

중간보기 “?”의 차별적 사용에 대한 한국과 미국 간 비교 문화적 검증: 혼합문항반응분석의 적용*

이 필 석 천 석 준 이 선 희†

University of South Florida, 심리학과

충남대학교 심리학과

본 연구는 한국과 미국의 응답자들이 평정척도 상의 중간보기 “?”를 서로 다르게 사용하는지에 대한 비교 문화적 검증을 실시하였다. 이를 위해 직무기술척도(Job Descriptive Index)에 대한 한국과 미국 집단(각각 932명)의 응답을 다집단 범주적 확인적 요인분석(Multi-Group Categorical Confirmatory Factor Analysis)과 혼합문항반응이론(Mixed-Model Item Response Theory)으로 분석하였다. 다집단 분석결과, 두 집단 간에 측정틀 동일성과 측정단위 동일성은 성립하였으나, 임계치 모수의 구조에서는 차이가 있는 것으로 나타났다. 혼합문항반응이론을 실시한 결과, 한국과 미국집단 모두에서 응답범주에 대한 반응양상이 다른 세 개의 하위 잠재집단이 존재하였다. 이 중 “?” 응답범주를 많이 선택하는 “?” 선호집단에 대한 보다 자세한 분석 결과, 미국 표본에서는 “?” 선호 집단의 비중이 매우 적고, 응답범주의 순차성 가정이 지지되지 않는 것으로 나타났다. 반면, 한국 표본에서는 “?” 선호집단이 다른 하위 잠재집단과 유사한 비중을 차지할 뿐만 아니라, “?”을 비교적 일정한 간격을 지니는 순차적 평정척도의 중간보기로 인식하는 것으로 나타났다. 본 연구의 결과를 바탕으로 중간보기 “?”의 사용에 대한 적용적 시사점 및 추후 연구 과제를 논의하였다.

주요어 : 평정척도 문항, 중간 응답범주, 측정동일성, 혼합문항반응이론, 직무기술척도

* 본 논문을 위해 귀중한 조언을 아끼지 않으신 심시위원들께 진심으로 감사드립니다. 또한 본 연구의 한국 표본자료를 수집해준 송준석과 미국 표본자료를 제공해준 Bowling Green State University JDI 사무실에도 감사드립니다.

† 교신저자: 이선희, 충남대학교 심리학과, 대전광역시 유성구 대학로 99

E-mail: sunhee_lee@cnu.ac.kr

심리측정 방법으로 매우 보편적으로 사용되는 리커트(Likert) 문항과 같은 평정척도(rating scale)는 응답자들에게 특정 잠재특성을 반영하는 진술문과 이 진술문에 대한 서로 다른 동의의 정도를 나타내는 순차적인 응답범주들을 제시한다. 응답자는 제시된 응답범주 중에서 해당 진술문에 대한 자신의 동의정도와 가장 일치하는 응답범주를 한개 선택하게 된다. 이렇게 선택된 응답범주는 점수화되어 해당 잠재특성상에서의 응답자의 상대적 위치를 추론하는 데 사용된다. 그런데 이러한 과정에서 가장 중요한 가정 중의 하나는 모든 응답자들이 제시된 응답범주들을 실제로 순차적으로 해석한다는 응답범주의 순차성 가정이다(Andrich & Schoubroeck, 1989).

그러나 제시된 응답범주의 중간 보기로 “?”, “확신할 수 없음(Not sure)”, 또는 “결정할 수 없음(Undecided)” 등이 사용되는 경우, 응답범주의 순차성 가정이 충족되지 않을 수 있다는 지적이 제기되어 왔다(예: Andrich, Jong, & Sheridan, 1997; DuBois & Burns, 1975; Gonzalez-Roma & Espejo, 2003). 예를 들어, Andrich 등(1997)은 “매우 비동의”, “비동의”, “확신할 수 없음 또는 결정할 수 없음”, “동의”, “매우 동의”로 구성된 5점 리커트 문항 자료를 분석한 결과, 중간보기에 해당하는 “확신할 수 없음 또는 결정할 수 없음” 범주가 중간 정도의 특질을 반영하지 않는다는 것을 발견하였다. 또한 Gonzalez-Roma와 Espejo(2003)는 중간보기를 “분명치 않음”, “?”, 또는 “중간(In-between)”으로 각각 제시하였을 때, “중간”의 경우만 응답범주의 순차성 가정이 충족될 뿐, 나머지 두 개의 제시방식에서는 이 가정이 충족되지 않음을 확인하였다.

그런데 이러한 중간보기의 문제를 보다 폭

잡하게 하는 것은 모든 응답자들이 제시된 응답범주를 동일한 방식으로 해석하지 않을 수 있다는 점이다(Brown & Maydeu-Olivares, 2013; Cheung & Mooi, 1994; Hernáñez, Drasgow, & Gonzalez-Roma, 2004). 예를 들어, Hernáñez 등은 “예”, “?”, “아니오”의 응답범주를 사용하는 16PF(Sixteen Personality Factor Questionnaire; Catell & Catell, 1995) 성격척도에 대한 미국인 표본의 응답 자료를 분석한 결과, 해당 표본 내에서 중간보기인 “?”을 자주 사용하는 집단과 “?” 보기를 선택하지 않고 양극단의 응답범주를 주로 선택하는 두 개의 하위 잠재집단이 존재한다는 것을 발견하였다. 또한, Maij-de Meij, Kelderman 그리고, van der Flier(2008)는 “예”, “?”, “아니오”로 구성된 암스테르담 생활사 검사(Amsterdam Biographical Questionnaire; Wilde, 1970) 응답자료를 분석한 결과 “?” 범주를 서로 다르게 사용하는 세 개의 하위집단이 존재한다고 보고하였다. 이처럼 응답자들이 동일한 중간보기를 서로 다르게 해석하여 사용한다면, 이러한 과정에서 나온 응답점수의 타당성은 심각한 문제를 가질 수 있다.

그런데 이러한 중간보기에 대한 선행연구들(예: Andrich et al., 1997; DuBois & Burns, 1975; Gonzalez-Roma & Espejo, 2003; Hernáñez et al., 2004; Hernáñez, Espejo, & Gonzalez-Roma, 2006; Kieruj & Moors, 2010; Kulas & Stachowski, 2009; Moors, 2008)은 모두 서구 표본을 이용하였기 때문에 이러한 결과가 비서구 문화권에서도 발생하는지 확인할 필요가 있다. 이와 관련하여, 한국과 미국 표본간의 측정동일성을 연구한 Riordan과 Vandenberg(1984)도 “공정도 부정도 아님(Neither agree nor disagree)”과 같은 중간보기가 한국집단에서는 긍정과 부정

의 사이에 위치한 “다소 긍정(Mild agreement)”으로 해석되어지는 반면, 미국집단에서는 별도의 범주인 “선택 없음(No option)”으로 해석될 수 있다는 가능성을 제기한 바 있다. 그러나 이제까지는 서로 다른 문화권의 표본을 이용하여 문화 간 중간보기의 차별적인 사용이나, 이와 관련된 하위 잠재집단의 구성이 문화에 따라 어떻게 다른지에 대한 경험적인 연구는 매우 부족한 실정이다.

이에 본 연구는 중간보기를 “?”로 사용하는 직무기술지표(Job Descriptive Index; Smith, Kendall & Hulin, 1969)에 대한 한국과 미국 표본의 응답 자료에 대한 혼합 문화반응이론(Mixed Model Item Response Theory) 분석을 바탕으로 1) 한국과 미국 표본 집단 내에 “?” 응답범주를 차별적으로 해석하고 사용하는 하위 잠재집단이 존재하는 지, 2) 만약 그렇다면, 두 문화집단 내 “?”를 자주 사용하는 잠재집단이 제시된 응답범주들을 순차적으로 해석하는지, 마지막으로 3) 이러한 하위 잠재집단이 두 문화집단에서 유사한 형태를 보이는지를 살펴보고자 하였다. 두 집단 간에 응답 범주의 사용상의 차이가 있는지를 비교하기 위해서는 측정하고자 하는 심리구인인 직무만족 요인들이 두 집단에게 동일한 의미를 가지는지를 먼저 확인해야 하기 때문에 우선 한국과 미국집단 간에 직무기술지표 척도의 측정동일성이 성립하는지를 점검하였다. 따라서 아래에서는 직무기술지표에 대한 한국과 미국 표본의 측정동일성 검증을 위해 사용된 범주적 확인적 요인분석(Categorical Confirmatory Factor Analysis)과 하위 잠재집단 분석을 위해 사용된 혼합문화반응이론에 대해 간단히 소개하였다.

측정동일성과 범주적 확인적 요인분석

측정동일성이란 관찰된 점수가 집단 소속에 따른 어떠한 영향도 없이 오직 잠재변인의 점수에 의해서만 결정되는 것을 의미한다. 즉, 동일한 수준의 잠재특질을 갖는 개인들이 특정한 응답을 선택할 확률이 소속 집단의 특성에 관계없이 동일할 때 집단 간 측정동일성이 존재한다고 할 수 있다(Raju, Laffitte, & Byrne, 2002). 따라서 집단 간 측정동일성이 성립하지 않는다면, 서로 다른 집단에 속한 개인들은 서로 다른 측정도구에 답한 셈이 되기 때문에 다른 집단에 속한 개인들 간 혹은 집단 간에 점수를 직접적으로 비교하는 것은 무의미하게 된다(Hui & Triandis, 1985; Wang & Russell, 2005).

측정동일성 검증에는 일반적으로 다집단 확인적 요인분석(Multiple Group Confirmatory Factor Analysis)이 사용되는데, JDI와 같이 “예”, “?”, “아니오”로 구성된 3분 척도는 선형적 요인분석이 아닌 범주적 요인분석이 적합하다고 알려져 있다(Millsap & Yun-Tein, 2004). 범주적 확인적 요인분석에서의 관찰변인(X_{ij})은 연속적 잠재반응변수(latent response variate, X_{ij}^*)와 임계치(threshold, v)와의 관계에 의해 결정된다(Wirth & Edwards, 2007). 이때 연속적 잠재반응변수(X_{ij}^*)란 문항 j 에 대하여 응답자 i 가 가지고 있는 잠재적 특질의 점수를, 그리고 임계치(v)는 각 응답범주를 구분하는 기준을 의미한다.

잠재반응변수가 정규분포(평균=0, 분산=1)를 따른다는 가정 하에, 이를 표현하면 식 1과 같다. 여기서 i 는 개인을, j 는 문항을 나타낸다. λ_j 는 요인부하값, ϵ_i 는 요인점수, 그리고 ϵ_{ij} 는 측정오차를 의미한다.

$$X_{ij}^* = \lambda_j \xi_i + \varepsilon_{ij} \quad (1)$$

식 2는 연속적 잠재반응변수(X_{ij}^*)와 임계치(v)의 함수로 관찰점수(X_{ij})를 정의한다. 여기서 v_{jc} 는 j 번째 문항에 있어 각 범주(C)를 구분하는 임계치이다. 이때, 임계치의 개수는(범주의 개수 - 1)로 정의된다.

$$X_{ij} = C, \text{ if } v_{jc} < X_{ij}^* \leq v_{j(c+1)} \quad (2)$$

다집단 분석을 통해 측정동일성을 평가하기 위해서는 다음과 같은 일련의 단계를 따른다(Millsap & Yun-Tein, 2004; Vandenberg & Lance, 2000). 첫째, 집단 간 요인 구조의 동일성을 검증하는 “측정틀 동일성(configural invariance)” 모형 단계이다. 이 모형에서는 집단 간에 어떠한 동일성 제약도 부과하지 않은 채, 요인 부하값, 임계치, 요인간의 변량 및 공변량을 자유롭게 추정한다. 따라서 이 모형은 각 요인에 대한 측정지표가 집단 간에 동일함을 검증하는 모형이다. 일단 측정틀 동일성 모형이 적절하다고 판단되면, 측정틀 동일성 모형은 둘째 단계의 모형인 “측정단위 동일성(metric invariance)” 모형과 비교된다. 측정단위 동일성 모형은 모든 요인부하값을 집단 간에 동일하게 제약한다. 즉, 이 모형은 각 요인과 요인들 간의 관계가 집단 간에 동일하다고 가정하는 모형이다. 측정단위 동일성 가정이 적절하다고 판단되면, 셋째 단계인 “임계치 동일성(threshold invariance)” 모형과 비교한다. 이 모형은 임계치를 집단 간에 동일하게 제약함으로써, 집단 간에 임계치가 동일하다고 가정하는 모형이다. 이 단계에서 비교 집단 간 임계치 동일성이 성립된다면, 두 집단의 응답자들이 해당 척도의 응답범주들을 동일하게 해석

한다고 볼 수 있다. 각 단계에서 이전 단계의 모형과 추가적인 제약을 가한 해당 단계의 모형의 합치도가 유의하게 다르지 않으면 해당 단계에서의 측정동일성이 성립하는 것으로 본다(Chen, 2007; Millsap & Yun-Tein, 2004).

혼합문항반응이론(Mixed Model IRT, MM-IRT)

전통적인 문항반응이론은 모집단 내 모든 응답자들이 동질적이라는 가정 하에 하나의 문항반응 함수(item response function)를 추정한다. 그러나 이러한 가정이 항상 성립하는 것은 아니며, 하나의 모집단내에서도 측정하고자 하는 심리특질과 문항 반응의 관계가 서로 다른 하위 잠재집단들이 존재할 수 있다(von Davier, Rost, & Carstensen, 2007). 혼합문항반응이론은 잠재집단분석(Latent Class Analysis)과 문항반응이론(Item Response Theory)을 결합시킨 모형으로, 단일차원의 척도에서 관찰된 응답 유형을 바탕으로 모집단 내에 존재하는 상이한 하위 잠재집단을 찾아내고, 하위 잠재 집단에 따라 서로 다른 문항 모수를 추정할 수 있도록 한다(Carter, Dalal, Lake, Lin, & Zickar, 2011)¹⁾.

최근 혼합문항반응이론을 적용하여 문항 응

1) 잠재집단분석(Latent Class Analysis)을 이용하면 JDI의 하위척도들에 대한 하위잠재집단 분석을 동시에 실시할 수 있다. 그러나 잠재집단분석은 임계치 모수의 순차성을 가정하기 때문에 임계치 모수들을 자유롭게 추정하여야 하는 본 연구에 적절하지 않다. 또한 JDI 하위척도들은 일반적으로 독립된 척도로서 사용되며 하위척도간 상관도 높지 않다는 점을 고려하여 본 연구에서는 각 하위척도별로 단일차원의 혼합문항반응분석을 적용하였다(예: Wang & Russell, 2005).

답유형에 따른 하위 잠재집단을 찾기 위한 연구가 활발히 진행되고 있다. 예를 들어, Eid와 Rauber(2000)은 독일의 한 기업의 종업원들을 대상으로 측정된 상사에 대한 만족 척도에서 양 극단의 응답범주를 선호하는 집단과 전체 구간의 응답범주를 고르게 선호하는 두 개의 하위 잠재집단을 확인하였다. 또한 Zickar, Gibby, 그리고 Robie(2004)는 미국 내 구직자 표본의 성격검사 응답 자료를 바탕으로 솔직한 응답을 하는 집단(no faking), 약간 허위응답을 하는 집단(slight faking), 그리고 심각하게 허위응답을 하는 집단(extreme faking)이 존재한다는 것을 확인하였다. Hernandez 등(2004)과 Maij-de Meij 등(2008)도 성격검사에서 중간응답에 대한 선호도와 기피도에 따라 한 개 이상의 하위 집단이 존재한다는 사실을 밝혀내었다.

이처럼 기존 연구들은 다양한 맥락에서 서로 다른 응답유형에 따른 하위 잠재집단의 존재와 그 특성을 알아보기 위해 혼합문항반응이론을 사용하여 왔다. 본 연구에서는 이 모형을 적용하여 한국과 미국 집단 각각에서 하위 잠재집단을 살펴보고, 이들이 두 문화집단에서 어떻게 다르게 나타나는지를 알아보고자 하였다. 즉, 각 문화집단 내 존재하는 잠재집단의 수와, 각 하위 집단의 크기 모수(class size parameter), 임계치 문항모수(threshold item parameter), 잠재특질모수(person parameter), 그리고 관찰점수(observed score)의 분포를 한국과 미국집단 간에 비교함으로써 한국과 미국 간에 “?” 응답범주와 관련된 차이를 보다 심도 있게 알아보고자 하였다.

기본적으로 혼합부분점수모형은 부분점수모형(Partial Credit Model; Masters, 1982)의 확장이기 때문에, 우선 부분점수모형을 간략히 살펴

볼 필요가 있다(식 3).

$$P(U_{ij} = h) = \frac{\exp(h\theta_i - \sigma_{jh})}{\sum_{s=0}^m \exp(s\theta_i - \sigma_{js})}, \quad (3)$$

$$\text{with } \sigma_{jh} = \sum_{s=1}^h \tau_{js}.$$

여기서, $P(U_{ij} = h)$ 는 m 개의 선택지를 가진 문항(j)에 대하여 개인(i)가 보기 h 를 선택할 확률을 나타내는데, 이 확률은 개인(i)의 잠재특질(θ)과 문항 임계치의 합과의 수리적 함수관계에 의해 표현된다. 이때, 모수 τ_{js} 는 잠재특질(θ)의 척도상에서 대한 문항(j)의 임계치(i) 위치를 나타낸다(Carter et al., 2011; Zickar et al., 2004).

부분점수모형은 Rasch계열의 다분 문항반응모형으로서 모든 문항 변별도가 동일하다고 가정한다. 또한 임계치에 대한 어떠한 제약도 하지 않기 때문에, 임계치 모수의 특성을 자유롭게 검증할 수 있다. 부분점수모형에서 문항의 임계치 모수는 각각의 범주를 구분하는 기준으로 작용하며, 만일 m 개의 범주가 있다면, $m-1$ 개의 임계치 모수가 있게 된다. 또한, 문항이 순차적인 범주형 자료라면, 두 번째 임계치는 첫 번째 임계치보다, 세 번째 임계치는 두 번째 임계치보다 더 높아야 한다. 만일 임계치들이 이러한 형태를 보이지 않는다면, 해당 응답범주가 순차적 자료가 아님을 시사한다(Eid & Rauber, 2000). 또한 임계치 간의 고른 간격은 각각의 범주가 일정한 간격을 가진다는 점을 시사하기 때문에 임계치 간의 간격 또한 중요하게 해석된다. 결과적으로, 임계치 간의 순서가 뒤바뀐다든지 간격이 고르지 않다면 고전적 심리측정학 가정이 위배되었을 가능성을 나타낸다(Rost, 1997; Zickar et

al., 2004).

부분점수 모형은 해당모델이 전체 표본에 동일하게 적용된다고 가정하므로 모든 집단의 구성원에게 동일한 임계치 모수들이 적용된다. 이에 반해 혼합부분점수모형은 전체 표본 집단이 몇 개의 하위 잠재집단($g=1, \dots, G$)으로 나뉠 수 있으며, 하위 잠재집단은 서로 다른 임계치모수와 특질모수(θ)를 지닌다고 가정한다. 혼합부분점수 모형은 다음과 같이 표현된다(식 4).

$$P(U_{ij} = h) = \sum_{g=1}^G \pi_g \frac{\exp(h\theta_{ig} - \sigma_{jhg})}{\sum_{s=0}^m \exp(s\theta_{ig} - \sigma_{jsg})}, \quad (4)$$

$$\text{with } \sigma_{jhg} = \sum_{s=1}^h \tau_{jsg}.$$

식 4에서 π_g 는 혼합비율 모수(mixing proportions parameter)라 불리며, 개인(i)이 g 번째 잠재 집단에 소속될 확률 값을 나타낸다. 혼합부분점수모형에서는 각각의 혼합비율 모수(π_g)가 고려되어 결국 응답확률이 집단에 따라 다르게 구해지는 것이다. 따라서 혼합부분점수 모형에서는 잠재집단별로 임계치 모수의 특성을 살필 수 있으며, 이를 통해 서로 다른 응답경향을 가지는 하위 잠재집단을 구별할 수 있다. 각 하위 집단의 크기 모수와 집단별 문항모수들은 기대최대화(Expectation-Maximization, EM) 알고리즘을 통하여 동시에 추정된다.

본 연구는 이러한 혼합부분점수모형의 특징을 이용하여 한국과 미국 표본 각각에 서로 다른 응답경향을 가진 하위 잠재집단이 존재하는지, 그리고 만약 한국과 미국에서 “?”를 자주 사용하는 하위 잠재집단이 존재한다면, 이 잠재집단의 특징이 한미 간에 어떠한 차이

를 보이는 지를 살펴보았다.

방 법

참가자 및 자료수집 방법

한국 표본은 932명의 직장근로자로 구성되었다. 다양한 조직에 근무하는 직장인 표본을 얻기 위해 한 인터넷 조사패널에 소속된 20세 이상에서 60세 미만의 직장인 420,000명 중 무선적으로 선택된 5,776명의 응답자에게 설문참여를 요청하는 이메일을 보냈다. 그중 1,624명이 해당 이메일을 읽었고, 정해진 설문기간 중에 참여한 932명(남=468명, 여=464명)의 자료를 본 연구에 이용하였다. 한국 표본집단의 연령 평균은 38.67세($SD=9.87$)였다.

미국 표본자료는 미국 볼링그린 주립대학교 JDI 센터로부터 제공받은 미국 규준집단(Lake, Gopalkrishnan, Sliter, & Withrow, 2010) 자료의 일부를 이용하였다. 규준집단 자료는 교육수준과 직종을 고려하여 미국 전 지역에서 18세 이상의 노동인력을 대상으로 추출된 1,458명의 인터넷 설문응답으로 구성되었다. 한미 집단 간 표본 자료의 대응성을 최대한 확보하기 위해, 우선 한국 표본의 연령대와 동일하게 미국 규준 집단 중 20세 이상 60세 미만의 표본을 추출하고($N=1,253$), 이 중 무작위 재표집(random resampling) 과정을 거쳐 추출된 932명(남=402명, 여=530명)의 응답을 본 연구의 자료로 사용하였다. 미국 집단은 평균 42.53세($SD=9.95$)이었다.

측정도구

JDI는 우수한 타당도와 신뢰도를 바탕으로 20개 이상의 언어로 번역되는 등 전 세계적으로 가장 널리 쓰이고 있는 직무만족 척도이다(예: Bowling, Hendricks, & Wagner, 2008; Kinicki, McKee-Ryan, Schriesheim, & Carson, 2002). Smith, Kendall, 및 Hulin(1969)에 의해 최초로 개발된 이래, 몇 차례의 개정을 거쳐 최근 2009년도 개정본(Lake et al., 2010)이 발표되었다. JDI는 직무자체(work; 18 문항), 보상(pay; 9문항), 승진(promotion; 9문항), 상사(supervision; 18 문항), 그리고 동료(coworker; 18 문항)의 다섯 개 하위 척도로 구성되어 있다. 각각의 문항은 한 단어의 형용사(예: 지루한)로 기술되어 있고, 응답자는 각 형용사가 자신의 직무를 반영하는지 여부를 “아니오”, “?” 그리고 “예” 중의 하나로 선택하게 되어 있다. 일반적으로 “아니오”는 0점, “?”는 1점, 그리고 “예”는 3점으로 채점된다.

JDI의 한국어 번역본은 일반적인 표준번역 절차(Brislin, 1970)를 따라 개발되었다. 그 과정에는 영어를 주언어로 사용하는 고등학교와 대학교를 나왔거나 미국에 10년 이상 생활한 3명의 이중언어자와 2명의 영어 원어민이 참여하였다. 구체적으로 이중언어자 2인이 독립적으로 JDI를 한국어로 번역한 것과 기존 JDI 번역본(Tak & Downey, 1991)을 토대로 토의를 거쳐 1차 번역안을 마련하였다. 이를 다시 제 3의 이중언어자가 영어로 역번역을 하여, 이 역번역본과 JDI 원문항을 영어 원어민 2명에게 의미의 동일성에 대한 검토를 받았다. 그 결과, 4개의 문항에 대해 원문항과 역번역된 문항 간에 의미 차이가 있는 것으로 지적받고, 이 4개의 문항은 1차 한국어 번역에 참여한 2

명이 토론을 통해 수정하여 최종 번역본을 만들었다.

분석

측정동일성 검증에 앞서, 한국과 미국 표본 집단 각각에 대해서 응답 빈도 비율을 확인하였다. 이어서, 양 국 표본 각각에 대해 5요인 모형이 적합한지를 알아보기 위해 범주적 확인적 요인분석을 실시하였다. 범주적 확인적 요인분석은 MPLUS 6.0(Muthén & Muthén, 2010)을 이용하였으며, 범주형 자료를 가장 적합하다고 알려진 평균 및 변량 조정 가중 최소자승(Weighted Least Squares Mean and Variance-adjusted, WLSMV) 추정법을 적용하였다(Brown, 2006).

측정동일성 검증을 위한 다집단 분석은 WLSMV 추정법과 쉼타 모수방법(Theta parameterization)²⁾을 사용하였으며, 구체적인 방법은 Hoffman(개인교신, L. Hoffman; 2012. 8. 12)이 제안한 방법을 따랐다. 각 단계별 측정동일성 모형을 살펴보면 다음과 같다. 우선 “측정틀 동일성” 단계에서는 한미 양 집단에서 모든 요인부하와 임계치 모수들을 자유추정하였다. 양 집단에 요인평균(factor mean), 요인분산(factor variance), 잠재반응변수(y^*)의 오차분산(residual variance)을 각각 0, 1, 1로 고정하였다. 다음 “측정단위 동일성” 모형에서는 양

2) CCFA의 모수 추정시 모형식별을 위해 쉼타 모수 추정법 또는 델타(Delta) 모수추정법이 사용된다. 쉼타 추정법은 연속적 잠재응답 변인(X^*)의 오차를 1로 제약하나, 델타 추정법은 이를 $1-\lambda^2$ 로 제약하는 방법을 쓴다. 이에 대한 보다 자세한 내용은 다음을 참조하기 바란다(Millsap & Yun-Tein, 2004, Muthén & Muthén, 2010).

집단간 요인부하를 동일하게 제약하였으며, 모든 임계치 모수들은 자유롭게 추정하였다. 이때 모형식별(model identification)을 위하여 미국집단에서는 요인분산을 1로 고정하였으나, 한국집단에서는 자유롭게 추정하였다. 요인평균은 양 집단 모두에서 0으로 고정하였다. 그리고 오차분산은 양 집단에 동일하게 1로 제약되었다. 마지막으로 “임계치 모수 동일성” 모형에서는 양 집단에서 모든 요인부하와 임계치 모수들을 동일하게 제약하였다. 모형식별을 위하여 미국 표본에서는 요인평균과 요인분산을 각각 0과 1로 고정하고, 한국 표본에서는 각각을 자유 추정하였다. 모든 오차분산은 양 집단에서 모두 1로 고정하였다. 이와 같이 각의 단계에서 잠재반응변수(y^*)의 오차분산을 모두 1로 고정하는 방식은 문항반응이론에서 모형식별을 위해 이용하는 방식과 동일하다.

단일 모형의 적합도는 CFI(Comparative Fit Index), TLI(Tucker-Lewis Index), 그리고 RMSEA(Root-Mean-Squared Error of Approximation)를 이용하여 판단하였다. 일반적으로 CFI와 TLI는 .90이상이면 좋은 적합도를 가진다고 해석한다. RMSEA은 0.06보다 작으면 좋은 적합도, 1.0보다 크면 부적합한 적합도라고 본다(Hu & Bentler, 1998). 모형간의 적합도 차이 검증은 CFI와 RMSEA의 차이값(ΔCFI 그리고 $\Delta RMSEA$)에 근거하였다. 특히, ΔCFI 는 표본 크기에 민감하게 반응하지 않는 장점을 지니며, ΔCFI 가 .01이 넘으면, 비교모형이 기저모형보다 통계적으로 유의미하게 모형 적합도가 떨어진다고 해석할 수 있다(Cheung & Rensvold, 2002). $\Delta RMSEA$ 또한 .01이 넘을 경우, 비교모형이 기저모형보다 모형 적합도가 낮다고 본다(Chen, 2007). 또한 MPLUS 프로그램의

DIFFTEST 옵션에서 WLSMV 추정법 사용시 제공하는 수정된 카이스퀘어의 비교검증값($MD\Delta\chi^2$)도 함께 보고하였다($MD\Delta\chi^2$ 에 대한 보다 자세한 내용은 Muthén & Muthén, 2010을 참조할 것).

한국과 미국 표본에서 JDI의 응답유형에 따른 하위 잠재집단이 존재하는지를 알아보기 위해서는 WINMIRA 프로그램(von Davier, 2001)을 사용하여 혼합부분점수 문항반응분석을 실시하였다. 혼합문항반응이론 분석에 앞서 각 하위 척도에 대한 단일차원 검증을 실시하였다. 혼합문항반응이론 분석에서 하위 잠재집단의 수는 CAIC(Consistent Aikake Information Criteria; Bozdogan, 1987) 모형 적합도 지수에 근거하여 결정하였다³⁾. 해당 표본의 잠재집단의 수는 CAIC가 감소하다가 증가하는 지점에 해당하는 모형의 집단 수로 결정한다.

마지막으로 한국과 미국 표본 간에 응답유형에 따른 하위 잠재집단의 차이가 있는지를 살펴보기 위해, 한국과 미국 표본 각각에서의 하위 잠재집단의 크기 모수, 임계치 모수, 그리고 잠재집단별 관찰점수 분포를 비교하였다.

결 과

한국과 미국의 응답범주별 응답 빈도 비교

한국과 미국 집단 간 각 응답범주별 응답빈도의 차이를 비교하였다. 모든 문항에서 한국 집단은 미국 집단보다 “?”에 대한 응답빈도가 높았다. 그림 1은 하위척도별로 한국과 미국 표본에서 3개의 응답범주에 대한 응답 비율을

3) BIC를 기준으로 하였을 때도 동일한 결과가 나타나 본 연구에서는 CAIC만을 보고하였다.

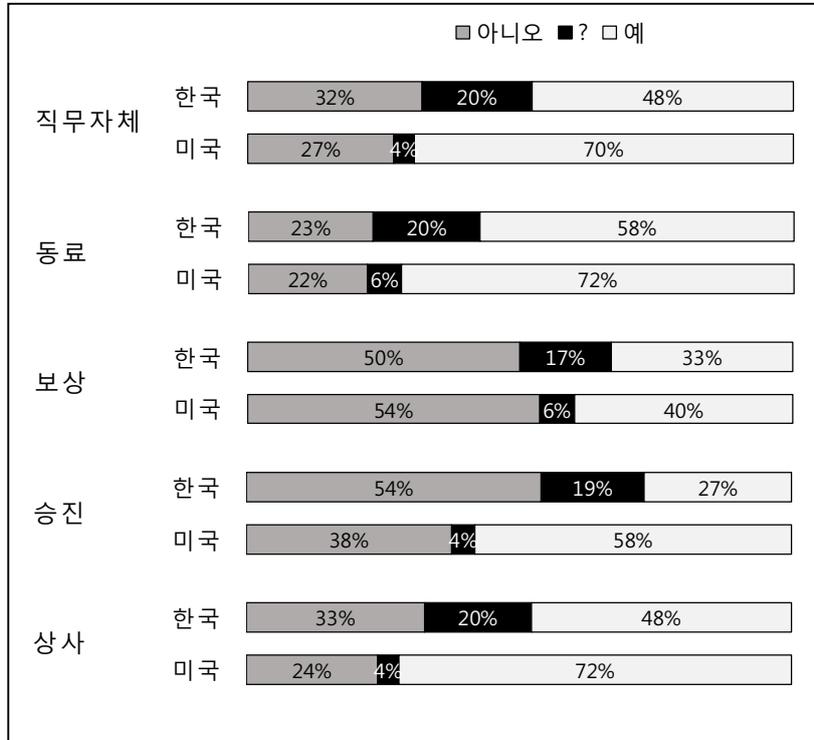


그림 1. 한국과 미국 표본 집단의 응답범주별 응답빈도

보여준다. 한국 집단의 “?” 범주 선택 비율이 미국 집단에 비해 직무자체, 상사, 그리고 승진척도에서는 각각 5.4배, 5배, 4.5배, 동료척도에서는 3.1배, 그리고 보상척도에서는 2.6배가 더 높게 나타났다.

범주적 확인적 요인분석(CCFA) 모형 검증

한국과 미국 간의 측정동일성을 검증하기 전에 우선 각각의 집단에서 5요인 구조(직무 자체, 보상, 승진, 상사, 동료)가 적합한지를 검증하였다. 각 5요인 구조의 첫 번째 문항을 참조지표(reference indicator)로서 1.0으로 고정하는 방식을 사용하였으며, 요인분산과 요인 간 상관은 각각 추정하였다. 표 1에 제시된 바와

같이, 미국의 경우 모든 적합도 지표에서 5요인 모형이 적합한 것으로 나타났다. 한국 표본에서는 CFI와 TLI가 기준에 다소 못 미치는 하였으나, RMSEA가 양호한 적합도를 나타냈다. 따라서 JDI 5요인 구조가 양 집단 모두에 적합한 것으로 판단하고, CCFA를 이용한 다집단분석을 실시하였다.

측정동일성 분석

표 2는 CCFA 다집단 분석을 통해 검증한 1) 측정틀 동일성 모형, 2) 측정 단위 동일성 모형, 3) 임계치 모수 동일성 모형의 전반적 적합도 지수와 모형 간 차이 검증 결과를 보여준다. 측정틀 동일성 모형(모형 1)의 전반적

표 1. 한국과 미국 단일집단 범주적 확인적 요인분석 (CCFA)모형 적합도 결과 비교

	χ^2	df	p	CFI	RMSEA (90% CI)	TLI
한국집단	7874.01	2,474	.001	0.89	0.048 (.047~.050)	0.89
미국집단	5227.13	2,474	.001	0.95	0.035 (.033~.036)	0.95

표 2. 한국과 미국 표본 집단 간의 측정동일성 검증 결과

모형	모형 적합도				모형 차이 검증			MD $\Delta\chi^2$ ^a		
	χ^2	df	p	RMSEA	CFI	Δ RMSEA	Δ CFI	$\Delta\chi^2$	p	df
1. 측정틀동일성	15,046.08	4,948	.001	0.04	0.93					
2. 측정단위동일성	15,334.38	5,015	.001	0.04	0.93	0.00	0.00	558.45	.001	67
3. 임계치동일성	15,475.84	5,154	.001	0.05	0.90	0.01	0.03	5097.42	.001	139

a. MD Δ 는 Mplus DIFFTEST의 결과치 임

적합도 지표가 양호하여($\chi^2(4,948)=15,046.08$, $p<.001$, RMSEA=.04, CFI=.93), 두 집단 간에 측정틀 동일성이 성립하는 것으로 판단하였다. 다음 단계인 측정단위 동일성 모형(모형 2)의 전반적 지표도 양호하였다($\chi^2(5,015)=15,334.38$, $p<.001$, RMSEA=.04, CFI=.93). Δ RMSEA와 Δ CFI를 이용한 적합도 차이 검증 결과, 모형 1과 모형 2의 적합도는 유의미한 차이가 없는 것으로 나타났다(Δ RMSEA=.00, Δ CFI=.00). MPLUS 프로그램에서 CCFA 모형간 차이검증 통계치로 제공하는 MD $\Delta\chi^2$ 는 유의한 차이를 보였으나(MD $\Delta\chi^2(67)=558.45$, $p<.001$), 본 연구의 표본 크기가 상당히 크며 표본 크기가 카이검증값에 영향을 준다는 점을 고려하여 MD $\Delta\chi^2$ 보다는 모형의 적합도와 Δ RMSEA 및 Δ CFI 결과를 바탕으로 해석하였다. 따라서 한미 집단 간에 측정단위 동일성이 성립한다고 결론 내렸다. 마지막으로 임계치 모수 동일성 모형(모형 3)의 전반적 적합도는 모형 1과 2에 비해 상대적으로 낮게 나타났다(χ^2

(5,154)=15,475.84, $p<.001$, RMSEA=.05, CFI=.90). 모형 3을 모형 2와 비교했을 때, MD $\Delta\chi^2$ 값이 유의미하였을 뿐만 아니라 Δ RMSEA과 Δ CFI도 각각 판단 기준을 넘어 두 모형의 적합도 차이가 유의하다고 결론 내렸다(MD $\Delta\chi^2(139)=5097.42$, $p<.001$, Δ RMSEA=.01, Δ CFI=.03). 이는 한국과 미국 표본 집단 간의 임계치 모수 동일성이 성립되지 않는 것, 즉, 문항의 관찰점수를 결정짓는 임계치 모수의 구조가 한국과 미국 집단에서 다르게 적용된다는 것을 의미한다.

한국과 미국 표본 집단 간의 임계치 모수 동일성이 존재하지 않은 것으로 판명됨에 따라, 각 문항에 대하여 한국과 미국 각각의 임계치 모수 구조를 보다 자세하게 살펴보았다. 그 결과, 전반적으로 한국 집단의 임계치 모수가 미국 집단의 임계치 모수보다 높게 나타났다. 즉, “아니오”와 “?”를 구분하는 값은 72개 문항 중 54개 문항(75%)에서 한국 집단의 임계치 모수가 더 높았으며, “?”와 “예”를 구

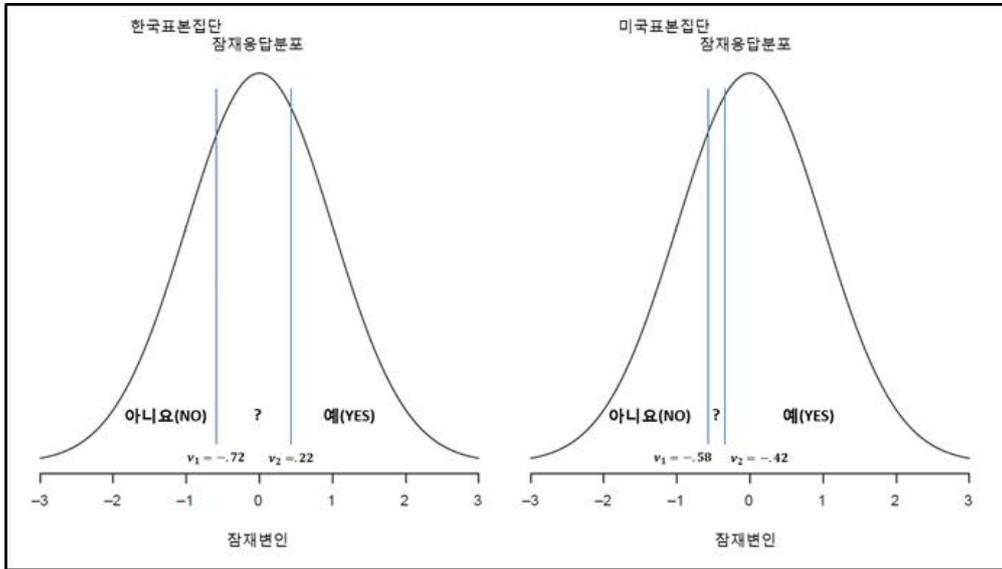


그림 2. 상사 척도의 10번 문항에 대한 한국과 미국 표본 집단의 임계치 구조 비교

분하는 값 역시 72개 중 66개의 문항(92%)에서 한국 집단의 임계치 모수가 더 높았다. 또 다른 흥미로운 차이는 모든 문항에서 한국 집단의 두 임계치 모수 간의 간격이 미국 집단의 모수 간 간격보다 더 넓게 나타났다는 것이다. 이러한 한미 두 집단 간의 임계치 모수의 구조 차이는 앞에서 살펴본 바와 같이 한국 집단의 “?” 범주 선택비율(23.3%)이 미국 집단(4.4%)보다 월등히 높았던 결과와 일치한다.

그림 2는 “상사” 척도의 10번 문항(“내가 얼마나 잘하고 있는지를 알려주는”)의 임계치 모수를 보여준다. 한국 집단의 v_1 와 v_2 는 각각 -.72와 .22인 반면, 미국 집단에서는 각각 -.58와 -.42로 나타났다. 따라서 -.60의 잠재특질 값을 가진 한국 근로자는 그의 잠재특질이 v_1 (-.72)보다 크고 v_2 (.22) 보다 작기 때문에 “?”를 선택하게 된다. 그러나 동일한 수준의 잠재특질을 가진 미국인 근로자는 그의 잠재특질이

v_1 (-.58)보다 작기 때문에 “?”가 아닌 “아니오”를 선택하게 된다. 같은 논리로, .10의 잠재특질을 가진 한국인 근로자는 역시 “?”를 선택하는 반면, 같은 수준의 잠재특질을 가진 미국인 근로자는 “예”를 선택하게 된다.

이러한 한미 간의 차이를 이해하기 위해서는 각 집단 내의 하위 잠재집단의 특성을 살펴볼 필요가 있다. 즉, 각 국가 내에서 “?” 응답범주를 선택하는 것과 관련된 하위집단의 특성과 비중이 다르다면, 이러한 차이가 전체 집단에서의 한미 간의 차이를 이해하는 데 도움을 줄 수 있기 때문이다. 따라서 아래에서는 혼합문화반응모형 분석을 통해 한국과 미국집단 각각에 하위 잠재집단의 특성을 자세히 살펴보았다.

혼합문화반응이론 분석 결과

혼합문화반응분석을 실시하기 전에 우선 각

하위 척도들의 일차원성 가정을 검증하기 위해 범주적인 탐색적 요인분석을 실시하였다. 일반적으로 탐색적 요인분석에서 첫 번째와 두 번째 아이겐(eigenvalue)값의 비율이 3배를 넘으면 문항반응이론의 일차원성 가정을 충족한 것으로 간주된다(Lord, 1980). 범주적 탐색적 요인분석 결과, 직무자체(한국=4.23배, 미국=6.51배), 보상(6.80배, 6.60배), 승진(5.01배, 9.03배), 상사(4.03배, 9.09배), 그리고 동료(5.91배, 6.90배) 등 모든 척도에서 이러한 기준을 넘었기에 각 하위 척도의 일차원성 가정이 충족되는 것으로 결론 내렸다.

혼합문항반응분석을 통해 한미 양 집단에서 각각 잠재집단 개수를 확인하기 위해 각 하위 요인별로 잠재집단의 수를 1개에서 5개까지로 하고 이에 따른 CAIC 지수의 변화를 살펴보았다(예: Carter et al., 2011). 표 3에서 보는 바와 같이, 전반적으로 잠재집단이 1개에서 2개, 2개에서 3개로 증가할 때는 CAIC 지수가 감소하다가 3개에서 4개로 변할 때는 CAIC 지수가

다시 증가하는 것으로 나타났다. 예외적으로, 한국의 상사 척도와 미국의 승진 척도에서는 4개의 잠재집단에서, 미국의 보상척도에서는 2개의 잠재 집단에서 CAIC가 감소하다가 다시 증가하였다. 그러나 이들 경우는 3개의 잠재 집단에서 발견된 CAIC값과의 차이가 미미하였으며, 잠재집단의 크기 모수와 범주간 응답확률 및 잠재집단별 관측점수 분포의 경향성을 보여주는 그래프를 종합적으로 고려했을 때 양국 모두 모든 척도에 대해 3개의 잠재집단을 가정하는 모형이 가장 적합하다고 결론 내렸다.

각 표본 집단에서 나타난 3개의 하위 잠재 집단의 특성을 살펴보면 다음과 같다. 우선 한국 표본에서는 긍정적으로 응답하는 집단, 부정적으로 응답하는 집단, 그리고 “?”을 주로 선택하는 집단으로 비교적 명확하게 구분할 수 있었다. 미국 표본의 경우는 5개의 하위 척도 모두에서 “?”을 선호하는 집단이 존재하였지만, 나머지 2개의 하위집단은 각 척도마

표 3. 한미 표본 집단의 잠재집단 수에 따른 CAIC 적합도 지수

집단	척도	잠재집단의 수				
		1	2	3	4	5
한국	직무	28438.90	26783.26	26453.16	26490.19	26531.71
	동료	26605.31	23638.70	23419.34	23331.27	23445.70
	상사	29199.04	26470.43	26280.47	26212.78	26305.36
	승진	13958.45	12789.48	12740.85	12768.89	12795.76
	보상	14128.97	13454.88	13414.19	13453.02	13538.09
미국	직무	18157.69	17926.26	17887.71	17978.27	18080.16
	동료	19724.85	18776.72	18652.77	18777.21	18957.66
	상사	18360.04	18113.21	18099.35	18332.05	18543.51
	승진	10597.41	10242.52	10114.73	10156.35	10226.70
	보상	10315.92	10229.01	10241.17	10318.41	10448.44

다 긍정적으로 응답하는 집단, 부정적으로 응답하는 집단, 전체응답을 고르게 선택하는 집단, 양극단을 선택하는 집단 중 두 집단의 조합으로 나타났다.

일단 각 표본에서의 하위 잠재집단을 3개로 결정한 후, 한국과 미국 집단의 하위 잠재집단들의 특성을 비교하기 위해 우선 각 잠재집단의 크기 모수를 살펴보았다. 그 결과, 한국과 미국 표본 집단 간의 가장 두드러진 차이는 “?”를 응답으로 선택할 확률이 높은 잠재집단(이후 “?” 선호 집단으로 칭함)의 분포비율에 있었다. 특히 상사(한국=42% vs 미국=14%), 직무자책(37% vs 15%), 보상(30% vs 9%) 등의 하위 척도에서는 한국집단이 미국집단과 비교하여 “?” 선호 집단의 비중이 월등히 높았다. 반면, 나머지 두 척도, 즉, 승진(한국=11% vs 미국=10%)과 동료(한국=11% vs 미국=16%)척도에 있어서는 그 비중이 두 나라에서 거의 유사하게 나타났다. 이처럼 미국 집단에서는 “?”을 선호하는 잠재집단의 비중이 다른 두 개의 잠재집단의 비중보다 매우 낮은 반면, 한국 집단에서는 다른 두 잠재집단과 전반적으로 유사하다는 특징을 알 수 있다.

본 연구의 주된 관심이 한국과 미국 집단 간에 “?”응답범주 사용상의 차이가 존재하는가를 보는 것이기 때문에, 한국과 미국 집단의 “?” 선호 집단의 임계치의 구조를 비교하였다. 한국의 경우, 모든 문항에서 첫 번째 임계치(v_1)보다 두 번째 임계치(v_2)가 더 높은 값을 가졌으며, 이 두 임계치 간의 간격이 비교적 고르게 나타났다. 반면, 미국의 경우에는 두 임계치 간의 간격이 매우 좁고 불규칙하였다. 그 예로서, 그림 3은 상사 척도에 대한 한미 표본 집단의 “?”을 선호하는 잠재집단의

임계치를 보여준다. 상단에 제시된 한국 표본에서는 첫 번째 임계치가 두 번째 임계치보다 일관되게 낮으며, 두 임계치 간의 간격 역시 비교적 고르게 나타났다. 반면, 하단에 제시된 미국 표본에서는 몇몇 문항(문항 1, 15, 17, 18)을 제외하고는, 두 임계치가 거의 겹치거나 첫 번째 임계치가 두 번째 임계치보다 오히려 높게 나타났다. 첫 번째 임계치가 두 번째 임계치보다 낮게 나타난 문항들에서도 그 간격이 한국 집단에 비해 매우 좁았다. 각 응답지가 순차적인 범주로 인식되었다면, 두 번째 임계치의 값이 첫 번째 임계치의 값보다 높아야 한다. 따라서 본 결과는 한국 표본의 “?” 선호집단은 “아니오”, “?”, “예”의 응답범주를 순차적으로 해석한 반면, 미국 표본의 해당 잠재집단은 이 세 응답범주를 순차적으로 인식하지 않았을 가능성을 보여준다.

마지막으로 한국과 미국 내의 각 하위 잠재집단들의 관찰점수의 분포를 비교해보았다. 만약 “?”를 “예”과 “아니오”의 중간 특질을 가진 사람들이 선택한다는 가정이 맞다면, “?”를 선호하는 잠재집단의 관찰점수의 분포는 중간점수의 빈도가 높고 양 극단점수의 빈도가 낮은 정상분포의 형태를 띠 것이다. 반면, “예” 또는 “아니오”를 선호하는 집단의 경우는 관찰점수의 분포가 각각 부적 그리고 정적으로 편포될 것이라고 예상할 수 있다. 마지막으로 양극단을 선호하는 잠재집단에서는 “U”자 곡선의 형태를 띠 것이다. 그런데 이러한 관찰점수의 분포형태는 세 개의 응답범주가 순차적으로 인식할 때만 가능하며, 응답범주의 순차성 가정이 위배된 경우에는 이상에서 기대하는 분포와는 다르게 나타날 것이다.

실제로 한국과 미국 집단내의 하위 잠재집단의 관찰점수의 분포를 확인해 본 결과, 한

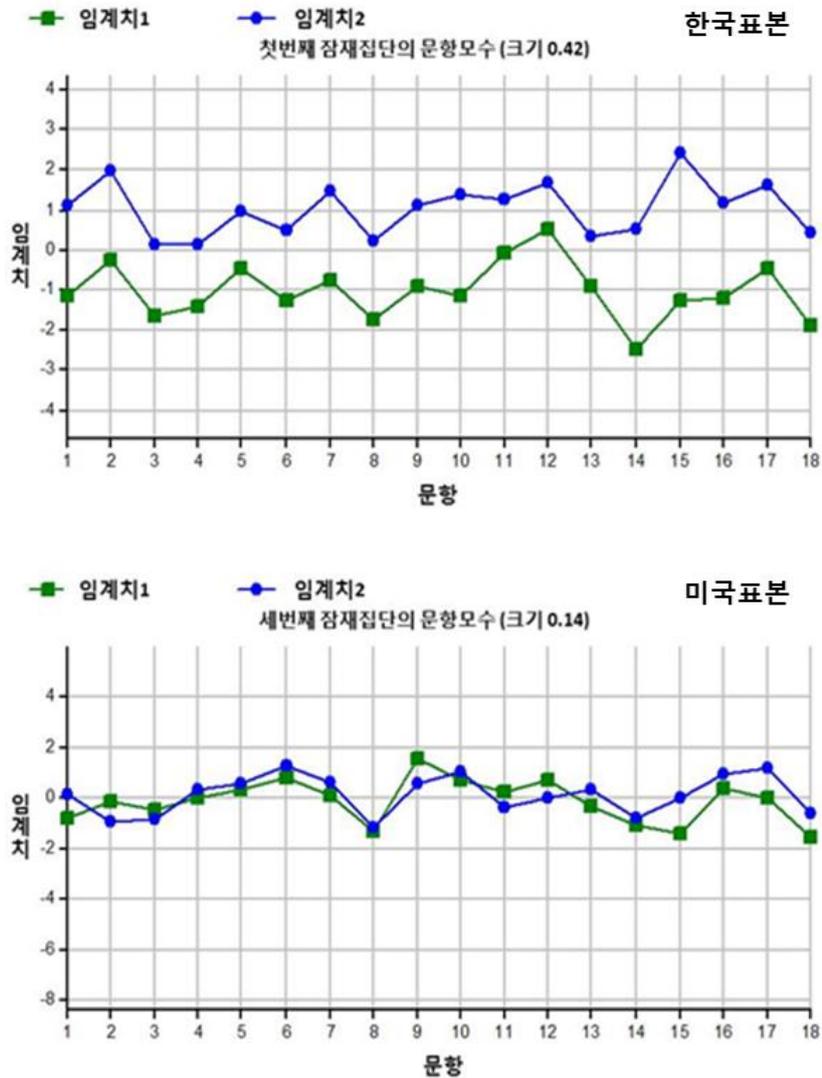


그림 3. 한미 간 “?”을 선호하는 잠재집단의 임계치

국과 미국 집단 간의 차이를 확인할 수 있었다. 그 예로, 그림 4와 5는 각각 한국과 미국 집단에 대해 세 하위 잠재집단별 직무자체 요인의 관찰점수 분포를 보여준다. 그림 4에 제시된 한국 표본의 경우, “?”응답을 선호하는 잠재집단 1(잠재집단의 크기=37%)의 관찰점수 분포는 정상분포를 따르고 있는 반면, 부정응

답을 선호하는 잠재집단 2(32%)는 정적으로 편포되어 있고, 긍정응답을 선호하는 잠재집단 3(31%)은 부적으로 편포되어 있었다. 이와는 달리 미국 표본(그림 5)에서는 확연히 긍정적 응답을 하는 경향을 지닌 잠재집단 1(53%), 비교적 고르게 응답한 잠재집단 2(33%), 그리고 “?”를 선호하는 잠재집단 3(14%) 모두가 부

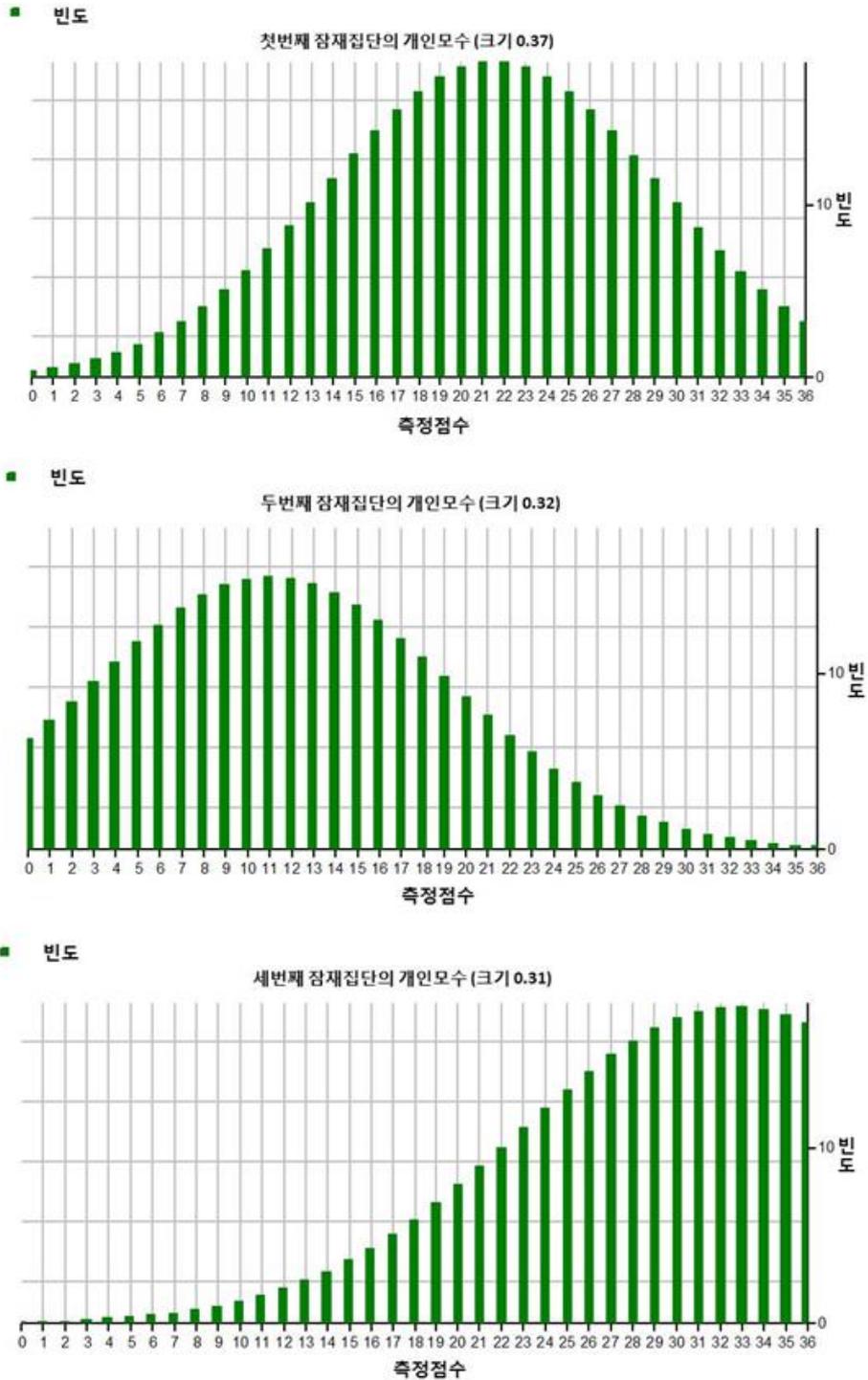


그림 4. 한국표본에서의 직무척도에 대한 잠재집단별 관찰점수 분포

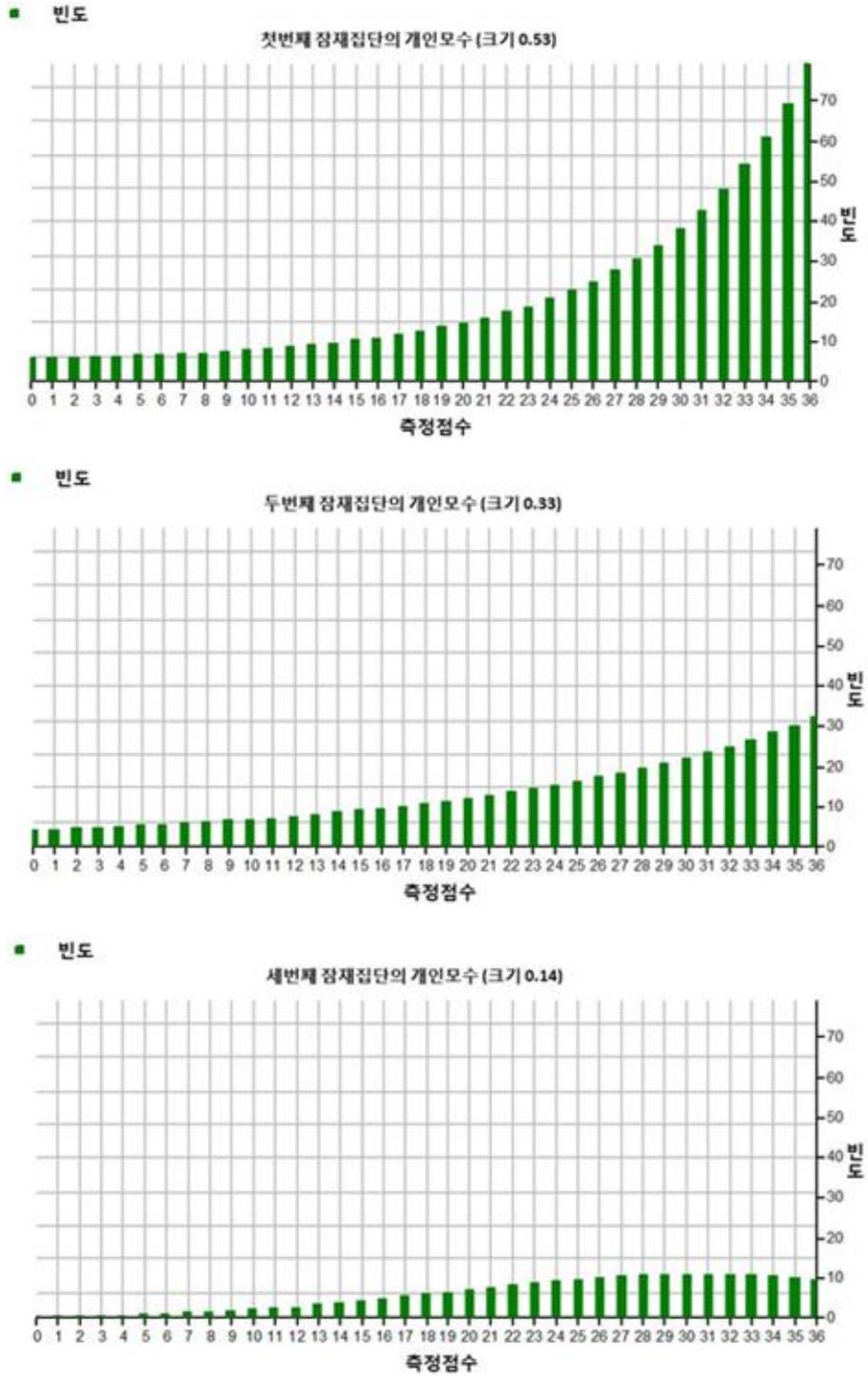


그림 5. 미국 표본에서의 직무척도에 대한 잠재집단별 관찰점수 분포

적 편포를 보이는 등 응답범주를 순차적으로 해석했을 경우와는 다른 분포형태를 나타냈다.

논 의

리커트 문항과 같은 평정척도 방식은 심리학을 비롯한 사회과학에서 매우 폭넓게 사용된다. 그런데 이러한 문항의 중간보기로 자주 사용되는 “?”, “확신할 수 없음”, 그리고 “결정할 수 없음” 등의 응답범주는 그 의미의 모호성과 해석의 자의성으로 인해 오랫동안 심리측정학적 논란의 대상이 되어왔다. 실제로 서구 문화권의 응답자들을 대상으로 한 선행연구들에서는 모든 응답자들이 이러한 유형의 중간보기를 순차적으로 해석하는 것이 아니라 경험적 증거를 제시하고 있다(DuBois & Burns, 1975; Gonzalez-Roma & Espejo, 2003). 이에 본 연구는 “?”를 중간보기로 사용하는 대표적인 척도인 JDI를 이용하여, 한국과 미국 표본 집단내의 모든 구성원이 “예”, “?”, “아니오”의 응답범주를 동일하게 해석하는지, 즉, 응답유형에 따라 서로 다른 하위 잠재집단이 존재하는지, 그리고, 이러한 하위집단의 특성이 “?”와 관련된 한국과 미국 응답자간의 응답경향 차이를 이해하는 데 어떠한 정보를 제공할 수 있는지를 살펴보았다.

JDI 척도에 대한 한국과 미국 표본의 측정동일성을 검증한 결과, 직무만족의 각 하위요인의 측정지표의 구성과 문항과 요인과의 관계는 한국과 미국 표본에서 동일하지만, 관찰점수를 결정하는 임계치 모수의 구조가 다른 것으로 나타났다. 특히 “아니오”와 “?”를 가르는 첫 번째 임계치와 “?”과 “예”를 가르는 두 번째 임계치 간의 간격이 있어서 미국집단

에 비해 한국집단의 간격이 넓은 것으로 나타났다. 이는 한국 응답자들은 미국 응답자들에 비해 “?” 응답범주를 선택할 가능성이 상대적으로 높은 것을 의미한다. 본 연구결과에서 나타난 두 임계치 간의 간격을 고려할 때, 한국 응답자들이 특히 “?”를 많이 선택했다기보다는 미국 응답자들이 “?”를 선택할 확률이 매우 낮았기 때문이라고 볼 수 있다. 이처럼 미국 응답자들의 “?”범주의 선택 비율이 다른 응답범주에 대한 선택비율에 비해 눈에 띄게 낮았던 것은 미국 응답자들의 양극단적인 직무만족도를 가지고 있기 때문일 수도 있지만, 또 다른 가능성은 그들이 “?” 범주를 평정척도의 순차적 범주상의 중간 수준이 아닌 별개의 선택지로 해석했기 때문에 나타난 결과라고 생각해볼 수 있다.

본 연구의 혼합문화반응분석 결과는 후자의 가능성을 지지한다. 혼합문화반응분석을 통하여 한국과 미국 집단 모두에서 “?”를 선호하는 하위 잠재집단의 존재를 확인했는데, 그 특성을 비교해보니 한국과 미국의 응답자들이 “?”를 서로 다르게 사용하는 경향이 나타났다. 우선 “?”를 선호하는 하위 잠재집단의 크기가 한국 표본에 비해 미국 표본에서 전반적으로 매우 낮게 나타났다. 이는 미국 응답자들이 한국 응답자들에 비해 “?” 범주를 선택하는 경우가 상대적으로 낮은 이유를 설명해준다. 나아가, 한국집단에서는 두 번째 임계치가 첫 번째 임계치보다 전반적으로 일관되게 높고 임계치 간의 간격이 넓은 반면, 미국 집단에서는 첫 번째 임계치와 두 번째 임계치 간의 간격이 거의 없거나 첫 번째 임계치가 두 번째 임계치보다 오히려 높은 경우가 발견되었다. 이러한 결과는 한국 응답자들은 “아니오”, “?”, “예”의 세 보기를 일정한 간격을 지니는

순차적 응답범주로 해석하는 반면, 미국 응답자들은 기본적으로 “?”를 선택하는 경우가 드물고 사용하는 경우에도 “?”를 “예”와 “아니오”의 중간지점이 아니라 “선택사항 없음(No option)”과 같은 별도의 명목 변인으로 해석했을 가능성을 지지한다.

이러한 혼합문화반응이론 분석의 결과들은 JDI에 대해 한국과 미국 간의 측정동일성, 특히 임계치 구조 동일성이 성립하지 않는 이유를 보다 자세히 이해하는 데 도움이 된다. 즉, 한국과 미국 간에 임계치 구조가 다르게 나타나는 주된 이유는 한국과 미국 내에 존재하는 “?”를 선호하는 하위 잠재집단의 비중과 특성이 다르기 때문이라고 할 수 있다. 다집단 범주적 확인적 요인분석 결과로는 한국 집단의 임계치가 미국 집단의 임계치보다 일반적으로 높고 한국 응답자들에 비해 미국 응답자들이 “?”를 선택하는 비율이 매우 낮았다는 것을 알 수 있었으나, 혼합문화반응이론 분석을 통해 이러한 차이의 상당부분이 “?”를 선호하는 하위 잠재집단의 구성비율과 특성의 차이에 의한 것임을 알 수 있었다. 집단 간의 측정동일성을 점검하는 가장 주된 목적이 집단 간의 관찰점수의 타당한 비교를 하기 위한 것이라는 점을 고려한다면, 이러한 결과는 측정동등성을 확보하기 위한 보다 효과적인 대안을 개발하는데 유용한 정보로 사용될 수 있을 것이다.

실제로 본 연구결과는 “?”를 중간보기로 사용하는 평정척도 문항의 채점이 우리가 상상하는 것 이상으로 복잡할 수 있음을 보여준다. 본 연구에서 사용된 JDI의 경우, 현행 점수체계에서는 “아니오”, “?”, 그리고 “예” 보기를 각각 “0”, “1”, “3”으로 채점한다. 이는 직무에 덜 만족할수록 “?” 보기를 응답할 확률이 높

다는 선행연구(Smith et al., 1969)를 근거로 채택된 방식이다. 그러나 본 연구결과에서 나타났듯이, 한국 응답자들의 상당수가 속해있는 “?” 선호 잠재집단이 “?”를 중간수준의 잠재특질로 해석하고 응답한다는 점을 고려할 때 “?”에 대한 응답이 0점에서 3점의 점수체계 중 불과 1점을 받는 것은 타당하지 않을 수 있다. 최근, Kulas와 Stachowski(2008)은 JDI의 “?”의 문제점을 지적하며, “?”에 대한 점수를 1점이 아닌 0점과 3점의 중간인 1.5점을 부여해야한다고 주장한바 있다. 한편, “?”를 중간이 아닌 제 3의 선택으로 해석할 가능성이 큰 미국 응답자들의 경우는 현재의 채점방식이나 Kulas와 Stachowski가 제안한 채점방식 모두 적합하지 않을 것이다. 이처럼 한국과 미국의 응답자들이 선택한 “?”가 서로 다른 직무만족도를 반영하는 상황에서는 동일한 방식의 채점을 적용한 점수를 바탕으로 두 집단의 직무만족도를 비교하는 것은 아무런 의미가 없다.

이러한 경우 응답경향성이 다른 두 집단의 점수를 직접적으로 비교하기 위해서는 두 집단의 문항 모수와 잠재특질을 동일 선상에서 치환시켜 비교하는 척도동등화(Linking, Kolen & Brennan 2004) 방법을 사용할 수 있다. 그러나 한국과 미국의 집단 내부에도 상이한 응답 특성을 지닌 하위 잠재집단이 존재한다는 점을 고려하면, 집단 내외 집단 간의 이중의 복잡한 척도동등화 과정이 요구된다. 따라서 보다 근본적이면서도 현실적인 해결책은 중간보기의 본래 취지에 부합하도록 “?” 보기를 명시적으로 중간수준을 암시하는 다른 보기로 대체하는 방법을 고려해봐야 할 것이다. 선행 연구들(예: Andrich et al., 1997; Gonzalez-Roma & Espejo, 2003) 또한 “불확실(Not sure)”, “상관 없음(Indifferent)”, “미결정(Undecided)”, “?”과 같

이 중간특질을 명확히 반영할 수 없는 표기의 사용을 지양해야한다고 주장한바 있다. 한국과 미국 간의 측정동일성의 위반이 한국과 미국 내에서 “?” 응답범주를 서로 다르게 해석하는 하위 잠재집단의 구성 비율의 차이와 “?”에 대한 해석의 상이성에서 기인하였을 가능성을 고려해 볼 때, 명확한 중간 보기의 해석을 유도하는 표기를 사용하면 집단간 그리고 집단내 측정동일성을 향상시키는 효과를 볼 것이라고 기대된다.

물론 본 연구의 결과는 JDI라는 특정한 검사에 대한 결과이기 때문에 향후 연구에서는 이러한 결과가 다른 측정도구에도 일반화될 수 있는지, 특히 3점 척도가 아닌 5점 또는 7점 등의 척도를 사용하는 문항에서도 이러한 결과가 나타나는지를 비교 문화적으로 검증하여야 할 것이다. 또한 본 연구의 결과가 “?”에만 국한되는지 아니면 유사한 종류의 중간보기로 간주되는 “확신할 수 없음(Not sure)”, “결정할 수 없음(Undecided)”, “분명치 않음” 등에도 적용되는지를 알아보아야 할 것이다.

비록 본 연구결과들은 한국과 미국의 응답자들이 “?”를 차별적으로 해석하고 사용한다는 증거를 제시하고 있지만, 향후에는 이를 보다 직접적으로 검증하기 위한 추후 연구가 필요하다. 예를 들어, 한국과 미국 집단 각각에 대해 중간보기를 “?”, “선택없음”, 그리고 “중간”으로 각각 다르게 표기된 세 개의 설문지를 동일한 개인들에게 일정한 시간적 간격을 두고 세 차례 반복 측정한다면, 실제 “?”보기의 해석이 각 집단별로 “선택없음”과 “중간지점” 중 각각 어디에 가까운지를 보다 직접적으로 확인해 볼 수 있을 것이다. 또 다른 방법은 “소리 내어 생각하기(think aloud)” 방법을 이용하여, 응답자들에게 각 응답범주를 선

택할 때마다 그 이유를 말로 표현하도록 함으로써 “?”을 실제로 어떻게 해석하고 선택하는지를 알아볼 수 있을 것이다.

한편, 향후 연구에서는 “?” 혹은 중간 응답지를 보다 자주 선택하는 경향에 영향을 미치는 변인들을 살펴보는 것도 흥미로운 연구가 될 것이다. 관찰점수와 잠재특질과의 관계가 중간보기를 다르게 사용하는 하위 잠재집단에 따라 상이하게 나타난다면, 이러한 하위 잠재집단에 대한 소속여부에 영향을 주는 선행 변인들을 찾아 볼 필요가 있다. 최근 Carter와 그의 동료들(2011)은 경영층에 대한 신뢰, 직무근속기간, 나이, 인종, 성별 등이 JDI자료에 대한 서로 다른 응답경향을 가진 하위 잠재집단 중 어떤 집단에 속하는지와 일정 정도 관련이 있는 것으로 나타났다. 향후 연구에서는 또 다른 성격적인 특질(예, 성실성, 개방성 등)이나 개인이 속해 있는 조직문화가 구성원의 “?”의 해석과 실제 사용에 어떻게 영향을 미치는지를 살펴보는 것도 흥미로운 것이다.

본 연구의 가장 큰 제한점은 한국 표본자료의 대표성과 미국 기준집단과의 대응성과 관련이 있다. 본 연구의 한국 표본은 특정 인터넷 패널에서 편의표집 하였기 때문에 한국 근로자를 얼마나 대표적으로 반영할 것인지에 대해서는 판단하기 어렵다. 또한 본 연구에서는 표본자료의 대응성을 위하여 미국의 기준집단을 한국 표본과 동일한 연령대에 맞추어 재표집하기는 하였으나, 그 외 다른 주요 변인(예: 교육수준, 직업 등)상에서 두 집단이 동일한지에 대한 보장을 할 수 없다. 따라서 향후 연구에서는 각 비교 집단의 대표성을 최대한 보장하면서도 집단 간 표본의 대응성을 고려할 수 있도록 자료가 수집되어야 할 것이다.

그럼에도 불구하고 본 연구는 평정척도의

중간보기로 자주 사용되는 “?” 응답범주의 해석이 문화 간에서 차이가 날 수 있으며, 이러한 차이가 각 문화 집단내의 하위 잠재집단의 특성에 기인할 수 있다는 것을 밝혔다는 점에서 의의를 가진다. 이러한 결과는 “?”를 사용하는 기존 척도들의 채점 및 서로 다른 문화권에 속한 응답자들 간의 점수 비교에 중요한 시사점을 제공한다. 이러한 문제의 보다 근본적인 해결을 위해서는 응답범주를 모든 사람이 동일하게 해석하고 사용할 수 있도록 명확하게 제시하는 방식으로 기존 척도를 수정하는 것이 필요하며, 동시에 이러한 원칙은 새로운 척도를 개발하는데도 반드시 참조되어야 할 것이다.

참고문헌

- Andrich, D., Jong, J., & Sheidan, B. E. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. In J. Rost & R. Langeherine (Eds.) *Applications of latent trait and latent class models in the social sciences* (pp.58-68). Munster, Germany: Waxman Verlag.
- Andrich, D., & Schoibroeck, L. V. (1989). The general health questionnaire: a psychometric analysis using latent trait theory. *Psychological Medicine*, 19, 469-485.
- Bowling, N. A., Hendricks, E. A. & Wagner, S. H. (2008). Positive and negative affectivity and facet satisfaction: A meta-analysis. *Journal of Business and Psychology*, 23, 115-125.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345-370.
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1, 185-216.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of Ipsative data in forced-choice questionnaires. *Psychological Methods*, 18, 36-52.
- Carter, N. T., Dalal, D. K., Lake, C. J., Lin, B. C., & Zickar, M. J. (2011). Using mixed-model item response theory to analyze organizational survey responses: An illustration using the Job Descriptive Index. *Organizational Research Methods*, 14, 116-146.
- Cattell, R. B., & Cattell, H. P. (1995). Personality structure and the new fifth edition of the 16PF. *Educational and Psychological Measurement*, 55, 926-937.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464-504.
- Cheung, K. C., & Mooi, L. C. (1994). A comparison between the rating scale model and dual scaling for Likert scales. *Applied Psychological Measurement*, 18, 1-13.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233-255.
- Dubois, B., & Burns, J. A. (1975). An analysis of the meaning of the question mark response

- category in attitude scales. *Educational & Psychological Measurement*, 35, 869-884.
- Eid, M., & Rauber, M. (2000). Detecting measurement invariance in organizational surveys. *European Journal of Psychological Assessment*, 16, 20-30.
- Gonzalez-Roma, V. & Espejo, B. (2003). Testing the middle response categories “Not sure”, “In between” and “?” in polytomous items. *Psicothema*, 15, 278-284.
- Hernández, A., Drasgow, F., & González-Romá, V. (2004). Investigating the functioning of a middle category by means of a mixed-measurement model. *Journal of Applied Psychology*, 89, 687-699.
- Hernández, A., Espejo B., & González-Roma, V. (2006). The functioning of central categories middle level and sometimes in graded response scales: does the label matter? *Psicothema*, 18, 300-306.
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424-453.
- Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology*, 16, 131-152.
- Kieruj, N. D., & Moors, G. (2010). Variations in response style behavior by response scale format in attitude research. *Journal of Public Opinion Research*, 22, 320-342.
- Kinicki, A. J., McKee-Ryan, F. M., Schriesheim, C. A., & Carson, K. P. (2002). Assessing the construct validity of the Job Descriptive Index: A review and meta-analysis. *Journal of Applied Psychology*, 87, 14-32.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer-Verlag.
- Kulas, J. T. & Stachowski, A. A. (2009). Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality*, 43, 489-493.
- Lake, C. J., Gopalkrishnan, P., Sliter, M. T., & Withrow, S. (2010). The Job Descriptive Index: Newly updated and available for download. *The Industrial-Organizational Psychologist*, 48, 47-49.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale NJ: Erlbaum.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Maij-de Meij, A. M., Kelderman, H., & van der Flier, H. (2008). Fitting a mixture item response theory model to personality questionnaire data: Characterizing latent classes and investigating possibilities for improving prediction. *Applied Psychological Measurement*, 32, 611-631.
- Millsap R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, 39, 479-515.
- Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality and Quantity*, 42, 779-794.

- Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*, 517-529.
- Riordan, C. M., & Vandenberg, R. J. (1994). A central question in cross-cultural research: Do employees of different cultures interpret work-related measures in an equivalent manner? *Journal of Management, 20*, 643-671.
- Rost, J. (1997). Logistic mixture models. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 449-463). New York: Springer.
- Smith, P. C., & Kendall, L., & Hulin, C. I. (1969). *The measurement of satisfaction in work and retirement: A strategy for the study of attitudes*. Chicago: Rand McNally.
- Tak, J., & Downey, R. G. (1991). Assessing construct validity of R-JDI in Korea. *한국심리학회지: 산업 및 조직, 4*, 87-91.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-69.
- von Davier, M. (2001). *WINMIRA2001: Windows mixed Rasch model analysis {Computer software and User manual}*. Kiel, the Netherlands: Institute for Science Education.
- von Davier, M., Rost, J., & Carstensen, C. H. (2007). Introduction: Extending the Rasch model. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications*. New York: Springer.
- Wang, M., & Russell, S. S. (2005). Measurement equivalence of the Job Descriptive Index across Chinese and American workers: Results from confirmatory factor analysis and item response theory. *Educational and Psychological Measurement, 65*, 709-732.
- Wilde, G. J. S. (1970). *Neurotische labilität gemeten volgens de vragen - lijstmethode*. Amsterdam: Van Rossem.
- Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current and future directions. *Psychological Methods, 12*, 58-79
- Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory. *Organizational Research Methods, 7*, 68-190.

1차원고접수 : 2014. 01. 21.

수정원고접수 : 2014. 06. 20.

최종게재결정 : 2014. 07. 08.

**Differential use of middle category “?” in
Job Descriptive Index Between Korean and American Samples:
Application of mixed-model item response theory**

Philseok Lee¹⁾

Seokjoon Chun¹⁾

Sunhee Lee²⁾

¹⁾Department of Psychology, University of South Florida

²⁾Department of Psychology, Chungnam National University

The present study investigated differential use of the middle category “?” option between Korean and American workers using the 2009 version Job Descriptive Index(JDI). For this purpose, a multi-group categorical confirmatory factor analysis and a mixed-model item response theory(MM-IRT) analysis were conducted across a Korean worker sample and American normative sample (N=932 for each). The multi-group analysis result supported configural and metric invariance across the groups. However, the threshold structure was not invariant across the groups. The MM-IRT analysis identified three latent subgroups within both Korean and American samples. Overall, the Korean group has larger subgroup class size favoring the “?” option than the American group. An inspection of item parameters revealed the threshold locations of the mixed partial credit model were frequently disordered for the American group, but not for the Korean group. The implications of the results and future research issues were discussed.

Key words : Likert item, middle category response option, measurement invariance, mixed-model item response theory, Job Descriptive Index