

구조방정식 모델에서 항목묶음이 인과 모수의 검정과 적합도 평가에 미치는 영향

조 현 철* · 강 석 후**

<요 약>

본 연구에서는 3개 일반모델(general models)의 실증분석을 통해 항목묶음(item parceling)이 구성개념 간의 인과관계를 나타내는 모수의 유의성 검정 결과 및 모델의 적합도 평가에 미치는 영향을 검토하였다. 연구 결과에 의하면, 개별항목을 적용한 분석과 비교할 때 항목묶음을 통한 분석을 적용해도 모델 인과모수의 검정통계량은 그다지 변하지 않으므로 유의성 검정 결과에도 변화가 없는 것으로 나타났다. 하지만 전반적 적합도지수의 측면에서는 RMSEA를 제외한 주요 모델 적합도 지수, 즉 GFI, AGFI, CFI 및 NFI의 값들이 상당히 개선되는 경향을 보였다. 주요 모델 적합도 지수들의 값이 높아진 것은 항목묶음을 이용하여 분석을 한 결과가 개별항목을 통한 분석의 결과에 비해 다변량 정규(분포)성의 개선 등으로 인해 높아진 것으로 해석된다. 하지만 항목묶음을 적용함에 있어서 주의해야 할 사항은 적용하기 전에 구성개념의 단일차원성(unidimensionality)을 우선적으로 검토해야 한다는 점이다. 본 연구에서는 항목묶음을 이용하여 분석을 할 경우 실제 구성개념간의 인과적 관계를 나타내는 모수의 유의성 검정과 모델의 적합도 지수들에 어떤 변화가 발생하는지를 세 개의 일반모델을 대상으로 파악하였다.

주제어: 구조방정식모델, 항목묶음, 적합도 지수 개선

I. 서 론

오늘날 심리학, 교육학, 사회학, 마케팅, 소비자행동 등과 같은 사회과학분야 대부분의 연구에서 실증연구를 수행함에 있어 연구자들이 측정하고자 하는 각종 개념(construct) 혹은 구성개념(construct)은 대부분 추상적인 것들이므로 이들을 하나의 측도(measure)로 직접 측정할 수 없는 실정이다. 따라서 하나의 구성개념에 대해 여러 설문항목을 적용하여 이 구성개념을 간접적으로 측정하는 것이 일반적이다(e.g., 고순화·이영선 2007; 김봉관·김태우 2003; 이지현·이동일 2005;

이태민·전종근 2003; 정기한·김대업 2002). 마케팅/소비자 행동 관련 연구에서 널리 사용되고 있는 다수항목 측정척도에 대해서는 Bearden and Netmeyer (1999)의 저서를 참조하기 바란다. 그런데, 다중회귀 분석이나 그 응용 기법들을 통해 이들 구성개념을 대상으로 분석해야 하는 경우에는 각 구성개념별로 다수의 설문항목을 통해 측정한 경우에도 이들을 합산(summing)하거나 평균(averaging)을 구해 단일 지표로 각각 통합하여 사용해야 한다. 이로 인해 각 설문항목에 내재된 측정오차를 분석에 전혀 반영할 수 없다(Bollen 1989; Hall, Snell, and Foust 1999). 구조방정식모델(structural equation

* 한양대학교, 교수 (chohyunc@hanyang.ac.kr)

** 한양대학교, 교수(shkang@hanyang.ac.kr)

modeling)이 가진 주요 이점 중의 하나가 바로 이 측정 오차를 분석에 반영하면서도 다수의 지표를 그대로 사용할 수 있을 뿐만 아니라 각 구성개념에 대한 측정항목들을 통합하여 두 개나 세 개 혹은 하나로 종합하여 사용할 수도 있다는 것이다. 그런데 각 구성개념에 대한 설문항목의 수가 많은 경우 이들 모두를 그대로 해당 구성개념의 지표로 사용하게 되면 모델의 주요 적합도 지표들이 관행적 수용기준인 0.90 이하로 낮아지는 현상이 빈번히 발생하게 되므로 '모델을 수용할 수 없다'는 심각한 문제점에 봉착하는 경우가 생길 수 있는데, 이러한 경우의 한 해결 방법으로 항목묶음(item parceling)이 널리 사용되어 오고 있다.

항목묶음이 최초로 Cattell(1956)의 연구에서 제시된 이래, 일반적으로 구조방정식 모델에서 학문 분야에 따라 차이는 있으나 널리 사용되어 오고 있는 실정이다. 항목묶음은 통상 측정지표가 다수의 항목으로 구성되어 있는 경우 이들 다수 항목을 구성개념별로 둘, 셋, (때로는 하나, 넷) 등으로 묶은 다음 각각 합산하거나 평균을 구해 새로운 항목묶음(parcel)으로 통합하여 사용하는 것이다. Bandalos and Finney(2001)의 고찰에 의하면 1989년부터 이 연구가 행해진 시점까지 *Journal of Educational Measurement*에서는 60%, *Journal of Educational Psychology*에서는 23%, *American Educational Research Journal*에서는 33%, *Applied Psychological Measurement* 25%, *Education and Psychological Measurement*에서는 18%, *Structural Equation Modeling*에서는 13%, *Journal of Marketing Research*에서는 9%

등으로 항목묶음을 이용하여 분석을 실시한 것으로 나타났다. 이러한 관심에 부응하여 항목묶음 자체를 주제로 다루고 있는 논문도 여러 학문 영역에서 상당수 발표되고 있다(e.g., Bandalos 2002; Bandalos and Finney 2001; Hall, et al. 1999; Little, Cunningham, Shahar, and Widaman 2002; Nasser and Wisenbaker 2003).

본 연구의 목적은 다음과 같다: (1) 널리 사용되어오고 있는 항목묶음 관행에 대해 찬성하는 견해와 반대하는 견해의 논점을 분석한다. (2) 항목묶음에 적용할 수 있는 각종 방법에 대해 살펴본다. (3) 항목묶음의 영향에 대해 기존의 실증적 연구는 주로 확인적 요인분석(confirmatory factor analysis) 모델을 대상으로 모델의 전반적 적합도 지수와 해(solution)의 추정 여부에 미치는 영향에 대해 고찰해 왔으나, 본 연구에서는 3개의 일반모델을 대상으로 인과관계를 나타내는 모수의 유의성 검정 결과와 모델의 전반적 적합도에 미치는 영향을 고찰하고자 한다.

II. 항목묶음에 대한 찬성과 반대 견해

항목묶음의 사용이 지지되는 근거와 한계점에 대해 Bandalos and Finney (2001) 및 Bandalos(2002)는 종합적으로 고찰하였는데 이를 요약하기로 한다. 우선, 항목묶음을 사용하는 이유는 다음과 같다: (1) 측도의 신뢰성이 증가한다는 점이다. (2) 보다 연속적이고 다변량 정규분포에 근접하는 분포가

언어 질 가능성이 높아진다는 것이다. (3) 모델이 간명(parsimony)해 지므로 상대적으로 작은 표본일 경우에도 적용이 가능해 진다는 점이다. (4) 모델의 적합도 지수들이 좋아진다는 것이다. 이러한 장점이 있음에도 불구하고 다음과 같은 문제점이 있는 것으로 Bandalos and Finney (2001) 및 Bandalos(2002)는 지적하고 있다: (1) 항목 묶음을 하게 되면 개별항목에 관한 정보가 상실된다는 점이다. (2) 항목묶음은 단일차원성이 확인된 경우에 한해 사용해야 하므로 단일차원성의 확인 없이 적용할 경우 문제점이 발생할 수 있다. (3) 모수 추정치와 요인적재치가 항목묶음에 포함된 특정 항목에 크게 의존할 가능성에 따른 문제점이 발생할 수 있다. (4) 원래의 요인구조가 애매해지거나 모델 모수 추정치에 편의(bias)가 발생할 가능성이 있다.

이들 가운데서 몇 가지 주요 사항에 대해 상세히 살펴보기로 한다. 첫째로, 개별항목을 이용하여 분석을 하는 경우에 비해 항목묶음을 이용하면 측도(measure)의 신뢰성이 더욱 향상되는 결과를 얻을 수 있으며, 좀 더 명확한 회전결과(rotational result)를 얻을 수 있다는 점이 지적되고 있다(Cattell and Burdsal 1975; Kishton and Widaman 1994). 집계 수준의 항목묶음과 개별항목을 비교한 결과 개별항목을 이용하였을 때 다음과 같은 단점이 있는 것으로 나타났다: 신뢰성의 약화, 공유치(communality)의 약화, 정규분포 가정의 위배로 인한 오류 가능성의 증가 등이다(Bagozzi and Heatherton, 1994; Kishton and Widaman, 1994; MacCallum, Widaman, Zhang and Hong, 1999).

둘째로, 최우법(maximum likelihood, ML)

은 특히 구조방정식모델에서 모델평가에 가장 일반적으로 사용되는 추정법의 하나인데(Yuan, Bentler, and Kano 1997; Nasser and Wisenbaker 2003), 이 추정법은 다변량 정규분포의 가정이 충족되어야 하므로 이를 어느 정도 이상 위배하게 되면 문제가 발생한다. 따라서 항목묶음을 사용하면 개별항목을 사용할 경우에 비해 비교적 연속적이고 정규분포에 가까운 특성을 지닌 분포를 얻을 수 있어서 다변량 정규분포를 전제로 하는 최우법의 적용이 가능해 진다(Nasser and Wisenbaker 2003; Thompson and Melancom 1996; West, Finch, and Curran 1995).

셋째로, 표본의 수가 적은 경우의 연구에 특히 유용하다고 지적되고 있다(Bagozzi and Edwards 1998; Bagozzi and Heatherton 1994; Hall et al. 1999; Yuan et al. 1997). 이는 추정 모수에 비해 표본 크기의 비율이 낮은 경우 추정할 모수의 수를 줄여주면 더 안정적인 모수 추정치를 얻을 수 있기 때문이다. Marsh and Hocevar(1988)는 측정항목과 응답자의 비율을 명확하게 고려해야 하며 이 비율이 낮은 경우 요인의 해에 대한 안정성(stability)이 저해되는 것으로 설명하고 있다.

끝으로, 항목묶음의 장점에 대해 연구자들이 일관되게 지적하고 있는 점은 모델의 전반적 적합도 지수들(goodness-of-fit indices)이 향상된다는 것이다(Bagozzi and Heatherton 1994; Bandalos and Finney 2001; Gibbons and Hocevar 1998; Hall et al. 1999; Little et al. 2002; Nasser and Wisenbaker 2003; Takahashi and Nasser 1996).

이런 장점들에도 불구하고 항목묶음에 대한 반론의 핵심은 다음 네 가지로 요약할 수 있다: 첫째로, 가장 중요한 지적사항은 항목

묶음에 사용될 특정 구성개념에 관련된 모든 개별항목이 단일차원성(unidimensionality)을 지니고 있다는 것이 전제되어야 한다는 것이다(Bandalos and Finney 2001; Cattell 1956; Hall et al. 1999). 그럼에도 불구하고 기존의 연구에서 항목묶음을 사용함에 있어 구성개념의 단일차원성에 대해 검토한 결과를 구체적으로 제시한 연구는 드물다고 지적되고 있다(Bandalos and Finney 2001).

둘째로, 항목묶음을 사용하면 원래의 요인 구조가 애매해지거나 가려질 위험이 있다는 점이다(West et al. 1995). 특히, 단일차원성이 확인되지 않은 상태에서 만약 개별항목들이 다차원적 구조를 가지고 있거나 "제삼의 구성개념으로부터 영향을 받고 있는 경우"(Hall et al. 1999)에는 항목묶음을 실시하면 심각한 문제점이 발생할 수 있다. 예컨대, 조직시민행동(organizational citizenship behavior, OCB)이나 *SERVQUAL*과 같이 구성개념이 단일차원이 아니라 다차원적 하위 요인들을 갖는 구조일 경우에는 이들을 구분 없이 항목묶음을 하여 사용하면 구성개념에 내재된 차원이 불명확해질 수 있으므로, 이런 다차원적 구성개념들의 경우 항목묶음을 하나로 하기보다는 하위 요인별로 각각 실시해야 할 것으로 본다.

셋째로, 모수 추정치와 요인적재치가 항목묶음에 포함된 특정 항목에 크게 의존할 가능성에 따른 문제점이 발생할 수 있다. 이 문제를 해결하는 하나의 방법은 특이한 이상치(異常值, outliers)가 존재하는지 검토해 보는 것이다(Bollen 1989 pp. 24-31 and p. 282).

넷째로, 필자들의 경험으로는 항목묶음이 2개 일 때에도 제대로 해(solution)가 구해

지는 경우가 많았으나, 학자에 따라서는 개별항목에 의해 추정된 해에 비해 항목묶음이 3개 미만일 경우 해가 구해지지 않거나(unconverged solutions), Haywood cases와 같은 부적해(improper solutions)가 생길 가능성이 증가하였다는 지적이 있다(Ding, Velicer, and Harlow 1995; Marsh, Hau, Balla, and Grayson 1998; Nasser and Wisenbaker 2003).

그런데, 항목묶음을 사용하여 얻을 수 있는 장점과 단점을 고려할 때 항목묶음을 이용할 것인가의 결정은 연구자가 자신이 수행하는 과학적 연구에 대한 철학적 입장과 연구목적에 따라 달라질 수 있다(Little et al. 2002). 가령 예측과 설명을 주된 목적으로 하는 연구에서는 단일차원성이 전제되는 경우 항목묶음이 허용될 수 있지만, 측도의 불변성(invariance)이나 항등성(equivalence)을 고찰하고자 하는 연구의 경우에는 원래의 연구결과와 후속연구 결과를 비교해야 하므로 후속연구에서만 항목묶음을 사용해서는 안 된다.

그리고 항목묶음 방법과 혼동하지 않아야 하는 기법이 있다. James, Mulaik, and Brett (1982)에 의해 제안된 2단계 모델링 접근법(a two-step modeling approach)이라는 것으로 잠재변수별로 좋은 지표들만 선별하여 남기고 가령 신뢰성을 저해하는 지표는 제외시키는 방법이다(참조: Anderson and Gerbing 1988; Schumacker and Lomax 1996, pp. 72-73). 항목묶음 접근법에서는 사용된 지표를 일반적으로 버리지 않지만, 2단계 모델 접근법에서는 검토과정을 거쳐 일부 지표를 버린다는 점이 근본적으로 다르다.

III. 항목묶음에 적용 가능한 방법

항목묶음 방법에 대해, Little과 그의 동료들(2002)은 크게 세 가지 기법을 제시하고 있다.

첫째는, 무작위 할당(random assignment) 기법으로 각각의 항목을 무작위로 어느 항목묶음 중 하나에 할당하는 방법이다. 항목의 수에 따라 둘, 셋, 혹은 가능하다면 넷의 항목묶음이 생성 가능하다. 물론 하나로 항목묶음을 실시하는 경우도 있다.

둘째는, 항목과 구성개념간의 균형(item-to-construct balance)을 이용한 기법으로 항목묶음들 간에 어느 정도의 균형 예컨대, 구성개념과 항목간의 관계(item-to-construct relation)를 간단히 계산하여 항목묶음들 간의 균형을 유지하게 하는 방법이다. 가령 요인 적재치를 기준으로 3개의 항목묶음을 구성할 경우 적재치가 가장 높은 항목 3개를 항목묶음에 차례대로 할당하고 그 다음으로 높은 적재치를 가진 항목 3개를 앞의 순서와는 반대의 순서로 항목묶음에 배당하는 방법으로 1차 선별과정에서 가장 높은 적재치를 할당받은 항목묶음은 다음 2차 선별과정에서 가장 낮은 적재치를 할당받게 된다. 항목이 다수라면 이러한 할당 과정이 계속된다.

셋째는, Little, Ottingen, and Baltes(1995) 및 Little and Warner(1997)가 제시한 선형적 설문지 구성법(a priori questionnaire construction)이라는 것으로, 예컨대, 3개의 구성개념이 각각 6개의 측정항목으로 이루어져 있다고 가정할 때 이 중 3개의 항목은 부정적인 방향으로 질문을 하고 나머지

3개의 항목은 긍정적인 방향으로 질문을 하여 하나의 부정적인 항목(단, 역전시켜 사용)과 긍정적인 항목을 1개의 쌍으로 묶어 사용을 하는 방법이다. 이를 통해 긍정적 항목과 부정적 항목에 기인한 편향을 상쇄할 수 있는 이점이 있다고 지적되고 있다.

Nasser and Wiesenbaker(2003)도 몇 가지 항목묶음 방법에 대해 언급하고 있다: (1) 무작위로 할당하는 방법이다(e.g., Marsh et al. 1998). (2) 항목 내용의 유사성과 요인구조를 기준으로 배정하는 방법이다(e.g., Nasser, Takahashi, and Benson 1997). (3) 요인구조의 유사성만을 기준으로 배정하는 방법이다(e.g., Yuan et al. 1997). Yuan et al. (1997)은 항목들을 평균(averaging)하여 항목묶음을 형성하는 방법을 제시하면서 평균하기 전에 주성분분석이나 탐색적 요인분석을 이용하도록 권하고 있으며, Cattell and Burdsal(1975)은 항목묶음을 형성하고자 할 때 회전시키지 않은 요인해(unrotated factor solution)를 사용하도록 제안한 바 있다. (4) 항목묶음을 왜도(skewness)를 바탕으로 실시하는 방법도 제안되어 있다(Nasser and Wisenbaker 2003). 가장 정(+의 왜도가 큰 항목과 가장 부(-)의 왜도가 큰 항목을 첫 번째 묶음으로 배정하고 두 번째로 정의 왜도가 큰 항목과 두 번째로 부의 왜도가 큰 항목을 다음 묶음으로 배정하는 등의 방법이다.

여기서 연구자들이 특히 유의할 점은 위에 제시된 항목묶음 방법 가운데 어떤 것을 적용하더라도 적용 전에 구성개념별로 측정항목들이 단일차원성을 지니고 있는지 검토하여 이것이 확보된 경우에 한하여 항목묶음을 사용해야 한다는 것이다(e.g., Bandalos

2002; Bandalos and Finney 2001; Kim, Brody, and Murry 2003).

토하고자 한다.

IV. 본 연구에서 세 일반모델을 통해 검토된 사항의 특징

앞에서 언급한 것처럼 항목목록에 대한 이론적인 개념적 연구는 상대적으로 풍부한 편이지만 실증적 연구는 부족한 실정이며, 또한 기존의 실증적 연구에서는 주로 확인적 요인분석을 통해 구성개념의 요인구조를 바탕으로 이들을 어떤 방법으로 항목목록을 할 것이며, 이것이 모수 추정(parameter estimation)과 모델의 적합도 지수(goodness-of-fit indices)에 미치는 영향에 초점이 두어져 있었다.

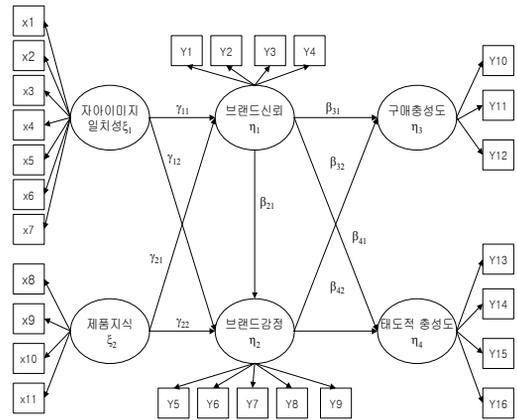
하지만 실제 항목목록을 사용하여 분석을 하는 대부분의 경우가 구성개념간의 관계를 파악하기 위해 구성개념간의 경로를 설정하지만, 이들 경로계수와 관련된 모수의 추정치 자체의 일정성이나 변동성 자체보다는 각 모수의 통계적 유의성이 일관되게 유지되는지의 여부에 더 관심이 있을 것으로 본다. 따라서 본 연구에서는 항목목록이 모수의 유의성 검정에 미치는 영향에 일차적 관심을 두고 있으며, 이어서 모델의 적합도에 미치는 영향에도 초점을 두어 고찰하고 있다.

따라서 본 연구에서는 항목목록의 수를 단계적으로 감소(즉, 각 항목목록에 포함된 지표의 수를 증가)시킬 때 각 구성개념간의 모수에 대한 유의성 검정 결과와 모델 적합도에 어떤 변화가 발생하는가를 일반모델에 해당하는 3개의 연구모델을 바탕으로 검

V. 실증연구

1. 연구모델 1의 항목목록이 모수 검정 결과와 적합도에 미치는 영향

본 연구에서는 항목목록이 연구가설과 관련된 모수의 검정 결과와 모델적합도 평가에 미치는 효과를 파악하기 위해 개별항목을 이용한 분석과 2개의 항목목록을 이용한 분석으로 구분하여 고찰하였다.



<그림 1> 연구모델 1

<표 1>에 나타난 바와 같이, 개별항목 분석과 2개 항목목록 분석을 비교할 때, 전반적으로 보아 t값이 약간 변동한 경우는 있으나 검정 결과가 변하는 정도의 변동은 8개 모수 어디에서도 나타나지 않고 전반적으로 일관된 검정결과를 보이고 있다. 여기서 한 가지 주목할 것이 있다. <그림 1>에서 β_{21} 이 나타나 있지만 ‘브랜드 신뢰-->

<표 1> 연구모델 1의 개별항목 분석과 항목묶음 분석 시의 모수 검정 결과 비교

	γ_{11}	γ_{12}	γ_{21}	γ_{22}	β_{31}	β_{32}	β_{41}	β_{42}
개별항목	0.835 (0.089)	-0.030 (0.094)	0.930 (0.087)	0.072 (0.085)	0.498 (0.065)	0.535 (0.063)	0.352 (0.052)	0.665 (0.058)
	9.437***	-0.320	10.631***	0.848	7.664***	8.451***	6.719***	11.412***
2개 항목 묶음	0.878 (0.084)	-0.025 (0.090)	0.932 (0.078)	0.060 (0.078)	0.594 (0.076)	0.440 (0.072)	0.356 (0.059)	0.652 (0.063)
	10.503***	-0.276	11.973***	0.771	7.787***	6.130***	5.991***	10.267***

*** p < 0.001 (단측검정)

[주] 1) 연구모델 1에서의 표본 크기(sample size) n = 193.

2) <그림 1>의 연구모델 1에서 적용된 측정 지표와 관련하여, 각 구성 개념별 결합신뢰성(CR: composite reliability)과 평균추출분산(AVE: average variance extracted)은 Fornell and Larcker (1981)의 공식에 의해 계산되었으며 구체적 수치는 다음과 같다: 자아이미지 일치성(CR = 0.87, AVE = 0.48), 제품지식(CR = 0.65, AVE = 0.32), 브랜드 신뢰(CR = 0.90, AVE = 0.70), 브랜드 감정(CR = 0.92, AVE = 0.71), 구매충성도(CR = 0.84, AVE = 0.64) 및 태도충성도(CR = 0.80, AVE = 0.51). 그리고 모든 지표는 표준화계수 값이 0.592와 0.957의 범위에 속하며 통계적으로 모두 유의하였다($p < 0.001$). 그런데, CR과 AVE에 대한 종합적 검토는 별도의 항을 마련하여 나머지 두 모델의 것들과 함께 검토하기로 한다.

브랜드 감정' 간의 경로계수에 해당하는 이 β_{21} 은 사전에 연구가설로 제시하지 않았다. 따라서 <표 1>의 추정 결과에서는 β_{21} 에 관한 검정 결과가 나타나 있지 않다. 앞으로 모델의 적합도 개선을 위해 수정지수(modification index, MI)를 참조하여 자유모수로 개방할 예정인 모수이다.

연구모델 1에 대한 전반적인 적합도 지수에 대해 살펴보기로 한다.

<표 2>는 연구모델 1에 대한 전반적인 적합도 지수의 변화를 설명하기 위한 것으로, 개별항목 분석 시에 비해 2개 항목묶음 분석의 경우가 GFI(goodness of fit index), AGFI(adjusted GFI), CFI(comparative fit index) 및 NFI(normed fit index)에 있어서는 전반적으로 적합도가 향상되었다. 다만,

예상한 것과는 달리 RMSEA(root mean square error of approximation)의 경우에는 그 값이 증가함으로써 이 측면에서는 적합도가 악화되는 것으로 나타났다. 그런데 CFI는 0.938, NFI는 0.921로 둘 다 0.90을 초과하고 있다. 하지만 GFI는 0.858, AGFI는 0.754로 둘 다 아직도 0.90에 미치지 못하고 있으며, RMSEA 역시 0.130으로 증가하여 수용하기 어려운 수준의 값이다. 따라서 '브랜드 신뢰 → 브랜드 감정'간의 경로계수에 해당하는 β_{21} 을 자유모수로 개방하기로 결정하는 모델 수정을 시도하였다. 이론적 관점에서 보아도 브랜드 신뢰가 쌓이면 브랜드 감정이 호전될 것으로 기대하는 것은 전혀 무리가 없다. 수정된 모델의 적합도는 다음과 같이 호전되었다: GFI가

<표 2> 연구모델 1의 개별항목 분석과 항목묶음 분석시의 적합도지수 변화

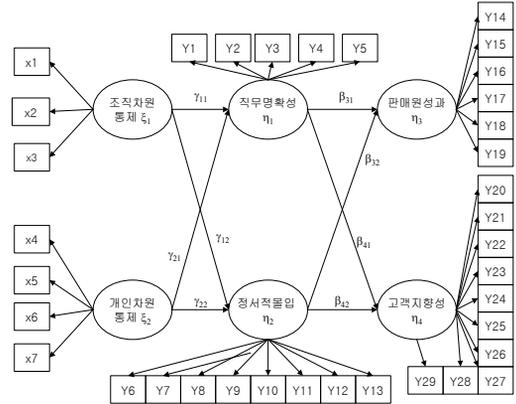
	GFI	AGFI	CFI	NFI	RMSEA
개별 항목	0.758	0.710	0.896	0.848	0.0919
2개 항목 묶음	0.858	0.754	0.938	0.921	0.130

0.918로 0.90을 초과하였고 NFI는 0.959로 0.95 이상으로 좋아졌으며, CFI는 0.976으로 0.95를 상회하였다. 또한 RMSEA도 개별항목 분석시의 0.0919에서 0.0297로 상당히 낮아져 적합도가 크게 개선되었다. 이 사례가 시사하고 있는 바는 항목목록이 적합도 개선에 기여할 수 있지만 결정적으로 적합도 개선에 기여하는 것은 이론에 근거한 지식을 충분히 활용하여 좋은 모델을 설계하는 것이라는 점이다.

2. 연구모델 2의 항목목록이 모수 검정 결과와 적합도에 미치는 영향

먼저 연구모델 2를 이용한 분석의 경우 위 모델에서 개인차원 통제, 직무명확성, 정서적 몰입, 판매원 성과 및 고객지향성은 단일차원의 구성개념이지만, 조직차원의 통제는 성과통제, 절차통제 및 전문성 통제의 세 가지 하부차원으로 이미 항목목록이 적용되어 있다. 따라서 구성개념이 다차원적

특성을 가진 조직차원의 통제를 제외한 나머지 구성개념에 대해서만 항목목록을 실시하였다.



<그림 2> 연구모델 2

연구모델 2의 모수 검정 결과를 나타낸 <표 3>을 보면 각 모수별로 t값이 약간의 변동을 보이는 경우는 있으나 유의성 검정 결과에 영향을 미칠 정도의 변동은 발견되지 않았다.

<표 3> 연구모델 2의 개별항목 분석과 항목목록 분석 시의 모수 검정 결과 비교

	γ_{11}	γ_{12}	γ_{21}	γ_{22}	β_{31}	β_{32}	β_{41}	β_{42}
개별 항목	0.231 (0.045) 5.178***	0.442 (0.040) 10.951***	0.274 (0.074) 3.690***	0.684 (0.065) 10.572***	0.527 (0.082) 6.468***	0.133 (0.043) 3.134***	0.294 (0.048) 6.115***	0.087 (0.020) 4.406***
2개 항목 묶음	0.217 (0.053) 4.089***	0.511 (0.045) 11.455***	0.311 (0.074) 4.216***	0.624 (0.060) 10.325***	0.649 (0.087) 7.464***	0.372 (0.068) 5.439***	0.388 (0.061) 6.355***	0.234 (0.047) 4.951***

* p < 0.001 (단측검정)

[주] 1) 연구모델 2의 표본 크기 n = 142.

2) <그림 2>의 연구모델 2에서 적용된 측정 지표와 관련하여, 각 구성 개념별 결합신뢰성(composite reliability, CR)과 AVE(average variance extracted)는 다음과 같다: 조직차원 통제(CR = 0.75, AVE = 0.53), 개인차원 통제(CR = 0.77, AVE = 0.47), 직무명확성(CR = 0.92, AVE = 0.71), 정서적 몰입(CR = 0.85, AVE = 0.42), 판매원 성과(CR = 0.82, AVE = 0.43) 및 고객지향성 (CR = 0.71, AVE = 0.35). 그리고 고객지향성의 2개 지표가 $p_s > 0.05$ 인 것을 제외하면, 그 외의 모든 지표는 표준화계수 값이 0.587과 0.941의 범위에 속하며 통계적으로 모두 유의하였다($p_s < 0.001$). 고객지향성에 대한 이 2개 지표의 경우 통계적으로 유의하지 않아 제외를 고려해야 할 후보이다. (단, CR과 AVE에 대한 종합적 검토는 별도의 항에서 다룰 예정임.)

<표 4> 연구모델 2의 개별항목 분석과 항목묶음 분석시의 적합도 지수 변화

	GFI	AGFI	CFI	NFI	RMSEA
개별 항목	0.671	0.624	0.757	0.711	0.110
2개 항목 묶음	0.899	0.836	0.932	0.912	0.0978

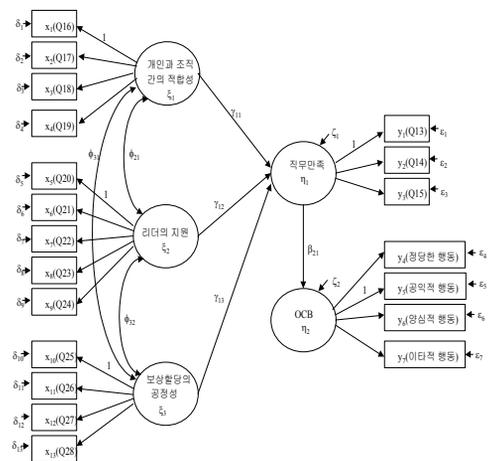
<표 4>는 연구모델 2에 대한 전반적인 적합도 지수의 변화를 나타낸 것으로, 개별항목 분석 보다는 2개 항목묶음 분석이 전반적 적합도 지수를 가리키는 GFI, AGFI, CFI 및 NFI 모두에 있어서 그 값들이 높아지는 경향을 보여 적합도가 향상되고 있음을 보여주고 있다. RMSEA의 경우에도 개별항목 분석시의 0.110에 비해 2개 항목묶음 분석에서 그 값이 0.0978로 약간 낮아지는 현상을 보였다. AGFI의 경우에는 수정모델에서 0.854로 개선되기는 하였으나 0.90에 약간 미치지 못하였다. 그런데 오래 전부터 GFI의 문제점이 지적되어 왔는데 최근 Sharma와 그의 동료들(Sharma, Mukherjee, Kurma, and Dillon 2005)은 GFI의 경우 잘못된 모델의 식별에 민감하지 못하며, 지표의 수를 줄여 주면 반드시 현저하게 상승하기 때문에 적합도 지수로 부적합하므로 사용하지 않아야 한다고 제안하고 있으며, *Journal of Marketing*과 같은 저명 저널에 실린 논문에서도 GFI와 AGFI를 전반적 적합도지수에서 제외시킨 사례가 발견되고 있다(e.g., Ganesan, Malter, and Rindfleisch 2005).

3. 연구모델 3의 항목묶음이 모수 검정 결과와 적합도에 미치는 영향

<그림 3>에서 나타난 것처럼 조직시민행동(Organizational Citizenship Behavior,

OCB)은 4개의 하부차원이 존재하는 다차원적 구조를 가지고 있어서 단일 차원성이 확보되어야 한다는 항목묶음의 전제에 위배된다. 따라서 조직시민행동을 제외한 나머지 항목에 대해서만 항목묶음을 통해 분석을 실시하였다.

연구모델 3의 분석결과인 <표 5>를 보면 먼저 모델의 분석을 개별항목 수준과 2개 항목묶음 수준으로 분석하였다. 모수 유의성 검정 결과에서는 γ_{13} 의 경우를 제외하면 모수 검정결과에 변화를 초래할 정도의 t값 변화는 나타나지 않았다. 단지, γ_{13} 의 t값이 't = 1.626, p > 0.05 (우측검정)'로 유의하지 않았던 것이 't = 1.677, p < 0.05(우측검정)'로 약간 상승하여 유의한 수준으로 바뀌었음을 알 수 있다.



<그림 3> 연구모델 3

<표 5> 연구모델 3의 개별항목 분석과 항목묶음 분석 시의 모수 검정 결과 비교

	γ_{11}	γ_{12}	γ_{13}	β_{21}
개별 항목	0.746 (0.104)	0.335 (0.070)	0.150 (0.092)	0.400 (0.047)
	7.159***	4.801***	1.626	8.555***
2개 항목 묶음	0.761 (0.111)	0.299 (0.066)	0.151 (0.090)	0.394 (0.046)
	6.871***	4.551***	1.677*	8.505***

* $p < 0.05$ (단측검정), ** $p < 0.01$, *** $p < 0.001$

[주] 1) 연구모델 3에서의 표본 크기 $n = 328$.

2) <그림 3>의 연구모델 3에서 각 구성개념별 결합신뢰성(CR: composite reliability)과 AVE(average variance extracted)는 다음과 같다: 개인과 조직 간의 적합성(CR = 0.84, AVE = 0.63), 리더의 지원(CR = 0.89, AVE = 0.62), 보상할당의 공정성(CR = 0.88, AVE = 0.64), 직무만족(CR = 0.78, AVE = 0.54) 및 OCB(CR = 0.88, AVE = 0.73). 그리고 모든 지표는 표준화계수 값이 0.392와 0.907의 범위에 속하며 통계적으로 모두 유의하였다 ($ps < 0.001$). (단, 상세한 것은 CR과 AVE에 대한 종합적 검토를 위한 항에서 다룰 예정임.)

<표 6>에서 살펴보면, 연구모델 3에 대한 전반적인 적합도 지수의 변화는 개별항목 분석 시에 비해 2개 항목묶음 분석의 경우가 모든 적합도 지수에서 향상되는 경향을 보이고 있다. 즉, GFI와 AGFI는 개별항목 수준에서는 적합도가 각각 0.881과 0.846으로 0.90에 미치지 못하였으나 2개 항목묶음 분석에서는 각각 0.956과 0.926으로 현저하게 상승하여 모두 수용할 수 있는 수준으로 호전되었다. 그리고 CFI도 0.925에서 0.976으로 더욱 호전되었으며, 또한 NFI의 경우 수용할 수 없는 수준인 0.886에서 0.952로 크게 증가하였다. 마지막으로 연구모델 3의 분석 결과에서는 RMSEA가 0.0724에서 0.0537로 낮아짐으로써 앞에서 고찰한 두 모델에서의 결과와는 달리 항목

묶음의 영향으로 호전되는 현상을 보였다.

4. 항목묶기 실시 전의 결합신뢰성과 평균추출분산에 대한 검토의 중요성

그런데 앞에서 다룬 세 모델의 경우 항목묶기에 초점을 맞춘 관계로 소홀히 한 점이 있다. 그것은 각 잠재변수별로 결합신뢰성(CR: composite reliability)과 평균추출분산(AVE: average variance extracted)을 구하고 이들에 대해 검토하지 못한 점이다. 통상 이들에 대한 검토를 먼저 하는 것이 바람직하다는 점을 지금부터 살펴보고자 한다. 그럼 CR과 AVE에 대해 중요한 사항을 먼저 검토하고, 이어서 세 모델에서의 항목묶기와 관련하여 AVE에 대한 검토가 매우 중요하

<표 6> 연구모델 3의 개별항목 분석과 항목묶음 분석시의 적합도 지수 변화

	GFI	AGFI	CFI	NFI	RMSEA
개별 항목	0.881	0.846	0.925	0.886	0.0724
2개 항목 묶음	0.956	0.926	0.976	0.952	0.0537

다 점을 종합적으로 검토하고자 한다.

각 잠재변수의 CR과 AVE는 Fornell and Larcker(1981)가 제시한 공식에 의해 계산되었는데, 이 두 공식을 풀어서 표현하면 다음과 같다(see Hair et al. 1998, p. 624).

결합신뢰성(CR) = (지표의 표준화계수의 합을 구한 후 제곱한 것) / [(지표의 표준화계수의 합을 구한 후 제곱한 것) + (지표의 측정오차들을 합산한 것)]

평균추출분산(AVE) = (지표의 각 표준화계수를 제곱한 후 합산한 것) / [(지표의 각 표준화계수를 제곱한 후 합산한 것) + (지표의 측정오차들을 합산한 것)]

위 식에 제시한 결합신뢰성 CR과 평균추출분산 AVE는 각 잠재변수별로 한 쌍씩 계산되며, 통상 CR에 대해서는 0.60 이상이라는 기준을, 그리고 AVE에 대해서는 0.50 이상이라는 기준을 적용해 오고 있다. 우선 CR에 대해 먼저 살펴보기로 하자. 그런데 Fornell and Larcker(1981) 및 Bagozzi and Foxall(1996) 등이 제시하고 있는 CR 계산 공식은 오차항 간에 상관(공분산)이 존재하지 않는 경우를 전제로 하고 있다는 점에 주목할 필요가 있다. 널리 알려져 있는 바와 같이 고전적 테스트이론(classical test theory)의 기본적 가정의 하나는 지표의 오차항 간에 무상관(無相關)이라는 가정인데, 주목할 점은 평행측도(parallel measure), (실질적) 타우등가측도 ((essentially) tau-equivalent measure) 및 동속측도(congeneric measure) 모두에서 상이한 지표들 간에 측정오차가 서로 무상관이라는 가정이 도입되

어 있다는 것이다(조현철 1999, pp. 176-183; Graham 2006; Raykov 2001a). 이처럼 CR에서 어떤 잠재변수에 있어서 지표의 오차항 간에 상관(공분산)이 존재하지 않는 것으로 전제하여 계산되는 경우에는 CR은 Cronbach's α 계수(Cronbach 1951)의 값과 차이가 거의 사라지게 된다(e.g., Novick and Lewis 1967; Raykov 2001a; Zimmerman 1972). (단, 엄밀히 말하면 Cronbach's α 계수는 실질적 타우등가측도를 전제로 함.) 한편, 지표의 오차항 간에 상관이 존재하는 경우 이들을 고려하여 계산하게 될 때에는 위의 CR 계산공식에서 분모에 오차항 간의 공분산에 해당하는 항이 추가되므로 이렇게 계산된 CR과 α 계수의 값은 서로 실질적 차이가 생겨나게 된다(e.g., Raykov 2001a, 2001b; Zimmerman 1972). 이로 인해 오차항 간의 공분산을 고려한 CR 쪽이 고려하지 않는 Cronbach's α 보다 그 값이 작아지게 된다. 이것을 달리 표현하면, Cronbach's α 는 지표들의 오차항 간에 상관이 존재하는 경우에도 이를 고려하지 않고 계산하게 되므로 '신뢰성이 과대평가 된다'는 것을 시사하고 있다. 그럼에도 불구하고 지표의 오차항 간에 상관이 존재하지 않는 것을 전제로 하는 Fornell and Larcker(1981)의 공식에 의해 CR을 계산하는 것이 일반화되어 있으나, 만약 측정 지표들의 오차항 간에 유의한 상관이 존재함에도 불구하고 이를 고려하지 않았다면 CR에도 신뢰성이 과대평가되는 결점이 존재하게 된다. 사실 Zimmerman(1972), Raykov (2001a) 등이 Cronbach's α 대신 CR을 권장하게 된 진정한 이유의 하나는 바로 오차항 간의 상관을 고려할 수 있다는 점 때문

이다. 그런데 구조방정식모델에서는 어떤 잠재변수에 있어서 지표의 오차항 간에 무상관이라는 가정을 통상적으로는 그대로 유지하지만 측정이론이나 측정된 지식을 바탕으로 오차항 간에 상관(단, 모델 상에서는 공분산)을 허용해야 할 필요가 발생한 경우에 한하여 이 무상관 가정을 완화할 수 있다. 하지만 타당한 근거 없이 단지 적합도 지수를 향상시킬 목적으로 오차항 간의 공분산(상관)을 허용해서는 안 된다.

그리고 CR을 계산하는 공식에서 통상 '표준화계수(standardized coefficient)'를 적용해야 한다고 지적되고 있는데, 여기서 가리키는 표준화계수는 잠재변수(구성개념)와 그 지표들 간의 관계를 가리키고 있는 경로계수의 표준화계수라는 점에 유의할 필요가 있다. 왜냐하면, LISREL이 제공하는 두 가지 표준화해(解) 가운데서 통상적인 표준화해(SS: standardized solution)에 있어서는 잠재변수 간의 관계를 나타내는 경로계수는 표준화되어 있으나, 각 잠재변수와 이의 지표들 간의 관계를 나타내는 경로계수는 (값의 변화는 있지만) 표준화되어 있지 않기 때문이다. 따라서 CR의 계산을 위해서는 (잠재변수 간의 관계만 표준화되어 있는 것으로는 곤란하며) 잠재변수와 이의 지표들 간의 경로계수도 표준화되어 있는 완전표준화해(SC: completely standardized solution)에서 제대로 표준화되어 있는 해당 경로계수(즉 completely standardized λ_x)를 사용해야 한다. 실제의 계산 결과에서 이들 간에 미미한 차이가 존재할지라도 이론적으로 타당한 표준화계수를 식별하여 사용해야 하기 때문이다.

그리고 AVE는 각 잠재변수가 자신의 지

표들에 의해 설명된 분산의 비율을 나타내고 있다(Fornell and Larcker 1981). 따라서 AVE에 대해 요구되는 0.5라는 기준은 바꾸어 말하면 50%가 오차에 의해 설명되고 있다는 것을 나타내므로, 50% 이상이 오차로 설명되어서는 곤란하다는 사고가 바탕이 되어 있다. 그런데, 같은 자료로 계산되는 경우 AVE의 값이 CR의 값보다 일반적으로 상당히 낮게나와 한층 더 엄격하여 이 0.5 이상이라는 기준을 모든 잠재변수에서 완전히 충족시키기란 용이한 일이 아닐 것이라는 예상을 할 수 있다. 이 점에 대해서는 *Journal of Marketing*과 같은 저명 저널에 실린 최근 논문 가운데서도 CR 혹은 Cronbach's α 는 제시하고 있으나 AVE는 제시하지 않은 논문들(e.g., Frels, Shervani, and Srivastava 2003; Homburg and Fürst 2005; Morgan, Kaleka, and Katsikeas 2004; Sirdeshmukh, Singh, and Sabol 2002)이 종종 발견된다는 점에서 간접적으로 살펴볼 수 있다. 따라서 'AVE \geq 0.50 기준'을 절대적 성격의 기준으로 간주하기 보다는 측정 지표를 정치화(精緻化)하기 위한 하나의 이상적 목표로 삼는 것이 바람직하다고 본다.

여기서 이 논문에서 제시한 세 연구모델에 대해 특히 AVE를 바람직한 수준으로 향상시키기 위한 방안을 검토해 보기로 한다. 앞에서 <그림 1>의 연구모델 1에 대한 CR과 AVE의 출력에 의하면, CR에 있어서는 하나가 0.65이고 나머지 다섯 개는 0.80-0.92 사이에 있으므로 모두 0.60을 초과하고 있어서 전혀 문제가 없으나, AVE의 관점에서 보면 자아이미지 일치성의 경우 0.48로 0.50에 매우 근접하므로 수용할 수 있는 수준은 되지만 제품지식의 경우에는

0.32로 0.50에 미치지 못하고 있으므로 이 제품지식의 지표들에 대한 면밀한 검토를 통해 이 값을 향상시킬 필요가 있다. 이 점에 대해서는 다음 연구모델 2에서 설명될 것이다. 나머지 네 개의 AVE는 0.51-0.81의 범위에 속하므로 수용할 수 있는 수준이거나 이를 초과하고 있다.

그리고 <그림 2>에 제시된 연구모델 2의 경우 CR은 0.71-0.92의 범위에 속하므로 전혀 문제가 없다. 그러나 AVE의 관점에서 보면, 개인차원 통제(AVE = 0.47)는 거의 0.50에 육박하고 있으나 정서적 몰입(AVE = 0.42)과 판매원 성과(AVE = 0.43)의 경우 0.50에 약간 못 미치고 있으며, 고객지향성(AVE = 0.35)의 경우 상당히 0.50에 미치지 못하고 있다. 특히 고객지향성의 경우 앞에서 언급한 바 있는 “유의하지 않은” 2개 지표가 AVE에 미치는 영향이 클 것으로 본다. 이러한 유형의 문제를 해결하는 하나의 유용한 방법은 LISREL output에서 (각 지표의 신뢰성계수의 하나에 해당하는) SMC(squared multiple correlation)를 참조하는 것이다. SMC는 단순회귀분석에 의해 제공되는 결정계수의 일종으로 각 지표가 잠재변수에 의해 설명된 변동(variation)의 비율에 해당하므로 이 값이 매우 작은 소수의 지표를 각 잠재변수별로 제외시키면 설명된 분산의 비율이 높아지게 되어 해당 AVE도 상승하게 될 것이므로 0.50 수준이나 그 이상으로 개선될 가능성이 생겨난다. 이는 지표를 항목묶이기 전에 2단계모델 접근법에 의해 보다 양호한 지표로 추린 다음 항목묶기를 해야 할 경우도 있다는 것을 보여주고 있다.

끝으로, <그림 3>에 제시된 연구모델 3

의 경우에는 CR이 0.78-0.88의 범위에 속하므로 양호한 편이며, 또한 AVE도 0.54-0.73의 범위에 있으므로 문제가 없다. 위의 검토결과를 요약하면, CR과 AVE 가운데서 AVE는 각 지표의 SMC와 함께 적용하면 불량한 지표를 골라내는 데 중요한 역할을 할 수 있으나 CR은 이러한 역할을 제대로 하지 못한다. 따라서 항목묶기를 실시하기 전에 잠재변수별로 AVE를 검토하는 것은 불량한 지표는 버리고 우량한 지표만 남겨서 더 좋은 항목묶기를 할 수 있게 해주므로 반드시 검토할 것을 권한다.

VI. 결 론

본 논문에서는 먼저 항목묶음 관행의 학문영역별 적용 실태, 항목묶음에 찬성하는 입장과 반대하는 입장의 논지를 각각 고찰하였다. 그리고 항목묶음의 방법에 대해서도 살펴보았다. 항목묶음을 사용하고자 하는 경우 장점들은 최대한 활용하고 단점들에 기인하는 오류는 범하지 않도록 연구자는 연구의 설계 시에 유의할 필요가 있다. 특히, 항목묶음 실시 전에 각 구성개념의 측정 지표들이 단일차원성(unidimensionality) 전제를 충족시키고 있는지를 사전에 검토하고 그 결과에 대해 연구결과의 한 부분으로 반드시 보고할 필요가 있다. 즉, 단일차원성의 검토 결과, 항목묶음을 실시한 목적, 방법, 절차와 그 영향 등에 관한 비교적 상세한 정보를 논문의 통계적 분석 항목의 일부로서 밝히는 것이 바람직할 것으로 본다.

그런데 기존의 항목묶음에 대한 실증적

연구는 주로 확인적 요인분석모델을 대상으로 하고 있으며, 구성개념들의 요인구조를 바탕으로 이들을 어떤 방법으로 항목묶음을 하며 이것이 모수 추정(해의 추정)과 모델의 적합도지수에 미치는 영향이 어떠한가에 분석의 초점이 두어져 있었다. 그러나 본 연구에서는 구조방정식모델을 이용하여 구성개념들 간의 관계를 고찰하고자 할 때 항목묶음을 이용하는 것이 모수의 추정 가능성 여부보다는 인과관계를 나타내고 있는 모수의 유의성 검정에 연구자들의 관심이 더 있을 것으로 판단하여, 이 점에 주된 초점을 맞추고 아울러 모델 적합도 지수의 변화에 미치는 영향도 함께 초점을 두어 세 개의 일반모델을 대상으로 고찰하였다.

이 논문에서 고찰한 결과를 종합하면, 전반적으로 보아 개별항목 수준의 분석과 항목묶음 수준의 분석 양자에 있어서 인과 모수의 검정통계량의 값에서 큰 변화가 없으므로 유의성 검정에 일관성이 있는 것으로 밝혀졌다. 그리고 모델의 전반적 적합도 지수에 있어서는 개별항목 분석 시 보다는 항목묶음 분석 시에 RMSEA를 제외한 GFI, AGFI, CFI 및 NFI에 있어서 항목묶음을 적용하면 이들 적합도 지수는 일관되게 향상되는 결과를 보였다. 하지만, 예상과는 달리 RMSEA의 경우에는 상반된 결과를 얻었다. 즉 RMSEA에 있어서는 적합도가 향상되는 사례와 악화되는 사례가 동시에 존재하였다. 그런데, 2개 항목묶음에서 해가 구해지지 않거나 부적해가 얻어지는 경우가 가끔 발생할 수 있는데(e.g., West et al. 1995), 이의 한 해결 방안으로는 지표(항목)의 수가 많은 경우 3개 이상으로 항목묶음을

을 하여 분석해보는 것이다.

그런데, 항목묶기에 들어가기 전에 결합 신뢰성(CR)과 평균추출분산(AVE)의 출력 결과 가운데서 특히 AVE가 낮은 잠재변수의 지표들(indicators)에 대해 신뢰성계수의 일종인 SMC를 이용하여 이 값이 매우 작은 소수의 지표를 추려내어 AVE를 개선시킨 다음 항목묶기에 들어가는 것이 유용하다는 점이 고찰되었다.

끝으로, 본 연구는 다음과 같은 한계점을 지니고 있다. 첫째, 본 연구에서 사용된 세 일반모델은 일부의 구성개념에서 지표의 수가 많지 않아서 4개의 항목묶음이나 3개의 항목묶음을 적용할 수 없었다는 한계가 있다. 둘째, 모수의 유의성 검정 결과와 관련하여 단지 세 모델에 대한 연구 결과로 이를 일반화하기에는 한계가 있다. 보다 다양한 모델에 대해 더 많은 연구가 행해질 때 일반화가 가능할 것으로 본다.

(논문접수일: 2007년 6월 16일)

(게재확정일: 2007년 9월 14일)

참고 문헌

- 고순화·이영선 (2007), 가상적 상황에서 의류제품에 대한 서비스회복이 고객만족과 구매 후 행동의도에 미치는 영향, 『소비자학연구』, 18 (1), 49-68.
- 김봉관·김태우 (2003), 브랜드자산 형성과정에 관한 연구, 『마케팅과학연구』, 11 (5월), 59-77.
- 이지현·이동일 (2005), 모델의 매력도 차원

- 이 모델 전형성과 의류광고 적합성에 미치는 영향, 『마케팅과학연구』, 15 (3), 49-65.
- 이태민·전종근 (2003), 유비쿼터스 접속성과 상황기반 제공성이 모바일 상거래 수용 의도에 미치는 영향에 관한 연구: 기술 수용모델 적용을 중심으로, 『경영학연구』, 33 (4), 1043-1071.
- 정기한·김대업 (2002), 내부마케팅과 시장지향성간의 관계에 대한 연구, 『마케팅과학연구』, 9 (5월), 19-46.
- 조현철 (1999), LISREL에 의한 구조방정식 모델, 서울: 석정
- Anderson, J. C. and D. W. Gerbing (1988), Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach, *Psychological Bulletin*, 103 (3), 411-423.
- Bagozzi, R. P. and J. R. Edwards (1998), A General Approach for Representing Constructs in Organizational Research, *Organizational Research Methods*, 1, 45-87.
- _____ and Gordon R. Foxall (1996), Construct Validation of a Measure of Adaptive-Innovative Cognitive Styles in Consumption, *International Journal of Research in Marketing*, 13, 201-213.
- _____ and T. F. Heatherton (1994), A General Approach to Representing Multifaceted Personality Constructs: Application to State Self-Esteem, *Structural Equation Modeling*, 1, 35-67.
- Bandalos, D. L. (2002), The Effects of Item Parceling on Goodness-of-Fit and Parameter Estimate Bias in Structural Equation Modeling. *Structural Equation Modeling*, 9, 78 - 102.
- _____ and S. J. Finney (2001), Item Parceling Issues in Structural Equation Modeling, in G. A. Marcoulides and R. E. Schumaker (Eds.), *New Developments and Techniques in Structural Equation Modeling*, 269-296, Mahwah, NJ: Lawrence Erlbaum.
- Bearden, W. O. and R. G. Netmeyer (1999), *Handbook of Marketing Scales: Multi-item Measures for Marketing and Consumer Behavior Research*, (2nd ed.), Newbury Park, CA: Sage Publications, Inc.
- Bollen, K. (1989), *Structural Equations with Latent Variables*, New York: John Wiley.
- Cattell, R. B. (1956), Validation and Intensification of the Sixteen Personality Factor Questionnaire, *Journal of Clinical Psychology*, 12, 205-214.
- _____ and C. A. Burdsal, Jr. (1975), The Radial Parceling Double Factoring Design: A Solution to the Item-v.s.-Parcel Controversy, *Multivariate Behavioral Research*, 10, 165-179.
- Cronbach, L. J. (1951), Coefficient Alpha and the Internal Structure of a Test, *Psychometrika*, 16, 297-334.
- Ding, L., W. F. Velicer, and L. L. Harlow (1995), Effects of Estimation Methods,

- Number of Indicators per Factor, and Improper Solution on Structural Equation Modeling Fit Indices, *Structural Equation Modeling*, 2, 119-144.
- Fornell, Claes and David F. Larcker (1981), Evaluating Structural Equation Models with Unobservable Variables and Measurement Error, *Journal of Marketing Research*, 18 (February), 39-50.
- Frels, Judy K., Tasadduq Shervani, and Rajendra K. Srivastava (2003), The Integrated Networks Model: Explaining Resource Allocations in Network Markets, *Journal of Marketing*, 67 (January), 29 - 45.
- Graham, James M. (2006), Congeneric and (Essentially) Tau-Equivalent Estimates of Score Reliability: What They Are and How to Use Them, *Educational and Psychological Measurement*, 66 (6), 930-944.
- Gibbons, B. C. and D. Hocevar (1998), Levels of Aggregation in Higher Level Confirmatory Factor Analysis: Application for Academic Self-Concept. *Structural Equation Modeling*, 5, 377-390.
- Hair, Joseph F., Jr., Rolph E. Anderson, Ronald L. Tatham, William C. Black (1998), *Multivariate Data Analysis*, 5th ed., Upper Saddle River, NJ: Prentice Hall.
- Hall, R. J., A. E. Snell, and M. S. Foust (1999), Item Parcelling Strategies in SEM: Investigating the Subtle Effects of Unmodeled Secondary Constructs, *Organizational Research Methods*, 2, 233-256.
- Homburg, Christian and Andreas Fürst (2005), How Organizational Complaint Handling Drives Customer Loyalty: An Analysis of the Mechanistic and the Organic Approach, *Journal of Marketing*, 69 (July), 95 - 114.
- James, L. R., S. A. Mulaik, and J. M. Brett (1982), *Causal Analysis: Assumptions, Models, and Data*, Beverly Hills, CA: Sage.
- Kim, S., G. H. Brody, and V. M. Murry (2003), Temperament Questionnaire and Measurement Invariance across Gender, *Journal of Early Adolescence*, 23 (3), 268-294.
- Kishton, J. M. and K. F. Widaman (1994), Unidimensional versus Domain Representative Parceling of Questionnaire Items: An Empirical Example, *Educational and Psychological Measurement*, 54, 757-765.
- Little, T. D., W. A. Cunningham, G. Shahar, and K. F. Widaman (2002), To Parcel or Not to Parcel: Exploring the Question, Weighing the Merits, *Structural Equation Modeling*, 9, 151-173.
- _____, G. Oettingen, and P. B. Baltes

- (1995), The Revised Control, Agency, and Means-Ends Interview (CAMI): A Multi-Cultural Validity Assessment Using Mean and Covariance (MACS) Analyses, *aterialien aus der Bildungsforschung*, 49, Berlin: Max Planck Institute.
- _____ and B. Wanner (1997), The Multi-CAM: A Multidimensional Instrument to Assess Children's Action-Control Motives, Beliefs, and Behaviors, *Materialien aus der Bildungsforschung*, 59, Berlin: Max Planck Institute.
- MacCallum, R. C., K. F. Widaman, S. Zhang, and S. Hong (1999). Sample Size in Factor Analysis, *Psychological Methods*, 4, 84-99.
- Marsh, H. W., K.-T. Hau, J. R. Balla, and D. Grayson (1998), Is More Ever Too Much? The Number of Indicators per Factor in Confirmatory Factor Analysis, *Multivariate Behavioral Research*, 33, 181-220.
- _____ and D. Hocevar (1988), A New, More Powerful Approach to Multitrait-Multimethod Analyses: Application of Second-Order Confirmatory Factor Analysis, *Journal of Applied Psychology*, 73, 107-117.
- Morgan, Neil A., Anna Kaleka, and Constantine S. Katsikeas (2004), Antecedents of Export Venture Performance: A Theoretical Model and Empirical Assessment, *Journal of Marketing*, 68 (January), 90 - 108.
- Nasser, F., T. Takahashi, and J. Benson (1997). The Structure of Test Anxiety in Israeli-Arab High School Students: An Application of Confirmatory Factor Analysis with Miniscales, *Anxiety, Stress, and Coping*, 10, 129-151.
- Nasser, E. and J. Wisenbaker (2003), A Monte Carlo Study Investigating the Impact of Item Parceling on Measures of Fit in Confirmatory Factor Analysis, *Educational and Psychological Measurement*, 63, 729-757.
- Novick, M. R., C. Lewis (1967), Coefficient Alpha and the Reliability of Composite Measurement, *Psychometrika*, 32, 1-13.
- Raykov, Tenko (2001a), Bias of Cronbach's Coefficient Alpha for Fixed Congeneric Measures with Correlated Errors, *Applied Psychological Measurement*, 25, 69-76.
- _____ (2001b), Estimation of Congeneric Scale Reliability Using Covariance Structure Analysis with Nonlinear Constraints, *British Journal of Mathematical & Statistical Psychology*, 54 (November), 315-323.
- Schumacker, R. E. and R. G. Lomax (1996), *A Beginner's Guide to Structural Equation Modeling*, Mahwah, NJ: Lawrence Erlbaum.
- Sharma, S., S. Mukherjee, A. Kumar, and

- W. R. Dillon (2005), A Simulation Study to Investigate the Use of Cutoff Values for Assessing Model Fit in Covariance Structure Models, *Journal of Business Research*, 58, 935-943.
- Sirdeshmukh, Deepak, Jagdip Singh, and Barry Sabol (2002), Consumer Trust, Value, and Loyalty in Relational Exchanges, *Journal of Marketing*, 66 (January), 15-37.
- Takahashi, T. and F. Nasser (1996), The Impact of Using Item Parcels on ad hoc Goodness of Fit Indices in Confirmatory Factor Analysis: An Empirical Example, Paper presented at the annual meeting of the American Educational Research Association, New York.
- Thompson, B. and J. Melancon, (1996), Using Item Testlets/Parcels in Confirmatory Factor Analysis: An Example Using the PPDP-78, Paper presented at the annual meeting of the Mid-South Educational Research Association, Tuscaloosa, AL (ERIC Document Reproduction Service No. ED 404 349).
- West, S. G., J. F. Finch, and P. J. Curran (1995), Structural Equation Modeling with Nonnormal Variables: Problems and Remedies, in R. H. Hoyle (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications*, (pp. 56-75), Thousand Oaks, CA: Sage.
- Yuan, K.-H., P. M. Bentler, and Y. Kano (1997), On Averaging Variables in a CFA Model, *Behaviormetrika*, 24, 71-83.
- Zimmerman, D. W. (1972), Test Reliability and the Kuder-Richardson Formulas: Derivation from Probability Theory, *Educational and Psychological Measurement*, 32, 939-954.

<Abstract>

The Effects of Item Parceling on Causal Parameter Testing and Goodness-of-Fit Indices in Structural Equation Modeling

Cho, Hyun-Chul* · Suk-Hou Kang**

The purpose of this article is to examine the effects of item parceling on the consistency of significance testing of the causal parameters with regard to the relationship between the relevant constructs, as well as the effects of the item parceling on the goodness-of-fit indices of LISREL's general models. Most of the researchers' major purpose of using structural equation modeling (SEM) is to test their research hypotheses associated with the causal parameters. Therefore, we investigated three general models of LISREL, rather than the frequently used confirmatory factor analytic (CFA) models by many other researchers. The results of the study showed that there was a high level of consistency in the calculated test statistics of causal parameters between the item-parceled solutions and the item-level solutions, and that the item-parceled solutions had better goodness-of-fit indices, such as GFI, AGFI, CFI, and NFI, than the solutions at the item level. However, in terms of RMSEA, there was no such tendency.

Key Words: structural equation model, item parceling, improvement of goodness-of-fit indices

* Professor, Hanyang University(chohyunc@hanyang.ac.kr)

** Professor, Hanyang University(shkang@hanyang.ac.kr)