

직무만족 측정의 동질성

박 동 건

아주대학교

심리학을 비롯한 사회과학, 교육학, 경영학 등에서의 연구에 외국의 적성검사, 인성검사 혹은 태도 질문지 등이 빈번히 활용되고 있다. 특히 외국의 심리검사나 태도 질문지가 번역, 변안 혹은 수정되어 실험연구, 현장연구 등에서 광범위하게 사용되고 있다. 이때 미묘한 오역 혹은 검사나 검사 문항에 의해서 측정되는 심리특질의 문화적 편이에 의한 측정의 비동질성의 문제는 대단히 중요한 연구 주제임에도 소홀히 다루어져 왔다. 본 연구에서는 비교언어적, 비교문화적 측정의 동질성을 평가하는 데 우수하다고 인정되고 있는 심리측정이론인 문항반응이론(Item Response Theory; IRT)에 근거한 통계적 절차를 직무만족측정의 동질성 연구에 적용해 보았다. 미국에서 직무만족측정을 위해 널리 쓰이는 직무기술지표(Job Descriptive Index; JDI)를 한글로 번역하고 원본과 번역본의 측정의 동질성을 개별 문항별로 판정해 보았다. 총 90문항중 30문항이 편의(biased)된 것으로 밝혀졌다. 원문항과 번역문항을 비교할 때 문항반응이론의 추정방법이 지닌 척도 미결정성 때문에 원검사와 번역검사의 측정원점과 측정단위를 갈래하는 척도 동등화(scale linking)가 필요하다. 편의된 문항이 포함되어 개별 척도가 오염되어 있는 상황에서의 척도 동등화는 문항반응이론의 수학적모형의 문항모수 추정을 왜곡시킬 가능성이 있기 때문에 척도 정화가 선행되어야 한다. 본 논문에서 적용한 척도정화(scale purification)절차의 필요성과 유효성이 입증되었다. 심리검사나 태도측정도구를 개발 혹은 번역할 때 문항반응이론에 근거한 추정과 동등화 방법이 유용하게 적용될 수 있음이 본 연구를 통해 실증되었다.

번역은 우주의 진화과정에서 잉태된 가장 어려운 과업이라는 다소 과장된 주장이 있지만 번역은 비교언어 및 비교문화연구과정에서 주요한 난관임에 틀림없다. 언어는 문화를 정의하는 특징이라고 일반적으로 간주되므로 문화적인 차이와 유사성에 관한 연구에서 언어의 탐구는 필수적이라고 할수있다. 그러나 심리특질의 문화적 유사성과 차이가 연구의 대상일 때 언어의 차이는 심리검사점수를 바탕으로 한 결론의 타당성을 저해하는 측정의 문제를 야기한다. (Candell & Hulin, 1987)

심리검사의 목적이 외국어 이해력의 평가가 아닌 심리검사는 수검자의 모국어로 실시되어야 한다. 그러나 번역의 과정에는 측정의 부동성의 문제가 필연적으로 수반된다. 즉, 원검사가 번역본과 동질적인지의 여부가 문제가 된다. 검사번역의 동등성 문제는 되번역(back-translation), 탈편중화(decentering) 등의 다양한 번역절차를 결합적으로 사용하고 이와 더불어 수검자의 응답을 바탕으로한 전통적인 통계분석을 행함으로써, 평가되고 보장될 수 있다고 주장한 연구자도 있다(Brislin, 1986). 이 단계는 필요한 단계이지만 측정의 동등성을 보장하기 위한 충분한 절차는 아니다 (Hulin, Drasgow & Parsons, 1983). 측정의 동등성(measurement equivalence)

* 이 논문은 1989년도 교육부 지원 한국학술진흥재단의 신진교수 학술연구조성비에 의하여 연구되었음. 필자는 현재 고려대학교 심리학과에 재직하고 있음.

은 관찰된 점수와 검사에 의해서 측정되는 잠재적 특질 간의 관계가 하위모집단에 관계없이 동일할 때 존재한다(Dragow, 1984). 검사의 번역의 경우, 원본과 번역본에 응답한 두 하위모집단(즉, 언어와 문화가 다른 두 집단)에 속한 수검자중, 검사가 측정하는 특성을 동일한 양만큼 소유하고 있는 각 집단의 수검자들이 집단에 관계없이 똑 같은 관찰점수를 얻을 때 원본과 번역본은 측정의 동등성을 나타낸다고 말한다. 동등한 검사를 사용하지 않는다면 검사점수가 집단의 차이를 진정 나타내는 것인지 혹은 측정상의 오류인지를 알 길이 없다. 문항의 수준에서 보면, 심리측정학적으로 동등한 문항은, 그 문항 혹은 문항들로 구성된 검사에 의해서 측정되는 심리특질에 있어서 동일한 양을 갖는 사람들이 여러 가능한 응답중 특정한 응답을 선택할 확률이 동일한 문항이다. 심리검사에서 각 문항에 대한 선택지가 응답 가능한 전부인 것이다. 만일 원 문항과 번역된 문항이, 그 문항에 의해서 측정되는 심리특질을 동등하게 보유하고 있는 사람으로부터 어떤 특정한 응답을 야기할 확률이 동일하다면, 원문항과 번역문항이 심리측정적으로 동등하다고 말할 수 있다. 문항의 의미의 동등성에 관한 이러한 정의에는 일어날 수 있는 반응의 범위가 한정되어있다. 행동적 반응을 근거로 하는 의미의 동등성에 대한 이러한 정의는 심리학적인 관점에서 의미가 있다. 동등성에 관한 이 같은 정의는 한정된 범위의 반응에 대한 것이고, 관찰할 수 없는 매개적인 반응은 제외된다. 가능한 반응들에 대한 제한과 매개적인 반응의 배제로 인해 연구자는 강력한 결과를 얻을 수 있다.

번역의 과정이 완벽하고 이와 더불어 치밀한 방법에 의해서 평가되게 되면, 인간과 세상사이의 여과기로서의 언어의 기능이 측정척도 문항의 의미에 대하여 갖는 효과에 대한 연구가 가능할 것이다. 앞으로 논의하게 될 심리검사 번역의 평가방법은 언어에 따른 검사문항이 갖는 의미의 동등성 여부를 가릴 수 있을 만큼 이론적으로 탄탄한 심리측정이론을 근거로 하고있다. 이 방법은 표본집단에서의 심리특질의 분포에 관하여 제한적인 가정을 하지 않는다. 이 방법

에 의해서 번역을 평가한 결과 상당수의 문항이 동등하다고 판단되면 언어학적으로는 상이하지만 동등화된 검사상에서의 점수에 바탕을 둔 비교언어학적인 연구는 하지없이 행하여 질 수 있음을 시사해 준다.

상당수의 문항이 동등하다고 밝혀지면 여러언어의 어휘의 추상적 의미가 상당히 중복될 수 있음을 주장할 수도 있다(Werner & Campbell, 1970). 다수의 문화에 걸친 어의적 보편성(semantic universal)은 비교문화 연구에 필수적이다. 어의적 보편성은 여러 종류 언어로된 심리측정학적으로 동등한 검사를 개발하는 데 심본 활용될 수 있다.

앞에서 언급한 측정이론은 문항반응이론(item response theory)이라고 흔히 불리는 이론으로, 이 이론을 번역된 검사의 분석에 적용하여 측정의 부등성의 문제를 해명하고 동시에, 문화, 언어 혹은 문화와 언어 모두에 기인한 차이를 발견할 수 있다. 문항반응이론은 검사가 측정하는 심리특질과 검사내 특정한 문항에 대한 수검자의 응답사이에는 어떠한 관계가 있음을 상정한다. 일반적으로 측정되는 특질을 많이 갖고 있을 수록 특정한 방향으로 응답할 확률이 증가한다.

문항반응과 측정되는 심리특질간의 관계를 나타내는 곡선을 문항특성곡선(item characteristic curve)이라고 부른다. 이렇게 단조적으로 증가하는 함수를 여러가지 수학적 모형으로 나타낼 수 있고, 그 수학적 모형은 상정되는 가정과 문항특성곡선을 표현하기 위해서 사용되는 모수의 가지수에 따라 달라진다. 번역된 문항의 문항특성곡선이 원문항의 문항특성곡선과 동일할 때 원문항과 번역문항은 동등한 것이라고 말할 수 있다. 문항특성곡선이 같지 않을 때 그 문항은 편향(biased)되었다고 말하기도 하고 상이한 문항 함수를 나타내고 있다고 말하기도 한다. 최근까지도 문항편의(item bias)라는 용어를 많이 사용하고 있지만, 변별적문항기능성(differential item functioning)이라는 용어가 가치 중립적이고 보다 정확한 용어라고 주장하는 사람도 있다(Holland & Thayer, 1988; Ellis, 1989). 이 논문에서는 종전의 관례대로 문항편의라는 용어를 사용

할 것이다.

본 논문은, 능력검사, 인성검사등과 같은 심리측정 도구의 번역의 평가에 관한 것이다. 심리측정도구는 심리특질의 개인차를 평가하기 위해서 제작된다. 이러한 특질은 원문화권에서 타당화과정을 거칠때 도태되지 않고 살아남아 인간의 행동을 이해하는데 중요한 개념인 경우가 대부분이다. 또한 그러한 심리특질은 번역본이 쓰여질 "목표 문화"에서도 중요성을 가지고 있다고 가정된다. 그렇지 않다면 번역을 시도할 이유가 없을 것이다.

Casagrande(1954)의 번역의 분류에 의하면 심리검사의 번역은 인종학적(ethnographic)번역으로 분류될 수 있다. 인종학적(ethnographic)번역에 있어서는 원어로 된 자극의 의미와 문화적 내용이 번역어에서 유지되어야 한다. 이러한 번역을 시도하는 번역자는 원어와 번역어에 대한 지식뿐만 아니라 원문화와 목표문화에 대해서 잘 알아야 한다. 번역자는 단어와 어구가 어느정도의 융통성내에서 번역문화에 들어맞는 지 검토하여 번역본에서 적절히 사용하여야 한다. 그러나 이러한 분류는 훌륭한 것은 아니다. 심리측정도구의 번역은 궁극적으로 번역문화권에서 번역된 도구에 의해서 측정이 시도되는 구성개념과 타 구성개념들과의 관계의 양태가 원문화권에서의 그 구성개념과 타 구성개념들과의 관계의 양태와 동일할때 완전하다고 보기 때문이다. 이러한 관점은 상이한 문화들 간에 공통부분이 존재함을 가정한다.

이 논문에서는 직무 만족을 측정하기 위해서 미국에서 널리 쓰이고 있는 직무기술지표의 개정판(Revised Job Descriptive Index 또는 R-JDI; Smith, Kendall, & Hulin, 1985)의 번역을 분석한 결과를 제시한다. R-JDI는 6개의 하위검사(work on present job, supervision, present pay, opportunities for promotions, coworkers와 job in general)척도로 구성되어 있으며, 하위검사내 각 문항에 대해서 동의, 부동의, 미결정 중에서 하나를 선택해서 답하도록 지시되고 있다. R-JDI는 본래 영어로 개발되고 타당화되고 표준화 된 것이다. 본 논문에서는 영어로 된 R-JDI가 한글로 번역되었다.

전통적 방법에 의한 번역의 분석

번역의 질을 평가하기 위한 전통적인 절차는 그리 많지 않다. 가장 흔히 사용되는 방법으로는 번역본을 원어로 되번역한 것과 원척도를 비교하는 것이다. 또 하나의 방법은 2개국어를 병용하는 사람들이 원척도와 번역척도에 응답한 것을 비교하는 것이다. 재번역과 2개국어병용자 분석을 다음에 간단히 서술한다(Hulin, Drasgow & Parsons, 1983).

(가) 되번역

이 방법은 세 단계로 구성된다. 원본 측정도구는 우선 번역언어로 번역된다. 번역된 도구는 다시 원래의 언어로 다른 번역자에 의해서 되번역(back-translation)된다. 되번역된 측정도구와 원래의 측정도구가 전단계에 관여하지 않은 사람에 의해서 비교된다. 차이가 지적되고 필요한 경우 문제가 있는 문항은 두 세번 더 위의 과정을 반복한다.

이 절차는 심리적 측정도구의 의미의 동질성을 확인하기 위하여 필요조건이지만 충분조건은 아니다. 이 절차에는 몇 가지 단점이 있다. 노련한 번역자는 일련의 추론과 통찰적 추측에 의해서 심히 잘못된 번역으로부터, 수용할 만한 수준의 되번역을 산출해 내는 경우도 있다. 두 번째 문제점은 되번역에 의해서 원래의 의미가 복원이 된다고 해서 번역언어로 된 문항이 의미가 있다고 입증할 수 있는 것은 아니라는 점이다. 그러므로 되번역은 번역의 분석에 있어서 필요한 첫 단계 일 수는 있지만 원본과 번역본의 언어적 유사성이 동등성을 보장하지는 못한다.

(나) 2개국어 병용자의 응답

성공적인 되번역후에 연구자는 원본과 번역본을 2개국어병용자에게 실시하여 응답을 하도록 할 수 있다. 표준화된 절차를 따라 두 검사에 대한 검사점수들을 산출하고, 점수들간의 상관계수를 계산하거나 혹은 다른 분석방법을 이용하여 원본과 번역본이 피험자들에 대해 유사한 정보를 산출하는 정도를 검토한다. 예를 들면, 피험자들이 양쪽 언어에서 동등한

정도로 능통하다면, 척도의 원본과 번역본에 나타나 있는 문항의 평균과 변량이 유사할 것이고 또한 피험자의 서열간의 상관의 높을 것이다. 좀 더 정교한 분석에서는, 피험자에 기인한 변량, 언어차에 기인한 변량과 피험자와 언어의 상호작용에 기인한 변량의 추정치를 구할 수 있다. 이 경우 마지막 두 변산원은 번역과정에서 도입된 체계적 오류나 무선적 오류임을 나타내 준다. 원본과 번역본에 의거한 점수들 사이에 강한 선형관계가 존재할 때 그 번역은 두 언어에서 동등한 척도를 산출해 냈음을 나타내는 것으로 해석하게 된다. 마찬가지로 언어 혹은 언어와 피험자 상호작용에 기인한 변량이 낮은 것은 재생력 높은 번역임을 나타내는 것이다.

번역의 질을 검토하는 이러한 방법이 지나는 문제점은 2개 국어 병용자를 피험자로 사용하는 데서 발생한다. 이들은 2개국어를 병용하기 때문에 단어의 의미를 단일 언어사용자와는 다르게 해석할 가능성이 있다. 그러나 대부분의 비교문화연구는 실질적으로 단일언어사용자에 초점을 둔다. 따라서 2개국어병용자들로부터 얻어진 결과를 단일언어사용자 집단에 일반화시키려 할 때 주의를 기울여야 한다.

고전검사이론에 의한 편의를 탐색

편의된 문항을 식별하기 위해서 종래에 사용되어온 방법은 고전검사이론에 바탕을 둔 것인데 집단간 문항반응을 비교하기 위해서 가장 빈번히 사용되는 절차는 다음과 같다: (가) 문항 찬택도 혹은 정답률 (P_i) (나) 문항반응과 검사총점간의 양분점 상관계수 (point-biserial correlation) (다) 피험자 평균점수 (라) 피험자 점수의 표준편차 (마) 문항간 공변량의 요인분석(Hulin, Drasgow & Parsons, 1983). 이 들중 앞의 네종류의 통계량은 집단 의존적(sample dependent)이다. 즉, 피험자의 심리특질의 분포가 상이한 집단을 비교하면 편의되지 않은 문항에 대한 위의 네가지의 통계량이 서로다른 값으로 산출된다. 예를들면, 문항이 언어학적인 측면과 심리측정학적인

측면에서 동등한 지를 평가할 목적으로, 태도문항의 경우에는 찬택 확률을, 능력문항에 있어서는 정답을 할 확률을 나타내는 P_i 값을 원본과 번역본에서 비교하는 경우, 비교되는 모집단들에서, 그 문항들이 측정하고 있는 특성의 분포가 동등하거나 사전에 알고 있음을 전제로 한다.

두 하위 모집단에서의 P_i 즉, 문항평균이 동일해야 한다는 엄격한 조건에 비해서 다소 완화된 가정으로 문항평균값의 서열이 원본과 번역본에서 동일해야 한다는 것이 있다. 다시 말하면 두 문화권에서 평가되고 있는 심리특질의 분포가 상이할 수도 있음을 인정하겠다는 것이다. 심리특질의 분포가 상이하더라도 문항이 정확히 번역되었다면 문항평균치들은 상대적으로 동일해야 한다. 한 집단에서 어렵거나 혹은 찬택도가 높은 문항은 다른 집단에서도 어렵거나 높은 찬택도를 가져야 한다. 반면에, 곤란도나 찬택도의 절대적 수준은 집단들에서의 심리특질 분포에 따라 차이가 있을 수 있다. 문항난이도의 서열이 집단에 관계없이 동일하다면 비편의된 것으로 본다. 이러한 정의는 Lord's Paradox(Lord, 1980; Hulin et al., 1983)라는 현상 때문에 그릇된 결론에 도달할 수 있다. 평균과 변별력이 모두 상이하여 비선형회귀선, 즉 문항특성곡선이 교차하는 두 문항은 집단별로 문항난이도를 산출하게되면 문항난이도 서열이 집단마다 달라질 수 있기 때문에 문항편의 여부를 그릇되게 판정할 위험성이 있다.

원본과 번역본에서의 문항과 총점간 양분점 상관계수의 차이가 원어에서 번역언어로의 전환에 따른 심리측정학적인 차이를 반영한다고 보고 이 상관계수의 차이를 편의의 증거로 내세우지만, 양분점 상관계수의 크기는 순수한 공변량 이외에 두 집단에서의 심리특질의 변산과 문항평균의 영향을 받는다. 문항-총점간 상관계수에 대해 집단이 미치는 영향력은 원본과 번역본에서 문항이 갖는 특성만큼 강할 수 있다. 문항특성에 관한 결론은 두 모집단의 특성이나 문항과 집단 간의 상호작용에 영향을 받아서는 안되며 오직 문항 특성과 관련된 사실에만 바탕을 두어야 한다. 이러한 단점의 개선을 모색한 고전검사이론에 바탕을

둔 몇가지의 방법이 제안된 바 있다(Berk, 1982)

문항반응이론에 의한 편협의 탐색

고전측정이론에 의한 통계량이 문항편의를 밝혀내는데 미흡하다는 인식하에, 표본 의존적이지 않음을 전제로 한 편협의에 대한 새로운 정의와 탐색방법이 모색되었다. 이 새로운 정의에 밑바침이 되는 개념이 문항특성곡선이다. 앞절에서 간단히 설명한 바와같이 문항특성곡선은 문항이 측정하는 심리특질에 대한 특정한 응답(태도측정문항에서는 찬택반응, 능력측정문항에서는 올바른 응답)의 조건확률을 나타내는 곡선이다.

문항특성곡선은 채점된 문항반응 U_i 를 종속변수로 하고 심리특질(θ)를 독립변수로 하는 비선형회귀로 개념화될 수 있다. 문항특성곡선의 높이는 심리특질 차원상에서 특정한 θ 값을 갖는 모든 사람에 대한 문항점수의 기대치이다. 심리특질차원 상의 특정한 θ 에 대응하는 문항특성곡선의 높이 즉 $P_i(U_i|\theta)$ 는 척도내의 여타문항의 영향을 받지 않으며 또한 표본에서의 의 분포에도 영향을 받지 않는다. 두 하위집단의 반응을 바탕으로, 한 문항에 대한 문항특성곡선을 도출했을 때 두 문항특성곡선은 적절한 선형전환을 통하면 동일하게 되는 표본독립적인 장점을 갖고 있다.

원어로된 문항과 번역된 문항에 대한 문항특성곡선을 비교 검토하여 문항의 의미의 동등성과 이러한 문항반응들로 구성된 점수의 동등성에 대해 판정을 내릴 수 있다. 문항특성곡선의 동등성은 피험자의 문항 응답으로부터 직접 추정할 수 있다. 두 집단의 문항 특성곡선이 일치하지 않게 되면 동등성이 없다고 볼 수 있다. 문항특성곡선 간의 피리의 정도와 방향을 살펴봄으로써 번역에 의한 편협의 원인을 짐작해 볼 수 있다. 문항특성곡선의 차이에 는 평균의 차이, 변별력의 차이, 추측방식에서의 차이를 들 수 있다.

문항반응과 심리특질의 관계를 수학적으로 나타내는 이론, 즉 문항특성곡선을 수학적으로 묘사하는 이

론 중에서 가장 중요한 이론이 문항반응이론(Item Response Theory)이다. 문항반응이론의 수학적 모형중 가장 일반적인 형태는 세 가지 종류의 문항모수를 담고 있는데 이 모수들은 특정한 반응을 할 확률과 θ 를 연결하는 S-형의 로지스틱 곡선(logistic curve)을 정의하게 된다. 문항반응이론의 로지스틱 곡선을 정의하는 공식은 다음과 같다(Lord, 1980).

(공식 1)

$$P_i(\theta) = C_i + \frac{(1 - C_i)}{1 + \exp[-D a_i(\theta - b_i)]}$$

여기서 $P_i(\theta)$ 는 문항 i에 대해 긍정적 혹은 올바른 반응을 할 확률이고 C_i 는 문항특성곡선의 하부 점근치이며 b_i 는 문항난이도 혹은 태도나 정서를 측정하는 문항일 때는 찬택을이다. b_i 의 척도(metric)는 θ 와 동일하며 a_i 는 문항특성곡선의 변곡점(inflexion point)에서의 기울기이다. 3-모수 문항 특성곡선에서는 $\theta = b_i$ 에서의 기울기가 $(1 - C_i) * (D/4) a_i$ 이다. 여기서 D는 문항특성곡선의 로지스틱 문항반응모형을 정규 오자이브 모형과 근사하게 하기 위하여 첨가되는 척도상수(scaling constant)로서 보통 1.702 값을 갖는다. 공식 1에서 C_i 값을 0으로 놓으면, 즉 문항에 추측해서 답하거나 척도 점수를 극대화 할 동기가 없다고 가정하는 경우 하부 점근치가 0이 되어 2-모수 모형이 된다. 잘 제작된 직무만족이나 태도를 측정하는 척도의 문항들의 c 값이 0에 가깝다는 것이 실증적 연구에서 밝혀진 바 있다(Hulin, et al., 1983).

문화적 차이에 의한 편협의

위에서 개괄한 문항반응이론에 근거한 절차에 의해서 문항이 동등하지 않다고 판정되는 이유는 번역의

오류때문일 수도 있지만, 오역은 문항특성곡선의 차이에 대한 여러 가능한 원인중 하나일 따름이다. 원본과 번역본 사이의 심리측정적 비동등성의 또 다른 원인일 수 있는 것은, 측정하려고 하는 심리특질 자체가 번역본이 사용될 문화에서 적합한지의 여부이다. 심리특질의 실질적 의미에 대한 비교문화적인 연구에 의해서 특질 자체가 문화적으로 적합한 구성개념인지의 여부가 밝혀질 수 있을 것이다. 이러한 연구가 이루어지기 위해서는 문제의 심리특질을 동등하게 측정할 수 있는 도구의 개발이 선행되어야 한다(Hulin, 1987).

추정되고 있는 심리특질이 어느 문화권에 속해있는 사람들에게나 의미가 있다면, 문화적 적합성에 관한 문제는 심리특질을 추정하는 데 사용되는 문항들에 국한된다. 예를들면, 술선성이라는 구성개념을 측정하는 검사의 특정한 문항이 언어학적으로는 번역이 잘 되었다고 하자. 그러나 그 문항이 번역본이 쓰여지는 목표문화에서는 어떤 상황에서도 볼 수 없는 행동의 발생 가능성에 관한 것이라면 목표문화에는 적합하지 않은 편이된 문항이다. 마찬가지로 어떤 문항은 언어학적으로는 번역이 잘 된 문항일 수 있지만 목표문화적 상황에서는 추정하고 있는 심리특질과는 하등 관련이 없는 개념의 존재 여부를 질문할 수도 있다. 그러므로, 상이한 문항특성곡선이 번역의 오류에 기인하는 것으로 결론을 내리기 전에 그 문항이 시사하는 특징이나 행동이 원문화에서는 측정하고자 하는 특질과는 유관하지만 목표문화에서는 그렇지 않은지를 주의깊게 검토할 필요가 있다. 이러한 개념과 특징을 편문화적(emic)개념이라고 말하며, 문화적으로 일반적인 범문화적(etic)개념과 대조된다. (Triandis & Brislin's study cited in Hulin & Mayer, 1986).

심리검사의 원본에 많은 편문화적 개념이 존재한다면 원본과 번역본 사이에 심리측정학적으로 동등하지 않은 문항이 존재하게 된다. 편이된 문항을 제거한 후, 번역이 잘 되었고 심리측정학적 동등성의 기준을 충족하는 문항을 이용하여 θ 를 추정한다. 이 상황에서는 검사의 원본과 번역본 모두로부터 그러한 문항

이 제거되어야 한다. 그러나 만일 번역 문항이 원문항과 심리측정학적으로 동등하지 않지만 문항특성곡선이 0이 아닌 기울기를 갖는다면 원 문항과 번역 문항은 양쪽문화에서 동일한 θ 를 측정하고 있다고 볼 수 있다. 편이된 문항이 모두 제거되어 문항이 축소된 검사가 θ 에 관하여 제공하는 정보는 모든 문항이 포함된 검사를 이용하여 θ 를 추정하는 경우보다 어느쪽 문화에서나 덜 할 것이다.

심리특질인 θ 에 관하여 보다 많은 정보를 갖고있는 검사에 바탕을 둔 비교문화적 연구가 보다 더 정확할 것이므로, 문화적으로 특유한 문항을 제거하지 않는 방안이 문항반응이론의 틀내에서 강구될 수 있다. 번역이 잘 된 범문화적 문항이 충분하다면, 편문화적 문항까지도 포함하는 검사를 제작하는 것이 가능하다. 중복되는 문항을 갖고 있는 검사들을 동일척도(metric)화 할 수 있는 문항반응이론의 특성을 심분 활용하여 편문화적 개념을 담고 있는 검사의 원본과 번역본이 동일한 척도상에서 동일한 θ 의 추정치를 산출해 낼 수 있도록 할 수 있다.

범문화적 문항과 편문화적 문항을 모두 담고 있으면서 양쪽 문화에 모두 적합한 척도를 개발하는 간단한 방법은 다음과 같다. 검사를 원어에서 목표어로 번역한 다음, 목표문화에 맞게 쓰여진 시안적 문항들을 첨가한다. 번역문항들과 시안 문항들을 합한 문항들에서 번역이 잘 된 범문화적 문항들을 결정한다. 번역이 잘 된 범문화적 문항은 양쪽문화에서 동일한 문항특성곡선을 가질 것이다. 다음 단계로, 원본과 번역본에서의 범문화적 문항들에 대한 응답들을 matched 언어본에 대한 반응인 것처럼 결합한다. 두 집단은 범문화적 문항에 대해 동일한 문항특성곡선을 갖기 때문에 이 단계가 가능하다. 편문화적 문항에 대해서는 해당 집단의 구성원들만이 응답하고 편문화적 문항을 받지 않는 집단의 응답자에 대해서는 그 문항들에 시간이 없어 반응하지 않은 것처럼 처리한다. 양 집단에 공통된 범문화적 문항에 대한 응답과 편문화적 문항에 대한 해당 집단의 응답을 결합한 자료를 사용하여 문항모수와 구성원의 심리특질을 추정한다. 심리특질값과 모든 문항모수 추정치들은 BILOG등의

모수 추정 프로그램에 의해 자동적으로 동일한 척도(metric)위에 놓이게 된다. 이러한 방식으로 제작되어 범문화적 문항과 편문화적 문항을 모두 포함하게 되는 심리검사가 범문화적 문항만을 담고 있는 검사에 비하여 보다 정확한 θ 추정치를 산출할 것임은 명백하다.

이러한 절차가 타당성을 갖기 위해서 필수적인 가정이 있다. 검사내의 문항들은 원본이든 번역본이든 해당 문화권 내에서는 일차원적이어야 한다는 것이다. 문화적으로 일반적인 개념을 측정하는 문항과 문화적으로 독특한 개념을 측정하는 문항 모두가 한 문화권내에서는 오직 하나의 심리특질만을 측정하는 것이어야 한다는 것이다.

방 법

피험자

서울과 수원에 위치한 6개의 제조회사(전자회사, 승용차 생산회사 및 가구제조회사 포함)의 근로자 390명에게 한글로 번역된 JDI를 배포하여 이중 355부가 회수되었다. 무응답이 많은 12명의 근로자를 제외한 343명의 자료가 분석되었다. 영어로 된 JDI에 대한 미국인 근로자의 응답은, 미국의 중부에 위치한 B 대학의 특정 심리학 학부과목의 과제의 일환으로서 학부학생들이 다양한 직종의 근로자로부터 수집한 것이다. 무반응이 많은 응답자를 제외한 511명의 자료가 최종적으로 분석되었다.

질문지의 번역

JDI에 대한 일차적 번역은 필자가 행하였다. 이 번역본을 3 사람에게 차례로 검토를 의뢰했다. 첫번째 사람은 미국에 13년간 거주하고 있는 재미 한국인이었다. 이 번역의 검토후에 A 대학 영문과에 재직하고 있는 한국인 영어학자에게 수정을 의뢰했다. A 대학 영문과 교수와 필자의 의견이 상치되는 문항에 대해서는 A대학 영문과에 재직하고 있는 미국인 강사와 의견을 교환하여 번역에 최종 마무리를 했다.

이는 많은 비교문화 연구자들이 요구하는 되번역의 절차를 따르지는 않았지만 그와 유사한 정도의 번역의 질을 달성했고 또한 보다 현실적이라고 필자는 판단했다. 각 문항에 대한 번역은 되도록 축어적으로 행하였고 한글이 어색한 부분에 대해서만 손질을 했다. 첫부분 의역은 원래의 언어가 의미하는 바를 축소 또는 확대하여 의미를 왜곡할 위험이 있기 때문이었다.

문항편의의 판별

JDI 문항에 대한 응답(예, 아니오 그리고 모름)은 본 연구에서 채택한 문항반응이론의 수학적 모형이 요구하는 대로 이분점수로 전환되었다. 긍정적인 응답이 정답인 문항에 대해 동의하거나 부정적인 응답이 정답인 문항에 대해 동의하지 않을 때 1점으로 채점되었다. 부정적인 응답이 정답인 문항에 대해 동의하거나 긍정적인 응답이 정답인 문항에 대해 동의하지 않았을 때 0점으로 채점되었다. [잘 모르겠다]는 반응도 또한 0으로 채점되었는데 그 이유는 JDI 문항에 대해 [잘 모르겠다]는 반응을 한 사람들의 직무 만족이 흔히 낮은 것으로 밝혀졌기 때문이다 (Smith, Kendall & Hulin, 1969). 또한 반응을 하지 않은 문항도 0점으로 처리했다.

각 집단에 대한 문항모수추정치 척도(metric)가 단위(unit)와 원점(origin)의 면에서 임의적이기 때문에 이 추정치의 비교가 가능하기 위해서는 공통의 척도(common metric)가 필요하다. 공통의 척도를 달성하기 위해서는 "비교" 집단에 대한 문항모수 추정치의 척도를 "기준" 집단의 척도로 전환해야 한다 (Linn, Levine, Hasting & Wardrop, 1981). 이 연구에서는 한국근로자를 기준 집단으로 하여 미국인 근로자의 모수 추정치를 전환했다.

문항편의 여부는 Lord(1980)가 제안한 χ^2 통계량을 계산하여 판정하였다(Hulin et al., 1983 pp 174-175). Lord의 χ^2 통계량은 a_i 와 b_i 모수 값에 차이가 없다는 동시적 검증을 하게 된다. 문항 i 에 대해서 $a_iA = a_iB$ 이고 동시에 $b_iA = b_iB$ 라는 가설에 대한 통계적 검증은, 행렬식으로 표현하면

$$z^2 = Vi' Ti^{-1} Vi$$

여기서 $Vi' = (biA - biB, aiA - aiB)$ 로서 하위집단 A와 B의 표본으로부터 산출한 문항 i 에 대한 모수 추정치의 차이에 대한 벡터이다.

Ti 는 $(biA - biB)$ 와 $(aiA - aiB)$ 의 절근적 표집변량(표준오차의 제곱)과 공변량으로 이루어진 2×2 행렬이다.

그리고 Ti^{-1} 는 Ti 의 역행렬이다. 이 검정 통계량은 2 df의 자유도를 갖는 z^2 분포를 따른다. z^2 통계량을 계산하는 데 필요한 표준오차와 표집공변량 추정치는 BILOG 프로그램이 산출한 것을 사용하였다. 두 집단의 문항특성곡선간의 차이가 통계적으로 유의미하다고 판명되었을때 그 부적인 결과에 대한 해석은 그리 쉽지 않다. 앞서 논의한 대로 문항특성곡선의 차이는 문항번역이 잘못되었기 때문일 수도 있고 혹은 문항의 내용이 원래의 언어와 문화에서는 당해 심리특질을 측정하는데 반하여 목표언어와 문화에서는 해당 심리특질과 무관한 것이기 때문일 수도 있다. 문항특성곡선간의 차이가 후자에 기인한 것이라면 편문화적 문항일 것이며 전자에 기인한 것이라면 재번역이나 가필을 할 필요가 있는 문항일 것이다.

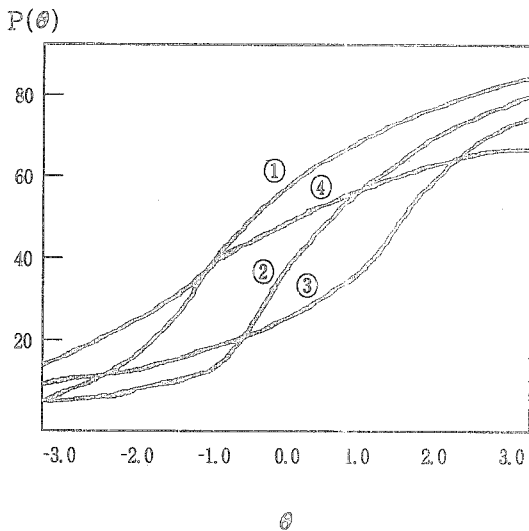


그림 1. 문항특성곡선

현재로서는 문항특성곡선의 피리가 어느 쪽 원인인지를 명확히 구별해 줄 수 있는 통계절차가 존재하지 않는다.

변별도 즉, 기울기가 달라서 문항특성곡선이 같지 않은 경우라면 편문화적 문항내용에 귀인할 수 있는 상황이라고 볼 수 있다. 예를들면, 심리특질이 $\theta = -1.0$ 인 사람의 $P(\theta)$ 와 $\theta = +1.0$ 인 사람의 $P(\theta)$ 의 차이가 원문화와 번역문화에서 다르다는 것이다. 즉, 그림 1의 ①과 ④의 차이이다. 이는 두 문항이 개인차를 동일하게 측정하지 않고 있으며, 따라서 동일한 것을 재고있지 않은 경우에 일어날 수 있는 현상이다. 이러한 문항에 대한 수검자의 응답은 원문화에서는 측정하려고 하는 심리특질의 추정에 기여하겠지만 목표문화에서는 별로 도움이 되지 못할 것이다. 문항 난이도 혹은 문항차택도의 차이는 문항내용자체의 유관성 여부보다는 번역의 오류에 기인한 것일 가능성이 상대적으로 더 높다고 볼 수 있다. 이 경우는 그림 1의 ①과 ②의 차이이다. 예를들면 심리특질이 $\theta = -1.0$ 인 사람의 $P(\theta)$ 와 $\theta = +1.0$ 인 사람의 $P(\theta)$ 의 차이가 원문화와 번역문화에서 똑 같아서 개인차를 동일하게 측정하기 때문에 동일한 것을 재고있다고 볼 수 있다. 그러므로 반응특성곡선의 차이가 편문화적인 이유때문이기 보다는 번역의 잘못에서 비롯된 것으로 볼 수 있다.

요인분석

JDI문항들은 총괄적(job in general)만족 척도를 제외한 다섯 종류의 하위척도를 통틀어 요인분석을 했을 때 "일반요인"에 대한 요인부하량이 그다지 크지 않은 것으로 밝혀졌다. (Parsons & Hulin, 1982). 이러한 결과는 JDI문항 전체로 판단할 때는 일차원적이라고 할 수 없음을 의미한다. 문항반응이론의 일차원성 가정에 적합한 자료가 되기 위해서는 각 척도를 별도로 분석할 때 가능해진다(Drasgow & Hulin, 1986). 각 척도를 별도로 요인분석했을 때 척도마다 하나의 일반요인을 도출할 수 있었다(Drasgow & Parsons, 1983). 즉 단일의 공통요인이 문항반응이론의 방법을 적용할 수 있을 만큼의 문

항변량을 설명했다.

본 연구에서는 원본과 한국본 JDI내의 각 하위척도가 일차원임을 가정한다. JDI 각 하위척도내의 문항을 SPSS를 이용하여 주축요인방식에 의해 요인분석(principal axis factoring)하고 일차원성 가정의 충족 여부를 판정하려고 시도했다. 축소된 문항간 양분점 상관계수 행렬에 대한 요인별 eigenvalue가 하위척도별로 표 1에 제시되었다.

표 1. 원본(한글본) JDI의 축소된 양분점 상관계수의 Eigenvalues

하위척도	Eigenvalues			
	1	2	3	4
WORK	6.14(6.85)	1.54(1.27)	1.06(1.06)	0.56
PAY	3.53(4.56)	0.57(0.60)		
PROMOTION	2.86(5.24)	0.88	0.57	
SUPERVISION	4.20(6.85)	1.50(0.97)	0.96(0.58)	0.77
COWORKERS	1.68(7.21)	1.47(1.09)	0.91(0.80)	0.64(0.72)
GENERAL	7.11(9.44)	2.23(1.61)	0.52	

첫번째 요인의 eigenvalue가 나머지 요인들의 eigenvalue에 비해서 상당히 큰 편이므로 JDI의 6개 하위 척도는 어느 정도 일차원성 가정을 충족하고 있는 것 같다. 또한 이 논문에는 제시되어 있지 않지만 각 하위척도 별로 요인별 요인부하를 살펴보면 첫번째 요인에 대한 문항의 요인부하가 여타요인의 요인 부하에 비하여 상대적으로 큰 것으로 미루어 첫번째 요인을 일반요인이라고 간주하고 일차원성 가정이 충족되고 있다고 필자는 잠정적인 결론을 내렸다. 2-모수모형과 3-모수모형 등의 문항반응이론의 수학적 모형은 일차원성의 가정이 다소 위배되도 강건(robust)하다는 즉, 문항모수의 추정이 매우 정확하다는 모의연구(simulation study)결과 (Drasgow & Parsons, 1983)와 본 논문의 요인분석결과를 중

합하여 필자는 JDI 각 하위척도에 문항반응이론의 통계 절차를 적용하는 것이 가능하다는 결론을 내렸다.

문항모수의 추정과 척도의 연결

본 논문에서 문항모수를 추정하기 위하여 사용된 방법은 주변최대우도추정방법(marginal maximum likelihood estimation procedure) (Bock & Aitkin, 1981)이고 이를 위해서 BILOG computer program을 사용하였다. 문항편의를 밝혀내는 데 주변최대우도추정방법에 의한 문항모수의 추정이 유용함이 실증된 바 있다(Drasgow, 1989; Thissen, Steinberg & Gerrard, 1986). Drasgow(1989)는 표본크기가 200에 불과하고 문항수가 겨우 5개인 열악한 상황에서도 2-모수 문항반응이론 모형의 주변최대우도추정치가 조건최대우도추정치(conditional maximum likelihood estimates)에 비해서 정확하다고 주장하고 있다.

비교되는 두 집단(이 논문에서는 한국인 집단과 미국인 집단)에게 자신들의 모국어판, 즉 미국인 집단에게는 영어로된 JDI를 시행하고 한국인 집단에게는 한글 번역된 한글본 JDI를 실시한 후 다음의 절차를 밟아 각 문항의 편의 여부를 결정했다(Lautenschlager & Park, 1988; Park & Lautenschlager, 1990).

1. 한국인 집단의 응답과 미국인 집단의 응답을 별도로 사용하여 두 집단으로부터 JDI 각 문항의 문항모수를 집단별로 추정한다.
2. 집단별로 추정된 문항모수 추정치는 상이한 척도(metric)상에 놓이게 됨으로 Divgi(1985)의 척도연결(linking)방법을 사용하여 두 집단에 대한 모수 추정치를 동일 척도상에 놓는다.
3. 연결되어 동등화된 모수추정치 즉, 두 집단의 난이도와 변별도 추정치를 사용하여 Lord의 α^2 값을 계산하고 각 문항의 편의 여부를 결정한다.
4. 단계 2에서 두 집단의 모수 추정치를 동일한 척

도상에 놓기 위해서 산출한 연결 방정식은 편이 된 문항의 모수 추정치가 포함된 자료에서 산출된 것이기 때문에 왜곡되었을 가능성이 있다. 따라서 단계 3에서 잠정적으로 편이된 것으로 판정된 문항의 모수 추정값을 빼고 나머지 문항의 모수 추정치만을 사용하여 연결방정식을 다시 산출한다.

5. 계산출된 연결방정식을 단계 1에서 추정된 집단별 문항모수추정값에 적용하여 α^2 값을 계산하

표 2. JDI 각 하위척도에 대한 α^2 값의 변화

문항	반복횟수				
	1	2	2	4	5
WORK ON PRESENT JOB					
1	1.8397	3.9277	3.9976	4.5504	6.3212
2	55.6595°	59.5058°	62.2368°	63.1708°	71.5066°
3	3.8327	14.8810°	7.5533	4.6542	2.7274
4	12.3626	2.7999	6.1729	8.4530	10.3988
5	14.3135°	40.0107°	25.7278°	19.7710°	16.1387°
6	18.2220°	17.6420°	21.1193°	23.9104°	28.7985°
7	10.4970	22.8215°	15.7551°	12.5760	10.3510
8	18.1894°	22.4481°	18.8177°	17.2958°	15.9501°
9	48.5937°	73.1748°	60.8854°	55.0057°	51.1526°
10	9.6321	26.8259°	16.2235°	12.0748	9.4881
11	22.1348°	30.3587°	28.0366°	27.4430°	28.5526°
12	1.6868	2.4702	1.7845	1.4492	1.1292
13	1.3021	3.9552	1.0001	0.3730	0.1033
14	4.2414	6.1001	5.2768	4.8460	4.5670
15	59.8709°	75.4179°	71.4570°	70.2636°	72.0791°
16	0.6611	5.0705	1.8540	0.8173	0.3065
17	7.9785	2.9199	4.7122	5.9955	7.2004
18	3.36372	1.31399	0.60818	1.25685	2.21714
PRESENT PAY					
1	31.5955°	13.008	8.676		
2	27.3795°	12.332	9.059		
3	18.2253°	5.176	2.732		
4	15.9063°	3.802	3.462		
5	9.5015	5.522	4.493		
6	0.9509	1.495	2.784		
7	17.0476°	45.219°	50.554°		
8	1.8698	5.315	6.330		
9	49.7137°	104.819°	114.705°		
OPPORTUNITIES FOR PROMOTION					
1	3.382	4.5002			
2	8.419	10.2049			
3	3.051	2.2830			
4	1.063	0.2205			
5	0.950	3.4856			
6	70.867°	47.3030°			
7	10.551	5.4182			
8	104.791°	74.6954°			
9	9.366	11.6764			

고 각 문항의 편이 여부를 다시 결정한다. 단계 4와 단계 5의 과정은 JDI내 각 하위 척도에서 편이되었다고 판정되는 문항이 전 단계와 후 단계에서 동일할

SUPERVISION				
1	5.2760	4.8491	5.5097	7.2064
2	8.4524	5.4701	3.3726	1.3879
3	1.7764	0.3768	0.1212	1.4847
4	1.9103	1.4666	1.4750	2.9459
5	1.0639	0.7448	0.9074	2.5164
6	27.4488°	31.9560°	37.8610°	46.4824°
7	1.3746	0.7049	0.6922	2.2594
8	10.7797	14.2879°	18.9527°	26.1608°
9	55.7184°	61.7905°	72.7029°	87.9865°
10	0.3626	0.2447	0.4119	1.4013
11	6.4015	10.7174	16.3801°	25.9020°
12	1.7057	0.7962	0.6118	1.2763
13	0.4569	1.1722	2.2928	4.9085
14	15.3474°	9.5203	5.4453	1.8711
15	46.7030°	41.6590°	37.6393°	32.4677°
16	19.3639°	25.2305°	32.6410°	44.1158°
17	18.7607°	14.6199°	11.0992	7.0995
18	3.37394	5.71439	8.21158	12.6721
COWORKERS				
1	13.7880°	15.2564°		
2	8.1792	6.8945		
3	4.1226	4.9303		
4	4.5589	3.7164		
5	0.5863	0.4926		
6	0.2820	0.5977		
7	11.0601	12.3160		
8	68.3473°	63.3709°		
9	5.9153	7.4886		
10	21.2063°	22.6936°		
11	34.2057°	31.7612°		
12	4.6628	7.1121		
13	24.0336°	19.1814°		
14	52.5335°	53.6790°		
15	6.1695	7.6473		
16	8.8105	10.5801		
17	0.9518	1.2315		
18	0.13769	0.37411		
JOB IN GENERAL				
1	41.0951°	44.8523°	51.1905°	
2	19.4800°	22.6366°	27.0354°	
3	49.5352°	39.7126°	33.3515°	
4	0.2284	0.1320	0.1092	
5	2.6121	4.4521	7.1913	
6	6.8031	5.0625	4.9434	
7	4.5529	2.8276	1.3023	
8	9.2655	10.1881	12.2684	
9	62.2482°	64.3084°	70.5265°	
10	16.9748°	11.2831	7.8806	
11	33.6145°	37.4145°	39.0036°	
12	7.7549	8.0955	9.8826	
13	12.6261	7.5019	4.2628	
14	2.6425	3.4323	4.0365	
15	13.0787	8.6559	5.8719	
16	0.2631	0.2825	0.3680	
17	23.6922°	27.9713°	31.4072°	
18	1.51283	1.54557	1.54582	

때까지 반복한다.

우리가 단계 2와 그 후의 반복과정에서 척도연결방정식(linking equation)을 산출하는 이유는 앞서 설명한대로 각 집단별로 문항의 모수를 추정할 때 척도(metric)의 원점과 단위가 임의적이기 때문이다(Lord, 1980).

표 2에 JDI 각 하위척도별로 반복추정과정에서 얻어진 χ^2 값이 순차적으로 제시되어 있다. WORK척도는 최종 χ^2 값을 얻기까지 연결계수를 5번, PAY척도는 3번, PROMOTION척도는 2번, SUPERVISION척도는 4번, COWORKERS척도는 2번, JOB IN GENERAL척도는 3번의 반복과정으로 산출했다. 전 단계에서 편이된 것으로 판정된 문항의 모수 추정치를 제거하고 난 후 잔여 문항의 모수치만으로 연결계수를 산출하는 과정을, 전단계와 후단계에서 동일한 문항들이 편이되었다고 판정될 때까지 반복해야 되는 이유를 극명하게 실증적으로 보여주는 예가 PAY척도의 경우이다. 첫번째 도출된 연결방정식을 사용하여 χ^2 값을 산출했을 때는 PAY척도의 9개 문항 중 67%에 해당하는 무려 6개의 문항이 편이된 것으로 잠정 판정되었다. 그러나, 편이된 것으로 잠정 추정된 문항을 제외한 3개의 문항의 모수 추정값으로 도출된 연결방정식으로 척도동등화를 이룬 후 χ^2 값을 계산했을 때는 단지 2개의 문항만이 편이된 것으로 밝혀졌다. 이때 편이된 것으로 밝혀진 문항을 제외하고 반복해서 χ^2 값을 산출했을 때 동일한 문항이 편이된 것으로 수렴되었다. 문항 9와 문항 7의 값이 각각 114.705와 50.554로서 대단히 편이된 문항들이기 때문에 이 문항들이 측정척도(metric)를 심히 왜곡시켜서 최초의 χ^2 값이 부정확하게 나왔을 것으로 짐작된다. 이는 본 논문에서 사용하고 있는 척도정화(scale purification)절차가 효력이 있음을 보여주는 것이다.

결 과

직무만족 측정 질문지인 JDI에 문항반응이론을 적

용하여 Lord의 χ^2 값을 얻은 결과가 표 2에 제시되어 있다. 원본의 문항의 문항특성곡선과 이에 대응하는 번역본의 문항의 문항특성곡선을 비교하여 원본과 번역본의 직무만족 측정의 동질성을 개별 문항별로 검토할 수 있었다. 별표가 표시된 문항이 한국인 집단과 미국인 집단의 문항특성곡선이 차이가 나는 문항이다. 별표 문항에 대한 두 집단의 문항특성곡선의 차이는 α 수준 0.001에서 통계적으로 유의미하다. 단일비교에 대한 α 수준을 0.001로 선정한 이유는, 본 연구전반의 관점에서 바람직한 α 수준과 연구에 포함된 문항의 수를 함께 고려해야하기 때문이다. McLaughlin(1986)의 모의연구(simulation study)에서도 α 수준을 낮추어 정하는 것이, 비편이된 문항이 편이된 문항으로 판정되는 실질적인 제 1종 오류율이 적절한 수준으로 유지됨이 입증된 바 있다. McLaughlin은 50문항에 1000명이 응답하는 상황을 상정하고 이러한 상황에 대한 100조(data sets)의 상이한 모의 자료를 생성한후 LOGIST computer program(Wood, Wingersky, & Lord, 1976)으로 각 조의 자료를 이용하여 50문항의 문항모수와 사람모수 θ 를 추정했다. 결과적으로 50문항에 대한 문항모수와 θ 추정치가 100조 있게 된다. 1조의 문항모수 추정치는 2조의 추정치와 비교하고, 3조의 추정치는 4조의 추정치와, 5조는 6조와 비교하는 방식으로 총 50회의 문항편이에 대한 분석을 하였다. Lord의 χ^2 통계량에 의해서 편이된 것으로 판정된 문항비율은 평균적으로 .12 이었다. $\alpha = .05$ 일때 50문항의 검증에서 제1종오류는 기대치 2.5(50 0.05)보다 큰 6 이었다. α 수준을 .01로 했을 때는 기대치보다 4배정도의 문항이 편이된 것으로 나타났고 $\alpha = .001$ 일때는 무려 10배에 달했다. McLaughlin의 연구 결과가 함의하는 바는 연구자가 정한 α 수준을 토대로 예상하는 것보다 월등히 많은 문항이 편이된 것으로 오판될 수 있다는 사실이다. 그러므로 전술한 통계적 논리와 McLaughlin의 모의연구결과를 감안하여 본 연구에서는 α 수준을 .001로 결정하였다. 임계치 $\chi^2 = 13.8$ 보다 작은 값을 갖는 비편이된 문항은 번역이 올바르게 된 문항이라

표 3. 한국과 미국인 표본의 JDI 최종 문항모수추정치

문항	한국인 표본		미국인 표본	
	B	A	B	A
WORK ON PRESENT JOB				
1 Fascinating	0.83508	1.25714	1.13780	0.98028
2 Routine°	0.49687	0.80735	1.60692	0.91116
3 Satisfying	0.53198	2.13318	0.41924	1.57857
4 Boring	-0.03067	0.76129	0.35472	1.13394
5 Good°	0.29184	1.81313	-0.15778	1.12573
6 Creative°	0.36768	1.30760	0.95046	1.06834
7 Respected	0.72231	0.94248	0.26033	0.73454
8 Uncomfortable°	0.77760	0.37531	-0.27990	0.57905
9 Pleasant°	1.68335	0.89262	0.22555	0.78714
10 Useful	-0.22694	0.97919	-0.64136	1.03211
11 Tiring° #	0.56776	0.31217	3.27731	0.23249
12 Healthful	2.79753	0.39839	2.27799	0.42753
13 Challenging	0.28721	1.24201	0.28512	1.31961
14 Too much to do	3.89057	0.14386	2.26134	0.13106
15 Frustrating°	-1.10426	0.43185	2.52366	0.26656
16 Simple	0.13270	0.84545	0.05057	0.80565
17 Repetitive	1.39731	0.62603	1.62110	0.85348
18 Gives sense of accomplishment	0.16474	1.39538	0.29399	1.69343
PRESENT PAY				
1 Income adequate for normal expenses	1.03118	1.15241	0.65980	0.98487
2 Fair	0.95097	1.19449	0.81361	0.69689
3 Barely live on income	0.82330	0.96174	0.57313	0.98323
4 Bad	0.47614	1.01682	0.44989	1.82177
5 Income provides luxuries	3.08255	0.70061	1.95976	0.83326
6 Insecure	-0.14185	0.69336	0.09200	0.62502
7 Less than I deserve°	0.66392	0.97317	1.53210	1.20703
8 Well paid	1.50690	1.18644	1.79911	2.14000
9 Underpaid° #	0.16459	1.36728	1.27698	1.75246
OPPORTUNITIES FOR PROMOTION				
1 Good opportunities for promotion	0.96952	1.15711	0.94625	2.12343
2 Opportunities somewhat limited	1.20752	0.48375	1.36206	0.87815
3 Promotion on ability	0.55338	0.71832	0.31615	0.82410
4 Dead-end job	-0.18520	0.84510	-0.19446	0.94498
5 Good chance for promotion	1.11055	1.35837	0.84541	2.21698
6 Unfair promotion° policy	0.43024	0.70684	-0.67576	0.52330
7 Infrequent promotions	0.60200	0.80472	1.02802	0.55187
8 Regular promotions°	0.36387	0.67470	1.66393	0.76485
9 Fairly good chance for promotion	1.22902	1.49486	0.76710	1.52024
SUPERVISION				
1 Asks my advice	-0.34198	0.79127	-0.23784	0.47615
2 Hard to please	0.25541	0.70734	0.07556	0.61638
3 Impolite	-0.43162	1.03377	-0.26608	1.10826
4 Praises good work	-0.07653	0.76144	0.12203	0.97235
5 Tactful	0.33770	0.76098	0.51260	0.90069
6 Influential°	-0.62134	0.54158	0.64181	0.78892
7 Up-to-date	0.24673	0.85937	0.40718	1.00732

8 Doesn't supervise° enough	-0.39519	0.62632	0.43061	0.67153
9 Has favorites° #	-0.09568	0.83911	1.19596	0.62219
10 Tells me where I stand	0.55183	0.54526	0.72867	0.60931
11 Annoying° #	-0.33137	1.09331	0.27198	1.21516
12 Stubborn	0.43859	0.75675	0.58702	0.63627
13 Knows job well	-0.46525	0.65278	-0.08280	0.78100
14 Bad	-0.18132	1.13741	-0.28108	1.34536
15 Intelligent° #	1.27822	0.53976	-0.26839	0.88066
16 Poor planner°	-0.28447	0.84466	0.57000	1.13454
17 Around when needed	1.09950	0.70855	0.64050	0.66922
18 Lazy	-1.04633	0.75197	-0.36413	1.04553
COWORKERS				
1 Stimulating°	-0.18014	0.25789	0.19236	0.68209
2 Boring	-0.57754	0.70647	-0.74244	1.01644
3 Slow	-0.52768	1.32504	-0.42618	0.94854
4 Helpful	-0.91814	0.77989	-0.98304	1.03060
5 Stupid	-1.04929	1.33617	-1.05806	1.16602
6 Responsible	-0.85790	1.09017	-0.73160	1.09680
7 Fast	-0.09386	0.68120	0.45810	0.58184
8 Intelligent° #	0.85764	0.67336	-0.55633	1.12459
9 Easy to make enemies	-0.66434	0.85226	-0.36841	0.70654
10 Talks too much°	-0.26554	0.97375	0.26075	0.54155
11 Smart°	0.25091	0.52367	-0.51345	1.37366
12 Lazy	-0.73901	1.18661	-0.42047	1.41136
13 Unpleasant°	-0.57021	1.04327	-1.00834	1.37722
14 Gossipy° #	-0.03979	0.84883	1.13024	0.57580
15 Active	-1.18149	0.54003	-0.55845	0.95846
16 Narrow interests	-0.26379	1.09575	0.08070	0.86159
17 Loyal	-0.13190	0.60063	-0.08452	0.76208
18 Stubborn	0.06172	0.68645	0.15418	0.67883
JOB IN GENERAL				
1 Pleasant° #	0.63615	1.54546	-0.17569	1.17932
2 Bad° #	-0.09578	1.14260	-0.58102	1.72771
3 Ideal° #	1.06346	0.87096	1.04665	0.98661
4 Waste of time	-0.83343	0.82508	-0.81107	0.89017
5 Good	0.08416	1.35159	-0.15559	1.58645
6 Undesirable	-0.45139	1.01186	-0.22659	1.50790
7 Worthwhile	-0.30846	1.28097	-0.18068	1.38482
8 Worse than most	-0.23101	0.83942	-0.82110	0.81374
9 Acceptable° #	0.42000	1.63321	-0.64623	1.16278
10 Superior	1.32057	1.08288	1.73988	1.02614
11 Better than most° #	1.63911	0.19405	0.36061	1.13163
12 Disagreeable	-0.01807	1.48397	-0.34208	0.95659
13 Makes me content	0.64466	1.40400	0.66365	1.26828
14 Inadequate	-0.14004	0.80371	-0.12397	1.18860
15 Excellent	1.21220	1.25661	1.41338	1.55913
16 Rotten	-0.77997	1.13686	-0.75734	1.27008
17 Enjoyable°	0.96832	0.71230	0.22541	1.72787
18 Poor	-0.52357	1.09770	-0.44137	1.41529

는 사실 이외에 한국과 미국의 양 문화권과 언어권에서 통용되는 보편적인 개념을 나타내 주는 문항이라고 말할 수 있다.

※ 2 값이 13.8보다 큰 값을 갖는 문항은 총 90문항 중 30문항이어서 약 33%가 편의된 문항으로 밝혀졌다. "Underpaid"(pay 적도; ※ 2 = 114.705)가 가장

큰 편의를 보여주고 있고 "Work on present Job" 척도의 문항 Challenging이 가장 높은 측정의 동등성을 달성한 것으로 밝혀졌다.

Lord의 κ^2 검정은 한국인과 미국인의 문항특성곡선 쌍 사이의 차이에 대한 전반적 유의도 검정이기 때문에 추측모수, 기울기 및 위치모수 쌍 중 어떤 쌍 사이의 차이에 대해서도 민감하다. 그림 1은 두 집단의 문항특성곡선 간의 차이를 평가할 때 얻을 수 있는 가능한 결과들을 나타내 주고 있다.

문항특성곡선 ①과 ②의 차이는 문항난이도 혹은 문항차택도에 기인한 측정의 비동등성을 나타내 주고 있고 ①과 ④의 차이는 변별도(기울기)의 차이에 기인한 측정의 비동등성을 나타내 준다. 또한 곡선 ①과 ③의 차이는 난이도와 변별도가 모두 차이가 나는 경우이다. 그림에 나타나 있는 문항특성곡선쌍은 모두 차이를 보이고 있지만 그 차이가 내포하고 있는 언어학적 의미와 문화적 의미에는 차이가 있다. 즉, 문항특성곡선의 기울기의 차이가 함축하고 있는 의미는 위치의 차이가 시사하는 의미와 다를 것이다.

문항난이도 차이에 기인한 문항특성곡선의 차이는 번역상의 문제와 연관되었을 가능성이 높고, 반면에 문항 변별도의 차이에 연유된 문항특성곡선의 차이는 편문화성을 시사해 주는 경우일 것이라는 추론이 있음을 앞에서 지적한 바 있다. 표 3은 문항특성곡선의 차이가 문화적인 편의에 기인한 것인지 혹은 번역의 오류때문인지를 추측할 목적으로 각 문항에 대한 문항난이도(b)와 문항변별도(a)의 추정값을 집단별로 제시하였다.

통계적으로 편의된 것으로 판정된 문항(표 3에 *로 표시) 중 대부분은 문항변별도 보다는 문항난이도에서 상대적으로 큰 차이를 보이고 있는 것 같다.

편의된 문항의 사후 내용분석을 하기 위하여 A 대학에 재직하고 있으며 미국에서 30여년간 체류한 적이 있는 심리학자에게 비편의된 문항의 번역에 대한 제검토를 의뢰했다. 번역이 적절하지 못하다고 지적된 문항은 12개의 문항(표 3에 #로 표시)이며 잔여 문항은 오역의 가능성이 적으므로 판명되었다. 예를 들어, intelligent라는 형용사는 supervision 과

coworkers 척도에 포함되어 있고 "지적이다"라고 번역했는데 내용적으로도 부적당한 번역인 것 같고 양 척도에서 모두 통계적으로 편의된 사실로 판단하여 오역임이 확실한 것 같다. 또한 supervision 척도의 통계적 편의 문항인 has favorites 는 '자기가 좋아하는 사람에게만 잘해준다' 로 번역했는데 '모든 일에 편파적이다'라는 번역이 원문의 뜻에 보다 충실하기 때문에, 오역에 의한 편의일 가능성이 높다. 통계적으로 편의된 문항중 내용적으로 볼때 오역일 가능성이 적은 18개 문항도 한국표본과 미국표본의 추정된 모수를 비교해 볼때 문항변별도 보다는 문항난이도에 의한 차이에 기인한 것이다. 그러므로 문항난이도와 변별도의 차이는 각각 오역과 문화적 상이성 때문일 것이라는 추리는 적어도 본 논문의 결과로는 명확히 입증되지 못한 셈이다.

논 의

본 논문은 심리검사의 원본과 번역본의 동등성을 문항을 분석단위로 하여 심리측정학적인 관점에서 탐구하고 있으며 측정의 동등성에 대한 정의는 문항반응이론에서의 문항반응에 대한 개념을 토대로 하고 있다. 심리검사문항의 번역이 종래의 번역이론에서는 심각하게 다루어지지 않은 부분이기 때문에 심리검사의 번역의 목적에 부합되는 번역 평가 절차가 요구되었다.

번역된 문항의 동등성에 대한 조작적 정의는, 측정하려고 하는 심리특질에서 동일한 양을 갖고 있는 사람들이 특정한 반응을 할 확률에 근거하고 있다. 동일한 양의 심리특질을 갖고 있는 사람들이 번역된 문항에 특정한 반응을 할 확률이 원문항에 동일한 반응을 할 확률과 같지 않을 때 그 문항은 동등하지 않다고 말할 수 있다. 올바른 응답을 하거나 찬택의 반응을 할 확률이 잠재적 심리특질에 대한 비선형회귀로 표현되는 문항특성곡선은 문항의 번역의 동등성의 평가를 위한 근거가 된다.

번역과 문화의 차이에 기인한 편의를 포착하기 위

하여 본 논문에서 사용한 방법이 갖는 강점은 측정되는 심리특질의 분포에 관한 특별한 가정을 하지 않는다는 데 있다. 문항반응이론의 특정한 모형이 어떤 모집단의 경험적 자료에 적합한 모형이라고 한다면 그 모집단으로부터 형성된 하위집단들에도 그 모형이 적합함이 수리적으로 입증될 수 있다. 문항특성곡선을 정의하는 문항모수는 동일모집단의 하위집단들에서는 적절한 선형전환의 과정을 거치면 이론적으로 불변이다. 이 불변성 때문에 연구자는 피험자 집단의 심리특질의 분포 형태와 관계없이 문항의 특성을 추정하고 비교할 수 있다. 그러므로 본 논문은 문항모수의 불변성을 이용하여, 특정문항에 있어서 비교되는 두 집단이 동일한 모집단에 속하는 집단이라는 가설을 검증한 것이다. 만일 두 집단이 문항반응이론의 특정한 모형이나 문항을 바탕으로 할 때 상이하다면 두 집단은 그 모형과 연구되고 있는 개념의 측면에서 동일한 모집단에 속하지 않는 집단이라는 결론을 내릴 수 있다.

반면에 고전점사이론적 방식으로는 문항의 동등성에 관한 전술한 방식과 같은 분석을 행할 수 없다. 예를들면, 고전점사이론에서의 문항난이도 P 는 특정한 표본 혹은 모집단 내에서 정의되며, 표본중에서 특정한 방식으로 응답한 사람의 비율을 나타낸다. 문항의 난이도가 표본의 특성이 변함에 따라 변한다는 바랍직하지 못한 모순된 상황이 일어난다. 마찬가지로, 고전점사이론의 문항변별력이나, 문항반응과 심리특질과의 관계는 모집단의 특성과 문항의 특성을 동시에 반영한다. 대부분의 번역의 목적이 번역집단에서의 심리특질치를 추정하는 것이기 때문에 문항모수가 우리가 추정하고자 하는 심리특질의 분포에 대한 가정에 영향을 받는다는 것은 문제가 아닐 수 없었다.

문항특성곡선의 괴리를 판정하기 위하여, 비교되는 두 집단으로부터 각각 추정된 문항모수치는 각기 다른 원점과 단위를 갖는 척도(metric)상에 놓이기 때문에 동일한 눈금을 갖도록 척도동등화를 반드시 행해야 한다. 그러나, 한 표본집단에 대해서 편의된 문항의 문항모수 추정치는 척도 동등화(scale linking)

과정에서 심각한 오류를 일으켜서 비편의된 문항을 편의된 문항으로 판정하게끔 오도하거나, 실제로는 편의된 문항을 비편의된 문항으로 판정하게끔 왜곡할 가능성도 배제할 수 없다. 이러한 왜곡의 가능성을 감소하기 위해서는 오직 편의되지 않는 문항만을 사용해서 척도동등화를 이루어야 한다. 이처럼 척도동등화를 시도하기 이전에 척도정화를 먼저 이루어내야 함을 인식한 연구자는 있었지만 정화를 위하여 제안된 절차가 논리적으로 완전하지 못했기 때문에 본 논문에서는 이를 보완한 절차를 적용했다. 본 논문의 결과는 이 보완된 절차의 효능을 입증하고 있다.

본 논문에서는 전체문항수의 2/3 가 언어학적인 측면에서 번역이 잘 되었고 동시에 문화적으로 보편적인 개념을 측정하고 있는 것으로 밝혀졌다. 이러한 사실만으로도 번역이 갖는 본유적인 어려움속에서도 범언어적이고 범문화적으로 동등한 측정을 달성할 수 있음이 입증되고 있다. 언어가 사람과 세상사이의 역파기의 역할을 하여 경험과 학습에 독특한 영향을 미치며 이로인해 다른 언어를 말하는 사람은 완전히 다른 세상에 살고 있어서, 비교언어적 연구는 근본적으로 불가능하다는 가설을 본 연구의 결과는 기각하고 있다.

편의된 문항의 발견에 논의의 초점을 맞추면 위와 반대되는 입장에 서게된다. 재생력이 뛰어난 번역이 기만 하면 범언어적, 범문화적 평가와 비교는 완벽하게 가능하다는 가설은 도전을 받게 된다. 본 연구에서와 같이 편의가 33%에 달하는 상황에서, 언어학적으로는 동등한 편이지만 심리측정학적으로는 만족스럽지 못한 검사도구를 근거로 비교언어, 비교문화적 연구를 하는 것은 타당하지 못할 것이다.

본 논문이 문항모수 추정과정에서 갖는 제한점은 각 하위척도가 일차원성의 가정을 벗어났을 가능성의 문제이다. 문항반응이론에 바탕을 둔 모수추정방법은 하위척도를 구성하고 있는 문항들이 오직 하나의 차원만을 측정하고 있을 때에만 정확한 추정치를 얻을 수 있다. 그러나 일차원성의 가정을 엄밀히 충족하는 자료가 존재할 수 있는지에 의구심을 나타내는 연구자도 적지않다. 또한 일차원성의 여부를 완벽

히 결정할 수 있는 통계적 절차도 존재하지 않는다. 요인분석방법이 일반적으로 쓰여지긴 하지만 각 문항의 응답을 기저하는 잠재적 특질이 정상분포를 이루고 문항간 사분상관계수(tetrachoric intercorrelations)를 요인분석할때만이 차원성을 정확히 결정할 수 있다는 제한점이 존재한다. 그러나 앞서도 말한 바와같이 일차원성은 실제의 자료를 분석할 때는 정도의 문제이지 여부의 이분적인 문제는 아니다. 또한 일차원성이 어느정도 위배되어도 문항모수추정은 비교적 정확하다는 연구 결과도 있다. 그러나, 모수추정치를 비교할때 어느 정도의 위배까지가 정당화될 수 있는지는 아직 명확히 밝혀지지 않고 있다. 이러한 상황에서 본 연구에서의 자료가 모수추정치의 비교를 저해할 정도로 다차원일 가능성도 배제 할 수는 없다.

심리측정학적으로 편의된 것으로 판정된 문항중에는 민족문화적 차이 때문만이 아니라 조직문화가 다른데서 기인된 편문화적인 문항도 있을 수 있다. 영어로 된 문항에 응답한 사람들이 속한 조직문화와 한글로 된 문항에 응답한 사람이 속한 조직문화의 차이 때문에 문항이 편의될 수도 있다는 것이다. 그리고 동일언어권과 동일문화권내에서도 응답자가 속한 조직의 문화적 차이에 의해서 문항편의가 발생할 수 있다. 따라서 본 논문에서 심리측정학적으로 편의된 것으로 판정된 문항중 오역의 가능성이 높은 문항이외의 문항중에는 미시적인 관점에서 조직문화적 차이때문에 편의된 것으로 나타난 문항도 있을 수 있다.

본 논문에서 편의되었다고 판정된 문항이 오역때문인지 혹은 측정되는 심리특질의 편문화성 때문인지 혹은 이 두 원인이 복합되어 편의되게 되었는데 지에 관하여 본 연구 결과만으로는 단언할 수 없다. 보다 정확한 원인을 밝히기 위해서는 되번역절차를 통해 미흡하다고 생각되는 번역을 수정하고 수정된 번역문항이 재차 통계적으로 편의되었다고 판정되면 문화적인 편향성에서 그 원인을 찾고, 편의되지 않은 것으로 밝혀지면 오역이 원인일 것으로 추정할 수 있을 것이다.

본 연구에서 적용한 문항반응이론에 근거한 절차는

이분문항 혹은 이분문항으로 전환할 수 있는 문항을 담고 있는 심리측정도구의 번역에 광범위하게 활용할 수 있는 유망한 기법이다. 따라서 외국의 심리측정도구를 많이 번역 혹은 번안하여 쓰고있는 심리학을 포함한 행동과학분야의 연구에서 자주 이용해야 할 기법이라고 판단된다.

참 고 문 헌

- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443-459.
- Brislin, R. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1, 185-216.
- Brislin, R. W. (1986). The wording and translation of research instruments. In W.J. Lonner & J.W. Berry (Eds.), *Field methods in cross-cultural research* (pp 137-164). Beverly Hills: Sage.
- Candell, G.L., & Hulin, C.L. (1987). Cross-language and cross-cultural comparisons in scale translation: Independent sources of information about item nonequivalence. *Journal of Cross-Cultural Psychology*, 17, 417-440.
- Casagrande, J. (1954). The ends of translation. *International Journal of American Linguistics*, 20, 335-340.
- Divgi, D. R. (1985). A minimum chi-square method for developing a metric in item response theory. *Applied Psychological Measurement*, 9, 413-415.
- Drasgow, F., & Parsons C. L. (1983). Application of unidimensional item

- response theory models to multidimensional data. *Applied Psychological Measurement*, 7, 189-199.
- Drasgow, F. (1984). Scrutinizing psychological tests: Measurement equivalence and equivalent relations with external variables are the central issues. *Psychological Bulletin*, 95, 134-135.
- Drasgow, F., & Hulin, C. L. (1986). *Assessing the equivalence of measurement of attitudes and aptitudes across heterogeneous subpopulations*. Unpublished manuscript. University of Illinois at Urbana-Champaign.
- Drasgow, F. (1989). An evaluation of marginal maximum likelihood estimation for the two-parameter logistic model. *Applied Psychological Measurement*, 13, 77-90.
- Ellis, B. B. (1989). Differential item functioning: Implications for test translations. *Journal of Applied Psychology*, 74, 912-921.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.
- Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory: Applications to psychological measurement*. Homewood, IL: Irwin.
- Hulin, C. L., & Mayer, L. J. (1986). Psychometric Equivalence of a translation of the Job Descriptive Index into Hebrew. *Journal of Applied Psychology*, 71, 83-94.
- Hulin, C. L. (1987). A psychometric theory of evaluations of item and scale translations: Fidelity across languages. *Journal of Cross-Cultural Psychology*, 18, 114-142.
- Lautenschlager, G. L. & Park, D. G. (1988). Item response theory item bias detection procedures: Issues of robustness, model misspecification and parameter linking. *Applied Psychological Measurement*, 12, 365-376.
- Linn, R. L., Levine, M. V., Hastings, C. N., & Wardrop, J. L. (1981). Item bias in a test of reading comprehension. *Applied Psychological Measurement*, 5, 159-173.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- McLaughlin, M. E. (1986). *Computing Lord's item bias statistic when abilities and item parameters are estimated simultaneously*. Master's thesis, Department of Psychology, University of Illinois at Urbana-Champaign.
- Park, D. G. & Lautenschlager, G. L. (1990). Improving IRT item bias detection with iterative linking and ability scale purification. *Applied Psychological Measurement*, 14, 163-173.
- Parsons, C. K., & Hulin, C. L. (1982). An empirical comparison of item response theory and hierarchical factor analysis in applications to the measurement of job satisfaction. *Journal of Applied Psychology*, 67, 826-834.
- Segall, D. O. (1983). *Test characteristics curves, item bias, and transformations*

to a common metric in item response theory : A methodological artifact with serious consequences and a simple solution. Unpublished manuscript, University of Illinois.

Smith, P. C., Kendall, L. M., & Hulin, C. L. (1969, 1985). *Measurement of satisfactions in work and retirement*. Chicago : Rand-McNally.

Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group-mean differences : The concept of item bias. *Psychological Bulletin*, 99, 118-128.

Wood, R. L., Wingersky, M. S., & Lord, F. M. (1976, June). *LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters*. Research Memorandum 76-6. Princeton, NJ: Educational Testing Service.

Werner, O., & Campbell, D. (1970). Translating, working through interpreters, and the problem of decentering. In R. Caroli and R. Cohen (Eds.), *A handbook of method in cultural anthropology*. New York: Natural History Press, 398-420.

Measurement Equivalence of a Translation of a Job Satisfaction Inventory into Korean

Dong-Gun Park

Ajou University

This study investigates the utility of a statistical procedure to evaluate the measurement equivalence of a translation of the Revised Job Descriptive Index, a job satisfaction inventory. An item analysis method based on item response theory is used to detect differential item functioning in the translation. Results of an IRT analysis procedure identified 30 out of 90 items as significantly biased. Importance of a proper scale linking procedure in this IRT procedure is emphasized. The implications of using an IRT analysis in examining the fidelity of translated psychological measurement instruments are discussed.