-47.

Chen, J. & Choi, J. (2009). A comparison of maximum likelihood and expected a posteriori estimation for polychoric correlation using Monte Carlo simulation. *Journal of Modern Applied Statistical methods, 8*, 337-354.

Christofferson, A. (1975). Factor analysis of dichotomous variables. *Psychometrika, 40*, 5-32.

Cudeck, R. & O'Dell, L. L. (1994). Applications of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. *Psychological Bulletin, 115*, 475-487.

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), 272-299.

Finch, H. & Monahan, P. (2008). A bootstrap generalization of modified parallel analysis for IRT dimensionality assessment. *Applied Measurement in Education, 21*, 119-140.

Forero, C. G., Maydeu-Olivares, A. & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte-Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling, 16*, 625-641.

Gorsuch, R. L. (1983). *Factor Analysis*, 2nd Ed. Hillsdale, NJ: Lawrence Erlbaum.

Guilford, J. P. (1981). Higher-order structure-of-intellect abilities. *Multivariate Behavioral Research, 16*, 411-435.

Harman, H. H. (1976). *Modern factor analysis,* 3rd Ed. Chicago, IL: Chicago University Press.

Harman, H. H. & Jones, W. H. (1966). Factor analysis minimizing residuals(Minres). *Psychometrika, 31*, 351-368.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.

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.

Humphreys, L. G. & Ilgen, D. R. (1969). Note on a criterion for the number of common factors. *Educational and Psychological Measurement, 29*, 571-578.

Jöreskog, K. G. & Sörbom, D. (1989). *LISREL 7: A Guide to the Programs and Applications (2nd Ed.)*. Chicago, IL.: SPSS Inc.

Jung, S. & Lee, S. (2011). Exploratory factor analysis for small samples. *Behavioral Research Methods, 43*, 701-709.

Jung, S. & Takane, Y. (2008). Regularized exploratory factor analysis. In K. Shigemasu, A. Okada, T. Imaizumi, & T. Hoshino (Eds.). *New trends in psychometrics* (pp. 141-149). Tokyo: University Academic Press.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *34*, 183-202.

Knol DL, Berger MPF. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research, 26*, 457-477.

Lee, C-T, Zhang, G., & Edwards, M. C. (2012). Ordinary least squares estimation of parameters in exploratory factor analysis with ordinal data. *Multivariate Behavioral Research, 47*, 314-339.

Lee, S. (2010). A review of CEFA software: Comprehensive exploratory factor analysis program.. *International Journal of Testing, 10*, 95-103.

Lorenzo-Seva, U & Ferrando, P. J. (2015). *FACTOR 10.3*. University of Rovira i Virgili, Spain.

Lorenzo-Seva, U. & Ferrando, P. J. (2015). POLYMAT-C: A comprehensive SPSS program for computing the polychoric correlation matrix. *Behavioral Research Methods, 47*, 884-889.

MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research, 38*(1), 113-139.

McDonald, R. P. (1985). *Factor analysis and ralated methods*, Hillsdale, NJ: Lawerence Erlbaum.

McDonald, R. P. (2005). Semiconfirmatory factor analysis: The example of anxiety and depression. *Structural Equation Modeling, 12*(1), 163-172.

Millsap, R. E. & Kim, H. (In press). Factorial invariance across multiple populations in discrete and continuous data. In P. Irwing, T. Booth, & D. Hughes (Eds.), *The Wiley Handbook of Psychometric Testing*. London: John Wiley & Sons.

Montanelli, R. G. Jr. & Humphreys, L. G. (1976). Latent Roots of Random Data Correlation Matrices with Squared Multiple Correlation on the Diagonal: A Monte Carlo Study. *Psychometrika, 41*, 341-347.

Mulaik, S. A. (2010). *Foundations of factor analysis*, 2nd Ed. Boca Raton FL: Chapman & Hall.

Muthen, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*(1), 115-132.

Muthen, B. O. (1993). Goodness of fit with categorical and other non-normal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 205-243). Newbury Park, CA: Sage.

Muthen, B. O., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished manuscript.

Muthen, L. K. and Muthen, B. O. (1998-2012). *Mplus User's Guide*, 7th Ed. Los Angeles, CA: Muthen & Muthen.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of comonents using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers, 32*, 396-402.

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling, 7*(4), 557-595.

Preacher, K. J. & MacCallum, R. C. (2003). Repairing Tom Swift's Electric Factor Analysis machine. *Understanding Statistics, 2*(1), 13-43.

Rummell, R. J. (1970). *Applied factor analysis*. Evanston, IL: Northwestern Univ. Press.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology, 15*(2), 201-293.

ten Berge, J. M. F. & Kiers, H. A. L. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. *Psychometrika, 56*, 309-315.

Thomson, G. H. (1934). Hotelling's method modified to give Spearman's g. *Journal of Educational Psychology, 25*, 366-374.

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Thurstone, L. L. (1935). *The vectors of mind.* Chicago: Univ. of Chicago Press.

Timmerman, M. E. & Loranzo-Seva U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods, 16*(2), 209-220.

Vandenberg, R. J. & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods, 3*(1), 4-70.

Wirth, R. J. & Edwards, M. C. (2007). Items factor analysis: Current approaches and future directions. *Psychological Methods, 12*(1), 58-79.

Yu, C-Y (2012). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes.* Unublished Doctoral Thesis. University of California, Los Angeles.

1차원고접수 : 2015. 08. 19.
수정원고접수 : 2016. 01. 28.
최종게재결정 : 2016. 02. 16.

한국심리학회지 : 일반
Korean Journal of Psychology: General
2016, Vol. 35, No. 1, 217-255

Exploratory Factor Analysis: How has it Changed?

Soonmook Lee Sungkyunkwan Univ.
Chang-Young Youn Daegu Univ.
Minhyung (Mina) Lee Sungkyunkwan Univ.
Sunho Jung Kyunghee Univ.

In the present study new developments in EFA(Exploratory Factor Analysis) that have occurred at the turn of the 21th century are discussed. New guidelines and an analysis of real data following the guidelines are given with practical comments. First, in a process of determining the number of factors, MRFA (minimum rank factor analysis) is recommended as the best method of Parallel Analysis (PA) instead of Horn's method (1965) and PA-PAFA (parallel analysis in principal axis factor analysis). Various fit indices such as CFI, RMSEA, and etc. allow us to consider "various psychometric criteria" before determining the number of factors as the indices are less sensitive to sample sizes than the conventional χ^2 statistic. In addition estimation methods that are applicable to categorical data (dichotomous or polytomous) have been developed so that item factor analysis can be readily performed for categorical data. Second, in a process of rotating factor structures, "simple" oblique structure can be easily computed just by minimzing the value of complexity function, and a "partially specified target" rotaion is also available adopting a target matrix whose elements are partially hypothesized by a researcher. Finally, in a process of interpreting factor structures, ESEM (Exploratory Structural Equation Modeling) will get prevalence in the near future as it allows us to free correlations between unique factors (measurement errors) and can produce more practical and interpretable factor stucutures. New guidelines on EFA are described in detail in the later part of this paper.

Key words : exploratory factor analysis, common factor analysis, exploratory rotation, target rotation, exploratory sturctural equation modeling

부 록

1. FACTOR 10.3의 사용(구글에서 Factor 10.3을 치면 탐색가능)

1) FACTOR 10.3 열어서 'Read Data'(free format 의 원자료.dat 는 "participant's scores" 선택, 정방행렬로 된 상관행렬.dat 는 "variance/covariance matrix" 선택)에서 자료(결측치가 있는 응답자는 지워지면서 분석됨)를 올리고 'Configure Analysis'를 클릭한다.
2) [Matrix analyzed]에서 원자료에서 피어슨상관 분석이면 'Pearson Correlation', 다분상관 분석이면 'Polychoric (tetrachoric) correlation' 선택(상관자료 입력시 'Pearson' 상관으로 인식)
   - [Procedure for determining the number of factors/components]에서 'Parallel Analysis (PA)' 선택
   - [Factors & Components]에서 'Minimum Racnk Factor Analysis (MRFA)' 선택
   - FACTOR 10.3은 발견법의 결과만 참고할 것이므로 요인수효를 최대한 (변수 수효의 1/3 이하일 것)으로 주고 회전은 아무 것이나 선택하고 넘어간다.
   - 'Ok' 클릭
3) 'Compute'를 눌러 분석을 실행한다.
4) 고유치 검사 및 누적분산비율은 결과 창에서 'EIGENVALUES OF THE REDUCED CORRELATION MATRIX'이란 이름으로 나타난다(본문 표 3에 제시). 자료에 따라서는 [Factors & Components] 에서 'ULS'를 선택해야 이 결과가 나온다. 'EXPLAINED VARIANCE BASED ON EIGENVALUES'(원상관행렬에 대한 것)의 결과를 보고하지 않도록 주의한다.
5) 평행성 분석의 결과는 결과 창에서 'PARALLEL ANALYSIS (PA) BASED ON MINIMUM RANK FACTOR ANALYSIS'이란 이름으로 나타난다(본문의 표 4에 제시). 이때 분석된 상관행렬이 다분상관인지 피어슨 상관인지 유의해서 문장을 볼 것. 전자이면 무선 고유치의 95백분위수를, 후자이면 무선 고유치의 평균을 경험고유치에 비교.

2. Mplus Syntax

<Syntax 0: 저장된 다분상관 행렬을 불러서 각 변수의 SMC 구하기>

```
DATA: File=Poly.txt; Type=correlation; NObservations=368;
VARIABLE: Names=V1-V18;
Analysis: Type=general; Estimator=ML;
MODEL: V1 ON V2-V18; (Syntax 0을 전체복사, V2∼V18을 차례로 종속변수로 해서 반복실시 )
OUTPUT: STDY; (표준화 해 아래 부분에 SMC (R-square)가 제공됨)
```

<Syntax 1: 요인수효 1-5에 대한 문항요인분석(결과는 표 5에 반영)>

```
DATA:        FILE=data.txt; ! free format이 아니면 format 문장 필요.
VARIABLE:    NAMES=V1-V18;     CATEGORICAL=V1-V18;
ANALYSIS:    TYPE=EFA 1 5;
             ROWSTANDARDIZATION=KAISER; !행 표준화
             ROTATION=GEOMIN(OBLIQUE); ! 탐색적 회전
```

```
                ESTIMATOR=WLSMV;
OUTPUT:         MODINDICES (ALL);
```

<Syntax 2: 요인수효 3-4에 대한 문항요인분석(결과는 표 6에 반영)>

```
DATA:           FILE=data.txt;
VARIABLE:       NAMES=V1-V18; CATEGORICAL=V1-V18;
ANALYSIS:       TYPE=EFA 3 4; (나머지는 Syntax 1과 동일)
```

<Syntax 3: 5번 문항 제외 후 3요인 모형에 대한 문항요인분석(결과는 표 7, 8에 반영)>

```
DATA:           FILE=data.txt;
VARIABLE:       NAMES=V1-V18; CATEGORICAL=V1-V18;
                USEVARIABLES=V1-V4 V6-V18;
ANALYSIS:       TYPE=EFA 3 3; (나머지는 Syntax 2와 동일)
```

<Syntax 4: 목표회전(결과는 표 7에 반영)>

```
DATA:           FILE=data.txt;
VARIABLE:       NAMES=V1-V18; CATEGORICAL=V1-V18;
                USEVARIABLES=V1-V4 V6-V18;
ANALYSIS:       TYPE=GENERAL; ROTATION=TARGET;
                (표본이 작거나 목표회전에서는 행표준화 안함) ESTIMATOR=WLSMV;
                !DIFFTEST=ESEM1.dat;(Syntax 5에 있는 ESEM 수정모형 1의 분석후 그것과의
                카이제곱차이검증을 할 때 느낌표를 제거하고 다시 돌림)
MODEL:
        F1 BY
        V1      V2      V3~0
        V4~0            V6
        V7~0    V8~0    V9
        V10~0   V11~0   V12
        V13~0   V14~0   V15~0
        V16~0   V17~0   V18~0 (*1);

        F2 BY
        V1~0    V2~0    V3
        V4              V6~0
        V7~0    V8      V9~0
        V10     V11     V12~0
        V13~0   V14~0   V15~0
        V16~0   V17     V18~0 (*1);
```

```
        F3 BY
        V1~0      V2~0      V3~0
        V4~0                V6~0
        V7        V8~0      V9~0
        V10~0     V11       V12~0
        V13       V14       V15
        V16       V17~0     V18 (*1);
```

각 요인의 문장 끝에 (*1) 의 의미는 F1, F2, F3 이 EFA 요인들(요인분산=1)로서 하나의 셋트가 됨.

```
OUTPUT:         MODINDICES(ALL);
```

<Syntax 5: ESEM(수정모형 1)>

```
DATA:           FILE=data.txt;
VARIABLE:       NAMES=V1-V18;
                CATEGORICAL=V1-V18;
                USEVARIABLES=V1-V4 V6-V18;
ANALYSIS:       TYPE=GENERAL; ROTATION=TARGET; ESTIMATOR=WLSMV;
                !DIFFTEST=ESEM2.dat;(Syntax 6에 있는 ESEM 수정모형 2의 분석후 그것과의
                카이제곱차이검증을 할 때 느낌표를 제거하고 다시 돌림)

MODEL:
        F1 BY
        V1        V2        V3~0
        V4~0                V6
        V7~0      V8~0      V9
        V10~0     V11~0     V12
        V13~0     V14~0     V15~0
        V16~0     V17~0     V18~0 (*1);

        F2 BY
        V1~0      V2~0      V3
        V4                  V6~0
        V7~0      V8        V9~0
        V10       V11       V12~0
        V13~0     V14~0     V15~0
        V16~0     V17       V18~0 (*1);
```

```
F3 BY
        V1~0      V2~0      V3~0
        V4~0                V6~0
        V7        V8~0      V9~0
        V10~0     V11       V12~0
        V13       V14       V15
        V16       V17~0     V18 (*1);

        V16 WITH V17; ! 측정오차 간 상관 추정

OUTPUT:         MODINDICES(ALL); TECH4;
SAVEDATA:       DIFFTEST=ESEM1.dat;(Syntax 4 의 DIFFTEST에서 이 파일 사용)
```

<Syntax 6: ESEM(수정모형 2)>

```
DATA:           FILE=data.txt;
VARIABLE:       NAMES=V1-V18;
                CATEGORICAL=V1-V18;
                USEVARIABLES=V1-V4 V6-V18;
ANALYSIS:       TYPE=GENERAL; ROTATION=TARGET; ESTIMATOR=WLSMV;
MODEL: (ESEM 수정모형1과 동일하고 다음의 문장만 한줄 추가)
        V1 WITH V2;

OUTPUT:         MODINDICES(ALL); TECH4;
SAVEDATA:       DIFFTEST=ESEM2.dat;(Syntax 5의 DIFFTEST에서 이 파일 사용)
```