

연구논문

성격검사의 문화 타당도 연구*

Cross-Cultural Validity of Personality Tests

손원숙**

Wonsook Sohn

이 연구의 주된 목적은 성격검사의 문화 타당도를 평가수준에 얼마나 요구해 하는 것이다. 차등기능을 가진 한 assay와 Sixteen Personality Factor (16PF) 성격검사와 PSBTTEST 고도화된 비교하면서 두 문화 집단 즉 한국과 미국간에 적용되는 가장 높은 성격검사의 문화를 이해함으로써 각 문화의 성격과 미국집단에서 적용된 내용을 비교함으로써 문화적 차이를 평가한다. PSBTTEST의 결과에 따르면 성격의 차이를 가장 정확히 평가할 수 있는 문화의 문화를 평가한다. 일반적으로 기능하고 있음이 나타났다. 따라서 이 검사 기술을 이용하여 집단 간 차이를 평가하는 방법을 같이 해 알 수 있어서 검사 결과가 보강되지 못할 것임에 따라 모든 연구가 필요할 것이다. 마지막으로, 차등기능을 가진 한 assay와 PSBTTEST의 문화를 평가하는 방법론에 대한 시사점을 논의하였다.

High fidelity translations of measurement instruments from one language to another are required for meaningful comparisons between cultures. A differential item functioning technique, PSBTTEST was used to assess the degree to which a personality test provided cross-cultural validity across U.S. and Korea. The Sixteen Personality Factor (16PF) questionnaire was administered in English to 844 American college students and in Korean to 538 Korean college students. Results of PSBTTEST classified 23 out of 51 items as significantly biased. Implications of this procedure for translation evaluation are discussed.

I. 서론

최근 국제학력비교나 비교문화연구에 대한 관심이 증가하고, 다양한 언

* 내용의 논리적 전개 뿐 아니라 문장 표현과 맞춤법에 이르기까지 세밀하게 피드백을 주신 심사위원 여러분께 진심으로 감사 드립니다.

** 이화여대 심리학과 시간강사(julie69@hanmail.net)

어로 이루어진 자격시험에 대한 수요가 급증하고 있다 (Hambleton, 1994). 이러한 추세에 맞추어서 한 나라의 언어로 개발된 검사를 다양한 언어와 문화에 맞게 번역하여 사용하려는 경향이 증가하고 있다. 예를 들어 다양한 인종을 보유하고 있는 기업에서 사원선발, 자격증 부여, 혹은 승진시험과 같은 중요한 장면에서 한가지 언어로만 구성된 검사를 실시한다고 가정하자. 외국인 사원의 검사 점수 결과는 실제 그 검사가 재고자 하는 능력보다는 그 검사를 구성하고 있는 언어능력수준을 반영하게 될 것이다. 따라서 이러한 편파를 배제하기 위하여, 검사를 적절하게 번역해서 사용하는 경향이 증가되고 있다. 이러한 시험들은 주로 사원들의 연봉, 승진, 그리고 시장성에 직접적인 영향을 주는 고부담(high stakes) 검사이기 때문에, 점수의 동등성 문제는 상당히 중요하다고 할 수 있다. 즉 각 언어버전 검사에서 나온 점수의 의미가 동일한지? 혹은 서로 비교 가능한 것인지? 등에 대한 전제조건을 검토해 보아야 한다. 이러한 과정을 통해서 각 검사의 결과를 신뢰할 수 있고, 공정하게 사용할 수 있을 것이기 때문이다. 미국의 경우, 특히 많은 전문인력을 필요로 하는 정보기술 (Information Technology) 분야에서 검사의 번역과 점수 동등성에 대한 연구가 두드러지고 있다. 예를 들어, 최근 컴퓨터 소프트웨어 회사인 NOVELL 이나, MICROSOFT 회사들은 여러 가지 다른 나라의 언어로 구성되어 있는 사원선발용 시험 혹은 자격시험에 대한 심리측정학적(psychometric) 검토를 활발히 하고 있다.

이처럼 연구자들이 검사를 번역하여 사용하는 상황에서 반드시 고려하여야 할 몇 가지 방법론적인 문제들을 다음과 같이 정리해 볼 수 있다 (Church & Lonner, 1998). 첫째, 점수의 동등성(equivalence)에 초점을 두면서 검사를 번안할 수 있는 방법들의 개발, 둘째, 비교 문화 혹은 비교 국가 자료를 사용하고 해석하는 방법들, 마지막으로 검사를 번안하는 지침의 개발과 사용에 관한 것이다. 이 중에서 첫 번째 문제는 바로 비교 문화연구의 가장 근원적인 쟁점이라고 볼 수 있는 동등성의 확립에 관한 것이다. 동등성은 검사 점수들이 서로 다른 집단에서 비교될 수 있는 측정수준을 말하는 것으로, 이런 동등성이 보장되지 않는다면 집단간 타당한 점수비교가 어려울 것이다. Drasgow (1984)는 이런 동등성을 문항반응

이론의 맥락에서 “관찰 점수와 그 검사가 측정하려고 하는 잠재속성간의 관계가 하위 집단들에서 동일할 때 동등한 측정이 얻어지는 것”이라고 정의 내렸다.

우리 나라의 많은 연구자들이 외국의 검사를 도입하여 사용하는 관행을 살펴보면, 두 가지 언어에 모두 능숙한 사람들을 이용하여 검사번역을 실시하고, 검사 점수들간의 동등성은 경험적으로 검증하지 않은 채, 당연히 두 검사들은 동등할 것이라는 가정을 받아들이고 있다. 그러나 국제검사위원회 (International Test Commissions) 지침에서도 밝혔듯이, 검사 개발자들은 여러 언어유형의 검사의 동등성을 형성하기 위하여 적절한 통계적인 방법을 체계적인 질적인 방법과 더불어서 사용하기를 권장하고 있다 (summarized by Hambleton, 1994). 번역과정이란 것은 동등한 검사를 만들기 위한 하나의 초기과정일 뿐이지, 이것이 두 검사간의 동등성을 보장하는 것은 아니라는 점을 강조하고 있다.

최근에 검사의 비교 문화적 동등성(cross-cultural equivalence)을 평가하기 위하여 사용할 수 있는 하나의 통계적인 방법으로 차별기능문항(Differential Item Functioning) 기법이 소개되고 있다. 차별기능문항(이하 DIF라고 부름)이란 한 검사 문항이 특정 배경변인이 서로 다른 두 집단 (예컨대 여자와 남자집단, 또는 백인과 흑인집단)에게 제시되었을 때, 그 검사 문항이 비교되는 두 집단에서 서로 다르게 기능 하는 현상으로 정의된다 (김신영, 2001 참고). 예를 들어, 영어로 구성되어 있는 原 성격검사를 한국어로 번역하였다고 가정하자. 이 경우 두 검사(영어판과 한글어판)는 비록 두 가지 다른 언어로 표현되어있다고 할 지라도 이 검사문항들은 기본적으로 동일한 내용 즉, 특성(trait)이나 구인(construct)을 측정하고 있음을 가정한다. 만약 이 검사문항들이 서로 다른 구인이나 특성을 측정한다면 검사 측정하고자 하는 구인을 서로 다른 정확성을 가지고 측정한다면 이 두 언어 버전의 검사는 편파적인 결과를 산출할 것이다. DIF 기법들은 번역의 질, 잠재변인과 문항들간의 관계, 그리고 서로 다른 집단간의 검사 점수의 동등성에 대한 정보를 얻을 수 있으며, 또한 문제를 나타내고 있는 문항들의 소재를 찾아낼 수 있다는 점에서 상당히 유용하게 응용되고 있다.

이 연구의 목적은 번역된 성격요인검사와 原 검사간의 문화간 타당도를 통계적인 방법을 통하여 평가해 보고자 하는 것이다. 즉 외향성을 측정하고 있는 성격요인검사 문항들이 문화와 언어가 다른 한국과 미국집단에서 동등성을 유지하고 있는지 경험적으로 평가하는 것이다. 이를 위하여 본 연구에서는 차별기능문항기법 중 다분문항을 다룰 수 있는 SIBTEST(Shealy & Stout, 1993)라는 프로그램을 이용하였다.

II. 차별기능문항

1. 차별기능문항의 정의

문항 혹은 검사 편파성 (Item or Test Bias)란 것은 검사결과가 과연 검사 문항 혹은 검사가 재고자 하는 특성에 대하여 정확하고 공정한 정보를 제공하는가에 관한 문제와 관련되어 있다. 최근에는 “문항 편파”라는 용어보다는 차별기능문항(DIF)이라는 좀 더 중립적인 용어를 사용하는 경향이 있는데, 즉 “공평하지 못함” 혹은 “좋지 못하다”는 그런 의미라기 보다는 다른 문항들과는 구별되는 기능을 하는 문항을 지칭한다고 볼 수 있다 (김신영, 2001).

주로 인지적인 검사에서 인종/문화 차별적 혹은 성 차별적 문항들을 걸러내는 용도로 사용되어왔던 차별기능문항 기법을 최근에는 검사 번역 장면에서도 활발하게 응용하고 있다. 서로 다른 언어 혹은 문화 집단에 속하는 개인들이 검사가 재고 있는 속성, 즉 동일한 능력, 성격, 혹은 태도를 가지고 있다면 이 두 집단 구성원들은 동일한 문항에 대하여 동일한 방식으로 (e.g., 긍정 혹은 부정) 응답하기를 기대한다. 그러나, 만약 그들이 동일한 방식으로 반응하지 않다면 그 문항은 DIF를 나타낸다고 할 수 있으며, 그 원인에 대한 심층적인 분석이 필요할 것이다. DIF의 개념을 보다 정확히 이해하기 위해서는 차별기능문항과 영향(impact)의 개념은 구별할 필요가 있다. 영향은 집단간 차이로서 재고자 하는 특성의 집단간 차이라고 설명할 수 있다. 즉, 서로 다른 특성을 가진 두 집단 구성

원들을 비교하여, 그들이 동일한 문항에 대하여 다르게 반응하는 경우, 이때 이 문항은 DIF를 나타내는 것이 아니고, 그 문항의 집단간 차이 혹은 영향이라고 부른다.

2. SIBTEST (Simultaneous Item Bias Test)

Shealy와 Stout (1993)는 DIF를 한 검사의 문항들에서 일차원성이 위배되면 나타나는 것으로 개념화 시키면서, Simultaneous Item Bias Test (SIBTEST) 프로그램을 개발하였다. 이 검사는 처음에는 오로지 이분문항만을 다룰 수 있는 검사로 출발하였으나, 현재 비일방적인 DIF (“Crossing DIF” 라고도 부름)를 추출할 수 있고, 다분 문항을 다룰 수 있으며, 두 개의 대응변수를 다룰 수 있는 프로그램들을 개발하고 있다. 이 SIBTEST의 이론적인 구조하에서 DIF를 다음과 같은 방식으로 정의하고 있다. 즉, 한 성취도 검사는 하나의 목표 능력 (target ability), θ 로 구성되어 있지만, 문항반응들이 하나 혹은 그이상의 잡음 결정인자 (nuisance determinants), η 에 의하여 결정될 때, DIF에 대한 잠재성이 나타난다고 한다.

이 SIBTEST는 검사 문항들을 두 개의 하위 검사, 즉 1) 대응(matching) 검사 혹은 문항과 2) 연구되어지는 검사 혹은 문항(a studied test or item)으로 분리한다. 각 준거집단과 초점집단에 속하는 피험자들은 그들의 대응 하위검사 점수에 따라서 J개의 집단으로 나뉘어 지며, 대응 하위검사에서 같은 점수를 가지고 있는 서로 다른 두 개의 집단 (준거 혹은 초점 집단)에 속한 피험자들을 각 연구문항에 대한 그들의 수행으로 비교하는 것이다. 이 검사에서는 다음과 같은 영가설과 대안가설을 평가하게 된다.

$$H_0: \beta_U = 0, \quad H_a: \beta_U \neq 0$$

β_U 라는 것은 하나의 문항이 연구되어지는 경우 일방적 DIF의 양을 나타내는 모수로서 다음과 같은 공식에 의하여 계산되어진다.

$$\beta_U = \sum_{j=1}^J p_j (Y_{Rj}^* - Y_{Fj}^*)$$

여기에서 P_j 는 하위집단 j 에서, 초점 집단 구성원의 비율을 나타내며, Y_{Rj}^* 와 Y_{Fj}^* 은 준거집단과 초점집단에 있는 피험자들의 연구 문항의 조정된 평균들이다(각 하위집단 j 에서). 즉, SIBTEST는 고전진점수이론에 근거하여 잠재변인 (e.g., 능력, 혹은 외향성)을 추정하고, 그 추정치 위에 문항 수행을 회귀시킨다. 각 대응변수 점수 수준에 따라서, 준거집단과 초점집단의 문항-진점수 회귀선의 차이에 대한 평균을 낸 것이다. 여기에서 유의미한 양수의 β_U 값은 준거집단에 유리하게 기능하는 문항이고, 유의미한 음수의 β_U 값은 초점집단에 유리하게 기능하는 문항이다.

특히, 대응변수의 선택은 차별기능문항 분석에서 중요하며, 이 SIBTEST의 이론적 구조하에서는 몇 가지 방법으로 대응변수 점수를 선택할 수 있다(Shealy & Stout, 1993). 첫째, 문항내용이나 고전 검사 통계치들을 검토하여 전문가적인 판단에 근거하여 대응변수를 구성할 문항을 선정할 수 있다. 그러나, 이는 흔히 영향(impact)과 DIF를 혼동할 가능성이 있으므로 주의가 요구된다. 두 번째로는 자동적인(automatic) DIF 분석을 실시하여, 자동적으로 각 문항을 DIF를 위하여 검토하고, 그 문항을 제외한 나머지 문항들을 대응변수로서 사용하는 것이다. 본 연구에서는 이 자동적인 DIF분석을 사용하였다.

III. 연구 방법

1. 연구대상

이 연구에서는 문화와 언어가 다른 두 개의 집단, 즉 한국 대학생 538명과 미국대학생 844명을 피험자로 사용하였다. 2000년 1월부터 5월에 걸쳐, 한국 표본은 한국의 두 개의 사립대학교로부터 얻어졌고, 미국 표

본 중 431명은 중부의 한 주립대학교에서, 나머지 413명은 미국의 <성격 및 능력 검사 연구소>로부터 얻어졌다. 이 연구에서 피험자를 대학생으로 국한시킨 이유는 비교문화연구에서 가장 중요한 문제라고 볼 수 있는 비교 가능한 표본의 선정을 위하여 가능한 한 같은 직업 군에 속하고, 비교적 비슷한 교육 수준을 가지고 있는 피험자를 선정하기 위한 것이었다.

2. 측정 도구

이 연구에서는 정상 성인을 위한 성격검사로써 가장 광범위하게 쓰여지는 검사 중 하나인 The Sixteen Personality Factor (16PF) Questionnaire (Cattell & Cattell, 1995)가 사용되었다. 이 16PF 검사는 한국에서는 염태호와 김정규(1990)에 의하여 표준화되어서 “성격요인검사”라는 이름으로 널리 사용되어 오고 있지만, 이 검사는 영어판 16PF와 비교해 볼 때 문항 수, 문항반응 양식, 그리고 문항내용 등 상당 부분 수정되어 있기 때문에 문항 단위에서 동등성을 알아보고자 하는 본 연구의 목적에는 적절치 않았다. 따라서, 본 연구에서는 기존의 성격요인검사 대신 Shaughnessy와 Kang(1998)이 번역한 한글판 16PF 검사를 수정하여 사용하였다. 16PF 검사는 세 가지 수준의 강제 선택 문항 양식을 사용하며, 추리척도를 제외한 나머지 16개 척도에서 “그렇다”, “잘 모르겠다 혹은 둘 다 그렇지 않다”, 그리고 “그렇지 않다”라는 방식의 반응양식을 가지고 있다. 이 16개의 주요요인들은 외향성, 불안감, 강인함, 독립성과 자기통제, 즉 5가지의 이차 요인들로 다시 묶여질 수가 있다. 이 이차 요인들 가운데, 본 연구에서는 51개의 문항으로 구성된 외향성 요인만을 살펴볼 것이다. 외향성 요인은 다섯 가지 하위척도로 구성되어 있는데, 즉, 온정성 (factor A: 11문항), 쾌활성 (factor F: 10문항), 사회적 대담성 (factor H:10문항), 내밀성 (factor N:10문항), 그리고 자기 의존성 (factor Q2:10문항)이다.

3. 분석절차

1) 한글판 16PF 검사

Shaughnessy와 Kang (1998)이 마련한 한글판 16PF 검사를 기초로 하여, 본 연구에서는 검사 번역가로서의 기준에 적절히 부합되는 두 명의 박사과정 학생들을 선정하여 이 한글판 16PF 검사의 번역 수준을 평가하도록 하였다. 여러 번의 편집 회의를 거듭하여, 이 두 명의 번역가로부터 최종 한글판 16PF 검사 문항들이 얻어 졌다.

2) 문항의 채점

16PF 검사 문항들은 미국 성격 및 능력 검사 연구소로부터 나온 검사 매뉴얼(Russell & Karol, 1994)에 근거하여 채점이 이루어졌으며, 정답으로 정해진 반응들은 “2점”을, 중립반응은 “1점” 그리고 나머지 반응들은 “0점”이 주어졌다.

3) 일차원성 (unidimensionality) 평가

차별기능문항분석에 앞서서 자료의 차원성(dimensionality)을 평가하였다. 검사를 구성하고 있는 문항들이 그 검사가 측정하고자 하는 능력, 성격, 혹은 태도와 같은 하나의 구성요인을 재고 있는지를 평가해 보는 것이다. 만약 문항들이 그 검사가 재고자 하는 구성요인 이외의 다른 특성을 재고 있다면 그것은 문항편파를 일으키는 하나의 원인이 될 수 있기 때문에 이 차원성 검사는 항상 필요한 절차이다. 각 인종 집단 별로 각 척도의 차원성을 평가하기 위해서 본 연구에서는 주성분 분석을 실시하였다. 각 인종집단별로, 외향성의 다섯 가지의 하위척도들은 개별적으로 평가되었으며, 먼저 PRELIS 2.0 (Jöreskog, & Sörbom, 1993)을 이용하여 다분상관계수 (polychoric correlations)를 바탕으로 주성분 분석이 실시되었다. Reckase(1979)에 따르면, 일차원성을 가지고 있는 검사는 하나의 주된 요인들로 구성되어 있고, 그 주된 요인의 고유값이 나머지 요인들의 고유값들보다 훨씬 커야 한다고 하면서, 적어도 그 주된 요인이 전체 분산의 20%이상을 설명해야 한다고 주장하였다.

4) 차별 기능 문항 분석

SIBTEST 프로그램 가운데 3가지 반응 옵션을 가지고 있는 성격요인검사 문항분석에 적절한 PSIBTEST(Polytomous Version of SIBTEST)을 사용하였다. PSIBTEST에서는 두 집단 즉, 준거집단 또는 초점 집단 모두에 대립하여 DIF를 나타내는 문항들을 구별하기 위하여 “e” 옵션과 함께 자동적인 DIF 절차를 이용하였다.

β 는 PSIBTEST에서 DIF의 양을 나타내는 계수로서 이분문항을 위한 지침에 따르면 영가설이 기각되었을 때, 만약 $|\beta| < 0.059$ 이면, DIF는 무시할 만큼 작은 정도 혹은 “A-수준의 DIF”; 만약 $0.059 \leq |\beta| \leq 0.088$ 이면, DIF는 중간 정도 수준 혹은 “B-수준의 DIF”, 그리고 만약 $|\beta| \geq 0.088$ 이면 DIF는 큰 정도 혹은 “C-수준 DIF” 라고 한다. 이 지침을 다분 문항에 적절하게 변환시키기 위하여서는, β 의 계수를 각 문항의 점수 범주의 수로 나누어 주면 된다. 예컨대, 성격요인검사는 세 개의 반응 옵션을 가지고 있고, DIF의 양을 알아보기 위하여 β 의 값을 “2”로 나누어 주면 된다. 본 연구에서는 이러한 DIF의 양을 산출한 후 C 수준의 DIF만을 통계적으로 유의미한 수준의 DIF로 간주하였다.

IV. 연구 결과

1. 기술 통계치(Descriptive Statistics)

척도별 신뢰도 계수를 각 인종집단별로, 그리고 다시 각 집단내의 성별로 살펴보았다. 미국표본에서는 .78의 중앙치를 가지고, .69-.87의 내적 일관성 계수 범위를 보였고, 한국표본에서는 .62의 중앙치와 함께 .54-.87의 범위를 가지고 있었다. 미국표본의 신뢰도 계수는 16PF 검사 매뉴얼에 제시된 신뢰도 계수값 (범위 .66-.86)들과 상당한 일치를 보여 주고 있지만, 한국표본에서는 비교적 낮은 수준의 내적 일관성을 보이고 있었다. 특히, 한국 남성표본의 신뢰도 계수가 낮은 것으로 나타났다.

2. 일차원성 (unidimensionality)

<표 1>에 주성분 분석의 결과가 요약되어져 있다. A, F, N 척도에서 집단간 설명된 분산의 양을 비교해 보면 한국표본에서 대체적으로 작은 설명량을 보이지만, 첫 번째 요인에 의해서 설명된 분산이 5개의 모든 척도에서 20% 이상으로 나타났으며, 첫 번째 요인의 고유값이 나머지 요인들의 고유값들보다 훨씬 큼을 알 수 있었다. Reckase(1979)의 기준에 따라, 각 하위척도는 두 개의 집단 모두에서 어느 정도 일차원성의 가정을 만족시킨다고 할 수 있었다.

<표 1> 주성분 분석의 결과

| | 미국표본(n =844) | | | 한국표본(n =521) | | |
|----|---------------------------------------|------------|-----------|--------------|------------|------------|
| | ^a PC_1 | PC_2 | PC_3 | PC_1 | PC_2 | PC_3 |
| A | 3.81 ^a (34.6) ^b | 1.62(14.7) | 0.97(8.8) | 2.57(23.4) | 1.36(12.4) | 1.28(11.7) |
| F | 3.88(3.88) | 1.15(11.5) | 0.92(9.2) | 3.00(30.0) | 1.37(13.7) | 1.04(10.5) |
| H | 6.08(60.8) | 0.96(9.6) | 0.59(5.9) | 5.96(59.6) | 1.00(10.0) | 0.64(6.4) |
| N | 4.69(46.9) | 1.20(12.0) | 0.92(9.2) | 2.93(29.3) | 1.34(13.4) | 1.11(11.2) |
| Q2 | 4.76(47.6) | 1.02(10.2) | 0.98(9.8) | 4.29(42.9) | 1.06(10.6) | 0.85(8.5) |

주: ^a고유값(Eigenvalues); ^b설명 분산량(% Variance); ^cPC=주성분(Principal Components)

3. 인종 집단에 따른 차별 기능 탐색

각 하위척도의 점수를 단일 대응변수로서 사용하였고, 연구문항(studied item)의 점수는 대응변수 점수 계산 시 제외시켰다. PSIBTEST에서 얻어진 결과는 <표 2>에 요약되어져 있다(유의도 수준 = .01).

총 51개의 외향성 문항 가운데 23개 문항(전체 45%)이 한국과 미국집단에서 차별적으로 기능하고 있음이 밝혀졌다. 이 중 11개 문항이 미국집단에, 그리고 나머지 12개 문항이 한국집단에 유리하게 기능 함을 알 수 있었다. 동일한 온정성의 정도를 가지고 있는 한국인과 미국인이 1번 문항

<표 2> 인종집단에 따른 PSIBTEST 분석 결과

| Factor A: 온정성 (11문항) | Factor F: 쾌활성 (10문항) | Factor H: 사회적대담성 (10문항) | Factor N: 내밀성 (10문항) | Factor Q2: 자기의존성 (10문항) |
|---|--|--------------------------------|---|---|
| 1번(미국) 33번(미국) 96번(한국) 159번(한국) 98번(한국) | 68번(한국) 100번(미국) 124번(한국) 70번(미국) | 73번(한국) 71번(미국) 105번(한국) | 47번(한국) 143번(미국) 148번(미국) 15번(한국) 84번(미국) 117번(한국) | 89번(미국) 121번(한국) 152번(한국) 123번(미국) 156번(미국) |

주: 각 DIF 분석결과, 각 문항에서 유리하게 기능하는 집단을 괄호 안에 명시함.

에 응답하였을 경우, 미국인이 한국인보다 이 문항에 대하여 더욱 온정적인 방향으로 대답하는 경향이 있었다. 또한 동일한 수준의 사회적 대담성을 가지고 있는 한국인과 미국인이 있다고 할 때, 73번 문항에 대하여 한국인이 더욱 사회적으로 대담성을 나타내는 방식으로 응답한다는 것이다. 위의 23개 DIF 문항들은 미국과 한국 집단에서 서로 다른 구인이나 특성을 측정한다든지 혹은 검사가 측정하고자 하는 구인을 서로 다른 정확성을 가지고 측정하고 있을 가능성을 시사하고 있다. 이와 같이 일부 문항에서는 문화간 타당성의 근거가 충분치 않으므로 이 외향성 척도의 점수를 이용하여 한국집단과 미국집단을 비교하는 연구는 타당하지 않을 것이다. 이러한 통계적인 차별기능문항 기법들은 왜 문항들이 두 문화그룹에서 차별적으로 기능하고 있는지, 그 원인에 대한 설명은 제공해 줄 수 없다. 따라서, 이러한 통계적인 분석 후에 질적인 방법으로 DIF 문항들의 원인을 평가 해 본 후, 집단간 동등한 점수 비교가 가능하도록 일부 문항을 수정하거나 삭제할 필요가 있을 것이다.

V. 논의

본 연구에서는 외국의 교육 및 심리검사를 도입하여 비교연구에 사용하

기에 앞서서, 검사 점수의 동등성 확립이 무엇보다도 중요하며, 이런 동등성을 평가하는 것은 질적인 방법 뿐 아니라, 통계적인 방법도 함께 수반되어야 한다는 점을 강조하고자 한다. 특별히, 차별기능문항 기법을 이용하여, 서로 다른 언어와 문화 집단에서 문항 수준에서의 동등성이 유지되고 있는지를 살펴보았다. 연구 결과를 요약하자면, 외향성 척도에서 총 45%의 문항이 차별기능을 나타내었고, 각 척도별로 살펴보자면, A척도의 45% 문항, F척도의 40%문항, H척도의 30%문항, N척도의 60% 문항, 그리고 Q2척도의 50% 문항이 DIF를 나타내었다. 이 결과는 중국어 버전과 스페인어 버전(Zhang & Yan, 2000)의 16PF(Ellis & Mead, 1998)를 다루었던 두 선행연구 결과와 어느 정도 일치되는 것으로 이 두 연구에서도 H척도에서 가장 적은 수의 DIF 문항을 발견하였고, N척도에서 가장 많은 수의 DIF 문항을 발견하였다. 따라서, H척도는 비교적 비교문화장면에서 사용하기 쉬운 문항편파가 없는 문항들로 구성되어 있다고 볼 수 있지만, 다른 척도들을 이용하여 서로 다른 언어 및 문화 집단간 비교를 할 때는 상당한 주의가 필요할 것이다.

본 연구의 결과가 시사하고 있는 몇 가지 점들을 논의하고자 한다. 첫째, 본 연구에서는 두 번의 단계를 거쳐서 영어판 16PF를 한글로 번역하였다. 그럼에도 불구하고, DIF 분석의 결과에 따르면 상당수 문항들이 두 집단에서 차별적으로 기능하고 있음이 밝혀졌다. 이는 검사개발자나 비교문화 연구자들에게 중요한 하나의 시사점을 제공해 주고 있다. 즉, 두 가지 언어로 된 검사들간의 동등성이라는 것은 단순한 번역의 과정을 통해서만 이루어질 수 없는 것이며, 보다 체계적인 질적인 방법과 양적인 방법이 사용되어야 한다는 점이다. 더불어서, 한 검사의 비교 문화적인 동등성의 확립은 단일 방법에 의해서 이루어 질 수 없는 것이고, 다양한 종류의 동등성을 다룰 수 있는 다양한 방법이 사용되어야 하는 연속적인 과정이라는 것이다. 우선, 검사를 번역하는 과정에서 상당한 시간과 노력을 투자하여야 하며, 일단 번역 된 검사를 비교문화연구에 사용하기에 앞서서 통계적인 방법과 질적인 방법을 이용하여 다각적으로 검사의 동등성을 검토해야 할 것이다. 더불어서, 연구자들은 검사번역의 초기단계에서 어떻게 검사들을 번역할 것인가의 문제에 대하여 신중히 검토해 볼

필요가 있다. 본 연구결과에서 DIF를 나타내고 있는 문항들을 두 명의 언어 전문가(bilinguals)에게 제시한 후 DIF의 원인을 분석하도록 하였다. 그들의 질적인 분석 결과에 따르면, 일부 문항 내용이 한국 문화에 적절하지 않은 경우와 번역의 어려움 등이 가장 중요한 DIF 원인이었다. 이는 검사번역의 중요성을 시사하는 것으로서, 효과적이고 체계적인 번역 기법의 개발의 필요성을 강조하고 있다. 특별히 성격검사와 같은 심리검사를 번역하는 문제는 심리적이고, 언어적이며, 문화적 고려가 모두 필요한 아주 어려운 과정이다. 이는 교육 및 심리검사를 번역하여 사용하는 것은 단순히 언어적인 번역만을 제공하는 것 이외에 그 검사가 재고자 하는 구성개념에 대한 이해를 바탕으로 그것을 각 문항에서 표현해 낼 수 있도록 하는 과정이 필요하다는 사실을 시사하고 있다. 앞으로 여러 가지 다양한 번역 기법들의 효과성을 검토하는 경험적 연구들이 진행되어야 하며 이를 통해 외국의 검사를 번역하여 자신들의 연구에 사용하려는 연구자들에게 도움을 줄 수 있을 것이다.

본 연구에서는 균일적인(uniform) DIF만을 추출할 수 있는 SIBTEST라는 프로그램을 사용하였다. 하지만, 문항수행률 수량화(비모수적 vs. 모수적)하고 대응변수를 결정하는 방법(잠재변인 vs. 관찰점수변인)에 따라 다양한 차별기능문항 기법들이 존재한다. 그리고 각기 다른 접근법을 가진 DIF 기법들은 약간씩 다른 결과를 낳는 경향이 있다. 따라서, 번역 장면에서 DIF기법을 응용하는 연구들은 가능하면, 서로 다른 접근법을 가진 적어도 두 개 이상의 DIF 기법을 상호보완적으로 사용하는 것이 바람직할 것이다.

마지막으로 검사 번역 장면에서 DIF 기법을 응용하는데 있어서, 몇 가지 한계점들이 존재한다는 점이다. 즉, 서로 다른 문화 혹은 언어집단에 속하는 개인들을 적절하게 대응시킬 수 있는 대응변수의 선택 문제이다. 전통적인 DIF 연구에서는 주로 검사의 총점을 대응변수로서 사용하여왔지만, 번역 DIF 연구에서는 이 총점이 적절치 않을 수가 있다는 점이다. 대체적으로 번역된 검사와 原 검사간에 동일한 신뢰도 수준을 유지하는 것이 어려우며, 검사를 번역하는 과정에서 서로 다른 난이도로 번역하는 체계적인 편파를 나타낼 수 있다는 점 때문이다. 따라서, 검사의 총점 대

신 기존의 DIF 연구에서 응용되어왔던 다변량 대응변수와 같은 새로운 대응변수의 모색이 필요할 것이다.

참고문헌

- 김신영. 2001. 《차별적 문항기능의 추출방법》 교육과학사.
- 염태호, 김정규. 1990. 《성격요인검사: 실시요강과 해석방법》 한국심리적성연구소.
- Cattell, R. B., and Cattell, H. E. 1995. "Personality structure and the new fifth edition of the 16 PF." *Educational and Psychological Measurement*, 55(6): 926-937.
- Church, A. T., and Lonner, W. T. 1998. "The cross-cultural perspective in the study of personality: Rationale and current research." *Journal of Cross-Cultural Psychology*, 29(1): 32-62.
- Dragow, F. 1984. "Scrutinizing psychological tests: Measurement equivalence and equivalent relations with external variables are the central issues." *Psychological Bulletin*, 95(1):, 134-145.
- Ellis, B. B., and Mead, A. D. 1998. "An application of the DFIT framework to assess the measurement equivalence of a Spanish translation of the 16PF questionnaire." Paper presented at the annual meeting of the International Congress of Applied Psychology, San Francisco, CA.
- Hambleton, R. K. 1994. "Guidelines for adapting educational and psychological tests: A Progress Report." *European Journal of Psychological Assessment*, 10(3): 229-244.
- Jöreskog, K., and Sörbom, D. 1993. *PRELIS 2: Users reference guide*. Chicago: Scientific software international.
- Reckase, M. D. 1979. "Unifactor latent trait models applied to multifactor tests: Results and implications." *Journal of Educational Statistics*, 4(3): 207-230.

- Russell, M. and Karol, D. 1994. *16PF Fifth Edition: Administrators manual*.
Champaign, IL: The Institute for Personality and Ability Testing, Inc.
- Shaughenssy, M. F., and Kang, M. H. 1998. "*Personality profile of giftedchildren: The 16PF Fifth Edition-A Comparative study of Korean and US Children.*" Unpublished manuscript.
- Shealy, R., and Stout, W. F. 1993. "A model-based standardization approach that separates true bias/DIF from group differences and detects test bias/DIF as well as Item bias/DIF." *Psychometrika*, 58: 159-194.
- Zhang, Z., and Yan, G. 2000. "*Differential Item Functioning in the 16PF Questionnaire: A cross-cultural comparison in items and traits.*" Unpublished paper.

