

이 연구는 다수의 소검사로 이루어진 검사에서의 요인구조 추출 및 검사의 신뢰도에 대한 것이다. 소검사(testlet)에 대한 측정적 관점에서 정의는 Lee, Brennan, 및 Frisbie(2000)에서 검사제작, 시행, 및 채점에 관련하여 제시되고 있으나 이 연구에서는 제작단계에 중점을 둔다. 제작시 형태상 실습과제, 시나리오 또는 문단이 있고 그 아래 한 묶음의 채점항목 또는 문항군이 있는 형식을 중심으로 한다(Wainer & Kiely, 1987). 상황이나 문단을 주고서 그 밑에 반응대안이 제시되는 사회적 능력검사, 실용지능검사(예: Wagner & Sternberg, 1993), 그리고 산업장면에서 널리 사용되는 모의과제 중심의 평가센터, 지필형 상황판단검사의 시나리오형(예: Motowidlo, Dunnette, & Carter, 1990)이 바로 이런 모습이다. 이런 검사들은 채점항목이나 문항에 맥락을 제시하는, 이른바 맥락의존적 검사에서 많이 발견되며 전통적 지능검사와 같은 탈맥락적 검사들에서도 문제를 영역별로 제작할 때 또는 독해력이나 지도 읽기 등에서 사용된다. 또한 소검사에서의 모의과제, 상황, 시나리오, 문단이 실제장면과의 가까운 정도 즉 현실충실도를 중심으로 할 때 그림 1과 같이 나누어 볼 수 있다.

이러한 소검사로 이루어진 검사를 사용해서

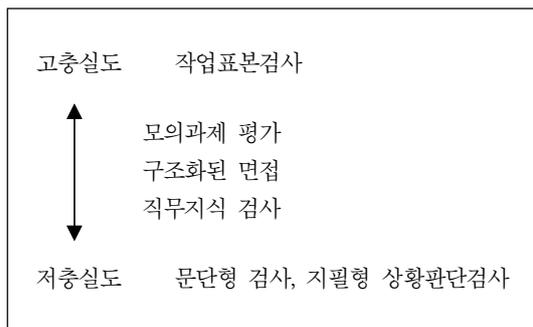


그림 1. 소검사로 이루어진 검사의 현실충실도에 따른 분류

자료가 수집될 경우 각 소검사내의 국지적 의존성(local dependence) 때문에 고전검사이론에서의 측정오차의 독립성 및 문항반응이론에서의 국지적 독립성(local independence)이 크게 위배될 수 있다. 국지적 독립이란 요인의 값이 고정될 때 그 요인을 채는 두 문항간에 상관관계가 없음 즉, 요인이 통제될 때(partial out) 문항간에 부분상관이 영임을 의미한다(Lord, 1980). 그러나 그 요인이 통제되어도, 상황이 있을 경우 문항간 부분상관이 0이 되지 않는다. 즉, 요인 이외에도 상황이 문항간 관계에 영향을 미치는 경우이다. 그럴 경우, 요인구조의 도출 및 신뢰도 산출에도 어려움이 있다. 요인분석을 선형과 비선형으로 나눌 때 전자는 일반적인 요인분석 즉 주성분분석과 공통 요인분석을 의미하고 후자는 문항반응이론을 의미한다. 문항반응이론에서 국지적 독립성을 명시적으로 요구한다면 선형적 요인분석에서는 암묵적으로 가정하고 있다(McDonald, 1985). 결과적으로 소검사로 이루어진 검사 자료에서는 검사가 무엇을 측정하는지 얼마나 정확하게 측정하는지를 알기 어렵다. 소검사에서 맥락을 제시하는 실제상황, 모의상황, 시나리오, 문단 등을 통괄해서 “상황”으로 부르기로 한다. 또한 ‘소검사로 이루어진 검사’를 소검사기반검사로 부르기로 한다.

소검사기반검사의 가장 큰 문제는 상황 자체가 채점자 또는 응답자에게 제공하는 자극의 역할을 넘어서, 각 문항에 대한 해석의 방식 및 응답의 유형을 제약하는 환경이 되면서 발생한다. 소검사내 문항간에는 이렇게 측정대상의 내용이 아니면서 측정에 미치는 영향의 동질성으로 인하여 문항간 관련성 증가를 가져온다. 측정이론에서 동일한 측정방법의 사용으로 인한 효과를 약해서 방법효과로 부른다. 상황에 의한 문항간 관련성 증가 역시 방법효과 일종이다. 따라서 상황으로 인한 방법효과를 제거하고 내용점수

를 산출하는 방식이 필요하게 되었다(예: Wang, Bradlow, & Wainer, 2001). 상황에 의한 방법효과는 요인이라는 잠재변수로 표현할 때 1개일 수도 있고 다수가 도출될 수도 있다. 현재의 방식에서는(예: Wang 등, 2001) 일단 전체검사내에 상황에 의한 방법요인이 1요인밖에 없다는 가정하에 시나리오 또는 문단의 효과를 제거하는 문항 반응이론의 수리모형을 사용하고 있다. 이 때의 문항반응이론에서 내용적 잠재변수는 1개로 가정되고 있다. 그러나 정작 소검사로 구성된 검사에서 몇 개의 내용요인이 존재하는지 그리고 그 추출은 어떻게 하는지에 대한 것은 아직 충분한 연구가 되어 있지 않다. 요인구조에 대한 주장이 있는 후에, 요인별로 Wang 등(2001)의 절차를 통해서 요인점수를 산출할 수 있을 것이다. 이 연구는 소검사가 반검사에서 요인구조의 추정 그리고 그러한 검사에서의 신뢰도 추정의 방법을 제시하고자 한다.

소검사로 구성된 검사를 분석하기 위해서는 소검사에서 주어지는 상황이 고정효과인지 무선효과(random effect)인지를 구분해야 한다. 하나의 상황이 상황의 집합(universe)으로부터 랜덤(random)하게 뽑혀 나온 것이라면 무선효과를 가지는 것이고, 이 상황효과는 상이한 검사에 걸쳐 반복되는 것이 아니므로 고정된 맥락으로 해석되지 않는다. 따라서 분석과정에서도 오차부분에 포함된다. 그러나 상황이 고정되어 모든 검사에서 반복되어 사용된다면(예: 특정의 구성개념을 측정하기 위한 상황이 동일하고, 그에 근거해서 묻는 질문만이 검사에 따라 달라질 경우) 그 상황은 고정효과를 가지는 것이고 분석에서도 고정효과를 가지는 요인(factor)이 된다. 이 때는 오차가 아니라 상황요인으로 분석될 수 있다(Lance 등, 2000).

예로서 산업장면의 평가센터에서 리더십을 측

정기 위한 실기평가에서, 시험관이 다수 있고 그들은 다양한 실기과제에서 뽑은 두 개의 상이한 리더십 실기(exercise)를 가지고 응시자를 시험한다고 하자. 각 시험관이 각 실기과제에 근거하여 1개씩 자유로이 질문을 던진다면 응시자의 점수에는 상황에 관계없이 발휘되는 리더십 역량과 각 리더십 실기상황의 효과 및 무선적인 측정오차가 포함된다. 이 때 리더십 실기상황이 많은 상황중에서 랜덤하게 뽑힌 상황이라면 개인응답자 점수에 반영되는 상황효과의 기대치는 0이 된다. 이 경우는 소검사 상황의 무선효과모형이다.

한편 두 개의 리더십 실기중 1개는 일대일 상황이고 다른 1개는 집단상황이며, 리더십 상황을 크게 나누어 이 두가지로 나누는 것이 관리목적상 타당하고 앞으로의 검사에서 계속 이러한 분류방식을 사용한다고 하자. 그렇다면 응시자의 점수에서 앞서의 무선효과모형과 다른 점이 발생한다. 즉, 각 리더십 실기상황은 고정되어 있으므로 응답점수에 반영되는 상황효과 역시 0이 아닌 상수가 되며 내용점수와 동일한 자격으로 해석되어야 한다. 즉, 소검사상황의 고정효과모형이다. 따라서 일대일 상황과제에서 점수는 “일대일 관계에서의” 리더십을 측정하는 것이고, 집단 상황과제에서의 점수는 “일대다수 관계에서의” 리더십을 측정하는 것이 된다. 이 해석에서 “일대일 관계에서의”와 “일대다수 관계에서의”는 리더십이라는 활동이 발생하는 맥락을 제공한다.

심리학에서 구성개념에 관련없는 나머지를 구성개념에 무관한 분산으로 처리하는 인지주의적 관점이 무선효과모형에서 잘 반영된다면, 구성개념이 발생하는 맥락을 제공하는 사회문화적 관점은 고정효과모형에 의해 잘 반영된다고 볼 수 있다. 따라서 상황에 의한 방법효과는 인지주의 관점에서 볼 때 방법편파(method bias)가 되고 사

회문화적 관점에서는 인간행동과 분리될 수 없는 환경 또는 맥락이 된다(예: 상황이 빠진 ‘상황 인지’는 정의되지 않는다). 동일한 경험자료에 어떤 모형이 존재하느냐 하는 것은 연구자의 연구 설계에 따라 가정되는 것이고 그 가정은 분석결과에 의해 지지되거나 기각될 것이다.

그림 1에서 상황의 고충실도로 갈수록 상황효과는 고정효과모형이 되고 저충실도로 갈수록 무선효과모형이 되기 쉽다. 예로서 평가센터에서는 고정된 실기과제가 있고 질문도 특정의 구성개념을 목표로 하여 제시되므로 상황효과는 고정효과변수이다. 그러나 지필형 상황판단검사의 경우 현장에서의 많은 중요상황에 근거하여 상황시나리오를 불특정 다수로 제작한다면 이 때의 상황효과는 무선효과변수이다.

이 연구는 소검사기반검사 자료에 대한 요인 분석에서 상황효과가 고정효과모형일 경우와 무선효과모형인 경우를 어떻게 구분해서 다루는가를 제시한다. 평가센터에서의 측정이 전자의 사례로 사용되고, 신입사원 선발장면에서 쓰이는 지필형 상황판단검사를 후자의 사례로 사용할 예정이다. 이 두 사례를 가지고 소검사기반검사에서의 요인구조 도출 및 신뢰도 산출의 방식을 제시하는 것이 이 연구의 내용이다.

연구문제

소검사기반검사는 심리학과 교육학의 두 분야에서 중요한 연구프로그램이 되어 있다. 교육측정에서는 단순정신능력보다 복합정신능력을 측정하려는 경향속에서 그 중요도가 인식되었고 사회심리나 산업심리 분야에서는 실사회 또는 실생활에서의 역량을 측정하려는 경향에서 사회적 지능, 실용지능, 상황판단력 등의 검사가 개

발되면서 부터라고 할 수 있다. 이러한 검사들의 기원은 1920년대로 거슬러 갈 수 있다(McDaniel, Morgeson, Finnegan, & Campion, 2001),

심리학에서 소검사기반검사의 구성개념에 대한 연구가 시작되었다가 어느 정도 방치된 상태인데, 역사적으로 보면 1960년대말부터 인사선발과 인적자원개발에 널리 쓰여온 평가센터(Assessment Center), 1990년대에 저충실도 시뮬레이션이란 이름으로 등장한 지필형 상황판단검사(Motowidlo, Dunnette, & Carter, 1990)가 소검사기반검사의 유행을 가져왔다. 전자에서의 상황과제들이 고정효과의 모형이라면 후자에서의 상황시나리오들은 무선효과모형이다. 평가센터에서 실시되는 소검사기반의 검사에서 오랜동안 내용차원에 대한 요인을 도출하고자 하였지만 과제중심의 요인만이 도출된 결과 평가센터의 검사에서 내용은 생각지 말자는 주장들이 있었다(예: Lowry, 1997; Robertson, Gratton, & Sharpley, 1987).

한편 이러한 영향을 받아서 지필형 상황판단검사의 연구에서는 요인분석을 시도하지도 않았다. 지필형 상황판단검사의 제작에서 내용요인들(예: 문제해결, 대인관계스킬, 설득력, 전술, 서비스지향성 등)을 측정하겠다는 의지를 가지고 중요사건들을 수집하여 시나리오를 만들고 그에 따른 문항들이 작성되지만 막상 사용시에는 대부분의 경우(예외: Wagner & Sternberg, 1993) 소검사내에서 최적/최악의 행동대안에 대한 선택을 요구한 후 그에 따라서 소검사별로 하나의 점수를 매긴 후 전체 소검사에 걸친 총점을 가지고 경험연구에 사용하고 있다. 따라서 신뢰도는 각 상황별 점수 즉, 소검사점수에 기초한 신뢰도가 보고된다. 내용요인 중심의 접근보다는 단순히 총점중심으로 하였고, 소검사기반의 요인분석이 실시되지도 않았다.

그러나 최근에 그런 검사들이 무엇을 측정하

느냐 하는 구성개념타당도의 질문이 크게 제기되면서 간접적으로라도 소검사기반검사의 구성개념을 파악하려는 움직임이 있어왔다. McDaniel, Morgeson, Finnegan, 및 Campion(2001)은 102개의 상관계수와 10640명에 기초한 메타분석 결과 지필형 상황판단검사에 준거타당도($\rho=.34$)가 있음을 보여주었고, 아울러 이 검사가 인지능력과 관계($\rho=.46$) 있음을 제시하였다. 이렇게 소검사기반검사의 총점을 다른 검사와 상관시키므로써 소검사기반검사의 구성개념을 간접적으로 파악하고자 한 연구는 많이 있다(예: Wagner & Sternberg, 1993; Chan & Schmitt, 1991; Oswald, Schmitt, Ramsay, Kim, & Gillespie, 2003; Wagner & Sternberg, 1993). 그러나 소검사내 문항들의 점수에 대한 요인분석을 통해서 직접적인 접근을 한 예는 아직 없다. 그렇다면 이러한 검사가 가지는 타당도는 측정되는 내용이 가지는 타당도가 아니라 단순히 소검사를 통한 측정방법이 가지는 타당도가 아닌가 하는 반론도 강하게 제기되고 있다(예: McDaniel, et. al., 2001; Weekley & Jones, 1999). 즉 무엇이 측정되는지는 잘 모르지만 그 방법만큼은 좋은 준거타당도를 가지고 있는 것으로 보자는 것이다.

한편 교육측정에서 보면 최근 15년 사이 소검사기반검사의 신뢰도를 중심으로 한 연구가 왕성하였다. 교육측정 분야에서는 소검사내의 관련성이 소검사간 관련성보다 클 경우에 발생하는 국지적 의존성(local dependence)을 감안하면서 어떻게 신뢰도를 산출하고 문항반응이론을 적용하느냐에 대한 연구가 중요한 연구프로그램이 되었다(예: Wainer & Kiely, 1987; Thissen, Sternberg, & Mooney, 1989; Wainer & Lewis, 1990).

Wainer와 Lewis(1990)는 국지적 의존성이 있을 때는 문단내 문항을 1개로 하거나, 문항간 의존성을 무시하고 문항기반의 신뢰도를 구하든가,

문단에 관련된 문항들만으로 하나의 점수를 산출해서 소검사기반의 신뢰도를 산출하든가 하는 세가지 가능성을 제시한 바 있다. 그러나 국지적 의존성이 있을 때 없는 것으로 간주하고 산출하는 문항기반의 신뢰도는 실제보다 큰 값이 되며(Sireci, Thissen, & Wainer, 1991; Wainer, 1995; Wainer & Thissen, 1996), 국지적 의존성이 크다면 각 소검사에서 하나의 점수를 산출해서 소검사점수기반의 신뢰도를 산출해야 한다는 주장도 있다(예: Frisbie & Druva, 1986). 또한 1요인모형의 문항반응이론의 적용을 위해 상황효과를 무선희효과로 취급하고, 구성요인에 무관한 부분으로 처리하는 모형이 개발되었다(Wang, Bradlow, & Wainer, 2001).

이제껏 제시된 심리학 및 교육학에서의 연구를 요약하면 본 연구의 연구문제가 도출된다. 우선 교육측정에서는 신뢰도에 대하여 보다 폭넓은 연구를 해왔다. 특히 소검사기반검사의 신뢰도는 문항점수기반의 신뢰도에 의해 과대추정되고, 소검사점수를 사용할 경우 혹시라도 일부 문항들 사이에 존재할 수 있는 국지적 독립성을 무시하므로써 정보의 상실이 되므로 신뢰도의 과소추정이 된다는 것이다(Yen, 1993). 따라서 보다 적절한 신뢰도 산출의 방식이 요구되고 있다.

그러나 교육학에서의 연구는 1요인을 가정하는 문항반응이론이 중심적인 도구가 되어 왔기에 다요인이 가능한 소검사기반검사에 대해서는 아무런 시사점이 없다. 한편 심리학 분야에서 널리 사용되어 온 소검사기반검사에 대한 연구는 구성요인에 관심을 가진 연구가 꾸준히 하나의 축을 이루면서 최근에는 평가센터에서의 검사에 내용요인과 상황요인이 동시에 존재한다는 결합모형을 제시하게 되었다(Lievens & Conway, 2001). 이것은 다특질다방법(MTMM: multitrait multimethod) 자료의 분석을 위해 확인적 요인분석이란 도구

를 사용하는 모형(예: Widaman, 1985)이 제시된 효과이기도 하다. Lievens와 Conway(2001)는 비록 내용요인의 분산과 상황요인의 분산간에 크기는 어느 한쪽이 크고 다른 한쪽이 작을 수 있으나 두 범주의 요인이 함께 도출됨을 보이고 있다. 이들의 연구에서 보면 상황효과가 훨씬 큰 경우도 있지만 내용요인이 전혀 도출될 수 없는 정도는 아니었다. 그러나 무선평과모형인 지필형 상황판단검사에 대해서는 아직껏 요인분석이 시도되지 않고 있는데 이것은 내용요인들이 평가센터에서와 같이 분명하게 설정되어 있지 않고 작업현장에서 중요사건들을 모아서 상황시나리오를 기술했기 때문에 그에 속한 문항들이 특별히 어떤 내용을 측정하는지 지정된 바가 없는 경우가 많기 때문이다.

검사 제작시 상황에 따른 문항들이 어떤 내용을 측정하는지에 대한 사전적인 범주화가 도입되면 평가센터 자료에서와 같이 MTMM자료로 간주하고 그에 대한 분석방법론을 적용할 수 있을 것이다. 그런데 소검사기반의 검사 자료를 MTMM자료에 준해서 분석하기 위해서는 일정한 논리가 필요하다. 즉, 이제껏 적용된 경우는(예: Lievens & Conway, 2001; Schneider & Schmitt, 1992; Conway, 1998) 모두가 고정효과모형을 암묵적으로 가정하였다. 그러나 소검사기반의 검사 자료에는 상황의 특성상 무선평과모형이 있을 수 있음을 연구자들은 간과하였다. 상황효과가 고정효과이면 상황요인은 더 이상 오차가 아니므로 신뢰도 산출시 진점수의 부분에 들어가게 되고, 무선평과모형이면 상황효과는 상황요인이 아니라 측정의 오차구조에 포함되므로 신뢰도 산출시 오차점수의 부분에 들어가게 된다. 따라서 고정효과/무선평과 개념의 도입을 통해 모형의 정확한 추정을 기하고 그에 따라 올바른 신뢰도의 산출을 하는 것이 필요하다.

자료수집의 설계가 MTMM자료에 유사하게 준비되고, 검사제작시에 상황에 대한 모형이 고정효과인지 무선평과효과인지를 안다고 해서 분석 준비가 된 것은 아니다. 수집된 자료내에 과연 상황효과가 있는지 아닌지가 검토되어야 한다. 검토결과 상황효과가 있을 것으로 기대되는 경우와 없을 것으로 기대되는 경우에 따라서 모형을 찾기 위한 경로가 달라진다. MTMM방식의 자료에 대한 모형찾기는 Widaman(1985) 이래로 확인적 요인분석의 사용을 통해 많은 발전을 이루어 왔으나 소검사기반의 자료에 대한 분석의 모형으로 충분히 적용된 상태는 아니며, 상황의 고정효과모형/무선평과모형까지 고려한 상태는 더욱 아니다.

이 연구에서는 우선 소검사기반검사에서 요인구조를 추정하기 위한 이론적 조망 및 요인구조의 모형찾기에 대한 조망을 한 후 이어서 실제 자료를 사용하여 모형찾기를 제시한다. 모형찾기의 우선 순서는 상황내 상관평균과 상황간 상관평균의 비교를 통해서, 상황효과가 존재하는 정도를 검토한다. 그 결과로 Widaman(1985)의 분석모형중 일부를 택하여 모형찾기의 경로를 정하고, 최적의 모형에서 상황효과를 고정효과모형과 무선평과모형으로 나누어 해석한 다음 그에 따라 신뢰도의 산출방식이 다름을 보이고자 한다.

소검사로 이루어진 검사의 요인구조

주어진 소검사기반검사에서 상황이 분명하고, 측정되는 내용요인에 대한 어느 정도의 범주화가 있다고 하자. 공통요인분석의 모형을 가지고 그러한 자료에서 요인구조를 추출하는 논리를 구하고자 한다.

검사점수 Z_j 에 내용요인(F)과, 측정의 상황으로 인한 특수성(S)이 잠재되어 있을 때 공통요인분석을 하는 표현은 식(1)과 같다(Harman, 1976, p.19).

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m + b_jS_j + e_jE_j \dots \dots \dots (1)$$

F는 공통요인이고 F간 상관이 가능하다. 그러나 S는 측정시 Z의 특수성(특수요인, specificity)으로서 F나 E에 상관이 없고, E는 측정치의 불완전함 즉, 랜덤오차를 나타낸다. Z, F, S, E는 표준화된(평균 0, 분산 1) 변수이다. 본 연구에서의 “상황”은 각 측정치의 특수요인 S에 속한다. 만일에 상황효과가 무선효과모형이면 식(1)에 있는 b_j 의 행동은 다음과 같다.

$$b_j = \gamma_0 + \gamma_1 w_j + r_j \dots \dots \dots$$

(2)

여기서 γ_0 는 각 “상황”에서의 응답자들 평균점수를 전체 상황에 대해 평균한 값

w 는 상황의 특수성을 예측하는 변수

γ_1 는 w_j 의 회귀계수

r_j 는 w_j 에 의해 설명되지 않는 랜덤오차

식(2)에서 γ_0 는 상수이다. 본 연구에서와 같이 상황에 대하여 예측변수 w_j 가 없다면 γ_0 를 0으로 만드는 재척도화(rescaling)를 거쳐서 b_j 는 e_j 와 같은 성질을 가지며 특수성은 E와는 독립인 그러나 또 하나의 랜덤오차가 된다(Raudenbush & Bryk, 2002 참조). 이 결과는 공통요인분석의 문헌에서 고유요인분산을

$$U_j^2 = b_j^2 + e_j^2$$

으로 표현할 수 있는 논리를 제공한다. 이 때 식(1)은 아래와 같이 된다.

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m + U_j \dots (3)$$

여기서 U는 고유요인(unique factor)이다.

식(3)의 모형을 구조방정식모형으로 검증하고자 할 때 소검사기반검사에서의 상황은 그 상황에서 측정되는 측정치의 고유요인(구조방정식모형에서 측정오차) 사이에 상관으로 나타난다. 이것은 상황이 무선효과변수 일때의 전형적인 모습이다.

한편 S_j 가 고정효과를 가진다면 어떤 “상황”에서도 특수성의 효과는 평균적으로 b_j 가 된다. 즉, 일반적인 회귀분석(고정효과모형임)의 논리와 같아진다. 식(1)에서 공통요인 F가 있고 요인계수를 고정된 값으로 제시한다는 것은 이미 고정효과모형을 가정하고 있는 것인데, S_j 가 고정효과모형을 가진다면 수학적으로 F와 동일한 자격을 가진다. 따라서 소검사기반검사 자료를 MTMM자료에 준해서 구조방정식모형으로 추정하고자 할 때 “상황”은 요인으로 표시되고 추정되어야 한다. 이 때 각 상황을 하나의 요인(M)으로 표시한다면 식(1)은 아래와 같다.

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + b_{j1}M_1 + b_{j2}M_2 + \dots + e_jE_j \dots \dots \dots (4)$$

식(4)에서 F와 M은 모두가 고정효과로서의 요인계수를 가지고 있으며, 공통요인이다. 만일에 이런 모형에 의해서 산출된 자료에 “탐색적인” 요인분석을 한다면 F와 M의 혼재로 인하여 내용요인과 방법요인을 구분하기 매우 어려울 것이다. 이 때 각 상황을 연구자가 알고 있으므로, 각 요인에 대한 지표변수들(indicators)이 하나의 상황을 나타내면 방법요인이고, 각 상황을 나타내는 지표변수가 한두개 정도밖에 안될 정도로

다수상황을 나타내면 다수상황에 걸쳐 수렴되는 내용요인으로 볼 수 있다. 물론 신중한 추론이 필요하다.

이 연구에서는 소검사기반검사의 제작시 연구자가 각 문항들을 통해서 어떤 내용을 측정할 것인지 어느 정도의 사전적 범주화가 된 경우의 자료에 대한 요인분석 및 신뢰도 산출을 보이고자 한다. 대부분의 소검사기반검사 자료를 MTMM 자료에 준해서 요인분석을 할 경우 내용요인과 상황이 잘 정교하게 정리된 모형이 아니므로 온전한 “확인적” 요인분석이 되지 않고 모형찾기라는 탐색과정을 거쳐야 한다. 일단 시초모형을 설정하고 최적모형을 찾는 과정을 모형찾기(specification search)라 한다. 즉 범주화된 내용요인과 상황을 마치 잘 구안된 구조처럼 시초가설모형으로 하여 주어진 자료에 단번에 검증해보려는 탐색적 분위기의 모형찾기를 하게 된다. 이때 식(1)로 표현되는 측정변수들의 자료에서 상황내 상관평균과, 상황간 상관평균 간의 비교가 모형찾기에 좋은 지침을 제공한다. 이 때 비교되는 상관들은 하나의 자료에서 구했으므로 독립된 상관이 아니라 어느 정도는 의존성이 있다. 따라서 엄밀하게는 의존성있는 상관계수간의 비교검증을(Steiger, 1980 참조) 해야할 것이지만 모형찾기의 목적은 탐색적인 측면이 많기 있으므로 연구자의 질적 판단으로도 족할 것이다.

검사제작시 상황효과를 고정효과모형으로 가정한 경우에는 상황요인 중심의 모형찾기를 한다. 그러나 상황내 상관과 상황간 상관의 비교에 따라 어떻게 모형찾기에 대한 함의가 달라지는지 보기로 한다. 상황내 상관평균에서 상황간 상관평균을 뺀 값을 “상관차이”로 부르기로 한다. 상관차이가 현저할 경우, 모형찾기 결과로 상황요인이 대체로 도출될 것으로 본다. 그 차이가 0에 가까울 경우 상황요인 중심의 모형찾기에서

검사내 상황의 수효에 비해 상황요인의 수가 현저히 감소될 것이 예상된다.

검사제작시 상황효과를 무선평과모형으로 가정한 경우에는 특질요인 중심의 모형찾기를 한다. 이 때 상관차이가 현저할 경우 고유요인간의 상관들이 유의할 것으로 예상된다. 즉, 하나의 상황에 대한 문항들에서 고유요인간의 상관들이 대체로 유의할 것이다. 그러나 상관차이가 0에 가까울 경우 이 고유요인간의 상관들이 유의한 경우가 대폭 감소될 것이 예상된다.

그러면 모형찾기에 대하여 보다 세부적으로 살펴본 후 분석 사례를 제시하기로 한다.

모형찾기

요인구조의 측정 및 모형찾기에는 MTMM자료 분석에 대한 Widaman(1985)의 모형분류가 좋은 참조가 된다. 그 중 일부를 이 연구의 목적에 따라 변형하여 표 1에 제시하였다. 우선은 측정변수간 상관자료에 아무런 구조가 없다는 모형 즉, 측정변수간 상관이 모두 영이라는 영모형을 검증한다. 이 모형이 기각되면, 다음은 상황요인 중심의 모형찾기를 할 경우, 표 1에서 방법효과 구조에 초점을 둔다. 내용요인(특질요인) 중심의 모형찾기를 할 경우 표 1에서 특질구조에 초점을 둔다.

상황요인 중심의 모형찾기

OTUM(no trait factor, uncorrelated method factors) 모형에서 출발해서 OTCM(no trait factor, correlated method factors)모형과의 비교를 한다. 여기서 주의할 것은 OTCM모형에 대한 대안으로서 ITUM(또는 1TCU)모형을 검증해 보아야 한다(CU모형

은 뒤에 오는 절에서 설명). Marsh(1989)는 OTUM 무엇인가 구성개념을 잴다는 수렴타당도가 있는 표 1. 소검사기반검사에 대한 분석모형들^a

↓ 특질구조	방법효과 구조 →				
	0M	1M	UM ≈ CU	CM ≈ CU' ^c	
0T	영모형	0T1M ^b	0TUM -	0TCM -	
1T	1T0M ^b	1T1M ^b	1TUM ≈ 1TCU	1TCM ≈ 1TCU'	
CT	CT0M	CT1M ≈ CTCU	CTUM ≈ CTCU	CTCM ≈ CTCU'	

^a Widaman(1985)의 분류를 변형한 것임. M=method factor, T=trait factor, UM=uncorrelated method factors, CU=correlated uniqueness, CM=correlated method factors.
^b 0T1M모형과 1T0M모형은 동치모형으로서 1개의 요인이 특질인지 방법효과인지 알 수가 없고, 1T1M모형 역시 두 요인중 각각 어느 것이 특질인지 방법효과인지 구분할 길이 없으므로 현실적으로 쓰이지 않는다.
^c CU'는 상황간에 고유요인상관이 맺어지는 모형으로서 CM모형에 대한 대안적 표현이다. 상황간에 고유요인 상관을 검증할 수 있는 점이 CU모형보다 장점이다.

과 같이 방법요인이 도입된 확인적 요인분석에서 수리적으로 적절한 결과가 있어도 방법요인들은 어느정도 특질분산을 나타내므로, 방법요인간 상관을 도입하고자 할 때는 그에 대한 대안도 검토할 것을 제시하였다. 즉 방법요인간 상관가능성은 특질요인의 존재를 시사할 수도 있다는 것이다. 이 점은 James, Demaree, Mulaik, 및 Ladd(1992)에서 상황은 준거와 관련이 있음을 보여주므로써도 지지되고 있다. 준거는 원래가 예측변수인 특질과 관련이 있도록 기대되는 것인데, 특질이 작용하는 환경인 상황과도 관련되므로, 상황이 특질을 반영할 수 있음을 보여주는 것이다. Conway(1996)와 Lievens와 Conway(2001) 역시 그 견해를 지지하고 있다. 대안을 비교한 결과 1T(one trait)의 모형이 0TCM모형보다 못하다면 방법효과밖에는 없는 모형이 된다.

만일에 0TCM모형에 대한 대안으로서 1T(1TUM 또는 1TCU)모형을 검토한 결과 1T모형이 선택되었다면 특질들에 대한 측정이 적어도 하나의 구성개념으로 수렴하는 것을 나타낸다. 즉,

것으로 본다. 다음은 이 1T의 모형을 원래 가정했던 k개의 특질(trait)이 있는 모형으로 확대하였을 때 후자의 모형이 더 합치도가 좋으면 각 특질은 변별타당도가 있는 것으로 본다. 이 때 특별한 가설이 없으면 k개 특질간에 상관이 있는 것으로 설정한다(CT모형). 물론 각 특질들의 분산이 유의한지 여부에 따라 일부의 특질은 수렴타당도가 없음이 밝혀질 수도 있고 특질간 상관계수가 1.0에 가까우면 그 두 특질간에는 변별타당도가 없는 것으로 해석하게 된다.

내용요인 중심의 모형찾기

원래 가정된 내용요인들을 기초로 CT0M(correlated trait factors, no method factor)모형을 설정해서 출발한다. 여기서 몇 개의 내용요인간에는 변별이 안되는 것으로 나타나 원래의 k개보다 적은 수효의 내용요인들로 구성된 CT0M모형이 될 수도 있다. 이 모형을 각 상황내 고유요인간 상관을 허용하는 CTCU모형과 비교해 볼 수

있다. 한번에 모든 상황에 대한 CTCU모형을 검토할 경우 추정상에 어려움이 있을 수 있다. 우선은 하나의 상황에 대해서 CU모형을 도입해서 검토해 본다. 이렇게 해서 각 상황별로 CTCU모형을 검증했을 때 합치도가 적절하면 처음의 CU모형에 다른 상황을 하나씩 더 CU모형의 형태로 추가하므로써 궁극적으로는 원래의 상황을 “모두” CU의 형태로 한 모형이(‘전체상황CU모형’으로 부름) 적절한지 검토할 수 있다.

그러나 각 상황에서 모두 CU모형이 유의한 합치도의 증가를 보이지 않거나, 부적절해가 나올 수 있다. 이 때는 앞단계의 모형인 CTOM으로 돌아가서 그 분석결과에서 고유요인간 상관가능성을 검토해야 한다. 이 검토는 이론적, 경험적(수정지수인 MI값 참조)으로 한다. 또는 처음 상황에 대한 CU모형은 적절해도 다음에 상황을 추가하는 어느 단계에서 더 이상 적절하지 않을 수 있다. 이 때 역시 전단계로(‘일부상황CU모형’) 돌아가서 다른 상황들에 대한 문항들에서 고유요인간 상관 가능성을 검토해야 한다. 전체상황CU모형, CTOM모형, 그리고 일부상황CU모형에서 이론적/경험적으로 고유요인간 상관가능성을 검토할 경우 의미있는 상관이 가능한 고유요인끼리만 선택적으로 자유모수화하므로써 간명한 CU모형이 될 수도 있고 상황간의 상관도 표현할 수 있음을 Marsh(1989)가 제시한 바 있다.

UM모형, CU모형, CU'모형

UM과 CU는 모수에 대한 표현을 다른 방식으로 한 대안적 표현(reparameterization)으로서 특질요인이 있을 때 방법요인이 도입되면 추정상의 문제가 자주 생기는 것에 대한 대안으로서 UM(상관이 안된 방법요인)모형 대신에 CU(방법내 문항들에서 고유요인간 상관)모형이 추천되고 있

다(예: 표1에서 1TUM과 1TCU, CTUM과 CTCU, CT1M과 CTCU). 실제로 UM모형과 CU모형은 “특질이 3개월 경우 서로가 수학적 동치가 되며 3개월을 초과하면 UM은 CU의 특수한 경우이다”(Marsh, 1989, p.341). 모형찾기 과정에서 CU모형을 추천하는데는 상당한 이유가 있다. 다특질다방법 자료에 대한 다년간 연구에서, 내용요인과 방법요인을 한꺼번에 구조모형에 설정할 경우 수렴이 안되거나 부적절 해를 가져오는 것이 밝혀졌다. Marsh와 Bailey(1991)는 255개의 실제 MTMM자료와 180개의 모의 MTMM자료를 분석한 결과 CU모형은 96% 이상이 수렴하였고, 방법요인을 설정하는 모형은 24% 이하의 수렴사례를 보였다.

상황의 고정효과모형이면 상황요인 중심의 모형찾기를 하는데 중간에 추정의 문제를 우회하고자 고유요인상관모형(예: 1TUM 대신에 1TCU, CTUM 대신에 CTCU)을 택할 수가 있다. 이 때는 방법요인을 명시적인 요인으로 설정하지 않고, 고유요인간 상관에 의해 표시한다. 따라서 분석결과, 고유요인간 상관의 정도에 따라 방법요인(상황요인)을 추론할 수 있게 된다.) 그러나 상황이 무선효과변수이고 내용요인 중심의 모형찾기에서는 상황내 고유요인상관을 설정하는CU모형을 사용하게 된다. 따라서 CTOM모형 다음에 CT1M의 대안으로서의 CTCU모형으로 간다. 다음에는 CTUM에 대한 대안으로서의 CTCU모형으로 간다.

1) 이 때 고유요인간 상관이 방법효과 때문일 수도 있지만 모형에서 누락된 내용변수일 수도 있다. 그러나 어떤 고유요인의 상관이 방법효과인지 누락된 내용요인인지는 상당정도 연구문제에 대한 지식에 근거한 질적 판단에 의존한다. 즉 충분히 이론적인 특질변수가 요인으로 설정된 모형의 분석에서 고유요인간 상관이 유의하면, 상황으로 인한 방법효과로 간주된다.

상황의 고정효과모형에서 상황요인 중심의 모형찾기를 할 때 하나의 상황으로 인한 방법요인의 형태도 단일하거나 복수일 수 있고 그 관계도 다양할 수 있으나 이 모든 것은 고유요인간 상관을 보고 해석가능하다(Kenny & Kashy, 1992). 고유요인간 상관계수들의 양상을 검토하면, 어떤 것은 유의하고 어떤 것은 유의하지 않은 것을 알 수가 있다. 하나의 상황내에서도 고유요인간 상관들이 두 개 또는 더 많은 수의 소집단으로 나뉘어 보고 그 상황내 방법요인이 복수임을 알 수가 있다(Marsh & Baily, 1991; Marsh, Byrne, & Craven, 1992).

그러나 모형찾기 과정에서 추정상 편리함을 가지고 있는 CU모형은 원래 상황간 상관을 표시하지 못하는 한계가 있다. 즉 상황간 상관이 없다는 비현실적 가정을 한다. 이 가정은 CU모형을 분석한 결과를 보고 검토가능하다. 또한 최종모형에서 고유요인간 상관관계의 유형을 보고 상이한 상황에 대한 문항들간에 상관이 있을 수 있는지 이론적, 경험적(예: 수정지수 참조)으로 검토한다. 예로서 상이한 상황에 속한 문항간에 상관이 가능한 곳을 고정모수에서 자유모수로 바꾸어서 추정해 본다. 이것은 상황의 무선효과모형일 경우 모형찾기의 자연스런 모습이지만, 고정효과모형일 경우 추정의 어려움을(예: ITCM, CTCM모형) 회피하면서 상황간 상관의(CM모형을 의미) 가능성을 추론해 주는(예: ITCU', CTCU'모형) 중요한 기능이 된다. 이렇게 CU모형의 제약이 해제된 모형을 통해서 CM모형을 대안적으로 추정하게 되며 이것은 표1에서 CU'라고 하였고 이것은 이미 다른 연구에서 제시된 바 있다(예: Marsh, 1989, 모형 4E'). 즉, CU'는 CU의 일반화된 형태이다. 이러한 실용성으로 인하여 고유요인상관 모형이 지금은 응용연구에서 정착이 되었다(Becker & Cote, 1994; Conway, 1996; Conway,

1998).

검사제작시 상황의 고정효과모형을 사용한 경우, 최종모형에서(CU모형이나 CU'모형) 고유요인간 상관들은 다시 상황요인으로 설정해서 요인을 해석하고 각 측정치의 신뢰도를 산출하게 된다. 추정상의 어려움을 우회하는 가운데 상황을 방법요인으로 표시하지 않고 고유요인상관모형을 사용했으나, 최종모형에서는 모형이 단순해졌으므로 상황의 고정효과를 명시적으로 나타내기 위하여 CU'를 방법요인모형으로 바꾸는 것이다. 이 절차를 부록에서 “CU' - UM 절차”라고 소개하고 있다. 결과로 얻는 방법요인모형(CTUM의 모양을 가짐)에서는, 상황간 상관까지도 CM모형으로 표시하는 대신에 독립적인 M요인으로 표시할 수 있다(부록1 참조).

끝으로 모형찾기의 과정에서 우선순위는 최적의 합치도보다는 내용요인구조 나아가서 상황효과 구조를 이해하는데 있음을 잊지 말아야 한다. 이제는 두 개의 실제자료에서 모형찾기를 하는 것을 보기로 한다.

소검사기반검사에서 모형찾기 사례

검사 자료

여기서 모형찾기를 위한 두 가지 접근의 응용을 보이기 위해서 두가지 자료를 분석하기로 한다. 하나는 검사제작의 설계상 상황의 고정효과모형이 기대되는 경우이고, 하나는 무선효과모형이 기대되는 경우이다. 고정효과모형이 기대되는 자료는 Schneider와 Schmitt(1992, p.37)의 상관행렬로서 세가지 특질(문제해결, 대인관계스킬, 주도성)을 네가지 상황(경쟁적 집단토론, 협동적 집단토론, 경쟁적 역할연기, 협동적 역할연기)에 의해

서 측정한(N=89) 경우이다. 이 때의 상황은 지필 형이 아닌, 평가센터에서의 충실도 높은 실습과제상황으로서 그 과제는 일정한 형태로 모든 지원자에게 반복되므로 상황효과는 고정효과로 볼 수 있다.

무선효과모형이 기대되는 자료는 M기업에서 문제해결력에 대한 다섯 가지 특질을(문제정의, 문제의 구조화, 정보수집, 해결안 제시) 측정하기 위해서 세 개의 지필형 상황을(상황1, 상황2, 상황3) 사용한 경우이다(N=163). 이 때의 상황은 조직 구성원들로부터 업무상 중요상황을 제시받을 것이다. 연구자들은 상황별로 다섯 가지 특질의 일부 또는 전체를 측정키 위한 이분채점(1, 0) 문항들을 개발하였다. 따라서 전집(universe)에서 랜덤하게 뽑혀진 상황들로 구성되므로 상황효과는 무선효과모형으로 볼 수 있다. 이 경우는 상황별로 적절한 수효의 문항들을 개발하며, 지필형 상황판단검사들이 대체로 이런 형태이다.

Schneider와 Schmitt(1992)의 상관자료에서 얻은 실습과제형 검사에서의 상황내/상황간 상관을 표 2에, M기업에서의 자료에 기초한 지필형 상황판단검사에서의 상황내/상황간 상관을 표 3에 제시한다.

이 자료의 분석을 위하여 LISREL 8.52(du Toit, du Toit, & Hawkins, 2001)를 사용하였고 Schneider와 Schmitt의 자료에는 추정법으로서 최대우도법(ML: Maximum Likelihood)을, M기업의 자료에는 사분상관자료에 적절한 대각선가중최소제곱법(DWLS: diagonal weighted least squares)을 사용하였다.

Schneider와 Schmitt의 자료분석

검사제작시 상황의 고정효과모형이 사용되었으므로 상황요인 중심으로 모형찾기를 할 것인

표 2. Schneider와 Schmitt(1992) 자료에서의 상황내/상황간 상관평균

	1.경쟁적 집단토론	2.협동적 집단토론	3.경쟁적 9
상황1	.69		
상황2	.22	.68	
상황3	.17	.21	.71
상황4	.22	.21	.35
상황내 평균=.72		상황간 평균=.23	차이=.4

데 표2에서 볼 때 상황요인이 충분히 존재할 가능성이 보인다. 일단은 OTUM모형으로 출발하였다. 전반적 합치도는 충분히 좋다고 할 수 없으므로 합치도 향상을 위한 모형수정의 여지가 있다. 수정지수(MI: modification index)를 보니 상황요인간 상관을 고정모수에서 자유모수로 할 것을 시사하는 값들이었다. 원래가 측정시 복수의 방법을 사용할 때 상관이 안되는 방법을 쓸 것이 가정되지만(예: Campbell & Fiske, 1959), 현실에서는 상관이 있을 수 있다. Schneider와 Schmitt의 자료에서 상황1과 2는 둘다 집단토론이라는 데서 상관이 가능하고(표2에서 평균 $r=.22$) 상황3과 4도 둘다 역할연기라는 데서 상관이 가능하다(표 2에서 평균 $r=.35$). 마찬가지로 상황1과 3은 경쟁적 상황이라는 데서 상관이 가능하고(평균 $r=.17$), 상황2와 4는 협동적 상황이라는 데서 상관이 가능하다(평균상관 $r=.21$). 따라서 이들간 상관을 고정한테 대하여 MI값이 큰 값으로 나오는 것은 기대할 수 있고 그로 인하여 OTCM모형을 검증해 봄직하다. 그러나 모형찾기 방식에서 제시한 바와 같이 상황간 상관은 나중에 CU' 모형으로 검증해 볼 수가 있고, 상황간 상관은 특질 분산을 설명할 수 있으므로 OTCM모형 대신에 ITCU모형을 먼저 합치시켜 보았고 그 결과가 표 4에 제시되어 있다.

OTUM과 ITCU간의 비교는 후자에서 합치도의

표 3. M기업 자료에서의 상관자료 및 상황내/상황간 상관평균

1. 문항간 상관				문항 ^a														
특질	문항	상황 번호	정답 률	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. 문제 정의	1	1	.40	1.00														
	2	1	.84	.23	1.00													
	3	2	.85	.33	.38	1.00												
2. 문제의 구조화	4	2	.88	.00	.30	.24	1.00											
	5	3	.79	.20	-.13	.17	.08	1.00										
	6	3	.58	.25	.26	.20	.27	.36	1.00									
3. 정보 수집	7	2	.31	-.04	.50	-.12	.10	-.03	.04	1.00								
	8	2	.71	.08	.24	.42	.34	.07	.15	.07	1.00							
	9	2	.31	.00	.23	.11	.41	.43	.43	.27	.34	1.00						
4. 분석	10	3	.55	-.02	-.12	.21	.29	.24	.28	-.06	.03	.16	1.00					
	11	3	.72	.11	.32	.09	.42	-.10	.17	.11	-.07	.44	.32	1.00				
	12	3	.65	.38	.25	-.18	.16	.17	.34	.07	-.02	.07	.45	.57	1.00			
5. 해결안 제시	13	1	.33	.21	.10	.31	.04	.32	.05	-.16	.06	.31	.17	.26	.21	1.00		
	14	2	.77	.20	-.08	.21	.22	-.13	.03	-.04	.08	.26	.06	.26	.00	.30	1.00	
	15	2	.41	-.11	-.08	.13	.02	-.11	-.04	-.07	.15	-.02	.14	.00	-.11	.05	.22	1.00

2. 상황내/상황간 상관평균

- 1) 상황내: 상황1($r=.18$), 상황2($r=.15$), 상황3($r=.28$), 평균 $r=.20$
- 2) 상황간: $r(1, 2)=.14$, $r(1, 3)=.17$, $r(2, 3)=.28$, 평균 $r=.14$
- 3) 상황내/상황간 평균 r 간 차이=.06

a 문항간 상관은 문항에 대한 채점이 응답자 능력을 인위적으로 이분하였으며 실제 지원자들의 능력분포 (총점의 분포는 편포도와 첨도가 0에 가까웠음)는 정규분포를 이룬다고 볼 수 있으므로 사분상관을 구하였다.

유의한 증가를 보여주고 있고, 여러 가지 상황에 걸쳐서 안정되게 도출되는 내용요인의 수렴타당도를 지지하게 된다. 이것은 OTUM에 비해서도 내용적으로 의미있는 모형이다. 다음은 과연 세 개의 특질요인에 변별타당도가 있는지 3개의 특

질이 있는 CTCU를 1TCU와 비교해 보기로 한다. 표 4에서 볼 때 CTCU는 1TCU에 비해 합치도가 유의하게 높지만 요인2의 분산이 유의하지 않고, 요인1과 3간의 상관이 0.92로 대단히 높다. 그렇다면 2요인(대인관계스킬)은 수렴되지 않는 요인

표 4, Schneider와 Schmitt의 자료분석

모형	모형표시	합치도	중요정보
0모형 (OTOM)	1) 특질요인 없음 2) 상황요인 없음 3) 추정법: ML	DF=66, $\chi^2=851.14(p=.00)$	
모형1 (OTUM)	1) 특질요인은 없음 2) 4개의 상황요인 3) 요인간 상관은 0으로 고정 4) 추정법: ML	DF=54, $\chi^2=104.97(p=.00)$ RMSEA=.11, ECVI=1.78 포화모형의 ECVI=1.77 NFI=.88, NNFI=.92 CFI=.94 RMR=.21 GFI=.83, AGFI=.75 전반적으로 개선의 여지 있음	<ul style="list-style-type: none"> • 요인간 상관에 대한 수정지수 (MI)가 상황요인간 상관을 시사하고 있음 • ITCU가 권고됨
모형2 (ITCU)	1) 1개의 특질요인 2) 상황요인은 없음 3) 상황요인 대신에 각 상황을 재는 세 개 측정변수의 고유요인간 상관을 자유모수로 함 4) 모형1과 같음	DF=42, $\chi^2=57.02(p=.06)$ RMSEA=.06, ECVI=1.46 포화모형의 ECVI=1.77 NFI=.93, NNFI=.97 CFI=.98, RMR=.065 GFI=.90, AGFI=.92 전반적 합치도 좋음 PNFI=.59, PGFI=.49	<ul style="list-style-type: none"> • 상황1과 상황4의 고유요인간 상관에 대한 MI값중 2개가 5.0을 넘는 값이 있음 • 상황1은 경쟁적 집단토론이고 상황4는 역할연기이므로 내용상 상관이 크게 기대되지 않음 • 현재 모형으로서도 합치도는 충분히 좋음
모형3 (CTCU)	1) 3개의 상관된 특질요인 2), 3), 4) 모형2와 같음	DF=39, $\chi^2=40.85(p=.39)$ PNFI=.56, PGFI=.46 모형확대로 인한 PNFI, PGFI의 하락이 심각하지 않음	<ul style="list-style-type: none"> • 모형2 : 모형3 $\Delta df=3, \Delta \chi^2=16.17 (p<.05)$ 합치도의 유의한 증가 • VAR(F2)=.02(t=.57) • COR(F1, F2)=.77 COR(F1, F3)=.92 COR(F2, F3)=.86 • 1개의 특질요인이 적절할 것임 • 모형2로 돌아감
모형4 (ITCU')	• 두 고유요인간 상관을 추가한 것 이외는 모형2와 동일	DF=41, $\chi^2=50.46(p=.15)$	<ul style="list-style-type: none"> • 모형2 : 모형4 $\Delta df=1, \Delta \chi^2=6.56 (p<.05)$ 합치도의 유의한 증가 • 자유모수가 된 추정치의 t값=2.51 • MI값에 더 이상 자유모수화를 검토할 대상이 없음
모형5 (CU' -UM)	• ITCU' 모형에 대한 전환 • 고유요인간 상관에 대한 CU' 를 모두 UM으로 바꾸어 대안적 모형으로 표시	DF=41, $\chi^2=50.46(p=.15)$ 모형4와 모든 합치도 동일	<ul style="list-style-type: none"> • 종결 • 이 모형에서의 모수추정치 보고

주. RMSEA: 개략화 오차평균, ECVI: 교차타당화 지수, NFI: 표준합치도, NNFI: 비표준합치도
CFI: 비교합치도, RMR: 원소간 평균잔차, GFI: 기준합치도, AGFI: 조정합치도,
PNFI: 간명도 반영한 NFI, PGFI: 간명도 반영한 GFI

이고, 나머지 요인1(문제해결)과 요인3(주도성) 각각은 여러 가지 측정방법에 걸쳐서 도출되는 수렴타당도가 있는 것으로 보이지만 둘간에 변별은 되지 않는다. 따라서 해석을 위한 모형은 보다 간명한 1TCU가 된다. Schneider와 Schmitt의 자료를 재분석한 Lievens와 Conway(2001)는 이 연구

항이란 점에서 상관이 있을 것으로 기대되므로 추가로 모형에 표시하고 추정기로 하였다. 이른바 ‘상황간 상관’이 설정된 것이고 표 1의 CU’ 모형에 해당한다. 이 새로운 모형은 모형4의 1TCU’ 모형으로서 합치도가 좋고 더 이상 자유모수로 할 곳이 없으므로 이것으로 모형찾기를

표 5. Schneider와 Schmitt자료의 표준화 요인계수(t값)

측정변수 ^a	요 인					특질요인	신뢰도	고유분산
	집단토론1	집단토론2	역할연기1	역할연기2	역할연기 ^b			
T1M1	.84(9.30)					.42(3.41)	.88	.12
T2M1	.66(6.51)					.25(1.95)	.50	.50
T3M1	.73(7.88)					.47(3.85)	.75	.25
T1M2		.80(8.25)				.21(1.67)	.68	.32
T2M2		.65(6.73)				.41(3.32)	.58	.42
T3M2		.87(9.52)				.32(2.54)	.86	.14
T1M3			.69(5.82)		.29(5.01)	.56(4.28)	.88	.12
T2M3			.46(3.84)			.68(5.40)	.68	.32
T3M3			.76(6.66)			.44(3.30)	.76	.24
T1M4				.83(8.87)	.29(5.01)	.48(3.56)	.97	.03
T2M4				.58(5.22)		.67(5.27)	.78	.22
T3M4				.69(6.47)		.59(4.55)	.82	.18

^a T1=문제해결, T2=대인관계스킬, T3=주도성,

M1=집단토론1, M2=집단토론2, M3=역할연기1, M4=역할연기2

^b 측정변수가 2개밖에 안되므로 식별(identification)을 위해 요인계수에 동일화 제약을 가하였다.

에서와 같은 위계적 분석을 하지 않았기에 매우 상관이 높은 세 개의 특질요인을 보고하고 있다.

모형2(1TCU모형)로 돌아가서 이론적/경험적으로 볼 때 자유모수로 해야할 곳이 있는지 검토한 결과 7번 문항(“문제해결”)을 측정, 경쟁적 역할연기 상황과 10번 문항(“대인관계스킬”)을 측정, 협동적 역할연기 상황)사이에서 그런 필요가 판단되었다(MI=5.82). 두 문항 모두 역할연기 상

관에 상황효과가 무선평과모형이면, 식(3)이 성립하며 모든 상황효과가 측정오차 분산에 들어 있는 모형4로서 종결된다. 여기서 각 측정변수를 내용요인들이 설명하는 부분이 각 변수의 신뢰도가 된다.

그러나 Schneider와 Schmitt연구에서 사용된 각 실습과제 상황은 모든 지원자에게 반복적으로

사용되어 고정효과모형이다. 따라서 모든 고유요인간 상관에 대하여 CU' - UM 절차에 따라 UM 모형을 도입하고 다시 계산하여 결과를 표 5에 제시한다. 표5를 보면 CU' - UM 절차의 결과로 4개 상황이 UM으로 표시되었다. 또한 상황3(역할연기1)과 상황4(역할연기2)간 상관이 요인으로 표시되었고 "역할연기"라는 요인이름이 주어졌다.

이렇게 설정된 모형5는 CU' - UM 모형으로서 합치도는 모형4와 동일하면서 상황효과에 대하여 그 안에서 특질요인이 작용하는 환경요인으로서의 해석은 물론, 각 측정변수의 분산을 설명하는데 특질요인과 함께 모형에(식4 참조) 표시가 되고 그 결과로 얻는 다중상관계수(R^2)가 그 측정치의 신뢰도가 된다. Schneider와 Schmitt의 자료를 재분석한 Lievens와 Conway(2001)는 4개의 상황간에 상관이 있다는 가정하에 처음부터 CTCM모형을 적용한 결과로 부적절한 해를 보여주고 있다.

표 5에서의 신뢰도는 각 측정치를 요인들에 회귀시켰을 때의 R^2 즉 다중상관계수(LISREL 출력에서 SMC로 제공)이다. 신뢰도에 대해서는 검사의 신뢰도 부분에서 다시 언급하기로 한다.

상황이 무선효과모형인 자료의 분석

M기업 자료에서는 검사제작시 상황의 무선효과모형이 사용되었고 표 3에서와 같이 상황효과가 기대되지 않으므로 특질중심의 모형찾기에서 고유요인간에 유의한 상관이 많지 않을 것으로 예상된다. 표 6에 제시된 바와 같은 모형찾기의 과정을 거쳤다.

우선 영모형은 기각되었고 모형1은 5개의 특질요인만을 설정한 CTOM모형으로서 F3(정보수집)이 주어진 측정치들에 의해서는 수렴되지 않

았다. 또한 이 요인과 다른 요인들과의 상관이 대체로 유의하지 않았으므로 모형에서 측정변수들과 함께 F3을 제거하기로 하였다. 모형2는 4개의 특질요인으로 된 CTOM모형으로서 좋은 합치도를 보였고 각 요인은 여러 가지 측정치에 의해 수렴함을 보였다. 요인간 상관은 대체로 유의했으나 변별이 안될 정도로 크지는 않았다.

이 상태에서 각 상황별로 상황효과가 수렴하는지 보기 위하여 1개 상황씩을 CU모형으로 표시하여 세 번의 CTCU모형을 시도하였으나 부적절한 추정치이거나 수렴하지 않음을 보여주었다. 따라서 상황간 관계가 있는지를 보기 위하여 모형2의 결과로 돌아가 검토하였다. 검토결과 고유요인간의 고정된 상관에 대한 MI값($=73.15$) 큰 것이 있었고 그것은 상황1과 상황3간의 상관을 시사하는 것이었다. 따라서 모형3은 그 고유요인간 상관을 자유모수로 하여 두 상황간 관계를 검증하는 CTCU'가 되었다. 이 모형은 모형2를 품는 모형이므로 합치도는 증가하였고 특히 χ^2 는 유의한 정도로 감소하였다. 이 모형의 분석결과에서 상황2와 상황3간의 상관을 시사하는 고유요인간 상관을 모형4에서 자유모수로 하였다. 그 결과는 유의한 정도로 χ^2 의 감소를 가져왔다. 이 상태에서 더 이상 모형의 수정을 할 것이 없으므로 모형찾기를 종결한다. 또한 이 자료에서 상황효과는 무선효과모형이므로 상황효과를 요인으로 표시할 필요가 없이 모두가 모형의 오차 부분에 반영된 상태로 놔두는 것이 적절하다. 따라서 CU' - UM 절차를 적용하여 각 측정치의 신뢰도를 다시 계산할 필요가 없다. 그러면 모형4에 대한 결과를 표 7에 제시한다.

표 7을 보면 각 특질요인에 대한 측정치의 신뢰도 및 고유요인 분산을 알 수가 있다. 12번 문항의 신뢰도는 매우 낮으므로 앞으로 문항 개선시에 제거하거나 상당한 수정을 해야할 것이다.

표 6. 상황효과가 기대되지 않는 자료의 분석

모형	모형표시	합치도	중요정보
0모형	1) 특질요인 없음 2) 상황요인 없음	DF=105, $\chi^2=844.51(p=.00)$	
모형1 (CT0M)	1) 5개의 특질요인 2) 상황요인은 없음 3) 특질요인간 상관 추정 4) 추정법: DWLS	DF=80, $\chi^2=526.92(p=.00)$ RMSEA=.06, ECVI=1.24 포화모형의 ECVI=1.48 NFI=.90, NNFI=.99 CFI=.99 RMR=.12 GFI=.94, AGFI=.91 전반적으로 좋은 합치도	<ul style="list-style-type: none"> • VAR(F3)=.04($t=.68$) • F3값은 다른 요인들과 대체로 유의하지 않은 상관 • 나머지 요인들간에는 수렴과 변별이 잘 됨 • F3(정보수집)를 제거하기로 함
모형2 (CT0M)	1) 4개의 특질요인 2), 3), 4) 모형1과 같음	DF=48, $\chi^2=313.21(p=.00)$ RMSEA=.05, ECVI=.81 포화모형의 ECVI=.96 NFI=.92, NNFI=1.00 CFI=1.00, RMR=.11 GFI=.96, AGFI=.94 전반적으로 좋은 합치도	<ul style="list-style-type: none"> • 각 요인의 분산이 유의(수렴) • 각 요인간 상관도 대체로 유의 • TE(9, 1)에 대한 MI=73.15 • 상황1과 상황3간 상관을 시사
모형3 (CTCU')	1) 4개 특질요인 2) 고유요인간 상관 1개를 자유모수화	DF=47, $\chi^2=269.95(p=.00)$	<ul style="list-style-type: none"> • 모형2 : 모형3 • $\Delta df=1, \Delta \chi^2=43.26 (p<.05)$ • 합치도의 유의한 증가 • TE(8, 4)의 MI값=7.66 • 상황2와 상황3간 상관을 시사
모형4 (CTCU')	1) 4개 특질요인 2) 고유요인간 상관을 추가로 1개 더 자유모수화	DF=46, $\chi^2=241.46(p=.00)$	<ul style="list-style-type: none"> • 모형3 : 모형4 • $\Delta df=1, \Delta \chi^2=28.49 (p<.05)$ • 합치도의 유의한 증가 • 더 이상 자유모수화 필요없음

상황1과 3간의 두 측정치간 상관($r=.56$)과 상황2와 3간의 두 측정치간 상관($r=0.39$)은 상관의 원인이 특별히 어디에 있는지를 알려주고 있다. 끝으로 특질요인간에는 기대되는 바와 같이 어느 정도의 상관이 있으나 충분히 변별되는 것으로 보인다. 그러나 각 요인의 분산을 보면 t 값이 비교적 충분히 크지 않다. 일반적으로는 $|t|>2$ 인 경우를 유의한 경우로 간주하고 있으나 통계적/실용적 이유²⁾로 인하여 이 연구에서는 해석의

대상으로 하였다.

2) t 값은 모수추정치에 표준오차로 나눈 값이며 대략적으로 표준정규분포를 따르기 때문에 $|t| \geq 2$ 이면 유의한 크기로 본다(Joreskog & Sorbom, 1989). t 값의 정확도는 표준오차의 정확도에 의존하는데, 최대우도법으로 추정할 때는 자료에서 다변량정규분포가 성립해야 표준오차가 정확한 값으로 나온다. 한편 현재 이 연구에서 사용되고 있는 추정법인 “DWLS를 사용할 경우의 표준오차는 대략적인 수치이다”(Joreskog & Sorbom, 1989, p.41). 뿐만 아니라

표 7. 상황효과가 기대되지 않는 자료의 모형추정치

순서	측정치		표준화 요인계수(t 값)					
	특질	상황	문제정의	문제구조화	분석	해결안 제시	신뢰도	고유요인분산
1.		1	.52				.27	.73
2.	문제정의	1	.51				.26	.74
3.		2	.67				.44	.56
4.		2		.45			.20	.80
5.	문제구조화	3		.40			.16	.84
6.		3		.72			.51	.49
7.		3			.58		.34	.66
8.	분석	3			.67		.44	.56
9.		3			.79		.63	.37
10.		1				.77	.59	.41
11.	해결안 제시	2				.42	.17	.83
12.		2				.08	.01	.99

고유요인간 상관: (1, 9)=.56, (4, 8)=.39, 괄호안은 측정치 순서

특질요인간 분산/공분산 및 상관(t 값)				
	1. 문제정의	2. 문제구조화	3. 분석	4. 해결안 제시
1. 문제정의	0.27(1.58)	.14(1.67)	.06(1.02)	.20(1.81)
2. 문제구조화	.58	0.20(1.29)	.14(1.73)	.08(.88)
3. 분석	.21	.55	0.34(2.18)	.18(2.04)
4. 해결안 제시	.50	.23	.40	0.59(1.40)
	1	2	3	4

(대각선 및 위: 분산과 공분산, 대각선 아래: 상관계수)

소검사로 구성된 검사의 신뢰도

LISREL 저자들은, 표준오차 산출시 정규분포의 가정 및 대표본의 가정을 가지고 있으므로 이러한 가정에 맞지 않는 자료를 다룰 때 “표준오차에 너무 집착하지 않을 것을” 사용자들에게 경고하고 있다(Joreskog & Sorbom, 1989, p.129). 표본이 대표본이 아니고(t 값이 작게 나옴), 추정법도 DWLS를 사용하고 있는 이 연구에서는 이상의 이유에 근거해서, 자유모수를 고정하는 t 값의 기준치로서 2.0이 아닌 낮은 값을 설정하였고, 중요한 내용요인의

이제껏은 소검사로 구성된 검사의 요인추출 및 각 측정변수의 신뢰도를 제시하였다. 무선희과모형의 경우 측정치의 분산가운데 각 특질이

구조이해가 중심이 되는 연구의 특성상 t 값이 좀 작아도 그냥 자유모수를 유지하고 해석에 참고하기로 하였다.

기여하는 부분이 신뢰도가 되었고, 고정효과모형의 경우 특질에 의한 분산과 상황적 특수성에 의한 분산을 합한 것이 그 문항의 신뢰도가 된다. 이제, 검사의 신뢰도를 추정하기로 한다.

검사의 신뢰도

검사의 신뢰도를 구하는 데는 전형적으로 고전 검사이론에 의한 방법과 구조방정식모형에 의한 방법이 있다. 고전검사이론에서는 하나의 검사가 있을 때 상황에 의한 국지적 의존성이나 오차상관이 없는 것으로 보고 각 문항을 독립적으로 취급하여 α 계수를 구하거나 각 상황의 점수를 구해서 상황점수를 가지고 α 계수를 구하는 것이 논의되었다. 그러나 전자는 상황자료에서 중복되는 정보를 모두 독립적으로 간주하여 신뢰도를 과대추정하게 되고, 후자는 상황효과가 적어 문항간에 국지적 독립성이 있을 수 있는 경우의 개별 문항이 제공하는 독립적인 정보들을 누락시켜서 신뢰도 과소추정의 위험이 있다. 구조방정식모형에서는 진점수분산을 직접 추정하는 방식을 사용하므로 고전검사이론에서의 관찰점수 중심의 방식과 다르다. Joreskog와 Sorbom(1989, p.79)은 문항에 대한 신뢰도로서 아래와 같은 공식을 제시하고 있다.

$$1 - \frac{\theta_{ii}}{\sigma_{ii}}$$

여기서 σ_{ii} 는 문항 i 의 분산이고, θ_{ii} 는 고유분산이다. 이것을 여러개 측정치에 대해서 즉, 검사에 대해 사용하기 위해 확대하여 행렬의 형태로 표시하면 아래와 같다(Lee, Dunbar, & Frisbie, 2001).

$$1 - \frac{\Sigma(\theta)}{\Sigma(X)} \dots\dots\dots (5)$$

여기서 $\Sigma(x)$ 는 각 문항의 분산/공분산행렬의 모든 원소를 합한 값을 나타내며, $\Sigma(\theta)$ 는 각 문항에 대한 고유요인의 분산/공분산행렬의 원소의 합을 나타낸다. 이 방법에서 오차간 독립성 위반을 수용하는 장점이 있고, 상황에 의한 문항간 국지적 의존성 역시 $\Sigma(x)$ 의 산출에 모두 감안되고 있다. 그러면 Schneider와 Schmitt 자료(상황효과 고정모형)와 M기업자료(상황효과 무선모형)에 대해 이상의 세가지 방식--문항점수기반, 상황점수기반, 그리고 구조방정식모형 방식--으로 검사신뢰도를 구하기로 한다.

문항점수기반의 α 계수는 아래의 공식으로 산출한다.

$$\frac{n}{n-1} \cdot \left(1 - \frac{\sum \text{VAR}(X_i)}{\text{VAR}(X)}\right) \dots\dots\dots (6)$$

n 은 문항수효, $\text{VAR}(X_i)$ 는 각 문항의 분산, $\text{VAR}(X)$ 는 총점의 분산인데 식(5)의 $\Sigma(X)$ 와 같은 값이다. 상황점수기반의 α 계수 역시 k 개의 상황이 있을 때 상황점수 Y_1 를 단위로 하여 위의 α 계수공식을 적용하며 공식은 아래와 같다.

$$\frac{k}{k-1} \cdot \left(1 - \frac{\sum \text{VAR}(Y_i)}{\text{VAR}(Y_1 + \dots + Y_k)}\right) \dots (7)$$

여기서 $\text{VAR}(Y_1 + \dots + Y_k)$ 는

표 8. 신뢰도 추정치

	구조방정식모형	문항점수기반
Schneider와 Schmitt 자료	.95	.85
M기업 자료	.72	.69

$\text{VAR}(X_1 + \dots + X_n)$ 즉 $\text{VAR}(X)$ 와 같고, 분자의 $\text{VAR}(Y_i)$ 는 i 번째 상황내에 있는 문항간 분산/공분산 행렬내 원소의 합이다. 이렇게 해서 구한 신뢰도의 추정치는 표 8과 같다.

표 8에서 보는 바와 같이 신뢰도는 상황점수를 기초로 구했을 때 가장 작고 문항점수를 기초로 할 때 좀더 크며, 잠재변수를 사용하는 구조방정식모형을 사용했을 때 가장 크다. 교육측정 장면에서는 문항점수기반의 신뢰도를 사용해 왔고, 한편 산업심리학에서는 지필형 상황판단검사에서 상황점수기반의 신뢰도를 사용해 왔다.

교육장면에서는 워낙 문항의 수효가 크고 검사를 대량 실시하는 가운데 능률을 위해서 문항점수기반의 신뢰도를 사용해 왔을 수가 있고, 산업심리학에서는 많은 경우 상황을 채점단위로 해왔으므로 상황점수기반의 신뢰도를 보고할 수밖에 없었다. 그러나 앞으로 “상황에 내재된(nested) 문항”이란 개념에서 볼 때 오차의 상관 및 국지적 의존성을 감안하는 구조방정식모형의 방식이 더 바람직하다.

요 약

이 연구에서는 소검사로 이루어진 검사에서 요인구조를 추출하고 신뢰도를 구하는 방향을 제시하였다. 요인구조를 추출하는데 사용되는 접근은, MTMM 자료의 분석방식에 준해서 하위 검사제작시 상황의 고정효과를 가정하는 경우와 그렇지 않은 경우(무선효과 가정)에 모형찾기의 경로를 달리하였다. 또한 모형의 가정에 따라 최종적으로 도달한 모형에서 CU' 부분을 UM으로 바꾸어 줄 것인지 여부를 결정하고 그에 따라 각 측정치의 신뢰도가 달라진다. 이 연구는 현재 국내외에서 널리 사용되어온 소검사기반검사의 구성개념타당화에 큰 출발점이 될 수가 있다. 예로서, 현재까지는 상황판단검사의 총점을 다른 변수들에 관계시키므로서 간접적으로 구성개념을 탐색해 왔으나(예: McDaniel 등, 2001; McDaniel,

Hartman, & Grubb, 2003), 이 연구에서의 접근은 직접적으로 구성개념을 검토하게 해준다.

요인구조 도출부분에서의 기여를 보면, 소검사기반의 검사자료에서 상황효과의 가정에 따른 분석전략의 구체화를 하였고 상황효과가 고정효과모형일 경우 상황의 특수성이 구성개념의 작용을 위한 환경변수로 해석되고 응답자에 대한 피드백에 포함되어야 한다는 것과 무선효과일 경우 전통적인 요인분석에서와 같이 상황효과는 고유요인에 포함됨을 제시하였다. 특히 CU'-UM 절차는, 검사제작시에 상황의 고정효과모형이지만 추정의 편리상 최종모형이 CU' 모형이 되었어도, 다시UM모형으로 바꾸어 주는 논리를 제시하였다. 실제로 이 논리는 개별 측정치의 신뢰도를 구하는데 있어서 Conway(1998) 및 그의 개선안인 Scullen(1999)에서 해결하지 못한 '전체모형속에서의 개별측정치 신뢰도' 추정을 제시하여 이전 연구에서 확립하지 못한 통계적 측면을 보완하고 있다.

소검사기반검사의 신뢰도 부분에서의 기여를 보면 상황점수기반의 신뢰도나 문항점수기반의 신뢰도가 가지는 문제점을 극복하기 위한 구조방정식모형 방식의 신뢰도 제시를 보여주고 있다. 그러나 소검사기반의 검사는 내용요인 국면, '상황'이라는 국면, 그리고 문항국면을 명시하여 신뢰도를 산출하는 일반화가능도이론을 이용해서 신뢰도('일반화가능도')를 구해볼 수도 있다. 구조방정식모형이 잠재변수를 이용하는 방식이라면 일반화가능도이론은 관찰변수를 사용하는 고전검사이론의 확대된 형태이다. 앞으로 두 방식에 의한 신뢰도 추정치의 이론적 경험적 비교가 필요할 것이다.

이 연구의 결과로 소검사기반검사에서 무엇이 측정되는지를 분명히 할 수 있고 다른 이론적 변수들과의 관계 및 준거를 설명하는 논리적 모

형을 세워 이론을 개발할 수가 있다. 또한 응답자들에게 구성개념에 근거한 피드백을 줄 수가 있고, 정의되는 역량에 따른 적절한 측정도구의 선정이 가능할 것이다.

끝으로 이 연구에서는 소검사기반검사 제작시 측정치들이 어떠한 내용요인을 측정하고자 하는지에 대하여 어느 정도의 범주화가 있는 경우를 가정하고 있다. 그러나 그러한 최소한의 범주화도 없이 제작된 소검사기반검사들도 많이 있으며, 그런 검사들의 경우 소검사 문항들이 상당정도 독립적이라는 가정이 있을 경우 Lee와 Kim (2002)의 방식을 사용해 볼 수 있다. 즉, 상황효과가 소검사내의 요인구조에 유일한 원인으로 보고 각 문항점수에서 그 요인점수를 제거한 잔차화점수를 전체검사에서 통합하여 잠재적 내용요인들을 구한 다음, 다시 원자료로 돌아가 이 내용요인들과 검사에서의 상황을 기초로 MTMM 분석의 방식을 사용하는 것이다. 그러나 이러한 Lee와 Kim의 방식은 소검사내 문항들간에 내용적 독립성을 어느 정도까지 확신해야 소검사내에서의 요인분석에서 상황효과와 구조만이 도출된다고 할 수 있는지에 대한 경험적 연구가 필요한 시점에 있다.

앞으로 심리학 여러 분야에서 사용되고 있는 지필형 상황판단검사와 같은 소검사기반검사의 제작시 미리 계획하여, 소검사내 문항수효를 제한하고 문항들이 측정하는 개념을 비교적 분명하고 상이하게 하므로써 가급적 MTMM자료에 준한 분석을 할 수 있도록 하는 것이 요인구조의 도출에 큰 도움이 될 것이다.

참고문헌

Becker, T. E. & Cote, J. A. (1994). Additive and

multiplicative method effects in applied psychological research: An empirical assessment of three models. *Journal of Management*, 20, 625-641.

Campbell, D. T. & Fiske, D. W.(1959). Convergent and Discriminant Validation by Multitrait-Multimethod Matrix. *Psychological Bulletin*, 56, 81-105.

Chan, D., Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82, 143-159.

Conway, J. M. (1996). Analysis and design of multitrait-multirater performance appraisal studies. *Journal of Management*, 22, 139-162.

Conway, J. M. (1998). Estimation and uses of the proportion of method variance for multitrait-multimethod data. *Organizational Research Methods*, 1, 209-222.

du Toit, du Toit, & Hawkins (2001). *Interactive LISREL: User's Guide*. Lincolnwood, IL: Scientific Software.

Frisbie, D. A., & Druva, C. A. (1986). Estimating the reliability of multiple true-false tests. *Journal of Educational Measurement*, 23, 99-105.

Harman, H. H. (1976). *Modern Factor Analysis*. Chicago, IL: The University of Chicago Press.

James, L. R., Demaree, R. G., Mulaik, S. A., & Ladd, R. T. (1992). Validity generalization in the context of situational models. *Journal of Applied Psychology*, 77, 3-14.

Joreskog, K. G. & Sorbom, D. (1989). *LISREL 7: A guide to the program and applications*. Chicago, IL: SPSS Inc.

- Kenny, D. A. (1979). *Correlation and Causality*. New York: Wiley.
- Kenny, D. A. & Kashy, D. A. (1992). Analysis of the Multitrait-Multimethod Matrix by Confirmatory Factor Analysis. *Psychological Bulletin*, 112, 165-172.
- Lance, C. E., Newbolt, W. H., Gatewood, R. D., Foster, M. S., French, N. R., & Smith, D. E. (2000). Assessment center exercises represent cross-situational specificity, not method bias. *Human Performance*, 13, 323-353.
- Lance, C. E., Teachout, M. S., & Donnelly, T. M. (1992). Specification of the criterion construct space: An application of hierarchical confirmatory factor analyses. *Journal of Applied Psychology*, 77, 437-452.
- Lee, G., Brennan, R. L., & Frisbie, D. A. (2000). Incorporating the testlet concept in test score analyses. *Educational Measurement: Issues and Practice*, 19, 9-15.
- Lee, S. & Kim, A. (2002). A Three-Step Approach to Factor Analysis on Data of Multiple Testlets. In S. Nishisato, Y. Baba, H. Bozdogan, & K. Kanefuji(Eds.). *Measurement and Multivariate Analysis*. Tokyo, Japan: Springer Verlag.
- Lievens, F. & Conway, J. M. (2001). Dimension and Exercise Variance in Assessment Center Scores: A large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology*, 86, 1202-1222.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- Lowry, P. E. (1997). The assessment center process: New directions. *Journal of Social Behavior and Personality*, 12, 53-62.
- Marsh, H. W. (1989). Confirmatory Factor Analysis of Multitrait-Multitrait Data: Many Problems & A Few Solutions. *Applied Psychological Measurement*, 13, 35-361.
- Marsh, H. W. & Baily, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: comparison of the behavior of alternative models. *Applied Psychological Measurement*, 15, 47-70.
- Marsh, H. W., Byrne, B. M., & Craven, R. (1992). Overcoming Problems in Confirmatory Factor Analyses of MTMM data: The correlated uniqueness model and factorial invariance. *Multivariate Behavioral Research*, 27, 489-507.
- McDaniel, M. A., Hartman, N. S., & Grubb III, W. L. (2003). Situational Judgment Tests, Knowledge, Behavioral Tendency. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology. Orlando, April, 2003.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., & Campion, M. A. (2001). Use of Situational Judgment Tests to Predict Job Performance: A Clarification of the Literature. *Journal of Applied Psychology*, 86, 730-740.
- McDonald, R. P. (1985). *Factor Analysis*. Hillsdale, NJ: Lawrence Erlbaum.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An Alternative Selection Procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75, 640-647.
- Oswald, F. L., Schmitt, N., Ramsay, L. J., Kim, R. H., & Gillespie, M. A. (2002). *Noncognitive predictors of college student success*. Unpublished manuscript.

- Raudenbusch, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd Ed. Thousand Oaks, Cal.: Sage.
- Robertson, I., Gratton, L., & Sharpley, D. (1989). The psychometric properties of managerial assessment centres: Dimensions into exercises won't go. *Journal of Occupational Psychology*, 60, 187-195.
- Schneider, J. R., & Schmitt, N. (1992). An exercise design approach to understanding assessment center dimension and exercise constructs. *Journal of Applied Psychology*, 77, 32-41.
- Scullen, S. E. (1999). Using confirmatory factor analysis of correlated uniqueness to estimate method variance in Multitrait-Multimethod matrices. *Organizational Research Methods*, 2, 275-292.
- Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28, 237-247.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87, 245-251.
- Thisen, D., Steinberg, L., & Mooney, J. A. (1989). Trace lines for testlets: A use of multiple-categorical models. *Journal of Educational Measurement*, 26, 247-260.
- Wagner, R. K. & Sternberg, R. J. (1993). *TKIM: The common sense manager user manual*. The Psychological Corporation.
- Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testings: A case for testlets. *Journal of Educational Measurement*, 24, 185-201.
- Wainer, H., & Lewis, C. (1990). Toward a psychometrics for testlets. *Journal of Educational Measurement*, 27, 1-14.
- Wang, X., Bradlow, E. T., & Wainer, H. (2001). Users Guide for SCORIGHT(version 2.0): A Computer Program for Scoring Tests Built of Testlets. Princeton, NJ: Educational Testing Service.
- Weekley, J. A. & Jones, C. (1999). Further Studies of Situational Tests. *Personnel Psychology*, 52, 679-700.
- Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, 9, 1-26.
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30, 187-213.
- 1 차원고접수 : 2003. 9. 15.
수정원고접수 : 2004. 6. 7.
최종게재결정 : 2004. 6. 10.
- 한국심리학회지: 일반
Korean Journal Psychology
2004. Vol. 23, No. 1, 51-76

Factor Structure and Reliability of a Testlet-Based Test

Soonmook Lee

Minwoo Kang

Sungkyunkwan University

When a test comprises of multiple testlets, it is very difficult to estimate factor structure and reliability of the data collected by the test. A testlet is defined by a scenario or passage and response alternatives following it. A testlet-based test suffers difficulties of estimating factor structure because the method factors are confounded with substance factors and of estimating reliability because the measurement errors are not independent and local independence is severely challenged. In the present study, we propose an approach of applying a series of confirmatory factor analysis to analyze the data prepared in the form of multitrait-multimethod data. The approach varies depending on whether the situation effect is fixed or random. In the specification search, we adopt correlated uniqueness model to avoid estimation problems when both trait factor(s) and method factor(s) exist in a model. The final model can be transformed to a correlated uniqueness model or a method factor model depending on whether situation effect is of random effect or fixed effect. In this process, we can estimate item reliability and in turn test reliability.

Key words : testlet, factor structure, reliability, situational judgment test, method effect

부록1. CU' - UM의 절차

식(4)로 돌아가서, 고정효과의 요인모형을 보기로 한다. UM모형이나 CM모형에서는 방법요인이 명시적으로 표시되므로 방법요인모형이라 하고, CU나 그것의 확대형태인 CU'는 방법효과를 고유요인간 상관으로 표시하므로 고유요인간 상관모형이라 하자. 고유요인상관모형에서 치명적 어려움은 각 측정치의 방법효과가 직접 주어지지 않는다는 것이다. 식(4)에서 F가 기여하는 분산을 구성개념분산(construct variance)이라고 할 때 M이 기여하는 분산은 방법분산(method variance)이다. 그리고 측정치와 요인의 분산이 1.0일 때, Z와 M간의 공유분산인 방법분산의 제공근은 Z와 M간의 상관계수로서 상황이 Z에 대해서 가지는 방법효과이다. $M \rightarrow Z$ 의 요인계수가 바로 방법효과이다.

Scullen(1999)은 고유요인상관모형에서 각 측정치에 들어있는 방법효과를 추정하는 방법이 필요함을 주장하였다. 그리하여 Scullen은 고유요인상관모형에서 하나의 방법에 대한 측정치들에서의 고유요인간 상관행렬을 빼내서 별도의 구조방정식모형에 넣고 1요인모형을 추정하면서 방법효과를 추정할 수 있음을 보여주었다. 만일에 하나의 방법에 대한 고유요인간 상관행렬내에 0이 없이 모두가 온전히 큰 상관으로 연결되어 있으면 1요인모형을 설정하는 것이 바람직하다. 그러나 0이 많을수록 몇 개의 구획(block)으로 나뉘고 각 구획별로 요인을 설정하는 다요인모형을 설정한다.

이 방법은 매우 실용적으로 보이지만 원래의 방법요인모형에서 추정하는 것이 아니므로 간접방법이라는 문제가 있다. 즉, 모형전체로서 합치도의 문제, 각 추정치의 표준오차의 부정확함 등이다. 따라서 저자는 Scullen의 방식을 보다 직접방식으로 바꾸기 위해 CU' -UM 절차를 개발하게 되었다. 즉 일단 최종모형에 도달하고 나면 모형이 많이 간명해지고 정보가 충분해지므로 CU'의 부분을 UM으로 바꾸어 주어도 추정상의 문제가 대폭 감소됨을 발견하였다. 특히 요인의 척도제공 방식으로서 어느 측정변수와의 요인계수를 1.0으로 주는 것 대신에 요인분산을 1.0으로 고정하는 방식을 쓸 경우 추정상의 문제는 거의 발생하지 않는다. 고유요인상관들을 방법요인으로 다시 표현할 때 대체로는 고유요인간 상관 모두를 UM으로 바꾸어도 되지만, 경우에 따라서는 그 상관들의 구획(block)별로 하나씩 UM으로 바꾸고 나머지 구획들은 그대로 CU' 형태로 놔두는 것이 추정상의 문제를 거의 확실하게 해결해 준다. 여기서 구획은 고유요인 서로간에 상관으로서 구획내 고유요인끼리는 어느 다른 고유요인과 상관이 되어 있으나 다른 구획내 고유요인과는 상관이 없는 것으로 정의한다. 예로서 아래와 같은 고유요인간 상관이 있다고 하자(번호는 측정치 번호).

(1, 4), (2, 4), (3, 5), (5, 6), (3, 7), (4, 7), (6, 8)

이들 고유요인간 상관은 하나의 구획이 되며 동시에 UM모형으로 바뀌므로써 합치도, 방법효과, 그리고 측정치의 신뢰도를 구할 수가 있다. 이 구획을 UM모형으로 바꾼 것이 그림2에 제시되어 있다. 만일에 하나의 구획을 동시에 UM모형으로 바꾸는데서 오는 추정상의 문제가 관찰되면 그림2에서의 M요인(동그라미) 하나씩에 대해서만 UM모형으로 바꾸어줄 때 각 측정변수의 R^2 가 증가하는 것이 바로 방법분산에 의한 증분이다. M요인들은 서로간에 상관이 없으므로 순차적으로 하나씩만 요인으로 설정해가면서 측정변수에 대한 R^2 에 기여하는 부분을 모두 더하면 상황효과로 인한 측정변수의 신뢰도 증가부분이 된다. 그림2의 Z변수들은 원래의 내용요인에 대한 지표변수(indicator)로서의 역할에 추가로, 방법요인에 대한 지표변수로서의 역할이 표시되면서, 고유요인간 상관에 대한 대안적 표시가 된다. 그림2에서 Z에 랜덤오차의 표시는 생략되었고 그들간 상관은 없다.

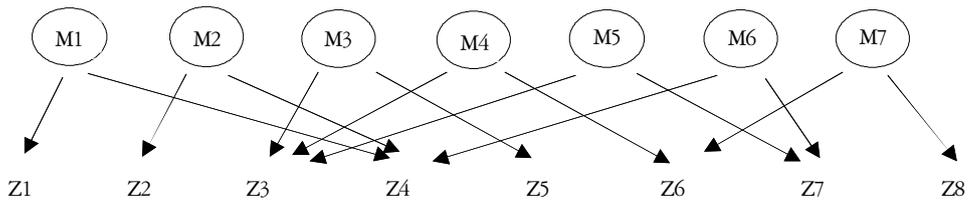


그림 2. 고유요인간 상관으로부터 방법요인의 설정

K C I