

인사조직분야 횡단연구의 타당성 평가 및 향후 제언: 인사·조직연구 게재 논문을 중심으로

박 원 우*
정 구 혁**
유 지 현**
최 동 원***
김 상 윤***

—〈목 차〉—

I. 서론	IV. 결과
II. 연구설계상 타당성의 의미와 기준	V. 연구설계상 타당성 증진방안
III. 연구방법	VI. 결론

연구방법론에서 타당성(validity)은 여러 분야에서 강조되고 있는 바, 대표적인 예가 연구설계에서의 타당성과 측정에서의 타당성이다. 측정에서의 타당성은 특정 개념이나 속성을 측정하기 위해 개발된 측정도구가 해당 속성을 얼마나 정확히 측정하고 있는가를 나타내는 개념이다. 그러나 연구설계에서의 타당성은 이보다 더 포괄적인 개념으로, 종속변수의 변화가 순수하게 연구자가 의도한 독립변수의 변화에 의한 것인지(내적 타당성), 연구설계에 있어서 변수의 조작적 정의 및 변수의 측정이 연구자가 의도하는 바대로 이루어졌는지(구성 타당성), 통계분석의 절차 및 결과 해석이 올바르게 이루어졌는지(통계적 결론 타당성), 연구를 통해 도출된 결과 및 관계가 다른 시간, 공간 및 상황, 대상에서도 일반화되어 적용가능한지(외적 타당성)를 의미한다. 따라서, 타당성이 엄격하게 고려되지 않은 연구는 그 결과가 현상을 잘 설명하거나 예측하지 못하는 등의 심각한 제 오류를 가져올 수 있다.

본 연구에서는 Cook & Campbell(1979)이 제시한 타당성의 세부 분류와 정의를 구체적으로 살펴보고, 각각의 타당성 정도를 판별하는 적절한 근거점과 핵심기준들(criteria)을 설정한다. 이를 바탕으로 최근 10년(2000년~2009년) 동안 인사·조직연구에 발표된 대부분의 실증연구들이 횡

* 서울대학교 경영대학 교수
** 서울대학교 대학원 경영학과 박사과정
*** 서울대학교 대학원 경영학과 석사과정

단연구(cross-sectional study) 방식으로 수행되고 있음에 주목, 연구들이 각각의 타당성 기준에 얼마나 부합하는가를 조사, 분석하여 국내 인사조직 연구에 있어서 타당성에 대한 고려 실태를 진단해본다. 나아가, 횡단연구 방식이 가지고 있는 태생적 한계에 대한 원론적 비판이 아닌, 횡단연구에 있어서의 타당성 인식 및 세부 증진방법을 살펴봄으로써 향후 한국 인사조직 연구에서 타당성 증대를 위해 고려해야 할 방향을 제시한다.

I. 서론

인간은 무엇이 옳은 것인가에 대해 끊임없이 탐구하며, 일련의 탐구결과를 타인에게 알림으로써 사실을 입증하고자 한다. 이 과정에서 연구자의 주장이 설득력을 얻기 위해서는 객관적인 입증이 필요하며, 과학적 연구방법론은 이러한 객관성을 증대시키기 위한 일련의 규칙이라 할 수 있다. 이러한 맥락에서 볼 때 과학적 연구방법론과 관련된 담론의 핵심은 연구설계상에 있어서의 타당성 증대에 있다고 볼 수 있다.

타당성은 크게 측정에서의 타당성과 연구설계에서의 타당성으로 나뉘볼 수 있다. 측정에서의 타당성은 해당 변수가 연구자가 살펴보고자 의도했던 개념이나 속성을 정확히 관측하였는지의 여부와 관련된 변수 측정상의 개념이다. 이는 내용 타당성(content validity), 기준 타당성(criterion-related validity), 구성 타당성(construct validity)으로 분류되며(Kerlinger & Lee, 2000) 연구 전반이 아니라 측정이라는 연구의 일부분에 국한된 개념이라 할 수 있다. 반면, 이보다 더 포괄적인 개념인 연구설계상 타당성과 관련하여 Campbell & Stanley(1963)는 타당성을 내적 타당성(internal validity)과 외적 타당성(external validity)으로 나누었으며, 이후 Cook & Campbell(1979)은 이 분류를 내적 타당성, 외적 타당성, 구성 타당성, 통계적 결론 타당성(statistical conclusion validity)으로 재차 분류하였다.

타당성의 중요성을 인식한 경영학 분야의 많은 학자들은 일련의 연구설계 타당성과 관련된 문제점들을 지적하고, 이를 증진시킬 수 있는 방법들을 다방면에 걸쳐 모색해왔다. 예를 들어, 내적 타당성을 증진시키기 위한 실험 설계 및 준실험설계에 대한 제안이 이루어졌으며(Highhouse, 2009; Grant & Wall, 2009), 외적 타당성에 대한 증진방안 및 관련 이슈들이 특히 실험연구와 관련된 맥락을 중심으로 심리학, 경영학의 각 분

과에서 다방면에 걸쳐 연구되었다(Anderson & Bushman, 1997; Berkowitz & Donnerstein, 1982; Dipboye & Flanagan, 1979; King & He, 2005; Lynch, 1982; Winer, 1999). 뿐만 아니라 구성 타당성의 의미 및 이와 관련된 연구방법론상의 문제점들 역시 꾸준히 제기되어 왔으며(Campbell & Fiske, 1959; Cronbach & Meehl, 1955; Drasgow & Miller, 1982; Schwab, 1980; Smith, 2005a; Smith, 2005b), 특히 최근의 많은 학자들은 나아가 구성 타당성 증진을 위한 대안적 방안을 제시하였다(Bagozzi, Yi, & Phillips, 1991; Westen & Rosenthal, 2003; Westen & Rosenthal, 2005).

하지만 타당성의 증진을 위하여 연구자들이 실질적으로 연구수행 측면에서 얼마나 많은 노력을 경주하였는가 살펴보면, 그 결과는 대체로 회의적이다. Mitchell(1985)은 당대의 조직연구 중 횡단연구들의 타당성 평가를 실시하여 당시의 연구들이 타당성 측면에서 다양한 문제를 내포하고 있음을 밝히고, 향후 개선방안을 논하였다. 하지만 경영학 연구방법론 현황을 1985-1987년과 1995-1997년의 두 기간으로 나누어 비교 분석한 Scandura & Williams(2000)의 연구결과에 의하면, Mitchell의 타당성 증진을 위한 제안은 사실상 무시되었으며, 시간이 경과함에 따라 연구의 타당성에 잠재적으로 위협이 되는 요인들은 오히려 증가한 것으로 나타났다.

이러한 동향은 한국 인사조직 분야 역시 크게 다르지 않은데, 이진규·박기찬(1993)은 한국 인사조직 연구 중 실증연구임에도 통계적 유의성 검증조차 수행하지 않은 논문이 많으며, 방법론상의 측정도구에 대한 신뢰도 및 타당성 검증 역시 소홀히 다루어지고 있음을 지적한 바 있다. 또한 실증연구 중 종단연구(longitudinal study)는 7%에 그친 반면 대다수의 연구가 횡단연구(cross-sectional study)였음이 한계점으로 지적되고 있고, 결과의 타당성을 보완할 수 있는 다양한 가설검증기법이 활용되지 않고 있는 것도 문제로 언급되었다(권석균, 2004).

이진규·박기찬의 연구와 권석균의 연구는 10년의 격차를 보이고 있음에도 불구하고, 두 연구 모두 공통적으로 한국의 인사조직 분야 실증연구들이 대부분 횡단연구에 편중되어 있음을 지적하고 있다. 특히, 권석균(2004)은 횡단연구 방식을 채택한 대부분의 연구자들이 이를 한계점으로 인정하고 향후 종단연구 방식의 검증이 필요함을 단지 언급할 뿐, 추가적으로 타당성 증진을 위한 노력을 하지는 않음을 비판하였다. 뿐만 아니라 Scandura & Williams(2000)의 분석결과를 살펴보면, 1980년대에서 1990년대로 오

면서 중단연구가 전체 발표된 연구들 중에서 차지하는 비중이 오히려 줄어들었음을 알 수 있다. 이러한 결과들은 실증연구의 횡단연구 편중 현상이 비단 한국에만 국한되지 않음을 시사한다.

본 연구는 지금까지 언급된 일련의 문제의식을 바탕으로 논의를 진행한다. 우선, 연구 설계 측면에 있어서 타당성의 정의는 무엇이며, 어떻게 세부적으로 분류될 수 있는지 알아본다. 또한 이렇게 세분화된 타당성들이 가진 각각의 의미와 중요성에 대하여 상세히 고찰하고, 더 나아가 현재 대부분의 실증연구들이 횡단연구 방식으로 수행되고 있음을 주목하여 횡단연구에 있어서 타당성 평가의 핵심기준들(criteria)을 설정한다. 일련의 과정을 통해 설정된 기준들을 활용하여 최근 10년간 *인사·조직연구*에서 발표된 횡단연구들의 현황을 조사하고 국내 인사조직 연구의 실태를 진단해 봄으로써 지금까지 올바른 방향으로 잘 이루어지고 있는 요소에는 어떤 것이 있으며, 앞으로 개선의 여지가 있는 측면에는 어떠한 것이 있는지 고찰한다.

II. 연구설계상 타당성의 의미와 기준

1. 연구설계상 타당성의 의미 및 분류 방식

연구설계상 타당성을 최초로 언급한 Campbell & Stanley(1963)는 타당성을 내적 타당성과 외적 타당성으로 나누었다. 내적 타당성은 종속변수의 변화가 순수하게 연구자가 의도한 독립변수의 변화에 의한 것인가의 여부로, 변수간의 관계가 순수한 인과관계(causal relation)인지를 판별하는 기준이 된다. 연구 수행 시 내적 타당성을 위협하는 요소로는 우연한 사건(history), 성숙효과(maturation), 시험효과(testing), 상호작용효과(interaction effect), 표본선정의 편중(selection bias) 등이 있으며(Campbell & Stanley, 1963), 다른 사건의 과급효과나 제 3의 변수를 고려하지 못한 경우에 내적 타당성이 훼손될 수 있다(Cook & Campbell, 1979). 한편 외적 타당성은 연구를 통해 도출된 결과 및 관계가 다른 시간, 상황 및 대상에도 일반화되어 적용가능한지의 여부를 나타내며, 일반화의 범위가 넓을수록 외적 타당성이 높다고 할 수 있다(Campbell & Stanley, 1963).

이후 Cook & Campbell(1979)은 타당성 분류를 내적 타당성, 외적 타당성, 구성 타당성, 통계적 결론 타당성의 네 가지로 재차 분류하였다. 여기에는 구성 타당성과 통계적 결론 타당성이 추가되었는데, 구성 타당성은 연구설계에 있어서 변수의 조작적 정의(operational definition) 및 실제 변수의 측정이 연구자가 의도하는 바대로 이루어졌는지 여부를 의미한다. 연구자들이 실제 연구하고자 하는 개념을 올바르게 대표하지 못하는 조작적 정의를 사용하는 경우 설령 연구의 가설검증이 성공하였다 하더라도, 이는 저자가 설명하고자 의도했던 관계와는 다른 관계를 나타내는 것이기 때문에 연구의 타당성이 훼손된다. 마지막으로 통계적 결론 타당성은 통계 분석의 절차와 결과 해석이 올바르게 이루어졌는가의 여부를 말한다. 즉, 통계적 검증력이 낮은 경우, 통계의 기본 가정을 위배한 경우, 그리고 측정값의 신뢰도가 낮은 경우 등의 문제가 발생할 때, 해당 연구에 있어서의 통계적 결론 타당성이 훼손된다(Cook & Campbell, 1979).

2. 연구설계 유형이 타당성에 미치는 영향

Schwab(2005)은 연구설계 유형을 분류함에 있어, 실험(experiment), 준실험(quasi-experiment), 설문연구(survey study), 현장연구(field study)로 나누고 있다. 그는 이 중 전자의 둘은 인과관계를 비교적 명확히 밝힐 수 있으나, 후자 둘은 인과관계를 명확히 할 수 없다는 단점을 가지고 있다고 주장하였다(Schwab, 2005). 또한, 다른 많은 학자들도 설문연구 또는 횡단연구가 내적 또는 외적 타당성 문제에 더 많이 노출될 수 있음을 지적해 오고 있다(King & He, 2005; Mitchell, 1985; Werner, Praxedes, & Kim, 2007). 그러나 연구의 특성으로 인해 실험, 준실험, 종단연구, 메타분석연구 등이 불가능하고 횡단연구를 할 수 밖에 없는 상황도 있다. 또한 비교적 쉽고 빠른 검증 과정 또는 여타의 이유로 인해 횡단연구 방식을 택하는 경우가 많은 것 역시 사실이다. 따라서 타당성을 높이기 위해서는 실험연구와 준실험연구, 혹은 종단연구와 메타연구 등 보다 다양한 연구방법의 도입, 활용이 보다 근원적인 처방이겠으나, 횡단연구 설계를 유지하면서 동시에 내적, 외적, 구성, 통계적 결론 타당성 각각을 제고시켜 전체 연구타당성을 증진시키는 방안을 살펴보는 것도 하나의 유효한 해결책이 될 수 있을 것이다. 본 연구에서는 횡단연구에 있어서 타당성의 기준이 될 만한 것이 무엇이고, 지금까지 국내 주요 논문의 현황은 어떠한지 알아본 후, 향후 연구들이 높은 타당성을 지닐 수 있도록 제언한다.

3. 연구설계상 타당성의 기준 설정

본 연구에서는, Cook & Campbell(1979)에 의해 제시된 타당성의 네 가지 분류법-내적 타당성, 구성 타당성, 통계적 결론 타당성, 외적 타당성-이 연구설계 관점에서의 타당성에 대해 가장 포괄적이고 자세한 분류를 제시한다고 판단, 이 분류를 사용하여 횡단연구의 타당성을 어떠한 기준으로 평가할 수 있고, 그 기준을 사용함에 있어서의 장점은 무엇인지 알아보도록 한다.

1) 내적 타당성

내적 타당성은 두 변수 간에 관계(relationship)가 관측되고 그것에 기반하여 인과관계를 추론할 경우 그 인과관계 추론이 얼마나 타당한지를 나타내는 개념이다(Cook & Campbell, 1979, p. 37). 즉, 연구가 내적 타당성을 가질 경우에만 연구자는 두 변수 간에 인과관계가 있다고 올바르게 추론하고 주장할 수 있다. 이 때 내적 타당성을 가지기 위해서는 연구설계상 대안이론배제의 메커니즘이 존재해야 한다(Cook & Campbell, 1979; Mitchell, 1985).

구체적으로 연구에 있어 내적 타당성은 특정한 변수 A의 변화가 진정 변수 B의 변화에 영향을 미친 것인가 하는 물음과 관련된다(Mitchell, 1985). 이러한 내적 타당성의 측면에서 Campbell & Stanley(1963)는 정당한 인과추론의 3요소를 제시한다. 첫째, 원인은 결과에 대해 시간적으로 선행해야 하며, 둘째, 원인과 결과는 공변(covary)해야 하며, 셋째, 해당 관계에 대한 대안적 설명이 없어야 한다. 사실, 위의 3요건을 하나의 독자적 연구로 만족시키는 방법은 일반적으로는 실험을 수행하는 수밖에 없다. 이러한 맥락에서 Campbell & Stanley(1963)가 내적 타당성을 위협하는 요인으로 꼽은 다양한 요소들이 기본적으로 실험연구, 혹은 준실험연구를 가정한 것이며 횡단연구에 대해서는 거의 언급하고 있지도 않다는 것은 주목할 만한 사실이다. 횡단연구는 변수 발생의 시간적 선행에 대한 고려도, 대조군의 설정도 없으므로 횡단연구 자체만으로는 종합적인 내적 타당성을 획득할 수가 없다(Mitchell, 1985). 결국 근본적인 처방을 위해선, 횡단연구를 넘어서 종단연구나 실험 및 준실험연구 등을 수행하고 그에 기반해 다양한 내적 타당성 위협 요소를 추가적으로 통제하는 노력을 경주해야 할 것이다.

하지만 경영 관련 연구의 특성상 실험이나 종단연구가 현실적인 제약, 혹은 윤리적 고

려 등으로 인해 불가능한 경우가 많은 것을 생각하면, 무작정 횡단연구는 내적 타당성을 보장할 수 없으니 수행하지 말라는 것은 비현실적인 요구일 수 있다. 나아가 횡단연구의 결과가 근원적으로 실험연구와 같은 수준의 내적 타당성을 가지지 못한다고 하여 그 결과나 인과의 추론 전체가 틀렸다고 일괄적으로 결론 내리는 것 또한 오류로 이어질 수 있다. 횡단연구는 인과관계에 대한 암시를 제공함으로써 향후 연구에 대한 예비단계로서의 기능을 수행할 수도 있고, 기존이론 검증 등의 목적에도 유용하게 쓰일 수 있기 때문이다. 또한 비록 인과관계를 결정적으로 보일 수는 없지만, 같은 횡단연구일지라도 연구 설계에 따라 상대적으로 내적 타당성을 증진시키는 것은 가능한 일이며 동시에 마땅히 추구되어야 할 목표라 하겠다.

(1) 통제변수

일반적으로 횡단연구에 있어서 Campbell & Stanley(1963)가 제시한 인과추론의 3요소 중 대안설명을 가능한 배제해야 한다는 목표는 통제변수의 추가로 추구된다. 이와 관련해 Mitchell(1985)은 횡단연구의 내적 타당성을 위협하는 가장 큰 요인 중 하나는 연구자가 관심을 가지는 두 변수 모두와 상관을 가지는 제 3의 변수를 무시하는 것이라 말하였다. 예를 들어, 교회의 수와 선술집의 수의 상관을 구한다면 아마도 큰 상관이 구해질 것이다. 그렇다고 하여 교회의 수와 선술집의 수 사이에 특정한 인과관계가 존재하는 것일까? 아마도 아닐 것이다. 위의 현상은 도시의 크기라는 제 3의 변수가 개입된 결과이다. 이 때 연구자가 도시의 크기라는 변수를 파악하지 못하고 교회의 수와 선술집의 수를 연결시키는 모종의 이론을 제시한다면 해당 연구는 내적 타당성의 측면에서 매우 큰 약점을 지닐 수밖에 없다(Mitchell, 1985). 이렇듯 통제변수의 포함은 연구의 내적 타당성을 증진시키는 주요한 방안이 된다(Schmitt & Klimoski, 1991). 하지만 통제변수를 고려한다고 해도 충분한 이론적 근거가 있어야 할 것이며 그렇지 않고 잘못 투입된 경우 제 3의 통제변수의 포함이 오히려 모델의 정합성 하락과 결과의 왜곡을 불러올 수도 있다(Becker, 2005). Meehl(1971)은 잘못된 통제변수의 사용을 당대의 사회과학연구에 있어 가장 흔한 오류로 꼽았으며, 이러한 잘못은 21세기에 들어서도 그다지 개선되지 않은 것으로 보인다(Becker, 2005). Spector, Zapf, Chen, & Frese (2000)는 무차별적인 통제변수의 포함이 연구대상이 되는 변수의 크기를 부당하게 줄여 결과적으로 2중 오류를 불러올 수 있음을 보였다. 또한 제 3의 통제변수가 연구대상인

두 변수 중 하나와만 상관을 지닐 때는 반대로 1종 오류의 가능성이 증가될 수도 있다 (Becker, 2005). 이렇듯 통제변수란 관습적으로 으레 포함해야 될 대상이 아닌, 연구 변수 이상의 충분한 근거와 숙고 끝에 포함되어야 할 대상이다. 이러한 측면에서 횡단연구의 내적 타당성을 평가할 때 통제변수의 포함에 대한 충분한 이론적, 경험적 근거가 제시되었는지를 평가해 보는 것이 필요하다.

(2) 역인과성

횡단연구는 원칙적으로 시간의 선행성을 고려하지 않기에 인과추론이 불가능하다(Mitchell, 1985). 하지만 많은 경우, 연구자가 미리 의도한 인과의 방향(causal direction)이란 것이 존재하며 그러한 특정 인과방향에 따라 논의가 진행되기 마련이다. 그렇다면 횡단연구는 시간의 선행성을 살펴보지 못하므로 이러한 인과의 이슈는 아예 다루지 않아야 하는 것인가? 오히려 연구자는 횡단연구의 한계점을 깨닫고 그럼에도 최대한 치열하게, 이론에 근거하여 합리적으로 자신이 생각하는 인과의 방향을 지지하여야 할 것이다. 횡단연구 자체는 인과의 방향을 말하지 못하지만, 그와 관련된 기존의 연구나 경험적 사례들을 들어 연구자는 역인과로 인한 대안설명의 가능성이 낮음을 보이고 나아가 연구의 내적 타당성을 증진시킬 수 있다. 또한 역인과로 구성되는 모델을 다양한 통계방식을 활용하여 검증해 보는 것도 가능할 것이다. 이러한 측면에서 횡단연구에 있어 역인과 문제를 고려하고 언급하였는지, 혹은 더 나아가 역인과 모델을 테스트하는 과정이 있었는지 살펴보는 것이 필요하다.

위의 과정에서 도출된 내적 타당성 관련 기준들은 아래와 같다.

- 통제변수를 모델에 포함시켰을 때 그것을 정당화하는 충분한 이론적 설명이 추가되었는가?
- 역인과성에 대한 고려는 있었는가?

2) 구성 타당성

연구를 수행할 때 구성개념(construct) 자체를 측정할 수는 없다. 따라서 연구자는 조작적 정의를 통해 해당 구성개념에 대응되는 측정 가능한 형태의 관찰치를 사용하여 연구를 진행한다. 이 때 과연 구성개념과 연구자가 이용하는 관찰치가 일치하는가 하는 의문을 가질 수 있다. 즉, 관측치의 변산이 해당 구성개념으로 인해 얼마나 설명되는지 질문을 해

볼 수 있고, 그 일치도를 나타내는 것이 구성 타당성이다(Kerlinger & Lee, 2000). 내용 타당성과 예측 타당성은 구성 타당성과 구별되는 개념이지만(Cronbach & Meehl, 1955; Guion & Cranny, 1982), 많은 연구자들은 구성 타당성이 더욱 포괄적이며, 예측 타당성과 내용 타당성을 포함하는 개념이라는 것에 동의한다(Mitchell, 1985).

연구의 측정수단이 연구자가 의도한 구성개념의 변산을 잘 감지하고 표현할 때, 비로소 구성 타당성이 높다고 할 수 있다(Cook & Campbell, 1979). 반면 특정 관찰치가 의도된 구성개념의 의미를 미처 다 담고 있지 않을 수도 있고 혹은 의도한 것보다 많은 것을 동시에 측정하고 있을 수도 있다(Schwab, 1980). 이에 따라 구성 타당성이 담보되지 않았을 때, A와 B라는 구성개념의 관계라고 해석된 것이 실제로는 A와 Y의 관계 혹은 X와 B의 관계, 심지어는 X와 Y의 관계를 반영하고 있을 수도 있는 것이다(Cook & Campbell, 1979). 이처럼 구성 타당성이 낮은 경우엔 관측치와 구성개념 사이의 불일치 문제로 인해 대안설명의 가능성이 제기될 수 있으며, 이러한 잠재적인 문제는 횡단연구에서 특히 두드러진다(Drasgow & Miller, 1982; Schwab, 1980).

(1) 수렴 타당성과 판별 타당성

Kerlinger & Lee(2000)는 구성 타당성을 확보하기 위해선 수렴 타당성(convergent validity)과 판별 타당성(discriminant validity)을 우선적으로 살펴보아야 한다고 말한다. 같거나 유사한 구성개념을 측정하는 관측치들은 서로 유사한 결과를 도출해야 한다는 것이 수렴 타당성 개념이고, 상이한 구성개념을 측정하는 관측치들은 서로 상이한 결과를 보여야 한다는 것이 판별 타당성 개념이다(Kerlinger & Lee, 2000). 수렴과 판별 타당성을 살펴봄에 있어 대표적인 방식은 Campbell & Fiske(1959)의 multitrait-multimethod matrix method(이하, MTMM)이다. 동일한 개념을 상이한 방식으로 측정했을 때 상관관계가 높게 나오면 수렴 타당성이 있다고 보고, 상이한 개념을 동일한 방식으로 측정했을 때 0에 가까운 상관관계가 나오면 판별 타당성이 있다고 본다. 이렇게 상이한 측정방식과 상이한 개념을 동시에 연결시켜 상관관계를 구해보는 MTMM 방식은 구성 타당성을 알아보기 위해 이상적이며 또한 가능하다면 항상 시도되어야 하는 방법이다(Kerlinger & Lee, 2000). 그러나 많은 경우 두 가지 이상의 측정방식을 연구에 도입하는 것은 노력과 비용 측면에서 지나치게 비경제적이거나 연구의 특성상 원천적으로 불가능할 수도 있다. 하지만 그런 상황에서도 연구자들은 구성 타당성 평가를 포기하기 보다는, 가능한 모든 노력을 다 기

을여야 한다(Kerlinger & Lee, 2000). 예를 들어, 요인분석을 이용하여 수렴과 판별 타당성을 구해볼 수 있으며(Kerlinger & Lee, 2000), 이전 연구의 해당 구성개념에 대한 타당성 측정 결과를 찾아보고 인용하는 것은 가능할 것 역시 가능하다. 이것은 MTMM을 사용한 경우에도 마찬가지인데, 비록 상이한 측정방식들을 사용하여 MTMM을 구성한다 하더라도, 그것이 개별적인 연구계획 하에서 하나의 연구집단에 의해서 수행된 이상, 각 측정 방식이 일정 크기의 공통분산을 공유할 수밖에 없고 진정한 multimethods가 될 수는 없기 때문이다(Mitchell, 1985).

(2) 동일방법편의

수렴 타당성과 판별 타당성에 대한 고려가 개별 변수 사이의 구성 타당성을 살펴보는 것이라면, 동일방법편의(common method bias)에 대한 검토는 연구에 쓰인 변수 전반에 공통으로 존재하는 잠재적인 구성 타당성 위협요인에 대한 고려에서 비롯한다. 예를 들어 A, B, C, D라는 구성개념 각각에 대해 구성 타당성이 구해질 수 있는데 그것이 위에서 제시한 수렴 타당성과 판별 타당성 검토이다. 반면 연구 자체가 가지는 특성이 각각의 A, B, C, D라는 구성개념 관측치에 전반적으로 어떠한 영향을 미쳤는지가 의문의 대상이 될 수 있다. 실제로는 완전히 상이한 구성개념들을 하나의 동일한 연구 혹은 방식에 기반하여 측정하였을 때 해당 관측치들이 0이 아닌 일정 크기의 상관을 가지게 될 수도 있는 것이다. 이렇게 동일방법편의는 전반적인 변수 상관에 직접적인 영향을 미칠 수 있으므로 특히 횡단연구에 있어 심각한 문제로 작용할 수도 있다(Fiske, 1982; Mitchell, 1985). 이러한 동일방법편의는 연구에 있어 매우 흔히 관찰되는 현상이며, 그 영향 또한 무시할 수 없다(Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; 박원우, 김미숙, & 정상명, 2007). 따라서 동일방법편의에 대한 고려 없이 개별 구성 타당성만 논한다면 숲을 보지 못하고 나무만 보는 형상이 될 수도 있다. 이러한 동일방법편의는 사전 연구설계를 통해 애초에 피하거나 혹은 통계적 방식을 통해 사후적으로 처리해야 한다(Podsakoff et al., 2003; 박원우 외, 2007).

위의 논의에서 도출된 구성 타당성 관련 기준들은 아래와 같다.

- (본 연구에서) 수렴 타당성과 판별 타당성을 구했는가?
- (이전 연구에서) 가져온 수렴 타당성과 판별 타당성에 대한 인용이 있는가?
- 동일방법편의를 고려하였는가? 단순히 언급만 하였는가, 사전 연구설계 혹은 사후

통계방식을 통해 처리하였는가?

3) 통계적 결론 타당성

Cook & Campbell(1979)에 의해 논의된 통계적 결론 타당성은 연구의 통계절차 및 결과가 얼마나 믿을만한가를 평가하는 기준이다. 구체적으로 말하자면 통계 과정에 내포된 불안정성이 클 때 통계적 결론 타당성이 낮다고 평가되며, 그러한 불안정성은 관측 방식이 가지는 낮은 신뢰도, 불완전한 표본 추출, 특정 통계방식이 가지는 한계, 혹은 잘못된 통계방식의 사용 등에 기인한다(Mitchell, 1985). 구성 타당성의 위협요인이 구조적 오류(systematic error)로 인해 발생한다면(Cronbach & Meehl, 1955) 통계적 결론 타당성 위협 요인은 주로 무선오류(random error)에 의해 발생한다(Cook & Campbell, 1976; Mitchell, 1985).

(1) 신뢰도

통계적 결론 타당성에 대한 다양한 위협 중 하나는 관측에 있어서의 낮은 신뢰도(reliability)이다. 신뢰도는 측정치가 얼마나 일관된 결과를 보여주느냐를 나타내는 개념이다. 이러한 신뢰도는 단지 측정항목(item) 사이의 일관성(내적 일관성)뿐만이 아니라 반복 측정, 상이한 환경, 상이한 관측자 상황 하에서도 과연 일관된 결과를 나타내는지 여부를 포괄한다(Cronbach, Glaser, Nanda, & Rajaratnam, 1972). 통상적으로 횡단연구에서 보고되는 신뢰도는 Cronbach's alpha인데 이것은 앞의 여러 신뢰도 개념 중 측정항목의 내적 일관성을 나타낸다. 하지만 내적 일관성은 신뢰도 개념의 일부일 뿐이며, 또한 가장 중요한 부분이라고 할 수도 없다. 항목간의 내적 일관성은 높지만 시간의 변화, 설문을 시행하는 관측자 혹은 매체 방식의 변화에 있어선 일관적이지 않은 측정도구가 있을 수도 있으며 연구자가 그로부터 도출된 관측값을 가지고 무비판적인 해석을 이끌어낸다면 잘못된 추론으로 이어질 수도 있다(Mitchell, 1985). 따라서 Cronbach's alpha뿐만이 아니라 검사-재검사 신뢰도, 관측자간 신뢰도, 동형검사 신뢰도(equivalent forms reliability) 등의 보고가 필요하다.

(2) 표본크기

표본크기(sample size) 역시 통계적 결론 타당성 평가에 있어 중요한 기준이 된다.

일반적으로 쓰이는 대다수의 통계처리 방식은 중심극한정리(central limit theorem)에 기반하는데, 이에 따르면 모집단의 분포와 상관없이 표집분포는 표본크기가 커짐에 따라 정규분포에 가까워지는 성질을 가지며, 이러한 성질 하에 다양한 통계방식의 전개가 정당화된다. 이 때 일반적으로 표본크기가 30 이상이 되어야 표집분포가 정규분포에 가까워지는 효과가 실질적으로 나타난다고 볼 수 있다(Kerlinger & Lee, 2000). 하지만 단순히 표본크기가 30이 넘는다고 하여 통계의 결과가 곧바로 정당화되는 것은 아니다. 모델에 포함된 변수의 수에 비례해 적절한 표본크기가 보장되지 않으면 비록 표본크기가 30을 넘더라도 상관연구에 기반한 연구들은 1종 오류에 취약해진다(Hair, Black, Babin, & Anderson, 2010). 예컨대 30개 변수를 가지고 상관을 구할 경우 435개의 상관이 도출되는데 이때 30개 모두가 실제로는 아무 상관없는 변수들일지라도 .05 유의도 기준에서 435개의 상관 중 20개는 유의하다고 잘못 판단할 위험이 존재하는 것이다. 물론 유의도 기준치를 줄이면 이 문제는 줄어들 수 있지만 그럴 경우 통계의 검증력(power) 역시 줄어들어 2종 오류가 증가할 위험이 있다. 따라서 표본크기를 늘려 1종 오류를 줄이는 수밖에 없는데, 이 때 일반적으로 변수 개수의 5배에서 8배 이상의 표본크기가 요구된다(Hair, Black, Babin, & Anderson, 2010).

(3) 복수의 표본추출

횡단연구에서 회귀를 이용한 분석은 매우 일반적이다. 추후의 결과 부분에서도 언급되었지만 10년간의 인사·조직연구 논문 중 상당수가 회귀분석을 사용하고 있다는 것을 본 연구를 통해서도 알 수 있다. 하지만 이렇듯 주로 쓰이는 회귀분석의 경우 과적합(overfitting)의 문제가 상당히 심각하다(Mitchell, 1985). 여기서 과적합이란 회귀분석을 통해 도출된 회귀계수가 연구자가 가진 특정 표본에만 적합하고 여타의 표본에 적용하였을 때엔 설명력이 확연히 감소하는 현상을 말한다. 회귀에 기반한 연구에서는 이러한 위험을 인지하고, 표본을 분리하여 회귀를 구해 도출된 결과를 원회귀식의 결과와 비교해보거나, 주 연구 표본으로 회귀식을 도출한 후 추가적인 표본에 해당 회귀식을 적용시켜보는 것과 같은 방식의 사후 검증과정이 필요하다(Mitchell, 1985). 따라서, 연구의 통계적 결론 타당성을 평가하기 위한 기준으로 복수의 표본추출(multiple sampling) 과정이 존재했는지 여부가 고려될 수 있다.

위의 논의에서 도출된 통계적 결론 타당성 관련 기준들은 아래와 같다.

- 신뢰도를 보고하였는가? 보고하였다면 Cronbach's alpha만 보고하였는가, 여타의 검사-재검사 측정 신뢰도, 동형검사 신뢰도 등도 보고하였는가?
- 해당 통계방식과 모델에 적합한 표본크기를 사용하였는가?
- 회귀분석을 사용한 경우, 복수의 표본추출을 통해 회귀분석에 내재한 과적합 문제를 극복하였는가?

4) 외적 타당성

외적 타당성이란 연구결과가 다른 시간, 상황 및 대상에도 적용될 수 있는지의 여부로 (Campbell & Stanley, 1963), 일반화 가능성이라고도 말할 수 있다. 일반화 가능성이 클수록 외적 타당성은 높은 것으로, 하나의 실험연구에 대해 다양한 상황과 집단, 측정도구를 사용하거나 메타분석을 하는 경우에 일반적으로 외적 타당성이 높다고 볼 수 있다(Mitchell, 1985).

사실 현장연구의 경우, 실험실 연구에 비해 일반화 가능성이 보통 높을 것이라 생각하지만 현장연구가 실험연구보다 그 자체로서 외적으로 더욱 타당하다고 할 근거는 없다 (Berkowitz & Donnerstein, 1982; Dipboye & Flanagan, 1979; Hunter & Schmidt, 2004; Willems & Howard, 1980). Mitchell(1985)도 설문연구의 외적 타당성이 실험연구 만큼이나 취약할 수 있음을 지적하며, 현장연구가 그 자체로 외적 타당성을 갖는다는 생각은 부적절할 수 있음을 시사하였다.

(1) 무작위 표본추출

먼저 횡단연구의 일반화 가능성은 무작위 표본추출(random sampling)을 통해 실현될 수 있다. 실험연구가 아닌 실증연구에서는 일반적으로 모집단으로부터 연구 표본을 선정하는 표본추출 과정이 필요하며 이 때 외적 타당성 문제가 발생할 수 있는데(Werner, Praxedes, & Kim, 2007), 연구결과를 모집단까지 확대할 수 있으려면 표본이 관심 모집단을 충분히 대표하고 있어야만 한다. 이 때 무작위 표본추출이란 표본들이 모집단으로부터 선정될 확률이 동일한 표본추출 방식으로(Kerlinger & Lee, 2000), 이의 사용은 표본의 체계적 편향을 막아주어 표본의 대표성을 높여 준다. Cook & Campbell (1979)은 무작위 표본추출을 시행하지 않은 경우 연구결과가 구조적 편의(systematic bias)를 갖게 되어 그 결과는 신뢰할 수 없고 그에 기반해 인과추론을 하기도 어려워진

다고 하였다. 따라서 연구가 무작위 추출 방식을 사용하였는지, 현실적으로 불가능한 경우라면 이를 극복하고자 어떻게 노력하고 고민하였는지를 살펴보도록 한다.

(2) 무응답 오류

연구의 외적 타당성을 평가하고자 할 때는 표본표출 방식과 더불어 무응답 오류(nonresponse error)도 고려해야 한다. 무응답 오류는 특정 연구조건에서 수집된 표본의 특이성에서 비롯된 것으로, 무응답자 비중과 변수의 응답자통계량과 모집단통계량 편차를 곱한 값으로 표현할 수 있다(Rogelberg & Luong, 1998). 즉, 응답자와 무응답자 간의 체계적인 차이로 인해 표본이 모집단을 충분히 대표하지 못하게 되는 경우 무응답 오류가 나타나는 것이다.

모집단에서 무작위 표본추출을 통해 대표성이 높은 표본을 취득한 경우라도 표본 전체가 응답한 것이 아니라면, 연구자는 표본 전체가 아닌 응답자 역시도 모집단을 대표하는지에 관심을 가져야 한다. 표본이 모집단을 대표해야 하는 것과 마찬가지로, 응답자가 표본 및 모집단을 대표하지 못해 무응답 오류가 발생한다면 연구의 외적 타당성은 떨어지게 된다(Sackett & Larso, 1992; Viswesvaran, Barrick, & Ones, 1993).

무응답 오류를 줄이기 위해 보통은 응답률을 높이고자 하는데, 설사 응답률이 높다 하더라도 여타의 요인들이 응답자와 무응답자 간의 차이를 유발하여 무응답 오류를 가져와 연구결과를 모집단에 적용할 수 없게 될 수도 있다(Roth & BeVier, 1998; Viswesvaran et al., 1993; Werner et al., 2007). 따라서 응답률에 상관없이 연구자는 무응답 오류를 분석하여 무응답 오류의 존재와 크기에 관한 정보를 제공할 수 있어야 한다(Roldberg & Stanton, 2007).

위의 논의에서 도출된 외적 타당성의 기준은 아래와 같다.

- 무작위 표본추출이 이루어졌는가? 아니면 이를 조심하고 우려하는 언급만 있었는가?
- 무응답 오류를 고려하였는가? 분석을 통해 오류가 없음을 결과로 제시하였는가? 아니면 오류 가능성에 대한 논의만 있었는가?

지금까지의 논의에서 도출된 내적 타당성, 구성 타당성, 통계적 결론 타당성, 그리고 외적 타당성에 대한 기준을 정리하면 <표 1>과 같다.

〈표 1〉 타당성 평가 기준

구분	판단 기준 (criteria)	판단 분류 (evaluation)
연구 유형	1. 연구의 유형은 무엇인가?	1. 횡단, 2. 종단, 3. 실험 & 준실험, 4. 메타, 5. 기타
내적 타당성	2. 통제변수를 모델에 포함시켰을 때 그것을 정당화하는 충분한 이론적 설명이 추가되었는가?	-.N/A, 0.no, 1.yes
	3. (횡단연구일 경우) 역인과성에 대한 고려는 있었는가?	-.N/A, 0.no, 1.단순 언급, 2.객관적 test
구성 타당성	4. (본 연구에서) 수렴 타당성과 판별 타당성을 구했는가?	-.N/A, 0.no, 1.yes
	5. (이전 연구에서) 가져온 수렴 타당성과 판별 타당성에 대한 인용(citation)이 있는가?	-.N/A, 0.no, 1.yes
	6. 동일방법편의를 고려하였는가?	-.N/A, 0.no, 1.단순 언급, 2.사전 설계, 3.사후 통계처리
통계적 결론 타당성	7. 신뢰도를 보고하였는가: alpha, 검사-재검사, 동형검사신뢰도	-.N/A, 0.none, 1.only alpha 2.alpha 외 1개 이상
	8. 과적합 등의 문제 해결을 위해 복수의 표본을 이용하였는가?	-.N/A, 0.none, 1.yes
	9. 통계적 결론 타당성을 위한 표본크기를 고려하였는가? 1) $n \geq 30$ 2) $n \geq \text{factor (variable)} * 8$ 3) SEM 사용시 최소 200개	-.N/A, 0.no, 1.yes
외적 타당성	10. 무작위 표본추출 사용 여부?	-.N/A, 0.no, 1.단순 언급, 2.test
	11. 무응답 오류를 고려하였는가?	-.N/A, 0.no, 1.단순 언급, 2.test

Ⅲ. 연구방법

본 연구의 분석대상은 2000년부터 2009년까지의 10년 동안 학술지 *인사·조직연구*에 발표된 192개의 연구논문들이다. 본격적인 연구에 앞서 일차적으로 해당 연구논문들

의 유형을 횡단연구, 종단연구, 실험 및 준실험연구, 메타연구, 기타의 다섯 가지로 분류, 조사하였다. 조사 결과 192개 연구 중 127개의 연구가 횡단연구 방식을 채택하였음이 드러났다. 먼저, 본격적인 조사에 앞서 임의적으로 12개의 연구논문을 추출하여 사전 조사를 실시하고 타당성 여부 평가 항목의 분류 및 설정에 있어서의 애매함, 주관성의 개입, 판별불가능성과 같은 잠재적인 문제점이 없는지 확인하였다.

이어, 127개의 횡단연구를 대상으로 위에서 설정한 10개의 기준에 의거, 각 연구들의 타당성 여부를 조사하였다. 분석결과의 코딩작업은 일차적으로 저자들이 각각 127개 연구들 중 일부를 전담하는 방식으로 이루어졌으며 분류가 애매한 경우 논의를 거쳐 결정하였다.

IV. 결과

먼저 타당성 기준 각각의 결과를 논의하기 전에 2000년부터 2009년까지 10년 간 인사·조직연구에 발표된 논문들의 연구설계 유형을 살펴보고자 한다. 본 연구의 분석대상이 되고 있는 횡단연구가 192편 중 127편(66%)으로 예상했던 바와 같이 가장 큰 비중을 차지하고 있는 것을 확인하였으며, 종단연구 논문이 17편으로 다음을 이었다. 실험 및 준실험연구 논문은 5편(3%) 있었으며 메타분석연구는 조직만족, 조직몰입과 성과에 관한 논문 1편이 있음을 확인하였다. 그 외 질적연구 등의 기타 유형의 논문은 41편으로 10년간 연구의 21%를 차지하였다.

횡단연구 방식의 논문을 각 타당성 기준에 따라 살펴본 결과는 <표 2>와 같다. 횡단연구 방식이 아니거나 횡단연구 방식이라도 연구목적과 특성에 따라 각 기준에 해당되지 않는 경우는 항목별로 분석에서 제외하였다. 예를 들어 과적합 기준에 있어서는 회귀분석 연구만을 그 대상으로 분석하였다.

먼저 내적 타당성 기준을 살펴보도록 하겠다. 첫 번째 기준인 통제변수의 사용에 있어 40%의 연구논문들이 통제변수를 포함시킬 때 별도의 이론적 설명을 하고 있지 않음을 확인하였다. 다음 기준인 역인과성 측면에선 34%(30편)가 역인과 가능성을 언급하고 있는 것으로 나타났으나 반수 이상인 65%는 이를 고려하고 있지도 않은 것으로 나타났다.

〈표 2〉 연구유형 및 타당성 기준 별 결과

1. 연구유형: 횡단=127(66%) 종단=17(9%) 실험 & 준실험=5(3%) 메타=5(1%)
기타=41(21%)

• 내적 타당성

2. 통제변수의 이론적 설명 여부: no=55(60%) yes=36(40%)
3. 역인과성 고려 여부: no=57(65%) 단순 언급=30(34%) 객관적 검증=5(1%)

• 구성 타당성

4. (본 연구에서) 수렴 타당성 및 판별 타당성 분석 여부: no=23(21%) yes=84(79%)
5. (이전 연구로부터) 수렴 타당성 및 판별 타당성 언급 여부: no=95(89%) yes=12(11%)
6. 동일방법편의 고려: no=38(35%) 단순 언급=36(34%) 사전 설계=25(24%)
사후 검증= 7(7%)

• 통계적 결론 타당성

7. 신뢰도 분석 여부: no=5(4%) only Cronbach's alpha=93(88%) alpha 외 추가 분석=8(8%)
8. 과적합 등의 문제 해결을 위해 복수의 표본 사용 여부: no=91(94%) yes=6(6%)
9. 표본크기 고려 여부: no=5(4%) yes=96(95%)

• 외적 타당성

10. 무작위 표본추출 사용 여부: no=72(64%) 단순 언급=29(26%) 사용=29(10%)
11. 무응답 오류 고려 여부: no=99(90%) 단순 언급=5(5%) 사후 검증=5(5%)

다음은 구성 타당성에 대한 기준이다. 먼저, 구성개념의 판별 타당성 및 수렴 타당성을 요인분석 등의 기법을 통해 통계적으로 분석한 논문은 84편으로 다수의 연구가 이들 기준은 충족하고 있었다. 반면 이전 연구에서의 구성 타당성에 대한 정보를 인용하여 비교한 연구는 12편(11%)에 그치는 것으로 나타났다. 이는 Mitchell(1985)이 1979년에서 1983년까지의 *Journal of Applied Psychology*, *Organizational Behavior and Human Performance*, *Academy of Management Journal*을 대상으로 해서 얻은 결과인 20%(26편)보다 낮은 수준으로 이전 연구로부터의 구성 타당성 인용에 더 큰 관심이 요구된다 할 수 있겠다. 각각의 연구표본을 이용해 수렴 타당성 및 판별 타당성을 분석하지 않은 연구 또한 23편(21%)이 확인되었다. 구성 타당성에 대한 또 다른 기준은 동일방법편의에 대한 고려이다. 연구설계를 이용하여 동일방법편의를 배제하고자 한

연구는 25편(24%), 설계에서 반영하지 못했으나 사후 검증을 통해 편의가 없음을 확인한 논문은 7편, 단순 언급만 한 논문은 36편(34%)으로 확인되었다. 박원우 외(2007)가 창간 이후 2005년까지의 인사·조직연구 게재 논문의 동일방법편의에 대한 인식수준을 알아보았을 당시 사전 통제를 한 논문이 총 3편에 그친데 반해, 본 연구에서는 사전 통제 논문이 25편으로 드러나 그간 동일방법편의 인식 수준에 발전이 있었음을 알 수 있었다.

통계적 결론 타당성의 첫 번째 기준은 신뢰도 분석으로 88%의 연구가 Cronbach's alpha만을 사용하고 있었으며 8편(8%)의 연구만이 알파 값 외의 다른 신뢰도 값을 추가적으로 제시한 것으로 확인되었다. 두 번째 기준인 회귀분석의 과적합성 고려에 있어서 복수의 표본을 사용한 연구는 총 6편이 있었다. 통계적 결론 타당성의 마지막 기준은 적절한 표본크기로 중심극한정리의 기준 등 앞서 제시한 보편적 기준에 비추어 볼 때 다수가 이를 충족시키는 것으로 나타났으며, 5편의 논문만 이를 충족시키지 못한 것으로 나타났다.

한편, 다른 타당성 기준에 비해 외적 타당성은 상당히 취약한 수준을 보였다. 먼저, 표본추출 기준을 보면 논문의 10%만이 무작위 표본추출 방식을 사용하고 있으며 나머지 다수는 임의 표본추출 방식을 채택하고 있는 것으로 나타났다. 연구의 한계 등으로 언급한 경우도 26%에 그쳐 과반수가 넘는 72편(64%)이 표본추출 이슈를 전혀 인식하지 못한 것을 확인하였다. 다음 기준은 무응답 오류에 대한 고려로 모든 연구가 전수 응답이 아님에도 불구하고, 인식 수준이 10%에 그쳐 무응답 오류에 대한 더 큰 관심이 필요함을 확인할 수 있었다. 같은 항목에 대해 Werner et al.(2007)이 2000년부터 2004년까지 경영학 분야의 대표적인 9개 저널에¹⁾ 게재된 논문들을 조사하였을 때 30%의 논문이 무응답 오류를 분석한 것으로 나타났는데, 당시 저자들이 이 결과를 가지고도 무응답 오류에 대한 인식 수준이 저조하다고 지적한 것을 감안하면 국내학계에서의 인식 제고가 시급하다 할 수 있겠다.

1) 9개 저널은 저널의 질에 따라 1군, 2군, 3군으로 구분하여 각 군별로 3개 저널이 선정되었다. 1군에서는 AMJ = *Academy of Management Journal*, ASQ = *Administrative Science Quarterly*, JAP = *Journal of Applied Psychology*, 2군에서는 HR = *Human Relations*, JOB = *Journal of Organizational Behavior*, JOM = *Journal of Management*, 3군에서는 JSBM = *Journal of Small Business Management*, GOM = *Group and Organization Management*, SAM = *SAM Advanced Management Journal*이 각각 선정되었다.

V. 연구설계상 타당성 증진방안

1. 내적 타당성 증진방안

앞서 고찰한 바와 같이, 내적 타당성이란 종속변수의 변화가 순수하게 연구자가 의도한 독립변수의 변화에 의한 것인지 여부를 의미하며, 이러한 인과관계를 입증하기 위해서는 원인의 시간적 선행, 인과요소의 공변, 대안설명의 배제와 같은 일련의 조건들을 충족시켜야 한다(Campbell & Stanley, 1963). 이에 의거하여 본 연구에서는 횡단연구의 대안설명배제 방안으로서 통제변수를 포함시킬 때 이에 관련된 이론적 설명이 제대로 이루어지고 있는지의 여부를 살펴보았으며, 인과추론에 있어 원인의 시간적 선행이 검증되었는지 살피기 위하여 역인과성에 대한 고려가 있었는지의 여부를 확인하였다. 이제부터는 앞서의 분석결과를 바탕으로, 연구의 내적 타당성 증진방안에 어떠한 것이 있을지 알아본다.

1) 연구설계 유형

연구결과에서 살펴본 것처럼, 지난 10년 동안 *인사·조직연구*에 게재된 연구논문 중 2/3에 이르는 숫자가 횡단연구 방식을 선택하였다. 이러한 경향은 실험연구가 갖는 외적 타당성과 일반화 가능성에 대한 의구심에서 기인하며, 이들의 가치가 경영학계에서 상대적으로 평가절하되는 분위기와도 무관하다 할 수 없다(Grant & Wall, 2009; Highhouse, 2009; Scandura & Willams, 2000). 하지만 비록 현장연구라 하더라도 현실적으로 대부분 임의적 표본추출에 의존하는 상황에서는 외적 타당성 측면에 있어서 실험연구에 비해 우위를 지닌다고 이야기하기 어려운 것이 현실이며, 이를 반영하듯 대부분 연구의 한계점으로 제한된 일반화 가능성을 스스로 지적하고 있다.

이러한 일련의 문제를 극복하기 위한 방안으로 Grant & Wall(2009)은 준실험설계에 대한 제안을 한 바 있으며, 앞에서 이야기한 것과 같이 여러 가지 연구설계 방식을 병행함으로써 주장의 타당성을 높이는 방안이 모색되고 있다(Scandura & Willams, 2000). 연구자들은 현장연구와 실험연구를 병행하거나 기존의 횡단연구에서 종단연구로 확장시키는 등, 다양한 방식을 통해 타당성 증진을 꾀할 수 있을 것이다.

2) 통제변수

본 연구결과에서도 드러난 바와 같이 대부분의 실증연구에 통제변수가 포함됨에도 불구하고 이러한 통제변수에 대한 이론적 설명이 기술된 경우는 분석대상의 절반에도 미치지 못하는 36편에 불과하였다. 이러한 결과는 통제변수를 대안의 배제라는 뚜렷한 목적을 가지고 이론적 근거를 기반으로 포함하였다기보다는, 상당수 연구자들이 심도 깊고 고려하지 않은 상태에서 기계적으로 통제변수를 포함시키고 있는 것은 아닌가 하는 우려를 불러일으킨다.

연구자들은 통제변수에 대한 충분한 이론적, 경험적 근거를 제시해주어야 하며, 이를 통해 각각의 변수가 연구에 가지는 의미를 독자들에게 충분히 설명해주어야 할 것이다. 이러한 작업은 횡단연구를 채택할 경우 그 속성상 근본적으로 직접적인 인과관계의 입증에 어렵다는 점을 고려해볼 때, 연구의 타당성 증진을 위해 더욱 강력히 요청되는 사안이라 할 수 있겠다. 앞에서 언급하였듯이 무차별적인 통제변수의 포함은 2종 오류의 원인이 될 수 있으며(Spector et al., 2000), 제3의 통제변수가 두 변수 중 하나와만 상관관계를 가질 때에는 반대로 1종 오류의 가능성이 증가할 수 있으므로(Becker, 2005), 통제변수는 연구에서 관성적으로 당연히 포함되어야 할 대상이 아닌 주 변수 이상으로 충분한 근거와 숙고를 통해 포함되어야 할 대상으로 인식되어야 할 것이다.

3) 역인과성

이미 언급한 바와 같이, 횡단연구는 원칙적으로 시간의 선행성을 고려하지 않기에 인과추론이 불가능하다(Mitchell, 1985). 하지만 그렇기 때문에 더더욱 연구자들은 자신이 생각하는 인과의 방향을 명확히 밝혀야 하며, 역인과성의 가능성이 낮음을 밝힘으로써 그들 자신의 연구결과가 타당함을 입증할 수 있어야 한다. 그러나 본 연구결과에 따르면, 2/3에 육박하는 연구들이 역인과성에 대하여 언급조차 하지 않았으며 나머지도 토의 및 한계점, 향후 제언에서 간단히 언급하였을 뿐, 객관적인 역인과성 검증이 이루어진 연구는 1%에 불과하였다. 특히, 한계점으로 역인과성에 대하여 언급한 경우도 대부분 이에 대한 대안으로 종단연구가 필요함을 기계적으로 언급하는 양상을 보였는데, 이는 진지한 한계점에 대한 고찰 및 숙고라기보다는 연구방식이 근본적으로 가진 한계점에 대한 변(辨)에 가까웠다.

하지만 횡단연구가 가지는 이러한 근원적 한계점에도 불구하고, 이를 극복하고자 노력

한 연구들도 소수지만 분명히 존재하였다. 경로분석이나 구조방정식모형 등의 통계적 기법을 활용하여 독립변수와 종속변수 간의 관계 방향을 상이하게 설정함으로써 각각의 적합도를 비교분석해 역인과성을 검증하는 방식이 있으며, 실제로 본 연구에서 역인과성을 객관적으로 검증한 것으로 분류된 연구들은 모두 구조방정식모형을 활용하였다. 하지만 통계적 사후 검증 외에도 횡단연구와 실험연구를 병행하는 것과 같이 상이한 연구유형을 결합시킴으로써 역인과를 고려하는 방안이 필요하다는 주장이 제기되고 있으며(Meijer, Paulien, Verloop, & Beijaard, 2002; Scandura & Willams, 2000), 실제로 이러한 방식을 활용하여 타당성 증진을 꾀하는 인사조직 분야의 연구들이 영미권 학술지에서 점차 많이 발간되는 추세를 보이고 있다. 이처럼 역인과성을 극복하기 위해선 국내에서도 하나의 연구주제에 대해 다양한 방식으로 접근하는 자세가 필요할 것이다.

2. 구성 타당성 증진방안

위에서 언급하였듯이, 구성 타당성을 확보하기 위해선 우선 수렴 타당성과 판별 타당성을 분석하여야 하며 나아가 동일방법편의를 극복하여야 한다. 구체적으로는 어떠한 방안들이 있는지 알아보도록 한다.

1) 수렴 타당성 및 판별 타당성 분석

우선, 수렴 타당성과 판별 타당성을 평가하는데 널리 쓰이는 방법 중 하나로 탐색적 요인분석(explanatory factor analysis, EFA)이 있다. 수렴 타당성과 판별 타당성이 담보된 경우, 만일 하나의 구성개념에 대한 다항목 측정도구(multi-item measure) 결과를 가지고 요인분석한다면 모든 항목은 하나의 공통된 요인으로 수렴되어야 하고 반대로 만약 상이한 구성개념들의 항목들을 요인분석한다면 다수의 서로 분리된 요인들이 도출되어야 한다(Brahma, 2009).

둘째로, 구성 타당성 측정에 대해 Campbell & Fiske(1959)가 언급한 원론적인 방법인 MTMM이 있다. 이는 앞서 설명한 탐색적 요인분석보다는 좀 더 엄격한 검증 방식으로, 동일한 개념을 서로 다른 방법으로 측정하였을 때 방법 간 상관관계가 높게 나오는 경우 수렴 타당성이 있다고 판단하고, 상이한 개념들을 동일 방법으로 측정하였을 때 개념 간 0 혹은 낮은 상관관계가 나온 경우에 판별 타당성이 있다고 판단한다(Campbell &

Fiske, 1959; Peter, 1981).

셋째, 확정적 요인분석(confirmatory factor analysis, CFA) 방식이 있다. MTMM의 경우 구성개념의 특성에 따라 다양한 방식으로 측정하는 것이 불가능할 수도 있어 MTMM의 적용 자체가 어려울 때가 있다. 이러한 제한적 적용 문제를 극복하기 위해 나온 방식이 CFA이다. EFA와 달리 CFA는 연구자가 직접 구성개념과 항목을 임의대로 연결시켜 모형의 적합도를 살펴보고, 나아가 다른 대안적 모형과 비교하는 것을 가능케 함으로써 좀 더 정밀한 구성 타당성 검증이 이루어질 수 있게 한다. 동일한 구성개념을 측정하였다고 기대되는 항목들을 하나의 요인에 할당한 경우의 모형적합도가 2개 혹은 다수의 요인에 할당한 경우의 적합도보다 통계적으로 유의미하게 크다면 수렴 타당성이 있다고 해석한다. 또한, 상이한 구성개념을 대표하는 항목들을 다수의 요인에 나눠서 할당하고 그 때의 적합도가 하나의 요인에 해당 항목들을 모두 할당한 경우나 연구자의 이론과 배치되는 대안적 요인구조를 구성한 경우의 적합도보다 유의하게 크다면 판별 타당성이 있다고 해석할 수 있다(Brahma, 2009). 물론, MTMM의 한계를 극복하기 위해 CFA가 제안되었지만, 이 역시 무선오류와 측정도구분산(measure-specific variance)을 분리해 낼 수 없다는 단점과 특질(trait)과 방법(method) 사이의 상호작용 검증이 불가능하다는 단점을 지니고 있다. 이에 Bagozzi et al.(1991)은 second-order CFA model, hierarchical CFA model, first-order multiple-informant multiple-item model, direct product model 등을 소개하고 있다.

마지막으로, 수렴 타당성과 판별 타당성을 평가하고 증진시키는 데에는 이러한 통계학적 방법 이외에, 사용된 측정도구의 이전 연구에의 적용 사례를 광범위하게 조사하여 분석하는 문헌탐색적 방식 또한 있을 수 있으며, 이는 통계적 방식과 함께 쓰여 서로의 약점을 보완하는 목적으로 활용될 수 있다.

2) 동일방법편의

박원우 외(2007)의 연구에 의하면, 동일방법편의의 문제는 연구설계 단계, 자료수집 단계, 그리고 사후 통계처리 단계에서 각기 서로 다른 방식으로 해결을 꾀할 수 있다.

첫째, 연구설계 단계에서 동일방법편의의 문제를 미연에 제거하는 방법에 대해 알아보자. 동일방법편의는 기본적으로 동일한 응답원천이나 측정환경 하에서 각 변수를 측정하였을 때 발생하는 문제이기에, 이러한 문제를 사전에 제거하기 위해서는 각 변수들을 각

기 다른 응답원천으로 측정하거나, 변수간의 측정시기 혹은 측정환경을 분리하여야 한다 (박원우 외, 2007). 그러나 연구설계 단계에서의 사전 제거 방법은 자기보고법이 가장 효과적인 측정방법일 경우(박원우 외, 2007) 혹은, 측정 시간과 비용의 제한을 많이 받는 경우(Podsakoff et al., 2003)에는 사용하기 어렵다는 단점이 있다.

둘째, 자료수집 단계에서는 설문지 작성 시 개별항목 또는 질문 전체의 흐름에 특별한 주의를 기울이는 방안이 있을 수 있다. 자세히 말해, 질문의 순서에 주의를 기울임으로써 일관성동기 같은 항목 간 영향을 최소화하고, 간결, 명확하고 구체적인 질문을 하도록 하며, 독립변수와 종속변수를 측정하는 척도의 종류와 척도점의 강도 등을 달리할 수 있다(박원우 외, 2007).

마지막으로, 통계처리 단계에서는 연구변수가 관찰변수인가 잠재변수인가에 따라, 또 잠재적으로 동일방법편의의 요인이 될 만한 변수를 따로 측정했는지 여부에 따라 다양한 통계기법을 상이하게 적용할 수 있다. 이는 박원우 외(2007)의 연구에서 자세히 밝히고 있으므로, 여기에서는 상황에 따른 통계처리 기법에 대한 분류를 재인용하는 선에서 마무리하고자 한다.

〈표 3〉 상황에 따른 적절한 동일방법편의 문제 해결기법

			특정 동일방법편의 측정여부	
			측정	비측정
변수의 수준	관찰변수		특정 편의변수의 편제거 기법	일반 편의변수의 편제거 기법 마커변수의 편제거 기법
	잠재 변수	non-MTMM	직접적으로 측정한 잠재방법요인 통제 기법	직접적으로 측정하지 않은 잠재방법요인 통제 기법
		MTMM	다특질측정 기법	일반 CFA 기법 CU기법 CDP기법

* 자료원: (박원우 외, 2007, p. 111 〈표 5〉)

동일방법편의의 해결은 연구설계 단계와 자료수집 시 측정단계, 그리고 사후 통계처리 단계 모두에 걸쳐 포괄적으로 검토하는 것이 세계적인 추세이므로 가능한 한 모든 단계에 걸쳐 동일방법편의 극복을 위해 각별한 주의를 기울여야 하겠다(Podsakoff et al., 2003; 박원우 외, 2007).

3. 통계적 결론 타당성 증진방안

1) 신뢰도

앞에서 설명하였듯 연구에 있어서 신뢰도는 내적 일관성(주로 Cronbach's alpha로 측정되는) 뿐만 아니라 여타의 다양한 의미의 일관성 개념을 포괄하며, 그러한 다양한 신뢰도 정보가 함께 제시될 때 더욱 풍부한 정보가 도출될 수 있다(Cronbach, Glaser, Nanda, & Rajaratnam, 1972). 여기서는 내적 일관성 방식과 구별되는 두 가지 신뢰도 측정 방식을 추가로 소개한다.

Magnusson(1967)은 신뢰도를 측정하는 다양한 방식을 제시하였는데 그 중 하나가 검사-재검사 신뢰도이다. 검사-재검사 신뢰도를 도출하기 위해 연구자는 두 번에 걸쳐 동일한 집단의 사람에게 동일한 측정방식을 적용하는데 이 때 두 번의 측정결과간의 상관인 검사-재검사 신뢰도이다. 이 검사의 목적은 시간에 걸친 측정의 일관성을 알아보는 것인데 이 때 두 번의 검사 간 간격은 검사의 특성과 측정 목적에 따라 상이할 수 있다. 검사 간 간격은 기억 효과가 사라질 것이라 기대되는 정도로 길게 설정되어야 하며 측정대상의 본질적 특성이 변화하지 않을 정도로 짧아야 한다. 예를 들어 아동 단어 능력 검사의 신뢰도를 측정할 때 검사-재검사 간격을 지나치게 길게 설정한다면 설사 실제로 해당 검사가 신뢰성을 가지고 있을지라도 검사-재검사 상관계수는 작게 나올 수밖에 없다. 아동의 단어능력 발달은 매우 빠르게 이뤄지는 것이 보통이기 때문이다.

검사-재검사 신뢰도는 꼭 두 번에 제한되어야만 할 필요는 없으며, 여러 번에 걸쳐서 측정을 시행하고 그 결과의 평균적인 양상을 살펴보는 것 또한 가능하다. 예를 들어 처음 시행에 있어서 측정치의 개인 순위를 관찰하고 두 번째, 세 번째 등의 시행에서도 여전히 개인 순위가 첫 시행 시와 비슷하게 유지된다면 해당 측정도구를 시간의 흐름에도 불구하고 신뢰할 수 있는 것이다(Kerlinger & Lee, 2000).

검사-재검사를 수행함에 있어 횡단연구라 하여 재검사를 수행하지 않는 것이 아니며 또한 재검사를 수행한다하여 곧바로 종단연구가 되는 것은 아님을 유념해야 한다. 종단연구에서 검사를 다시 시행하는 것은 변화를 관찰하기 위한 목적이지만, 횡단연구의 검사 신뢰도를 위한 재검사는 검사도구가 대상의 일관된 특성을 포착하기에 적절한가를 알아보기 위한 것이다. 당연히 이 때 위에서 언급했듯이 목적에 알맞은 정도의 시간간격이 요구되고 이 간격은 종단연구가 요구하는 시간간격과는 매우 다를 수밖에 없다. 따라서

횡단연구를 시행한다 하여 검사-재검사 신뢰도를 배제하는 것이 정당화되는 것은 아니다.

소개할 신뢰도 중 다른 하나는 동형검사 신뢰도이다. 동형검사 신뢰도를 구할 때는 하나의 동일한 항목 목록에서 두 개의 대등하지만 동일하지는 않은 개별 측정도구가 개발된다. 각 측정대상자들은 그 두 개의 측정 모두에 참가하게 되며 결과적으로 각 개인은 두 개의 측정치를 보고하게 된다. 이 때 그 두 개 점수의 상관이 바로 동형검사 신뢰도이다. Cronbach's alpha의 경우 기 확정된 하나의 측정도구 내부의 일관성을 검토하는 반면 이 방식은 잠재적으로 사용 가능한 전체 항목 집단의 총체적 일관성 혹은 유효성을 검토한다. 이 방식은 또한 위의 검사-재검사 신뢰도와 달리 두 번의 측정에 있어서 다른 항목을 이용하므로 기억 효과의 영향에서 비교적 자유롭다. 하지만 여전히 고려해야 할 문제점들이 있는데 두 번의 시행에 있어 먼저 시행한 측정에 비해 나중에 시행한 측정의 경우 피로(fatigue) 등의 문제가 생길 수도 있으므로(Kerlinger & Lee, 2000), 상황에 따라 각 개인 피험자 별로 측정도구 적용순서의 균형화전략이 필요할 수 있다.

2) 표본크기

표본은 너무 작게 설정될 경우 믿을만한 통계적 결론 타당성을 제공하지 못하고 너무 크게 설정될 경우엔 효율적인 연구에 있어 걸림돌이 된다. 따라서 자신의 연구목적과 방식, 대상의 특성에 따라 적절한 표본크기를 설정하는 것이 매우 중요하다. 일반적으로 요구되는 표본크기는 위에서 언급했듯이 무작위 표본추출이 전제될 때 최소 30이 넘어야 하며 변수가 많아질수록 늘어나야 한다. 또한 SEM이나 HLM등 다양한 통계적 방법론에 따라 요구되는 표본크기도 각기 다르다. 하지만 기본적으로 적용할 수 있는 하나의 가이드라인을 Kerlinger & Lee(2000)가 제공하고 있다. 그들이 제시한 식은 아래와 같다.

$$n = \frac{z^2 \sigma^2}{d^2} \quad (1)$$

이 때 n은 요구되는 표본크기를 뜻하고 Z는 α 값에 대응되는 표준점수이다. 예를 들어 해당 연구에서 설정하고자 하는 α 값이 .05라면 Z는 1.96이 된다. σ 는 모집단의 표준편차로서 모집단의 모수값을 알지 못하는 경우 과거의 연구를 바탕으로 추정된 값을 적용하게 된다. 이것은 하나의 연구를 수행할 때 이전 연구의 결과를 고려해야만 적절한 표

본크기를 정할 수 있다는 것을 의미한다. d 는 연구자가 감수할 수 있는 결과치의 실제값으로부터의 편차로서, 예를 들어 평균값을 가지고 연구를 할 때 연구자가 실제 모집단의 평균값에서 표본의 평균값이 .2 정도 차이 나는 것은 인정하고 넘어갈 수 있다고 생각한다면 이 때 d 는 .2로 설정하게 된다.

그러나 위의 식은 어떤 상황에서나 적용할 수 있는 것은 아님을 명심해야 한다. 각 연구방법론에 따라 고려되어야 할 상황들은 상이할 것이며 구체적인 분석방식마다 각기 독자적인 기준들이 존재할 것이다. 중요한 점은 표본 추출을 시행할 때 경험법칙(rule of thumb)에 기대지 말고 합리적인 근거에 의거하여 적절한 표본크기를 도출해야 한다는 것이다.

3) 복수의 표본추출

앞에서 언급하였듯이 회귀분석에 기반한 분석 방식들은 과적합 문제에 특히 취약하다. 따라서 회귀분석 결과의 신뢰성을 검증하기 위해서는 단지 하나의 표본구조만으로는 부족한 것이 사실이다. 이러한 문제를 해결하기 위해 복수의 표본추출이 요구된다. 이 때 복수의 표본추출이란 이미 가지고 있는 표본의 구조를 변형하여 여러 번 검증해 본다는 의미의 복수일 수도 있고, 모집단에서 여러 번 표본추출을 거친다는 의미의 복수일 수도 있다. 물론 연구자의 상황이 허락한다면 이 두 가지의 복수 표본추출을 모두 적용할 수도 있겠다.

우선 손쉽게 사용할 수 있는 방법 중의 하나는 hold-out 표본 방식이다. 이 방식에 따르면 연구자는 자신이 가진 표본 중 일부를 hold-out 표본으로 설정하고 나머지 표본에 기반하여 회귀식을 도출한다. 그리고 이 회귀식을 가지고 미리 설정해 두었던 hold-out 표본의 데이터를 예측해보고 이때의 회귀 결정계수 값(R_c^2)을 검토한다. 이 때 도출된 R_c^2 이 회귀식의 기존 R^2 에 비해 유의하게 작으면 연구자가 가진 회귀식의 과적합을 의심하게 된다. R_c^2 의 크기를 가늠하기 위해 Nicholson(1960)이 도출한 Herzberg식을 이용할 수 있다. 만약 원래의 회귀식이 타당하다면 hold-out 샘플에 적용하여 구한 R_c^2 이 아래의 Herzberg식의 결과값보다 커야 한다.

$$\widehat{R}_c^2 = 1 - \left(\frac{N-1}{N-K-1} \right) \left(\frac{N+K+1}{N} \right) (1-R^2) \quad (2)$$

이 때 N 은 전체 표본의 크기이고 K 는 hold-out 표본의 크기이다. Hold-out 표본에 포함될 관측치들을 정하는 방식은 다양할 수 있지만 전체 표본에서 무작위 표출하는 것이 가장 일반적이다.

둘째, 표본을 재구성하기 위한 또 다른 방식으로 bootstrapping이 있다. 위의 hold-out 방식이 원래 가지고 있는 표본을 단순히 분리하여 이용하는 것과 달리, bootstrapping 방식에서는 N 의 크기를 가진 원 표본에서 N 번의 복원추출을 통해 원 표본과 같은 크기를 가진 새로운 표본을 도출하고 이 방식을 여러 번 반복한 결과 만들어진 여러 개의 새로운 표본을 가지고 분석을 진행한다. Bootstrapping을 통해 얻어진 다수의 표본들로부터 각각 회귀식을 도출하였을 때 일관된 회귀식들을 도출해 낸다면 과적합의 위험이 작다고 판단할 수 있다(Efron & Gong, 1983).

셋째, 교차검증(cross-validation) 방식이 있다. 교차검증 방식에서는 가지고 있는 표본을 둘로 나눈 후 각각의 반 쪽 표본에 대해 회귀식을 따로 구한 후 도출된 회귀식으로 다른 쪽의 데이터를 예측해 본다. 이를 통해 도출된 결정계수 값으로 과적합을 가늠하게 되는데 표본을 나누기 전의 R^2 에 비해 교차 검증을 통해 얻어진 R^2 가 유의하게 작다면 과적합을 의심하게 된다.

앞의 방식들은 이미 가지고 있는 표본의 구조를 변형시켜 복수의 하위표본을 구성하고 그를 통해 과적합을 검사하는 방식들이다. 하지만 연구자의 여건이 뒷받침된다면 아예 모집단으로부터 표본추출을 추가로 시행하여 회귀식을 따로 구해보고 원 표본의 회귀식과 비교해보는 방안이 가능하다. 이 방식은 단순히 통계적 결론 타당성뿐만 아니라 외적 타당성 또한 증진시켜주는 방식이라 하겠다.

4. 외적 타당성 증진방안

본 연구결과에 따르면 외적 타당성에 대한 고려가 다른 타당성에 대한 고려에 비해 특히 취약하였다. 이에 외적 타당성을 증진시키기 위해 어떠한 노력을 기울여야 할지 알아보도록 한다.

1) 무작위 표본추출

연구자는 적절한 표본추출 방식을 통해 표본의 대표성을 확보하여 연구결과의 일반화

가능성을 높일 수 있다. 이를 위해 무작위 표본추출이 권장되는데 이는 모집단으로부터 표본의 각 요소를 동일한 확률로 선택하기 위해 사용하는 방식이다(Kerlinger & Lee, 2000). 무작위 표본추출을 실행하기 위해 연구자는 우선 모집단을 구체적으로 정의하고 이를 통해 적절한 샘플프레임을 확보하여야 한다. 여기서 샘플프레임이란 모집단 구성원 전체가 기재된 일종의 목록을 말한다. 표본추출 확률을 동일하게 유지하려면 표본단위에 대한 확률을 알 수 있어야 하는 만큼 무작위 표본추출에는 모집단에 대한 정보와 그에 기반한 샘플프레임이 반드시 필요하다.

샘플프레임을 확보한 후에는 난수표 등을 이용하여 무작위 방식으로 샘플프레임에서 표본을 추출하도록 한다. 적절한 샘플프레임을 선정했다 하더라도 무작위 추출을 하지 않고 모집단에서 연구자가 접근하기 쉬운 대상을 표본으로 선정하거나 혹은 자발적 응답자를 지원받는 식으로 표본을 선정하게 되면 대표성이 확보되지 않아 외적 타당성이 저해된다.

이처럼 무작위 표본추출 방식이 외적 타당성을 높이는 가장 이상적인 표출 방식임은 확실하나 종종 시간이나 비용 등의 문제로 사용되지 않을 수 있으며(Cook & Campbell, 1963), 실험연구가 아닌 이상 실제 조직에서 무작위 표본추출을 하기가 쉽지 않기 때문에 사실상 인사·조직 분야의 연구에서 무작위 표본추출 방식이 흔하게 이용되지는 않는다. 본 연구결과에 따르면 무작위 표본추출을 채택한 연구는 11%로 나타났는데 이러한 수준은 비단 국내뿐 아니라 해외 저널에서도 마찬가지로, Drabeck, Braitto, Cook, Powell & Roger(1982)의 연구에서는 12%, Short, Ketchen, & Palmer(2002)의 연구에서는 15% 미만의 수준을 보고한 바 있다.

하지만 무작위 표본추출을 채택하지 않은 연구의 경우 그 결과의 신뢰성이 떨어지고 인과적 추론도 어려워지는 만큼(Cook & Campbell, 1979), 연구자는 무작위 표본추출을 실행하고자 최대한 노력해야 할 것이다. 또한 불가피하게 임의표출 등의 비확률 표출을 채택하는 경우에는 해당 표본추출 방식 선정에 대한 이유를 논문에 반드시 포함시켜야 할 것이다(Kerlinger, 1986).

2) 무응답 오류

무응답 오류는 응답자와 무응답자 사이에 유의한 차이가 있을 때 나타나는 문제로 이에 대한 평가는 표본의 대표성에 있어 중요한 정보를 제공해 준다(Werner et al.,

2007). 응답률이 70%에 달하는 연구라 해도 무응답 오류로부터 완전히 자유로울 수는 없으며 무응답 오류의 존재와 수준 및 방향에 관한 정보가 없다면 연구자는 응답률이 높다는 이유로 무비관적으로 연구결과를 받아들여서도, 응답률이 낮다는 이유로 무조건 결과를 무시해서도 안 된다(Roldberg & Stanton, 2007). 따라서 외적 타당성 확보를 위해서는 응답률에 상관없이 무응답 오류를 평가해야 할 것이다.

Roldberg & Stanton(2007)은 무응답 오류를 검증하는 방법으로 모두 9가지를 제시하였으나 여기서는 보편적으로 사용되는 3가지 방식을 소개하기로 한다. 첫째는 기록 데이터 분석(archival analysis) 방식이다. 전체 설문표본을 포함한 기록데이터를 활용하는 방법으로, 인사기록이나 멤버십정보 등이 기록데이터에 해당한다. 응답자와 무응답자의 기록을 비교하여 차이를 확인하는데 이 때 차이가 존재한다고 해서 반드시 무응답 오류가 발생하는 것은 아니며, 해당 기록데이터의 차이가 이론적인 견지에서 연구의 관심이 되는 설문응답과 체계적으로 관련이 있을 때만 오류에 해당한다.

둘째는 사후접근(follow-up approach) 방식이다. 사후접근 방식은 응답자와 무응답자 일부를 무작위로 선정해 재설문 조사하는 방식으로, 사후표본으로 수집된 자료는 관심변수의 응답자와 무응답자간의 의미 있는 비교를 가능하게 한다.

셋째, 웨이브분석(wave analysis) 방식이다. 시한 내에 회수된 설문과 이후에 회수된 응답자 설문을 비교하는 방식으로, 무응답자나 모집단 자료를 수집할 수 없는 경우 사용될 수 있는 방식이다. 초반 응답자와 후반 응답자간에 차이가 있으면 무응답 오류가 존재할 수 있음을 나타낸다.

이 외에 응답자와 무응답자의 체계적 차이를 파악하기 위해 개인의 연구주제에 관한 관심도 및 연구참가의 자발성 등을 설문항목으로 포함시키는 방식도 있다.

위에서 제시한 무응답 오류에 관한 구체적인 검증 없이는 표본추출이 얼마나 연구결과 일반화 가능성에 영향을 미쳤는지를 판단할 수 없으므로 연구자들은 위의 방식들을 통해 무응답 오류 분석을 실시하고 그 결과를 제시해야 하며, 학술지 심사위원(journal reviewer) 또한 이의 실행 여부를 심사기준으로 포함시켜야 할 것이다(Werner et al., 2007).

VI. 맺음말

지금까지 본 연구는 연구설계상 내적 타당성, 구성 타당성, 통계적 결론 타당성 및 외적 타당성의 정의를 알아보았고 그에 따라 타당성 평가의 기준이 될 수 있는 항목을 선정, 이를 바탕으로 *인사·조직연구*에 지난 10년간 발표된 논문에 대한 평가를 실시하였다. 평가 결과, 표본크기 및 수렴 타당성 및 판별 타당성에 대한 고려 등과 같은 항목에 있어서는 대다수의 논문들이 충분한 인식과 대처를 보이고 있었으나, 무작위 표본추출과 무응답 오류 등과 같은 항목에 대해서는 절대 다수가 언급조차 하지 않고 있음을 알 수 있었다.

이러한 결과를 바탕으로, 향후 횡단연구의 타당성 증진을 위해 다음과 같이 제언한다.

- 통제변수를 포함할 경우 이에 대한 충분한 이론적, 경험적 근거를 제시해야 한다. 무차별적인 통제변수의 포함이 2중 오류를 낳을 수 있고(Spector et al., 2000), 제 3의 통제변수가 연구대상인 두 변수 중 하나와만 상관을 지닐 때는 반대로 1중 오류의 가능성이 증가될 수도 있으므로(Becker, 2005) 통제변수란 당연히 포함해야 할 대상이 아닌 연구변수 이상의 충분한 근거와 숙고 끝에 포함되어야 할 대상으로 인식해야 한다.
- 횡단연구의 경우 반드시 역인과성에 대한 고려를 위해 사후 검증이나 추가적인 문헌연구 등을 실시해야 한다. 실험, 준실험, 종단, 메타연구에 비해 횡단연구는 인과성추론에 있어 취약하므로, 특별한 주의와 추가적인 검증과정이 필요하다.
- 수렴 타당성과 판별 타당성에 대해서는 기본적으로 요인분석으로 검증하고, 추가적으로 이전 연구에서 구성 타당성 정보를 인용하도록 한다.
- 동일방법편의에 대한 고려는 세계적인 추세이므로 반드시 극복하려 노력한다.
- 회귀분석을 실시할 경우 과적합을 방지하기 위해, 복수의 표본을 사용한다.
- 표본의 대표성을 보장하기 위하여, 지금까지 암묵적으로 시행해 온 임의표출이 아닌, 무작위 표본추출을 하도록 한다. 또한, 무응답 오류에 대해서도 충분히 고려함으로써 외적 타당성을 증진시켜야 한다.

본 연구에서 도출한 평가기준이 모든 연구자, 검토자, 편집자에게 적용되는 기계적이고 일률적인 가이드라인일 수는 없으며 소개된 타당성 증진방안들을 모든 연구에 무조건 적

용할 수도 없을 것이다. 그러나 본 연구의 결과를 유념하고 소개된 타당성 증진방안들을 일차적으로 고려하는 것은 연구의 전반적 타당성 향상을 위한 첫걸음이 될 수 있을 것이며, 각 연구자들은 이를 바탕으로 개개의 연구목적에 합치하는 타당성 증진방안을 모색해야 할 것이다.

참 고 문 헌

- 권석규. 2004. 인사조직분야의 연구방법론: 현황과 과제. 「한국인사·조직학회 발표논문집」(한국인사조직학회), 213-227.
- 박원우, 김미숙, 정상명, & 허규만. 2007. 동일방법편의(Common Method Bias)의 원인과 해결방안. 「인사·조직연구」(한국인사조직학회), 15(1): 89-133.
- 이진규 & 박기찬. 1993. 한국 인사관리 연구의 동향과 과제. 「인사·조직연구」(한국인사조직학회), 2(1): 75-137.
- Anderson, C. A., & Bushman, B. J. 1997. External validity of "trivial" experiments: The case of laboratory aggression. *Review of General Psychology*, 1(1): 19-41.
- Bagozzi, R. P., Yi, Y., & Phillips, L. W. 1991. Assessing construct validity in organizational research. *Administrative Science Quarterly*, 36(3): 421 - 458.
- Becker, E. T. 2000. Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. *Organizational Research Methods*, 8: 274-289.
- Berkowitz, L., & Donnerstein, E. 1982. External validity is more than skin deep: Some answers to criticisms of laboratory experiments. *American Psychology*, 37(3): 245-257.
- Brahma, S. S. 2009. Assessment of construct validity in management research: A structured guideline. *Journal of Management Research*, 9(2): 59-71.

- Campbell, D. T., & Fiske, D. W. 1959. Convergent and discriminant validation by the multitrait - multimethod matrix. *Psychological Bulletin*, 56(2): 81-105.
- Campbell, D. T., & Stanley, J. C. 1963. *Experimental and quasi-experimental designs for research on teaching*. Chicago: Rand McNally.
- Cook, T. D., & Campbell, D. T. 1976. The design and conduct of quasi-experiments and true experiments in field settings. In M. Dunnette (Ed.), *Handbook of industrial and organizational psychology*, 223-326. Skokie, IL: Rand McNally.
- Cook, T. D., & Campbell, D. T. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin.
- Cronbach, L. J., Glaser, G. G., Nanda, H., & Rajanratnam, N. 1972. *The dependability of behavioral measurements: Theory of generalizability for score sand profiles*. New York: Wiley.
- Cronbach, L. J., & Meehl, P. E. 1955. Construct validity in psychological tests. *Psychological Bulletin*, 52(4): 281-302.
- Dipboye, R. L., & Flanagan, M. F. 1979. Research settings in industrial and organizational psychology: Are findings in the field more generalizable than in the laboratory? *American Psychologist*, 34(2): 141-150.
- Drabeck, T. E., Braito, R., Cook, C. C., Powell, J. R., & Rogers, D. 1982. Selecting samples of organizations: Central issues and emergent trends. *Pacific Sociological Review*, 25: 377 - 400.
- Drasgow, F., & Miller, H. E. 1982. Psychometric and substantive issues in scale construction and validation. *Journal of Applied Psychology*, 67 (3): 268-279.
- Fiske, D. W. 1982. Convergent-discriminant validation in measurements and research strategies. In D. Brinberg & L. H. Kidder (Eds.), *Forms of validity in research*, 77-92. San Francisco: Jossey-Bass.
- Grant, A. M., & Wall, T. D. 2009. The neglected science and art of

quasi-experimentation: Why-to, when-to, and how-to advice for organizational researchers. *Organizational Research Methods*, 12(4): 653-686.

Guion, R. M., & Granny, C. J. 1982. A note on concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied Psychology*, 67: 239-244.

Hair, J. F., Jr., Black, W. C., Babin, B. J., & Anderson, R. E. 2010. *Multivariate data analysis*. Upper Saddle River, NJ: Pearson Prentice Hall.

Highhouse, S. 2009. Designing experiments that generalize. *Organizational Research Methods*, 12(3): 554-566.

Hunter, J. E., & Schmidt, F. L. 2004. *Methods of meta-analysis*. Beverly Hills, GA: Sage.

Kerlinger, F. N. & Lee, H. B. 2000. *Foundations of Behavioral Research*. New York: Harcourt College Publishing.

King, W. R., & He, J. 2005. External validity in IS survey research. *Communications of the Association for Information Systems*, 16: 880-894.

Lynch, J. G., Jr. 1982. On the external validity of experiments in consumer research. *Journal of Consumer Research*, 9(3): 225 - 239.

Magnusson, D. 1967. *Test theory*. Reading, MA: Addison-Wesley.

Meehl, P. E. 1971. High school yearbooks: A reply to Schwartz. *Journal of Abnormal Psychology*, 77: 143-148.

Meijer, P. C., Verloop, N., & Beijaard, D. 2002. Multi-method triangulation in a qualitative study on teachers' practical knowledge: An attempt to increase internal validity. *Quality & Quantity*, 36: 145-67.

Mitchell, T. R. 1985. An evaluation of the validity of correlational research conducted in organizations. *Academy of Management Review*, 10(2): 192-205.

- Nachmias, C., & Nachmias, D. 1981. *Research methods in the social sciences*. New York: St Martin's Press.
- Nicholson, G. E. 1960. Prediction in future samples. In I. Olkin., S. Ghurye, W. Hoeffding, W. Madow, & H. Mann (Eds.), *Contributions to Probability and Statistics*, 322-330. Stanford, CA: Stanford Union Press.
- Peter, J. P. 1981. Construct validity: A review of basic issues and marketing practices. *Journal of Marketing Research*, 18: 133-145.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88: 879-903.
- Rogelberg, S. G., & Stanton, J. M. 2007. Introduction: Understanding and dealing with organizational survey nonresponse. *Organizational Research Methods*, 10: 195 - 209.
- Sackett, P. R., & Larson, J. R., Jr. 1992. Research strategies and tactics in industrial and organization psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial organizational psychology*, 419-489. Palo Alto, CA: Consulting Psychologists Press.
- Scandura, T. A., & Williams, E. A. 2000. Research methodology in management: Current practices, trends, and implications for future research. *Academy of Management Journal*, 43(6): 1248 - 1264.
- Schmitt, N. W., & Klimoski, R. J. 1991. *Research methods in human resource management*. Cincinnati, OH: South-Western.
- Schwab, D. P. 1980. Construct validity in organizational behavior. In L. L. Gummings & B. Staw (Eds.), *Research in Organizational Behavior*, 2:3-34. Greenwich, GT: JAI Press.
- Schwab, D. P. 2005. *Research methods for organizational studies*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Short, J. C., Ketchen, D. J., & Palmer, T. B. 2002. The role of sampling in

- strategic management research on performance: A two-study analysis. *Journal of Management*, 28(3): 363 - 385.
- Smith, G. T. 2005a. On the complexity of quantifying construct validity. *Psychological Assessment*, 17(4): 413 - 414.
- Smith, G. T. 2005b. On construct validity: Issues of method and measurement. *Psychological Assessment*, 17(4): 396 - 408.
- Spector, P. E., Zapf, D., Chen, P. Y., & Frese, M. 2000. Why negative affectivity should not be controlled in job stress research: Don't throw out the baby with the bath water. *Journal of Organizational Behavior*, 21: 79-95.
- Viswesvaran, C., Barrick, M. R., & Ones, D. S. 1993. How definitive are conclusions based on survey data: Estimating robustness to nonresponse. *Personnel Psychology*, 46: 551-567.
- Werner, S., Praxedes, M., & Kim, H. 2007. The reporting of nonresponse analyses in survey research. *Organizational Research Methods*, 10(2): 287-295.
- Westen, D., & Rosenthal, R. 2003. Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology*, 84(3): 608-618.
- Westen, D., & Rosenthal, R. 2005. Improving construct validity: Cronbach, Meehl, and Neurath's ship. *Psychological Assessment*, 17(4): 409 - 412.
- Willems, E. P., & Howard, G. S. 1980. The external validity of papers on external validity. *American Psychologist*, 35: 387-388.
- Winer, R. S. 1999. Experimentation in the 21st century: The importance of external validity. *Journal of the Academy of Marketing Science*, 27 (3): 349-358.

An Evaluation of Validity of Cross-Sectional Research in Organizational Science and Recommendations for Future Research: Based on Studies in Korean Journal of Management

Won-Woo Park*
Goo Hyeok Chung**
Ji Hyun Yoo**
Dongwon Choi***
Sangyun Kim***

ABSTRACT

In research methodology, validity is one of the most frequently mentioned concepts, validity in terms of measure and validity in terms of research design particularly. The former is a concept of how exactly researchers measure their constructs or attributes and thus, it is related to the degree to which the measures correctly indicate the attributes. More comprehensively, however, the latter, validity in terms of research design, means how exclusively variations in dependent variable(s) are interpreted by those in independent variable(s) in such a way that researchers originally designed and predicted (i.e., internal validity), whether operational definitions and measurements of variables coincide with constructs presented by the researchers (i.e., construct validity), whether the analytic procedure for results is conducted in the statistically correct

* Professor. College of Business Administration. Seoul National University

** Doctoral student. College of Business Administration. Seoul National University

*** Master student. College of Business Administration. Seoul National University

manners (i.e., statistical conclusion validity), and how generalizably other researchers or practitioners can apply the findings to other settings such as different time, places, and subjects (i.e., external validity). Accordingly, unless an empirical study is rigorously conducted in terms of validity, it may spawn critical errors (e.g., the findings can't explain or predict the phenomena in the correct ways).

To improve validities of future research in organizational science, this study is aimed to specify and define validities, based on Cook & Campbell's (1979) research, and it develops generally-accepted scientific criteria with respect to validity. Specifically the authors evaluate them with the criteria noting that most studies published in Korean Journal of Management during the last decade (2000-2009) were cross-sectional. Moreover, not only avoiding non-constructive critique for cross-sectional studies but also suggesting the detailed methods for better validity, the authors present recommendations to improve validity of future researches in the human resources and organizational behavior literature in Korea.