

전기자료(Biodata) 문항의 가중치 부여체계간의 타당도 연구: 분석집단크기에 따른 비교연구*

박 동 건[†] 전 인 식
고려대학교 심리학과

본 연구에서는 전기자료 문항의 경험적 척도화 및 요인 척도화 방법들간의 타당도 연구를 수행하였다. 보험 판매사원 4813명을 대상으로 154문항의 전기문항을 실시하였다. 분석 대상 인원 에 따라 대(4118명), 중(835명), 소(393명)로 나누어 각 집단에 경험적 척도화 방법(등가중, 회귀가중, WAB, Stokes, 상관계수 방법)과 요인 척도화 방법(등가중, 회귀가중 방법)을 실시하여 타당도 계수의 변화 정도를 알아보았다. 연구 결과 집단의 크기가 커질수록 교차타당화에서 타당도가 안정적으로 나타났다. 대집단에서 경험적 척도화 방법의 등가중, WAB, Stokes, 상관계수 방법 과 요인 척도화의 등가중 방법은 교차타당화시에 타당도가 안정적인 것으로 나타났다. 그에 비해 회귀가중 방법은 경험적 척도화와 요인 척도화 모두에서 교차타당화시에 타당도의 감소가 유의미하게 나타났다. 그러나 회귀가중 방법은 경험적 척도화의 개발 및 교차표본에서 다른 방법들보다 높은 타당도를 보였고, 세 집단의 개발표본에서 얻은 타당도 계수들간의 차이가 나타나지 않았다. 또한 회귀가중 방법이 다른 가중치 부여 체계에 비해 가장 손쉽게 이용할 수 있는 방법이기 때문에, 향후 연구시 회귀가중 방법의 이용을 고려해야 할 것이다. 끝으로 본 연구의 결과가 폭넓게 수용되기 위해 후속연구의 필요성과 본 논문의 제한점에 대해 논의하였다.

주요어 : 전기자료, 타당도연구, 가중치 부여체계, 교차타당화

전기자료 문항은 개인들에게 최근에 일어난 일들 중 문항과 관련하여 전형적인 행동이나 경험을 회상하거나 보고하도록 하는 것이다. 전기자료의 기본적인 가정은 개인의 과거 경험과 행동을 통해 미래의 행동을 예측할 수 있다는 것이다. 전기자료는 과거의 행동이 미래의 행동을 모두 예측한다는 것을 의미하지 않고, 이전 경험에 대한 정보가 미래의 행동을 예견하는데 다른 것보다 예측할 가능성이 더 높다는 것을 뜻한다(Mumford

* 본 논문은 2001년 고려대학교 특별연구비에 의해서 연구되었음.

[†] 고려대학교 심리학과, 02-3290-2066, sykhpark@kebi.com

& Owens, 1987).

전기자료는 일관성 있게 다양한 유형의 준거를 예측하고, 다양한 직무에서의 성과를 예측하는데 사용되고 있다(Hunter & Hunter, 1984).

전기 자료를 선발절차의 일환으로 많이 사용하고 있는 산업중의 하나가 보험 업계이다. 보험 업계에서는 1930년대 초반부터 생활설계사의 선발시에 전기자료를 사용하여 왔다. 우리나라에서는 생활설계사라는 직업이 주로 연고를 중심으로 판매가 이루어지고 있기 때문에 사회적으로 인정받지 못하는 직업으로 인식되고 있다. 또한, 영업적 역량을 갖춘 일부 설계사의 경우는 일반인의 상상이상의 연봉을 받고 있지만 대부분의 설계사들은 연고가 떨어지면 성과가 급격히 떨어져 기대했던 것만큼의 보상을 받지 못한다. 상품의 속성이 무형의 것이므로 고객을 이해시키기 어렵기 때문에, 주변 사람들 이외의 고객을 개척하는 것이 용이하지 않다(Mcmanus & Kelly, 1999). 그러한 생활 설계사의 특징 때문에 약 50%에서 80%정도가 거절에 대한 두려움과 낮은 성과로 인해서 1년 이내에 스스로 직업을 포기한다. 생활설계사의 보수는 일반적으로 상품 판매에 따라 결정되지만, 초기 교육에 있어서는 대부분의 회사가 금전적 지원을 하고 있어 신입 생활설계사가 업무를 수행하지 못하고 회사를 떠날 경우 회사의 비용 손실은 막대하다. 따라서 회사는 직무에서 성공할 가능성이 높은 지원자를 찾아내기 위해 타당한 선발 도구를 사용하는 것이 매우 중요하다.

전기자료의 타당도 연구

전기자료 문항들은 대부분 이질적이기 때문에 각각의 문항들은 낮은 문항간 상관을 보인다(Kavanagh & York, 1972; Owens, 1976). 따라서 내적 일관성 신뢰도는 대체적으로 낮은 편이다. 반면, 많은 연구에서 검사-재검사 신뢰도는 .60에서 .90로 매우 높은 편이다(McManus & Mitchell, 1987).

신뢰도에 대한 이러한 일관적인 결과와는 달리 타당도에 대한 연구들은 다양한 결과를 내놓고 있다.

McGrath(1960)는 100명의 신용 우수자와 100명의 신용 불량자를 대상으로 WAB 방법과 등가중 방법을 적용

했는데, 두 방법 사이에 타당도 차이는 없었으며, 교차타당화에서는 WAB 방법이 약간 더 높았다(WAB: $r=0.58$, 등가중: $r=.54$).

이와 비슷하게, Mitchell과 Klimoski(1982)의 부동산 영업 사원 689명을 대상으로 한 WAB 방법과 요인 척도화 방법의 타당도를 비교하는 연구에서, 개발표본에서 얻은 타당도는 WAB 방법이 더 높았으나, 교차타당화시에 타당도가 더 크게 감소되었다. 연구자들은 WAB 방법이 감소 폭은 컸으나 교차 타당도 역시 더 높기 때문에 연구 목적이 최대의 예측력을 갖는 것이라면 경험적 척도화 방법을 쓰는 것을 권장하였다(WAB - 개발: .59 교차: .46, 요인 척도화 - 개발: .36 교차: .36).

그러나, Lecznar와 Dailey(1950), 그리고 Malone(1978)은 WAB 방법이 처음에는 더 높은 타당도를 산출하지만, 교차타당화에서는 상관이나 회귀 방법이 더 높은 타당도를 보이며, 타당도의 감소도 더 적다고 주장했다. 또한 Malloy(1955)는 275명을 대상으로 수행한 이직을 준거로 한 연구에서 WAB 방법과 회귀가중 방법을 비교했는데, 교차타당화에서 WAB 방법의 타당도의 감소가 더 크게 일어났다는 결과를 얻은바 있다.

Fuentes, Sawyer와 Greener(1989)는 656명의 항공기 조종사에 대한 연구에서 경험적 척도화, 요인 척도화, 개념적 척도화의 타당도를 비교했는데, 교차타당도는 경험적 척도화($r=0.31$)가 다른 두 방법(요인 척도화: $r=.27$, 개념적 척도화: $r=.16$)보다 높았으며, 교차타당화에서 유의미한 타당도의 감소를 보인 것은 요인 척도화 뿐이었다.

Dalessilo와 Crosby(1991)는 9,079명의 보험판매사원을 대상으로 WAB 방법과 요인 척도화 방법, 그리고 WAB 방법으로 가중치가 부여된 상태에서 요인분석을 한 새로운 혼합된 방법을 비교했다. WAB 방법에서 .16, 요인 척도화에서 .15 그리고 혼합된 방법에서는 .18의 교차타당도를 얻었다. 타당도의 감소는 WAB 방법과 혼합된 방법에서 유의미했으며, 요인 척도화 방법에서는 유의미한 감소가 나타나지 않았다.

국내 연구 중에 강혜련(1996)은 5922명의 보험 판매사원을 대상으로 WAB 방법을 수행하였는데, 타당도 계수는 개발표본에서 .30, 교차표본에서 .23로 유의미한 감소를 보였다.

한편, 이종구, 윤창영, 임대열(1998) 등은 최근 연구에서, 식품매장판매사원 300명을 대상으로 경험적 척도화 방법(WAB, 회귀가중, 등가중)과 요인 척도화 방법(회귀가중, 등가중)의 타당도 비교를 수행하였다. 연구결과 경험적 척도화 방법이 개발표본에서 더 높은 타당도를 보였고, 타당도의 감소는 요인 척도화 방법에서 더 적게 나타났다. 경험적 척도화 방법은 개발표본에서 WAB 방법($r=.36$)보다 회귀가중($r=.53$)과 등가중($r=.39$)방법에서 더 높은 타당도를 산출하였다. 교차타당화에서는 WAB 방법($r=.07$)이 급격한 타당도 감소를 보인 반면, 회귀가중($r=.27$)과 등가중($r=.27$)은 더 안정적이었으며 두 방법 간에는 타당도 차이가 나타나지 않았다.

기존 타당도에 관한 연구들은 각기 다른 크기의 표본에 대해 실시된 것이기 때문에 어떤 방법이 더 효과적이라고 평가하기는 힘들다.

따라서, 본 연구에서는 그 동안 이루어졌던 전기 자료의 타당도 연구가 개발표본의 크기가 달랐기 때문에 각기 불일치한 결과를 나타내었다고 전제하고, 집단의 크기에 따른 경험적 척도화 방법과 요인 척도화 방법의 타당도의 변화를 알아보고자 한다.

전기 자료 척도화의 방법

과거의 행동이나 경험을 회상하는 문항에 대한 응답은 개인의 전형적인 행동과 경험을 설명해 줄 수 있는 유용한 자료가 된다. 다양한 과거의 행동이나 경험이 미래의 행동을 예측하도록 하기 위해서는 문항들을 의미 있는 유형으로 결합시켜야 한다(Mumford & Owens, 1978).

수행을 예측하는데 유용하다고 증명된 전기 자료의 척도화 방법은 여러 가지가 있다. 그 대표적인 방법으로는 경험적 척도화(empirical scaling), 개념적 척도화(rational scaling 또는 construct-oriented scaling), 요인 척도화(factorial scaling), 하위 집단화(subgrouping)의 4가지가 있다. 문항 개발 방법에 따라 척도화와 타당도 산출 방법이 달라질 수 있으므로, 문항 개발과 척도화, 그리고 타당도 과정은 밀접한 관련을 가지고 있는 경우가 많다.

본 연구에서 사용된 경험적 척도화 방법과 요인 척도화 방법에 관하여 간단히 설명하면 다음과 같다.

경험적 척도화(경험적 점수화, empirical keying) : 경험적 척도화는 전통적으로 가장 보편적으로 사용되어 왔으며, 직무 수행이나 이직 등 다양한 준거의 예측시에 일관성 있게 높은 타당도를 보여온 방법이다. 이 방법은 우수 집단과 부진 집단을 변별해 줄 수 있는 전기자료 문항들을 선별하고 가중치를 주는 척도화 절차이다. 그러나 이 방법은 구성 개념들에 대한 설명이 부족하고, 폭넓은 이론의 뒷받침 없이 수행의 준거를 예측한다는 제한점이 있다(Mumford & Stockes, 1992). 또한 특정 표본에서 특정 준거에 대한 예측력을 최대한 확보하는 절차이기 때문에, 표본이 달라졌을 때 타당도가 일반화되기 힘들고, 준거를 어떻게 정의하느냐에 따라 사용이 제한될 수 있다는 단점을 가지고 있다(Mumford & Owens, 1987, Thayer, 1977)

요인 척도화(Factorial Scaling) : 이 척도화 방법은 예측력은 있지만 결과에 대한 과학적인 해석이 불가능한 경험적 척도화의 단점을 보완하기 위해 개발된 방법이다. 이 방법은 요인분석을 통해 전기자료 문항들에 잠재해 있는 의미 있는 차원들을 추출해 척도를 구성한다. 요인 분석을 전기 자료에 응용한 것은 많은 양의 자료를 줄이고자하는데 그 목적이 있다. 다른 성격 질문지들처럼 전기 자료도 많은 문항들을 포함하기 때문에 요인분석을 통해 측정된 차원들을 더 잘 이해 할 수 있다. 그러나, 전기자료 문항들은 독립적인 사건에 대해 묻는 경우가 많으므로 서로 이질적이어서 공통된 요인을 추출하기 힘든 경우가 많으며(실제 추출된 요인들도 전체 변량의 일부만을 설명한다), 추출된 요인들도 연구자의 주관에 따라 여러 가지로 해석될 수 있다는 문제점을 가지고 있다(박영석, 1995; 이종구 등, 1998). 그러나 전기 자료의 중요 차원과 수행과의 상관을 이해함으로써 특정 차원을 대표하는 문항을 추출해 내는 것이 가능하다면 좀더 효과적인 방법이 될 것이다.

문항 분석 및 가중치 부여

가중치는 일반적으로 준거 수행 수준과 응답된 문항을 비교해서 산출되며, 문항과 준거 수행 수준 사이의 관계 정도에 따라 가중치를 부여하는 방법은 다양하다.

본 연구에서는 경험적 척도화 방법과 요인 척도화 방법이 시도되었다. 경험적 척도화 방법에는 문항수준의 분석과 선택지 수준의 분석이 있다. 그 중 등가중과 회귀가중 방법은 문항 수준의 분석이고 WAB, Stokes, 상관계수 방법은 문항 선택지 수준의 분석방법이다. 각각에 대해 살펴보면 다음과 같다.

등가중(Unit weighting; equal weighting) 방법 : 등가중 방법은 모든 문항에 동일한 가중치를 부여하거나 아예 가중치를 부여하지 않는 방법으로, 원점수를 그대로 사용하는 방법이다. 이 방법은 검사문항과 준거와의 상관관계를 통해 .05 수준에서 유의미한 관련성이 있는 문항만을 선별하여 분석에 사용한다.

본 연구에서 등가중 방법은 준거인 급여평균과 문항수준에서 상관계수를 산출하여, .05 수준에서 유의미한 문항들만을 분석에 사용했다. 대집단의 경우는 총 154문항 중 78문항이 유의미한 상관을 보였으며, 중집단의 경우는 38문항이 유의미한 상관을 보였다. 그리고 소집단의 경우는 12문항이 준거와 유의미한 상관을 보였다.

회귀가중(regression weighting) 방법 : 이 방법들은 전기자료 문항과 준거가 연속적이고 서열화 될 수 있는 수량 척도 일 때, 문항과 준거 사이의 상관이 선형에 가까운 형태를 보일 때 적합한 방법이다(Hogan, 1994). 회귀가중 방법은 통계적으로 준거와 유의미한 관계를 갖는 문항들만을 다중회귀와 같은 방법을 사용하여 준거에 대한 가중치를 부여한다.(Gandy, Outer-bridge, Sharf, & Dye, 1989; Lecznar & Dailey, 1950; & Malone, 1978).

때로 제한된 수의 표본을 사용하기 때문에 표본과의 비율을 고려해 분석에 사용될 문항의 수를 줄여야 할 경우가 있다. 이 경우에는 보통 준거와의 상관이 .05 수준에서 유의미한 문항들만을 보유한다(Hogan, 1994).

본 연구에서 회귀가중 방법은 준거인 급여평균과 문항의 상관계수를 산출하여, .05 수준에서 상관이 통계적으로 유의미한 문항들을 선별한 후 단계적(stepwise) 회귀분석을 실시하여, 각 문항에 회귀계수(β)를 가중치로 부여하였다. 그 다음 원점수와 회귀계수를 곱해 얻은 가중치의 점수로 준거에 대한 예측 타당도 계수를 산출하

였다. 이 방법의 적용 결과 예측력을 갖는 문항은 대집단의 경우 21문항, 중집단의 경우는 8문항이었다. 그리고 소집단의 경우는 8문항이 예측력을 보였다.

WAB(Weighted Application Blank) 방법 : 전통적인 WAB 방법에서는 집단을 상위 27%와 하위 27%로 나누어 두 집단의 반응 빈도를 백분율(%)로 표시한 후, 두 집단의 반응 빈도의 차이를 계산하여 'Strong(1926) 가중치 표'(Strong's Tables of Net Weights for Differences in Percents)를 통해 순가중치(net weight)를 구한다(England, 1971; Stead & Shartle, 1940, Guion, 1965). Strong(1926)의 순가중치 표를 이용할 때는 빈도 차이의 절대값에 해당하는 값을 표에서 찾아 상, 하위 집단간 반응 빈도의 차이가 없으면 0의 가중치를 부여하고, 상위 집단에서 더 큰 빈도를 보인 선택지에는 양(+)의 값을, 하위 집단에서 더 큰 빈도를 보인 선택지에는 음(-)의 가중치를 부여한다.

이와 같은 방식으로 전체 문항 선택지 별로 가중치를 부여한 WAB 점수를 가지고 준거와 상관을 구해 .05 수준에서 유의미한 문항만으로 WAB 점수 총합계를 구해 준거와의 상관계수를 산출하여 예측 타당도 계수를 산출한다.

WAB 방법으로 모든 문항에 가중치를 매긴 결과 대집단에서는 123문항, 중집단에서는 72문항, 그리고 소집단에서는 28문항이 준거와 유의미한 상관을 보였다.

Stokes 방법 : 이 방법은 WAB 방법과 같이 선택지 별로 가중치가 부여되는 방법이다. 문항의 선택지들과 준거와의 상관계수를 구해 유의미한 상관계수는 방향성에 따라 1 또는 -1의 가중치를 부여하고 유의미한 상관을 보이지 않는 선택지에는 0의 가중치를 부여한다. 그 이후의 절차는 WAB 방법과 동일하게 문항 전체 선택지 별로 가중치가 부여된다. 부여된 가중치문항을 가지고 준거와 .05 수준에서 유의미한 문항만을 가지고 준거와 유의미한 문항의 총합을 통해 타당도를 산출하는 방법이다(Gandy & Dye, 1994).

Stokes 방법으로 문항들에 가중치를 부여한 결과 대집단에서는 100문항, 중집단에서는 52문항, 그리고 소집단에서는 35문항이 준거와 유의미한 상관을 보였다.

상관계수 방법 : WAB 방법이나 Stokes 방법과 마찬가지로 선택지 별로 가중치를 준다. WAB 방법에서 Strong(1926)표를 이용하는 이유는 과거에는 선택지와 준거와의 phi상관계수를 산출하는 것이 어렵기 때문이었다. 그러나 작금에는 상관계수를 구하는 것이 용이하기 때문에 WAB의 원래의 취지를 따르는 것이 낫다고 생각되었다. 또한 준거집단을 상위 27%와 하위 27%로 분석하게 되면 정보의 손실을 가져오기 때문에 상관계수를 구하여 가중치를 주는 것이 바람직하다고 필자는 판단하였다. 그 다음 절차는 WAB과 동일하다. 이와 같은 방법을 적용하여 문항들에 가중치를 부여한 결과 대집단에서는 87문항, 중집단에서도 87문항, 그리고 소집단에서는 이 보다 적은 59문항이 준거와 유의미한 상관을 보였다.

교차 타당화

경험적 척도화와 요인 척도화 모두 특정 표본에 대해 예측력을 최대화하기 때문에 표집오차에 의해 타당도가 과대추정 될 우려가 있다. 따라서 독립적인 표본에 대한 교차타당화 절차가 필수적으로 요구된다. 하나의 표본만을 사용할 수 있다면 애초에 그 하나의 표본을 개발표본과 교차 타당화를 위해 따로 떼어놓은 표본(hold-out sample)으로 무선적으로 나누어서 연구를 진행해야 한다. 여기에 대해서는 여러 가지 주장이 있지만, England(1961)는 교차타당화 표본보다 개발표본이 1/3 정도 더 많은 수로 구성되어야 한다고 한다. 개발된 점수를 교차타당화 표본에 적용시켜 봄으로서, 타당도의 감소를 파악해서 실제 예측력에 대한 더 나은 추정치를 얻을 수 있을 것이다(Hogan, 1994).

방 법

검사도구 : 검사 문항은 London House(1980)에서 제작한 *Experience and Background Inventory*, Glennon과 Owens(1966)가 개발한 *A Catalog of History Items*, Owens' *Biographical Questionnaire*(BQ;Owens & Schoenfeldt, 1979), 그리고 LIMRA(1937)의 *Career Profile* 등의 문항을 참조해 총 154 문항으

로 검사를 구성했다. 154문항은 학교생활의 적응, 학업 성취, 지적/문화적 추구, 내외향성, 리더십, 상담자 역할, 경력개발 관심, 종교활동, 스포츠활동, 독립성, 부모의 사회경제적 지위, 성취동기, 설계사 업무의욕, 부모의 지지/관심, 과학과목에 대한 흥미, 기술활용 및 적용, 자신감, 부모의 통제, 설득력, 사교성, 본인의 경제적 지위, 직업 선택시 주위의 지지, 주변에 설계사 존재 등으로 구성되어 있다.

표 1. 소집단 준거의 평균 및 최대, 최소값

급 여(단위:원)	평 균	최소값	최대값	표준편차
개발표본	947293	2900	5639260	847661
교차타당화표본	1298611	5500	8485666	1166271

표 2. 중집단 준거의 평균 및 최대, 최소값

급 여(단위:원)	평 균	최소값	최대값	표준편차
개발표본	1028532	4381	5518218	899088
교차타당화표본	1058222	5500	5828746	963728

표 3. 대집단 준거의 평균 및 최대, 최소값

급 여(단위:원)	평 균	최소값	최대값	표준편차
개발표본	109024'	361	1101675	983675
교차타당화표본	105119'	350	563926	901165

자료수집 : 국내 B 생명보험회사 보험 판매사원(생활설계사) 4813명을 대상으로 1999년 2월부터 2000년 7월까지 3회에 걸쳐 설문을 실시하였으며, 전체 집단에서 대 집단의 개발표본은 2,881명(70%) 교차표본은 1,237명(30%)으로 구성하여 총 4,118명, 중집단의 개발표본은 592명(70.9%) 교차표본은 243명(29.1%)으로 총 835명, 소 집단의 개발표본은 282명(71.8%) 교차표본은 111명(28.2%)으로 총 393명으로 무작위로 나누어 분석을 실시 하였다. 분석에 포함된 보험 판매사원의 재직기간은 6개월 이상으로 급여가 안정기에 들었다고 판단되는 사람들 이었다.

표 4. 가중치 방법과 집단의 크기에 따른 타당도 변화

가중치 부여방법	개발표본		교차표본		개발:교차 Z값
	타당도	N	타당도	N	
소집단					
등가중	.517**	208	.251**	90	2.46*
회귀가중	.552**	213	.228**	91	3.02**
WAB	.625**	196	.280*	83	3.35**
Strokes	.658**	195	.219*	84	4.29**
상관계수	.731**	183	.309**	72	4.29**
요인(등가중)	.238**	273	.147	109	.86
요인분석(회귀)	.238**	273	.147	109	.86
중집단					
등가중	.432**	401	.206**	159	2.67**
회귀가중	.463**	423	.274**	180	2.45*
WAB	.573**	353	.277**	140	3.70**
Strokes	.562**	385	.238**	153	4.03**
상관계수	.608**	357	.226**	141	4.78**
요인분석(등가중)	.182**	490	.182**	203	0
요인분석(회귀)	.264**	490	.180*	203	1.06
대집단					
등가중	.307**	1760	.248**	767	1.38
회귀가중	.448**	2037	.369**	891	2.41*
WAB	.400**	1664	.348**	745	1.34
Strokes	.357**	1600	.284**	750	1.76
상관계수	.405**	1653	.346**	759	1.59
요인분석(등가중)	.213**	2443	.145**	1050	1.95
요인분석(회귀)	.267**	2454	.184**	1054	2.29*

준거 : 본 연구에서 쓰인 준거는 급여평균으로 하였다. 생활설계사의 급여에는 단지 새로운 보험판매만 포함하고 있는 것이 아니라 기존 계약이 파기되지 않아야만 받을 수 있는 유지 수당, 그리고 새로운 생활설계사를 모집함으로써 얻을 수 있는 모집 수당이 포함되어 있다. 즉 우수한 생활설계사란 보험 판매, 기존 계약 유지, 증원을 모두 잘해야 하는 것이다. 그러므로 본 연구에서는 검사에 참여한 생활설계사들 각자의 재직기간동안(1998년 10월부터 2001년 1월까지) 받은 급여의 평균을 준거로 삼았다.

분석 및 결과

본 연구에서는 표본의 크기에 따른 가중치 부여체계 간 타당도 비교를 위해 먼저 크게 대, 중, 소 세 개의 집단으로 구분하였다. 각 집단마다 각각 7가지 가중치 방법을 적용하여 전기 자료에 가중치를 부여하고, 집단 간 타당도와 각 가중치 방법별 타당도를 비교하였다.

각 집단은 7:3의 비율로 개발표본과 교차표본으로 나누었으며 준거와의 상관을 통해, 타당도 계수를 산출하였다. 개발과 교차표본에서 산출한 타당도 계수들의 차이 검증을 위해서 각 상관 계수들을 Fisher's Z로 변환하여 유의미성을 보았다.

각 가중치 방법과 집단의 크기에 따른 타당도의 변화는 표 4와 같다. 각각의 가중치 부여 방법에 대한 자세한 논의는 다음에 제시하였다.

등가중 방법

그림 1에서 보는 바와 같이 등가중 방법의 경우 대 집단 개발표본의 타당도(.307)와 교차타당화 집단의 타당도(.248)는 유의미한 차이를 보이지 않았다. 하지만 중 집단과 소집단에 있어서는 모두 유의미한 차이를 보였다. 구체적으로 중집단에서 개발표본의 타당도 계수는 .432였고 교차타당화 계수는 .206으로 유의미한 차이 ($Z=2.67, p<.01$)를 보였으며 소집단의 경우에도 개발표본(.517)과 교차표본(.251)에서 유의미한 차이($Z=2.46, p<.05$)를 보였다.

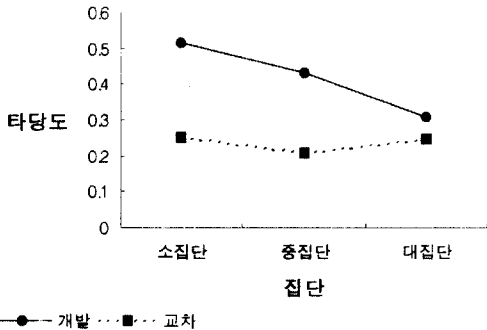


그림 1. 등가중

회귀가중 방법

회귀가중 방법이 집단 크기에 따라 어떻게 변화하는지는 그림 2에 나와 있다. 회귀가중 방법의 경우에는 집단 크기에 상관없이 모두 유의미한 차가 있었다.

대집단의 경우 개발표본의 타당도(.448)와 교차표본의 타당도(.369)는 통계적으로 유의미하였다($Z=2.41, p<.05$). 중집단(개발:.463, 교차:.274)과 소집단(개발:.552, 교차:.228)에서도 모두 유의미한 차이를 보였다(중집단: $Z=2.45, p<.05$, 소집단: $Z=3.02, p<.01$).

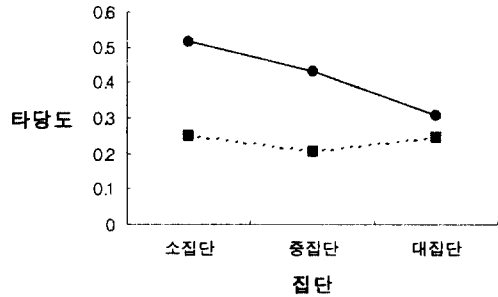


그림 2. 회귀가중

WAB 방법

그림 3에서 개발표본과 교차타당화 표본을 대상으로 하여 집단 크기별로 타당도 분석을 한 결과 WAB은 다른 방법과 비교하여 개발표본에서 일관되게 높은 타당도 계수(대집단:.400, 중집단:.573, 소집단:.625)를 산출하였다. 교차타당화에 있어서도 다른 방법들에 비해 비교적 높은 타당도 계수가 나왔다(대집단:.348, 중집단:.277, 소집단:.280). 대집단에서 개발과 교차표본에서의 타당도 감소는 유의미하지 않았지만 중집단과 소집단에서는 유의미한 차가 나타났다(중집단: $Z=3.70, p<.01$, 소집단: $Z=3.35, p<.01$).

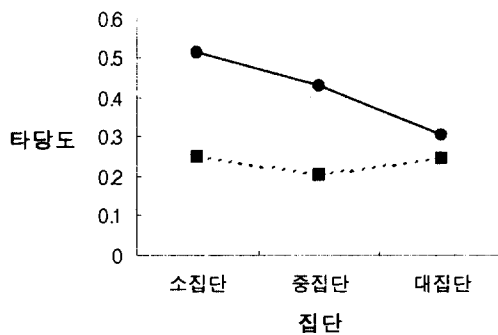


그림 3. WAB

Stokes 방법

본 연구에서 Stokes 방법은 그림 4에서 보는 바와 같이 개발표본에서 높은 타당도 계수(대집단: .357, 중집단: .562, 소집단: .658)를 보이지만 교차타당도에서는 다른 방법에 비해 상대적으로 낮은 타당도 계수(대집단: .284, 중집단: .238, 소집단: .219)를 보인다. 대집단을 제외하고 중, 소집단에서 개발표본과 교차타당화 표본의 두 상관 계수의 차이가 유의미하게 나타났다(중집단: 4.03, $p < .01$, 소집단: 4.29, $p < .01$).

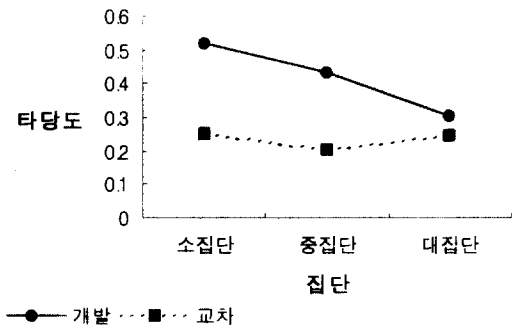


그림 4. Stokes

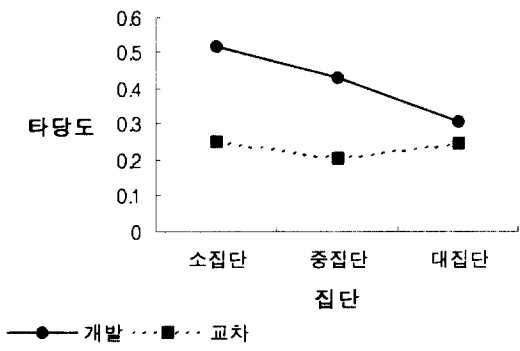


그림 5. 상관계수

상관계수 방법

그림 5에 집단크기에 따라 상관 계수 방법으로 구한

타당도 계수를 제시하였다. 대집단에서의 타당도 계수 (개발: .405, 교차: .346)의 감소는 유의미하지 않았지만 중 집단과 소집단에서의 상관계수(중집단: 개발: .608, 교차: .226, 소집단: 개발: .731, 교차: .309)는 다른 방법과 비교해 보았을 때 개발표본과 교차표본의 타당도 계수의 차가 상대적으로 다른 방법들보다 유의미하였다(중집단: $Z=4.78, p < .01$, 소집단: $Z=4.29, p < .01$).

요인 척도화

그림 6과 7은 요인 척도화 등가중 방법과 회귀 가중 방법을 사용하였을 때 나온 결과이다. 이 방법들은 다른 가중치 방법들에 비해 개발표본과 교차타당화 표본에서 모두 낮은 타당도 계수가 산출되었다. 하지만 소집단, 중집단, 대집단으로 하여 집단 크기별로 타당도 분석을

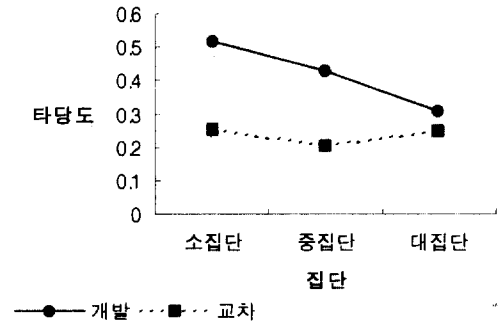


그림 6. 요인분석(등가중)

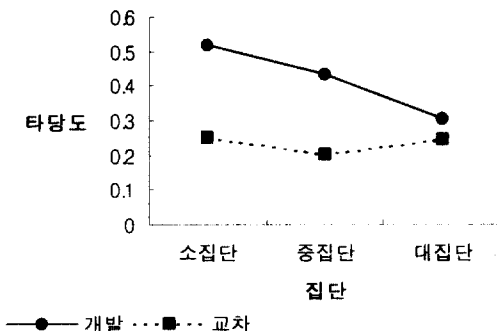


그림 7. 요인분석(회귀)

한 결과 WAB은 다른 방법과 비교하여 개발표본에서 일관되게 높은 타당도 계수(대집단:400, 중집단:573, 소집단:625)를 산출하였다. 교차타당화에 있어서도 다른 방법들에 비해 비교적 높은 타당도 계수가 나왔다(대집단:348, 중집단:277, 소집단:280). 대집단에서 개발과 교차표본에서의 타당도 감소는 유의미하지 않았지만 중집단과 소집단에서는 유의미한 차가 나타났다(중집단: $Z=3.70, p<.01$. 소집단: $Z=3.35, p<.01$).

대집단, 중집단, 소집단 각각에서 7가지 가중치 부여 방법의 차이는 그림 8, 9, 10을 보면 쉽게 이해 할 수 있다.

그림 8의 대집단의 경우 개발표본에서 회귀>상관계수>WAB>Stokes>등가중>요인분석(회귀)>요인분석(등가중) 순으로 높은 타당도 계수가 산출되었고 교차타당화에 있어서는 회귀>WAB>상관계수>Stokes>등가중>요인분석(회귀)>요인분석(등가중)의 순으로 높은 타당도를 보인다. 어느 방법에서나 대집단의 경우는 중집단이나 소집단보다 개발표본과 교차타당화 표본의 타당도 계수 차에 있어서 안정성을 보여주고 있고 교차타당화 표본에서의 타당도 계수는 중집단이나 소집단에 비해 높은 경향성을 보여주고 있다.

중집단과 소집단의 경우에 개발표본에서 가장 높은 타당도를 보인 것은 상관계수 방식이고 요인 척도화를 제외하고 가장 낮은 타당도를 보인 것은 등가중방법이다. 또한 등가중, 회귀, WAB, Stokes, 상관 계수 방법에 있어서 중집단과 소집단의 경우는 모두 개발표본과 교

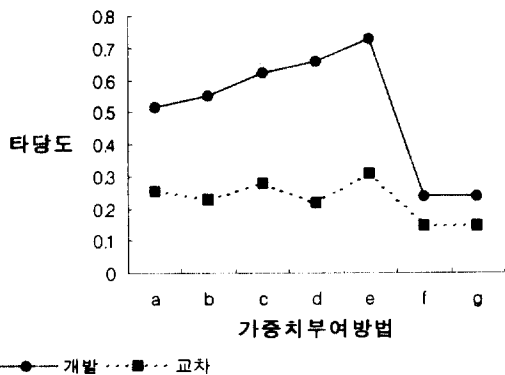


그림 8. 대집단

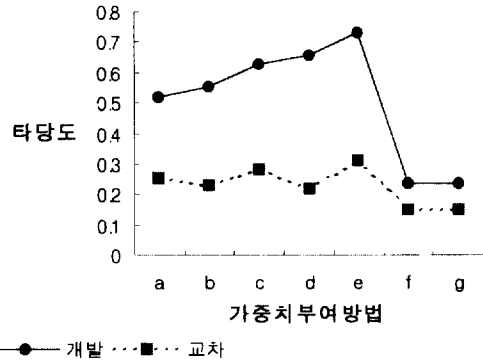


그림 9. 중집단

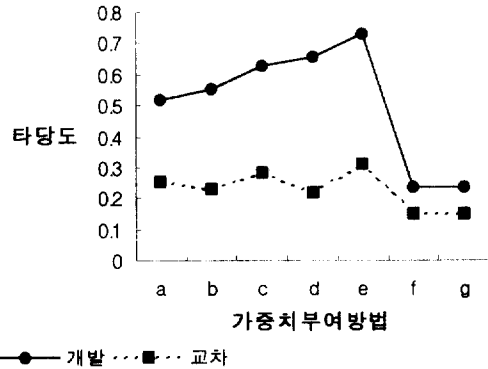


그림 10. 소집단

차표본에 유의미한 차이가 있었다.

대집단과 중집단, 대집단과 소집단간의 상관계수의 차이검증은 표 5에 제시되었다. 개발표본에서는 경험적 척도화의 회귀가중 방법, 요인 척도화의 방법들에서 타당도 계수의 차이가 나타나지 않았고, 경험적 척도화의 다른 방법들에서는 집단간의 타당도 계수의 차이가 유의미하게 나타났다. 반면 교차표본에서는 모든 가중치 방법들간의 타당도 계수의 차이가 통계적으로 유의하게 나타나지 않았다.

<그림 8, 9, 10 공통> a. 등가중 b. 회귀가중 c. WAB d. Stokes e. 상관계수 f. 요인척도화(등가중) g. 요인척도화(회귀가중)

표 5. 집단간 타당도 계수의 상관 차 검증(Z)

	대vs중 (개발)	대vs중 (교차)	대vs소 (개발)	대vs소 (교차)
등가중	-2.61**	.53	-3.46**	0
회귀가중	-.35	1.29	-1.83	1.38
WAB	-3.93**	.89	-4.06**	.65
Stokes	-4.6**	.11	-5.52**	.27
상관계수	-4.76**	1.42	-5.95**	.31
요인 (등가중)	.73	-.47	-.42	0
요인 (회귀)	0	.07	.41	.4

논 의

본 연구는 전기자료 문항의 경험적 척도화 및 요인 척도화 방법들간의 안정성을 비교하고, 전기자료의 여러 가지 가중치 부여체계간의 타당도를 비교하기 위해 수행되었다.

본 연구 결과에 의하면, 첫째, 상대적으로 표본수가 적은 중집단, 소집단에서는 경험적 척도화의 방법들이 교차타당도에서 유의미한 타당도 감소가 나타났지만, 표본수가 많은 대집단일 경우에는 경험적 척도화 방법에서 회귀가중 방법만이 유의미한 타당도의 감소를 보였을 뿐 다른 방법들은 교차타당도에서도 타당도가 안정적인 것으로 나타났다. 이와 같은 결과는 많은 선행연구에서 나타난 타당도의 차이가 연구에 사용된 표본의 크기에 인한 것일 수도 있음을 시사해준다.

둘째, 대집단의 개발표본에서 경험적 척도화 방법들간의 타당도는 문항 수준에서 가중치를 부여하는 방법과 선택지 수준에서 가중치를 부여하는 방법들간의 차이는 없는 것으로 나타났다. 중집단, 소집단인 경우에는 선택지 수준에서 가중치를 부여하는 방법들이 상대적으로 타당도가 높은 것으로 나타났으나, 교차타당도가 더 크게 감소되었다. 표본수가 적은 경우에는 개발표본에서 문항의 선택지 별로 가중치가 부여되는 기법(WAB, Stocks, 상관계수방법)에서 나온 타당도 계수를 현장에

적용하는데는 상당한 주의가 요구된다.

셋째, 대집단에서 경험적 척도화 방법 중 회귀가중은 교차타당도의 감소가 유의미하게 일어났음에도 불구하고 개발표본, 교차표본에서 상대적으로 가장 높은 타당도를 확보할 수 있었고, 세 집단의 개발표본에서 얻은 타당도 계수들간의 차이가 나타나지 않았다. 비록 교차표본에서 경험적 척도화 방법들간에 유의미한 차이는 없었으나 가장 손쉽게 이용할 수 있는 방법이 회귀가중 방법이기 때문에 향후 회귀가중 방법의 이용에 주목할 필요가 있다.

넷째, 요인 척도화 방법이 경험적 척도화 방법에 비해서 대, 중, 소집단 모두에서 타당도 계수가 낮은 것으로 나타났다. 요인 척도화 방법은 전기자료 타당도 연구에서 예측력을 최대한 확보하는 동시에 심리학적인 이해를 도모할 수 있는 이상적인 측면이 있지만, 전기문항 자체의 속성상 다른 검시들과 달리 요인구조로 묶이기 힘들어 타당도의 크기가 작은 것이 큰 단점이라 할 수 있다.

다섯째, 요인 척도화의 회귀가중 방식에서 대집단에서는 교차타당도의 감소(개발 : $r=.268$ 교차 : $r=.184$)가 통계적으로 유의미했으나, 중, 소집단에서 교차타당도의 감소(중집단 - 개발 : $r=.264$ 교차 : $r=.180$, 소집단 - 개발 : $r=.238$ 교차 : $r=.147$)가 유의미하지 않았다. 상관계수의 감소 정도는 대, 중, 소집단에서 비슷하지만, 대집단에서만 유의미한 교차타당도의 감소가 나타나는 것은 통계 검증시 고려되는 표본의 크기 때문인 것으로 보인다.

지금까지 경험적 척도화 방법은 타당도의 일반화 가능성 측면에서 비판을 받아왔다, 즉 경험적 척도화 방법은 특정한 표본에서 특정 준거에 대해 개발되기 때문에 타당도가 상황의 변화에 따라 불안정하다는 것이다 (Hunter & Hunter, 1984). 그러나 본 연구 결과에서 볼 수 있듯이, 가중치 개발단계에서 충분히 큰 수의 표본을 확보할 수 있다면 이러한 문제점은 많이 극복될 수 있을 것이다. 또한 <표 4>에서도 볼 수 있듯이 소집단과 중집단의 경우 개발과 교차 표본에서 문항 선택지 수준의 가중치 방법은 문항 수준의 방법보다 유의미한 감소의 경향성을 보이고 있다. 그러므로 생활사 검사에 가

중치를 부여할 경우, 문항 수준의 가중치 방법을 사용할 것인지 혹은, 문항 선택지 수준의 가중치 방법을 사용할 것인지는 표본의 수에 맞추어 선택해야 할 것이다. 즉 소집단과 중집단의 경우 개발 표본에서 문항 선택지 수준 방법으로 가중치로 얻을 경우, 타당도를 과대 추정할 우려가 있다는 점을 고려해야 한다.

본 연구의 결과가 폭넓게 수용되려면 몇 가지 문제 점을 보완하는 체계적인 후속 연구가 필요하다.

첫째, 본 연구에 사용된 전체 154문항에서 준거와 유의미한 52문항을 선택하여 요인 척도화 방법을 실시하였기 때문에 요인의 수가 검사제작시 고려했던 23개 보다 줄어들었고, 각 요인별 문항의 수 또한 줄어들었다. 그로 인해 요인 척도화의 설명력이 상대적으로 적게 나타났을 가능성이 있기 때문에, 다음 연구에서는 이러한 문제점을 보완해야 할 것이다.

둘째, 본 연구에서는 '급여평균'을 준거로 사용하였다. 이 준거는 연구 목적으로 사용하기에는 문제가 없었으나, 전기자료를 회사에서 선발도구로 사용하기 위해서는 직무수행을 보다 정확하게 반영할 수 있는 준거를 실무에 종사하고 있는 전문가들과 연구자가 함께 개발해야 할 것이다.

셋째, 본 연구에서는 다소 많은 154문항을 사용하였고, 모든 문항에 정확하게 응답한 비율이 약 60%정도였다. 즉 대부분의 보험 판매 사원들인 주부들에게는 154 문항이 부담이 되어, 무성의하게 답변했을 가능성도 있다. 그 결과 각 집단의 가중치 방법에 따라 각기 다른 표본수가 나타나는데 이는 pairwise 방식을 사용하여 통계처리를 하였기 때문이다. listwise 방식을 사용할 경우 각 집단의 가중치 방법마다 동일한 표본수를 유지할 수는 있지만 각 피험자마다 서로 다른 문항들에 대해 무응답을 한 경우가 많기 때문에 분석할 가치가 있는 자료도 삭제되어 너무 많은 자료의 손실을 가져오게 된다. 이런 문제 때문에 연구자가 원자료를 관찰하여 일관되게 무성의한 자료를 제외한 후 통계처리를 하였다. 추후 연구에서는 이와 같은 문제들에 대처하기 위해 적절한 수의 문항으로 검사를 구성하고, 전기자료 문항 외에 무성의 척도를 포함시켜 무성의하게 응답된 사례를 분석에서 제외시켜야 할 것이다.

넷째, 본 연구는 여러 척도화 방법에 따른 타당도의 안정성 비교와 타당도 계수간의 비교에 주 관심이 있었기 때문에 개별문항의 해석 및 요인에 대한 해석을 고려하지 못했다.

참고문헌

- 강혜련 (1996). 경험적 척도화에 의한 전기자료(Biodata)의 타당화 연구. *한국심리학회지: 산업 및 조직*, 9(2), 1-16.
- 박영석 (1995). 이직자의 생활사. *한국심리학회지: 산업 및 조직*, 8(1), 75-102.
- 이종구, 윤창영, 임대열 (1998). 전기자료를 이용한 영업 사원의 수행예측: 척도화 및 가중치 부여 체계간의 비교 1998년 한국 산업 및 조직 심리학회 추계 학술 대회 발표 논문집. 한국 산업 및 조직 심리학회.
- Dalessio, A. T., & Crosby, M. M. (1991). *Validity or understanding: A comparison of three biodata scoring methods*. Paper presented at the 6th annual conference of the Society for Industrial and Organizational Psychology, St. Louis.
- England, G. W. (1971). *Development and use of weighted application blanks*(Bulletin No. 55). Minneapolis: University of Minnesota, Industrial Relations Center.
- Fuentes, R. R., Sawyer, J. E., & Greener, J. M. (1989). *Comparison of the predictive characteristics of three biodata scaling methods*. Paper Presented at the 97th annual convention of the American Psychological Association, New Orleans.
- Gandy, J. A., & Dye, D. A., & MacLane, C. N. (1994). *Federal Government Selection: the individual Achievement Record*. In G. S. Stokes, M. D. Murnford, & W. A. Owens(Eds.), *Biodata handbook*. Palo Alto: CPP Books.
- Gandy, J. A., Outerbridge, A. N., Sharf, J. C., & Dye, D. A. (1989). *Development and initial validation of the Individual Achievement Record(IAR)*. Washington, DC:

- U.S. Office of Personnel Management.
- Glennon, J. R., Albright, L. E., & Owens, W. A. (1966). *A Catalog of life history items*. Greensboro: Center for Creative Leadership.
- Hogan, J. B. (1994). Empirical keying of background data measures. In G. S. Stokes, M. D. Mumford, & W. A. Owens(Eds.), *Biodata handbook*. Palo Alto: CPP Books.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Kanvanagh, M. J., & York, D. R. (1972). Biographical correlates of middle managers' performance, *Personnel Psychology*, 25, 319-332.
- Lecznar, W.B., & Dailey, J. T. (1950). Keying biographical inventories in classification test batteries. *American Psychologist*, 5, 279.
- Malloy, J. (1955). The prediction of college achievement with the life experience inventory. *Educational and Psychological Measurement*, 15, 170-180.
- Malone, M. P. (1978). Predictive efficiency and discriminatory impact of verifiable biographical data as a function of data analysis procedures. (Doctoral dissertation, Illinois Institutes of Technology). *Dissertation Abstracts International*, 4520-B.
- McGrath, J. J. (1960). Improving Credit evaluation with a weighted application blank. *Journal of Applied Psychology*, 44, 325-328.
- McManus, M. A., & Mitchell, T. W. (1987). Test and retest reliability of the Career Profile. Hartford : Life Insurance Marketing and Research Association. Minneapolis : Personnel Decisions Research Institute.
- McManus, M. A., & Mary L. Kelly. (1999) Personality Measure and Biodata : Evidence regarding their incremental predictive value in the life insurance industry, *Personnel Psychology*, 44
- Mitchell, T. W., & Klimoski, R. J. (1982). Is it rational to be empirical? A test of methods for scoring biographical data. *Journal of Applied Psychology*, 67, 411-418.
- Mumford, M.D., & Owens, W. A. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement*, 11, 1-31.
- Mumford, M.D., & Stokes, G. S. (1992). Developmental determinants of individual action : Theory and practice in the application of background data measures. In M.D.Dunnette (Ed), *Handbook of industrial and organizational psychology* (2d ed., vol. 3, pp. 61-138). Palo Alto, CA: Consulting Psychologists Press.
- Owens, W. A. (1976). Background data. In M. Dunnette (Ed.), *The handbook of industrial and organizational psychology* (pp. 609-644). Chicago: Rand-McNally.
- Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology*, 35, 1-62.
- Schmidt, F. L., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Metaanalyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407-422.
- Siegel, L. A. (1956). A biographical inventory for students: Construction and Standardization of the instrument. *Journal of Applied Psychology*, 40, 5-10.
- Stead, N. H., & Shartle, C. L. (1940). *Occupational Counseling techniques*. New York: American Book.
- Strong, E. K. (1926). An interest test for personnel managers. *Journal of Personnel Research*, 5, 194-203.
- Thayer, P. W. (1977). Somethings old, somethings new. *Personnel Psychology*, 30, 513-524.

1차 원고 접수: 2001. 3. 11

최종 원고 접수: 2001. 5. 18

A Comparative Study of Validity of Biographical Data Weighting Methods by Sample Size

Dong-Gun Park In-Sik Chon
Korea University

The purpose of this research was to examine the validity of empirical scaling methods and factor scaling methods of biographical items . 154 biographical items were administered to 4813 insurance salespersons. The subjects were divided into three groups: large(4118 salespersons), middle(835), and small(393) group. For each group, empirical scaling methods (unit weighting, regression weighting, WAB, Stokes and correlation coefficient) and factor scaling methods (unit weighting, regression weighting) were employed to examine the variation of the validity coefficients. The results revealed that the larger the sample, the more stable the cross-validation. It was also found that in large groups, empirical scaling methods (unit weighting, WAB, Stokes, correlation coefficient) and factor scaling methods (unit weighting) showed stable validity in cross-validation. The regression weighting, however, showed meaningful decreases in validity when cross-validated for both the empirical scaling method and the factor scaling method. Nevertheless, the regression weighting was more valid than the other weighting methods for the development of empirical scaling and for the cross validation sample. Furthermore, there is no statistically significant variation with respect to the size of validity coefficient in the developing samples for the three groups. Because regression weighting is more easily available than any other weighting method, it should be considered to use the regression weighting method for future research. Lastly, the limitations of this study and future research directions were discussed.

keywords : *biographical data, validity study, weighting method, cross-validation*