

소표본에서 오는 전기자료 타당도의 불안정성 개선 방법 모색*

이 종 구†

대구대학교 심리학과

본 연구에서는 텔레마케터의 수행을 보다 잘 예측하기 위한 전기자료 문항의 구성과 함께 경험적 점수화 방법을 적용했을 때 타당도의 시간적 안정성을 검토하고 소표본에서 오는 타당도의 불안정의 개선 방법을 모색하고자 하였다. 이를 위해 성장기 경험과 입사전 최근 경험을 확인하기 위한 219문항으로 이루어진 전기자료 검사를 L사의 텔레마케터들을 대상으로 1(432명), 2(316명), 3(229명)차에 걸쳐 반복 실시하였다. 1차자료를 대상으로 219문항 중 88문항에 대해 WAB 가중방식에 의해 전기자료 검사 총점을 구하여 검사-재검사 신뢰도를 계산한 결과 .70으로 비교적 양호하게 나타났으나 교차타당화 자료는 물론 재검사 자료에 대한 타당도 감소가 현저하게 나타났다. 타당도의 안정성 확보를 위해 1, 2차 자료를 통합하여 가중치를 재계산한 결과, 교차타당화 자료와 재검사자료에 대한 타당도 감소가 둔화되는 것으로 나타났다.

주요어 : 전기자료, 타당도, 소표본, WAB 가중

* 이 논문은 2002년도 대구대학교 학술연구비 지원에 의한 논문임.

이 논문에 대해 꼼꼼한 수정과 유익한 조언을 해주신 익명의 세 분 심사위원께 진심으로 감사드립니다.

† 교신저자 : 이 종 구, gooya@daegu.ac.kr

본 연구는 다음의 두 가지 목적을 가지고 시도되었다. 첫번째 목적은 전기자료(biographical data: 일명 biodata)의 타당도를 높이기 위하여 수행과 관련된 문항 목록을 구성하는데 있으며, 둘째는 전기자료검사의 검사-재검사 신뢰도를 확인하고(시간적 안정성) 산업현장의 실시 여건상 비교적 소수의 표본 자료만을 이용할 수 밖에 없을 때 안정된 타당도를 확보하는 방안을 모색하는데 있다.

‘미래 행동의 가장 좋은 지표는 과거의 행동이나 경험’이라는 것이 전기자료를 이용한 수행예측에서의 가장 기본적인 전제이다(Owens, 1976; Wernimont & Campbell, 1968). 이 전제는 이론적으로 사람들이 시간에 걸쳐 대체로 일관된 행동을 보일 것이라고 가정하는 행동 일관성 원리(behavioral consistency principle)에 바탕을 두고 있다(Dean, 1999). 그러나 입사전과 입사후에 경험하는 행동영역이 크게 다를 수 있기 때문에 과거 경험 중 입사후 수행에 필요한 경험만으로 전기자료 문항 목록을 구성하는데는 한계가 있다. 기본적으로 전기자료 검사는 행동 일관성 원리를 따르고는 있으나 미래의 행동과 동일한 과거 수행 측정치만을 가지고 미래 행동을 예측코자 하는 것은 아니다. 이에 대해 Wernimont와 Campbell(1968)은 ‘행동의 표본(sample)’과 ‘행동의 신호(sign)’를 구분하였다. 행동의 표본은 미래의 행동을 예측하기 위하여 미래 행동 영역과 공통되는 수행영역에서 뽑은 과거 행동들이다. 행동의 신호는 준거 수행영역의 행동과 동일하지는 않으나 인과적으로 관련되거나 상관이 높을 것으로 가정된 영역으로부터의 행동들이다. 행동 일관성 원리는 사실상 전기자료 문항을 수행예측에 이용하는 유일한 이론적 바탕으로 행동의 표본 이외에 행동의 신호까지 검사 문항으로 포함시키는데 대한 근거까지는 제공해주고 있지는 않지만 실무적으로 전

기자료 검사의 문항 목록은 대체로 행동의 표본과 신호 모두를 사용하고 있다(Dean, 1999).

여러 연구들에서 전기자료와 다양한 준거들간에 .30에서 .40 사이의 비교적 높은 타당도가 일관되게 보고되고 있다(Hunter & Hunter, 1984; Nickels, 1994a, 1994b). 24개 조직을 대상으로 한 Carlson, Scullen과 Schmidt(1999)의 연구에서는 평균 .53의 높은 타당도가 보고된 바 있다. 전기자료 검사의 타당도에 관한 이러한 연구결과는 선발도구 평가의 가장 기본적인 요건인 예측력을 충분히 충족시키고 있음을 시사한다.

1922년 영업수행 예측을 위해 전기자료가 사용된 이래 영업직을 대상으로 가장 많이 사용되어 왔다. 그러나 Carlson 등(1999)의 연구에서 하나의 조직에서 개발되고 가중치가 부여된 전기자료 검사를 24개 조직의 7,334명의 관리자 표본에 대해 교차 타당화를 시도한 결과, 타당도의 평균이 .53이며 표준편차가 .05로 여러 조직에 걸쳐 타당도 차이가 크지 않은 것으로 나타났다. 이와 함께 여러 인구통계적 특성이 다른 집단들 간에도 타당도 차이가 크지 않은 것으로 나타났다. 물론 Carlson 등(1999)의 연구는 비록 조직이 다르긴 하지만 모두 관리자들로 조직간에 공통적으로 필요한 역량이 많이 포함되어 있을 가능성이 있어서 전기자료의 타당도가 여러 직무에 걸쳐 일반적임을 확신할 수는 없지만 일반화의 가능성 또한 높은 것으로 보인다.

전기자료와 직무수행간 높은 타당도에도 불구하고 전기자료 검사가 국내외의 기업의 인사담당자들에게서 폭넓게 수용되고 있지 못한 가장 큰 이유로 다음 두 가지를 들 수 있다. 매우 높은 준거 타당도의 유혹에도 불구하고 실무자들이 전기자료 검사의 도입을 꺼리는 첫 번째 이유는 이 검사의 학술적 지지 기반이 취약하다는데 있다. 전기자료 검사가 왜 수행을 잘 예측해 주는가를

납득하기 전에 선발 도구로서의 채택을 꺼리게 되는 것이 조직의 일반적인 경향이다. 즉 전기자료에 관한 연구들은 아직 문항의 내용과 수행측정치 간의 명확한 이론적 연결고리를 아직까지는 찾아내지 못하고 있다(이종구, 2001). 이의 부분적인 해결책으로 우선 수행에 영향을 주는 것으로 가설화된 심리적 구성개념(예, 사회적 관계, 성격, 지적 자원, 동기 및 태도 등)을 확인하고 이러한 구성개념의 영향이 반영된 전기자료 문항이 개발되어야 한다.

두 번째 이유는 전기자료의 타당도는 시간에 걸쳐 꾸준히 감소하는 경향이 있다는 데 있다. 특히 사회적 바람직성을 향한 반응 편향이 낮고 타당도 유도집단에서의 타당도가 가장 높게 보고되고 있는 경험적 점수화의 한 방법인 WAB(weighted application blank) 가중 방식에서 더욱 그러하다(Dunnette, Kirchner, Erickson, & Banas, 1960). 이의 이유로 Mumford와 Owens(1987)는 시초 점수화에 사용된 외적 준거 및 준거집단 특성의 변화 또는 조직 내외(인사정책, 노동시장조건 등)의 변화를 들었다. 따라서 전기자료 검사의 보다 폭넓은 확대를 위해서는 우선 수행관련 심리적 구성개념이 반영된 문항개발이 이루어져야 하며 시간적 안정성을 지닌 가중치가 유도되거나 시간변화에 따라 지속적인 타당도의 안정성을 확보하기 위해 경험적 점수화의 가중치의 재산출이 요구된다.

전기자료의 문항구성

전기자료와 수행간의 이론적 연결고리를 찾는 데 도움을 줄 수 있는 방법 중의 한가지는 문항 선정 단계에서 연역적인 방법 사용의 비중을 높이는 것이다. 이를 위해 특정 직무 수행에 필요한 KSAOs(knowledge, skills, abilities, motivation,

and other personality characteristic) 또는 핵심역량(competency)을 확인하고 이들에 영향을 주는 것으로 판단되는 전기자료 문항을 수집한 후, 수행 예측력이 높음과 동시에 수행에 영향을 주는 기존의 심리적 구성개념과 관련성이 높은 문항을 선별해내는 방법을 생각해 볼 수 있다. 물론 이러한 방법이 전기자료 검사의 이론화 부재 문제를 해결해 주지는 못하지만 예측력을 높일 수 있다는 장점 외에 과거 행동과 미래 수행간의 이론적 연결고리를 추론하는데 도움을 줄 수 있다.

전기자료의 문항구성에 특별한 원칙이 있는 것은 아니다. Mumford와 Owens(1987)는 전기자료 문항들은 원칙적으로 과거행동과 경험을 묻는 것이어야 하며, 문항과 반응선택지들은 구체적이고 간결한 것이 좋으며, 과거와 현재행동, 가치관, 태도 및 의견을 묻는 문항은 일반적으로 수용되는 것을 사용하도록 권고하였다. 또 직무분석 자료는 직무수행영역 정의에 도움을 주고 이는 준거개발과 함께 준거수행과 관련성이 있는 것으로 보이는 배경(background)과 경험을 추론하는데 상당한 도움을 주기 때문에 직무분석자료의 이용을 강하게 권고한 바 있다.

전기자료 검사는 여러 직무 그리고 다양한 수행준거를 예측하는데 활용되어 왔다. 그 중에서도 영업 수행의 예측에 가장 많이 이용되고 있는데 그 첫 번째 이유는 타 직무에 비해 수행준거가 명확하다는 점이다. 대부분의 직무에서 수행준거가 정확하지 않은 국내에서는 더욱 그러하다. 두 번째는 영업직의 업무수행에서 요구되는 다음과 같은 특성들이라고 볼 수 있다. 영업사원의 업무 수행에서 가장 두드러지게 요구되는 특성은 자율성(autonomy)과 거절(rejection) 경험에 대한 대처이다(Vinchor, Schippmann, Switzer, III., & Roth, 1998). 우선 영업사원은 집과 회사 밖에서 감독자 없이 업무를 수행한다. 따라서 상당한 자

울권이 있으며 자신의 직관력과 설득력에 의지하여 술선해서 일해야 한다. 상황 압력에 의한 행동 규제가 적은 이러한 상황적 여건은 자신의 과거 경험의 표출을 용이하게 한다. 둘째, 다른 직종과는 달리 영업직은 대인 접촉 기회가 많고 수도 없이 거절당해 보기 때문에(파는 경우보다 못 파는 경우가 훨씬 많다), 이에 대한 감내력 또는 낙천성이 요구되고 성취를 위한 지속적인 노력이 요구된다. 이러한 영업직의 특성은 일반 인지 능력보다 동기, 성격 특성 및 전기 자료가 영업수행을 더 잘 예측하게 해주는 원인이 될 수 있다. Vinchur 등(1998)의 메타분석 결과에서도 성격 5요인 중 성실성(conscientiousness)의 하위 요인인 성취지향성(achievement-orientation)과 전기자료가 영업사원의 수행을 가장 잘 예측하는 것으로 나타났다. 따라서 영업직의 KSACOs에 바탕을 두고 문항이 수집되고 수행에 영향을 주는 것으로 알려진 심리적 구성개념(성실성, 외향성 등)과 관련성이 높은 문항들을 선별해 낸다면 이전 연구에 비해 수행 예측력이 더 높은 전기자료 문항 목록을 구성해 낼 수 있을 것이다.

전기자료의 시간적 안정성

앞서 전기자료가 지닌 두 번째 문제로 시간경과에 따른 타당도 감소 문제를 언급한 바 있다. 전기자료의 타당도와 타당도의 안정성 문제를 논의하려면 이에 앞서 전기자료 문항의 척도화 방법에 관한 이해를 필요로 한다. 전기자료 문항은 4가지 척도화 방법 - 개념적 척도화(rational scaling), 경험적 점수화(empirical keying), 요인척도화(factorial scaling), 및 하위집단화(subgrouping) - 에 의해 점수화될 수 있다(강혜련, 1996; 박영석, 1995).

4가지 척도화 방법 중 가장 보편적으로 오랫동안 사용되어 왔고, 가장 실재적이고, 높은 타당도

를 확보할 수 있는 방법은 경험적 점수화 방법이다. 그러나 이 방법은 이론에 근거하지 않아 준거와 예측변수간의 관계에 대한 이해가 어렵고, 표본의 변화나 시간 경과에 따라 타당도가 다소 불안정하다는 문제점을 안고 있다. 경험적 점수화 방법은 외적 준거의 변화, 집단이 달라짐 또는 조직 내외적 환경이 변화함에 따른 타당도 감소가 점수화 방법 중 가장 큰 것으로 보고되고 있다(이종구, 2001; 이종구, 윤창영, 임대열, 1999; Mitchell & Klimoski, 1982, 1986).

식품매장 판매사원 300명을 대상으로 한 이종구 등(1999)의 연구에서 경험적 점수화 방법의 타당도가 가장 높게 나타났으나 교차타당화에서 가장 큰 타당도 감소현상이 나타났다. 그러나 이들의 연구는 문항 및 가중치 개발 집단과 교차타당화 집단의 표본수가 각각 194명과 104명으로 경험적 점수화 방법을 이용하기에는 지나치게 표본의 수가 적다는 문제를 지니고 있다. 한편 박동건과 전인식(2001)은 보험판매사원(생활설계사) 4813명을 대상으로 검사를 실시하고 이 중에서 대표본(4118명), 중표본(835명), 소표본(393명)을 뽑아 이들 자료를 각각 7대 3의 비율로 개발집단과 교차타당화 집단으로 분할한 후 타당도 변화를 알아보았다. 이들의 연구 결과, 경험적 점수화 방법을 이용했을 때 평균적으로 가장 높은 타당도를 보였고 타당도 감소의 폭도 가장 크게 나타났으나 감소된 후의 타당도도 다른 가중치 체계보다 높게 나타나 한번의 교차타당화 결과만 놓고 본다면 경험적 점수화가 가장 효율성이 높은 것으로 보인다. 또 표본의 크기가 클수록 경험적 점수화의 타당도 감소가 적게 일어났는데, 이들의 연구에서 대표본일 경우 개발집단과 교차타당화 집단의 타당도가 각각 .400과 .348로 상당히 안정적으로 나타났다. 이와 관련하여 9017명을 대상으로 전기자료 검사를 개발한 Carlson 등(1999)의

연구 결과, 전기자료의 타당도의 안정성이 매우 높게 나타났음은 물론 한 조직에서 개발된 전기자료 검사를 타 조직에서도 그대로 이용할 수 있음을 보여주었다. 따라서 대표본을 이용할 경우 경험적 점수화의 두 가지 문제 중 타당도 감소와 관련된 문제는 표본의 수를 크게 함으로써 어느 정도 해결되는 것으로 보인다. 그러나 연구 대상으로 삼은 조직의 규모가 작을 경우 적은 표본수로 인한 타당도 감소 문제는 여전히 나타날 것이다.

경험적 점수화의 해석상의 문제와 소표본일 때의 타당도 감소 문제를 부분적으로 해결해줄 수 있는 대안으로 요인 척도화, 개념적 척도화, 하위 집단화 방법이 제안되었으나 이 중 경험적 점수화와 유사한 맥락에서 자주 시도되는 방법은 요인 척도화 방법이다(Karas, 1997). 요인 척도화 방법을 적용하면 첫째, 개념적 척도화보다는 못하지만 준거 수행의 차원을 찾아낼 수 있고 준거-예측변수의 차원간의 관계에 대한 해석이 가능해진다. 둘째, 타당도의 불안정성 문제도 부분적으로 해결될 수 있다. 한 예로 Mitchell과 Klimoski (1986)의 부동산 영업직을 대상으로 한 연구에서, 경험적 접근에 비해 요인적 접근의 타당도 감소가 덜 일어난다는 결과가 보고된 바 있다.

요인척도화와 WAB 가중 방식의 타당도 감소

문제에 대해 소표본을 사용한 이종구 등(1999)과 박동건과 전인식(2001)의 연구 결과를 비교한 것이 그림 1이다. 요인척도화는 비교적 소표본을 대상으로 한 이종구 등(1999)의 연구 결과 개발집단과 교차타당화 집단간 타당도 차이가 작은 대신 개발집단의 타당도가 경험적 점수화에 비해 낮고, 대, 중, 소 표본을 비교한 박동건과 전인식(2001)의 연구에서 대표본의 경우 요인척도화의 타당도가 경험적 점수화에 비해 개발 및 교차타당화 집단 모두에서 낮게 나타났다. 또 이 연구에서는 타당도 감소도 경험적 점수화에서 오히려 더 적게 나타났다.

박동건과 전인식(2001)의 연구에서 소표본의 경우 경험적 점수화를 사용했을 때 개발집단과 교차타당화 집단간의 타당도 차이(.63→.28)는 요인척도화(.24→.15)보다 크게 나타났으나 이종구 등의 연구와는 달리 경험적 점수화시의 타당도가 요인척도화시의 타당도보다 두 집단 모두 더 높게 나타났다.

이상의 연구 결과에서 가장 바람직한 방법은 대표본을 대상으로 한 경험적 점수화 방법이다. 또 경험적 점수화 중에서도 전기자료 문항의 특성이 대개 명명 또는 서열 척도 수준에서 제작되기 때문에 회귀나 등가중 방법보다 WAB 가중 방법이 보다 적절할 수 있다. 그러나 특정 회사

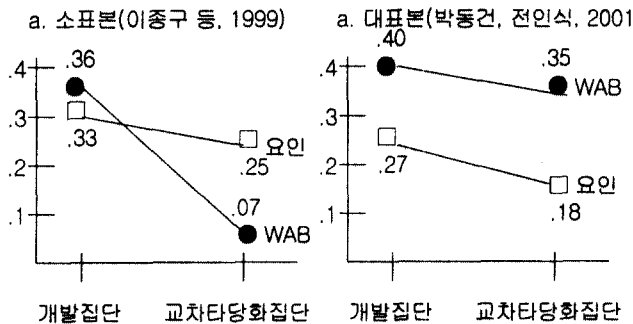


그림 1. 소표본과 대표본의 타당도

를 대상으로 전기자료 검사 제작시 회사 구성원이 많지 않아 대표본을 구성하기 어려운 경우가 많다. 이때는 대개 여러 회사를 대상으로 대표본을 추출하여 제작하는 것이 권장사항이기는 하지만 국내에서 이런 형태로 검사를 제작하기란 대단히 어렵다.

경험적 점수화에서 점수부여 체계를 개발하는 데는 개발집단의 사례수가 대단히 중요하다(Hogan, 1994). 충분한 사례수가 확보되어야 수행에서의 우연적 차이를 제거하고 준거집단의 수행차이를 정확하게 변별해 줄 수 있다. 우연에 의해 오는 예측력의 과장된 증가는 예측 변수의 수에 따라 증가하고 사례수가 증가함에 따라 감소하기 때문에 예측변수대비 사례수의 비율을 최대화 시켜줄수록 안정적인 수행예측을 가능하게 한다.

Nunnally(1978)는 사례수와 문항수의 최소 비율 5 대 1 또는 10 대 1 정도되기를 권장하고 있는데 이를 기준으로 한다면 100문항을 이용할 경우 500에서 1000명 정도의 사례수를 요구한다고 보면 된다. 그러나 이 기준은 item-keying 방법을 이용할 때의 기준이고 option-keying 방법을 이용할 경우(예, WAB 가중) 개별 문항 단위로 가중치가 유도되는 것이 아니라 개별 문항의 반응선택지마다 가중치가 유도되기 때문에 이때 총 예측변수의 수는 모든 문항의 반응선택지의 수와 같아진다. 예컨대 문항의 수가 100개이고 각 문항의 선택지가 3개이면 고려해야 하는 총 예측변수의 수는 300개인 셈이 되어 Nunnally가 제안한 사례수 기준을 충족시키려면 1500명에서 3000명의 사례수가 필요하게 된다. 국내 단일 회사에서 이 정도의 사례수 확보가 가능한 회사는 많지 않다.

본 연구에서는 전기자료 검사의 신뢰도 검토를 통해 이러한 문제의 부분적 해결 방법을 생각해 보고자 한다. 전기자료 검사에 포함된 문항들은 비교적 서로 이질적이어서 개별 문항간 상호상관

은 대체로 낮은 편이다(Kavanagh & York, 1972). 따라서 내적 일관성계수는 비교적 낮게 나타난다. 검사-재검사 신뢰도를 보고한 연구의 수도 많지 않은데 신뢰도의 범위는 대략 .60에서 .90 사이인 것으로 알려져 있다(Hogan, 1994). 만약 한번의 반응을 토대로 다른 반응을 예측하는 것보다 여러 번의 반응을 통합하여 다른 반응을 예측하는 것이 예측 오차를 줄일 가능성이 높다면 2회 이상의 반응자료를 통합하여 가중치를 계산하는 것이 예측 오차를 줄이는 한가지 방법이 될 수 있을 것이다. 본 연구에서는 소표본에서 계산된 가중치의 불안정성을 부분적으로 해결하기 위한 한가지 방법으로 동일집단을 대상으로 전기자료 검사를 반복적으로 실시한 자료를 통합하여 가중치를 계산하고 이를 토대로 타당도의 안정화를 시도하고자 한다.

연구방향

보다 다양한 원천으로부터 문항의 이론적 해석에 도움을 줄 수 있는 방법으로 전기자료 문항을 수집하고 선별한 후 반복 측정된 자료를 통합하여 경험적 점수화의 가중치를 계산했을 때의 타당도와 이의 안정성을 검토하고자 시도된 본 연구의 목적을 상술하면 다음과 같다. 본 연구의 첫 번째 목적은 우리문화에서 접할 수 있는 환경과 경험할 수 있는 환경과 행동들을 나타내고 이론적으로 해석하기 쉬운 전기자료 문항을 수집하는데 있다. 국내 연구에서 이용하고 있는 전기자료 문항은 대부분 외국자료를 번안한 것이다. 전기자료는 문항자체가 성장환경 및 경험에 관한 것으로 이는 문화에 따라 상당한 차이가 있다. 따라서 한국인을 대상으로 한 문항 개발이 절실히 요구된다. 이를 위해 영업직의 KSAOs에 바탕을 두고 문항을 수집하되 수행에 영향을 주는 것

으로 알려진 성격 특성들(성실성, 외향성 등)과 관련성을 토대로 문항을 선별하고자 한다. 본 연구의 두 번째 목적은 한번의 반응을 토대로 가중치를 계산하는 것과 두 번의 반응을 토대로 가중치를 계산했을 때의 타당도의 안정성을 비교하는데 있다.

연구 1: 전기자료 문항개발¹⁾

기본적인 인간본성, 대인관계, 관심사 등 가치의 중요도는 문화마다 차이가 크기 때문에 보편적으로 사용할 수 있는 전기자료 검사를 구성하려면 문화간에 차이를 보이는 문항을 제외시켜야 하고(Whitney & Schmidt, 1997) 특정 문화 내에서 사용하려면 그 문화에 적절한 문항 구성을 필요로 한다. 이와 함께 전기자료 검사의 취약점인 전기자료 문항과 미래수행간의 이론적 연결고리를 찾는데 도움을 주기 위해 문항의 초기 선별과정에서 수행에 영향을 주는 것으로 밝혀진 심리적 구성개념과의 관련성이 고려되어야 할 것이다.

연구 1은 전기자료 문항개발과 선별에 관한 연구이다. 문항개발과 선별과정은 이론화에 도움을 줄 수 있는 다음 3가지에 주안점을 두고 실시하고자 하였다. 첫 번째 주안점은 문항개발을 위한 면접 과정에서 영업 사원에 대한 직무분석 결과에서 나온 KSAOs를 이용하는 것이고, 두 번째는 면접대상자로 영업사원은 물론 발달심리학 전공자를 대상으로 하여 KSAOs를 더 잘 반영해주는 자료를 얻고자 함에 있다. 세 번째는 문항선별 과정에서 영업수행과 관련성이 높은 심리적 구성

개념과 수집된 전기자료 문항 측정을 동시에 실시하여, 심리적 구성개념과의 관련성을 토대로 선별해 내고자 한다. 선별의 기준으로 삼은 심리적 구성개념은 성격 5요인 중에서 영업수행과의 관련성이 가장 높게 보고되고 있는 외향성(extraversion)과 성실성(conscientiousness)(Vinchur et al., 1998), 그리고 혁신성향을 대상으로 하였다. 본 연구에서 독창성을 포함한 혁신 성향과 전기자료 요인간의 관계를 알아보려고 하는 이유는 'MBC의 다큐멘터리 성공시대' 자료의 내용분석을 하는 과정에서 성공한 사람의 특성 중 가장 많이 나타난 내용이 성실성과 독창성을 포함하는 혁신적 사고 및 혁신적 사고에 대한 추진력이었기 때문이다.

방법

문항개발

본 연구에서 전기자료 문항수집은 성장기 경험과 입사전의 경험으로 나누어 이루어졌다. 성장기 경험에 관한 자료는 면접(영업사원, 발달심리학 전공자)과 'MBC 다큐멘터리 성공시대(1회-96회: 1997.11.23~1999.5. 23)'에 나온 인물 중 기업가 26인의 자료에 대한 내용분석을 통해 수집하였다. 내용분석은 각 인물의 성장기 환경과 경험과 관련된 내용을 추출하고 문항화하였다. 우선 국내 모 식품회사 영업사원을 대상으로 한 집단면접(focus group interview: 이하 FGI라고 함)을 통해 영업사원의 과업(task)과 이 과업 수행을 위한 KSAOs를 확인하였다. 4~7명씩 FGI에 참여하였으며 전체 참여인원은 24명이었다. 이후 영업사원(25명)과 대학원생(아동발달 전공 7명, 기타 심리학 전공 9명)에게 앞서 영업사원을 대상으로

1) 연구 1에서 선정된 문항의 다수는 이종구(2001)의 연구에서 이미 사용된 바 있다. 그러나 보다 많은 문항을 포함시키기 위해 문항의 선정기준을 달리하였다.

한 FGI에서 확인된 과업과 KSAOs를 제시하고 이와 관련된 생활사적 정보(성장환경과 경험)를 수집하였다. 성공시대의 내용분석과 면접을 통해 수집된 전기자료 문항은 모두 396개였다. 여기에 대학생을 대상으로 하여 주로 초중고 시절의 환경과 경험을 묻는 문항들로 이루어진 Owens(1976) 전기자료 검사 114문항을 추가하여 총 510개의 문항 목록이 구성되었다. 이렇게 수집된 510개의 문항에서 내용적으로 거의 동일한(3인의 평정자가 모두 동의) 113문항을 제외한 397개 문항이 최종 평정에 이용되었다.

비교적 최근 경험인 입사전 경험과 환경을 측정하는 문항은 국내 S 보험사에서 10명의 우수사원 면접을 통해 수집하였다. 면접과정에서 보험판매사원의 행동 특성 중심으로 개발된 금융연구소의 세일즈 역량을 제시하고 가급적 이와 연관된 사회경제적 배경 및 경험을 보고하게 하였다. 이와 함께 LIMRA career profile의 문항을 참조해서 문항수정과 함께 동일 내용의 문항을 제외시켰다. 수집된 문항은 총 200문항이었다.

질문지 구성과 조사

성장기 경험에 관한 전기자료 문항을 선별하기 위해 제작된 질문지는 397개의 전기자료 문항, Hurt와 Joseph(1977)의 혁신성 척도 20문항, Kirton(1976)의 적응자-혁신자 척도 32문항, 성격 특성 문항은 Costa와 McGrae의 240문항을 포함하여 총 689 문항으로 이루어져 있다. 689 문항의 질문지는 628명(남자 361명, 여자 267명)의 대학생을 대상으로 수업시간에 실시되었다. 조사대상자들에게 연구의 중요성을 강조하고 보너스 점수로 동기부여를 한 후 질문지를 작성케 하였다. 모든 문항은 5점 척도 또는 5개의 선택지로 구성되어 있다. 질문지에 대한 응답은 90분에서 120분 정

도 소요되었다. 입사 전 최근 경험에 관한 200문항의 질문지는 S 보험사의 보험판매사원 700명(우수사원 350명, 부진사원 350명)을 대상으로 이루어졌다.

자료분석

모든 자료는 SAS 6.12판을 이용하여 분석하였다. 성장기 경험에 대한 자료는 후속 타당화 연구와 일관성을 유지하기 위해 628명의 자료 중 여자 267명의 자료만을 대상으로 분석하였다. 397개의 전기자료 문항 각각과 외향성의 하위척도인 활동성과 낙천성, 성실성 요인, Hurt-Joseph 척도의 혁신성 요인, Kirton 척도의 독창성요인간의 상관을 계산하였다. 최근 경험에 대한 자료는 문항 중 비수량 척도는 우수-부진 집단과 문항반응간의 χ^2 검증을, 수량척도에 대해서는 우수-부진 집단을 독립변수로, 개별문항에 대한 반응을 종속변수로 하여 t-검증을 실시하였다.

결 과

성장환경과 영업수행에 영향을 주는 것으로 알려진 심리적 구성개념(활동성, 낙천성, 성실성, 혁신성, 독창성)간의 상관분석 결과를 토대로 이들 구성개념 중 1개 이상과 유의한 상관을 보인 문항 중 .15이상인 117문항을 최종 타당화 연구에 사용할 문항으로 선정하였다.²⁾ 최근 경험을 측정 한 200문항 중 56개의 문항이 우수집단과 부진집

2) $r=.15$ 이상이라는 기준을 적용한 이유는 심리적 구성개념과의 상관이 유의한 문항의 수는 200개가 넘어 이들 문항을 모두 수용하기 어려워 .15라는 다소 임의적인 기준을 사용하였다. 전기자료 문항들은 문항간 상관이 낮아 전통적인 문항분석과 요인분석을 통한 문항 선별은 적절하지 않다.

단간 t-검증과 χ^2 검증 결과 .10 수준에서 유의한 차이를 보였다.

연구 II: 타당도의 시간적 불안정성 검토

연구 2에서는 연구 1에서 선별된 전기자료 문항에 우수 텔레마케터 면접을 통해 추출된 문항을 추가한 검사 목록의 검사-재검사 신뢰도를 확인하고 소표본에 대해 경험적 점수화 방법을 적용했을 때의 타당도의 불안정성 문제를 부분적으로 해결하기 위한 방법을 모색하고자 하였다. 이를 위해 2차례의 응답결과를 토대로 가중치를 계산하고 이 가중치를 적용한 타당도가 보다 안정성있는 수행 예측을 해주는지를 검토하고자 한다.

질문지 구성

본 연구에서 최종적으로 사용한 전기자료 문항은 총 219문항으로 이중 117 문항은 성장기 경험을 측정하는 문항이며, 56문항은 영업직 수행에 영향을 주는 것으로 알려진 입사 이전의 배경과 경험을 측정하는 문항들이며, 46 문항은 우수 텔레마케터 면접을 통해 수집된 문항이다. 우수 텔레마케터 면접은 업적평가를 기준으로 상위 15%에 해당하는 텔레마케터들을 대상으로 그룹별로 이루어졌다(그룹 당 8명씩 4그룹). 인터뷰에서 나온 내용 중 업무수행과 관련된다고 진술한 46개의 입사전 경험을 문항화 하였다. 준거자료로는 회사에서 제공한 업적평가 자료를 이용하였다.

자료수집 및 조사대상자

국내 L사의 텔레마케터를 대상으로 2001년 12월에서 2002년 2월까지 3차에 걸쳐 자료가 수집

되었다. 검사지 작성은 근무시간 중에 통제된 방에서 10-20명씩 집단적으로 이루어졌으며 1, 2, 3차의 검사 실시 간격은 약 1개월이었다. 3차에 걸쳐 수집된 자료의 수와 조사대상자의 평균 연령 및 재직기간은 표 1에 제시되어 있다. 최소 재직기간은 6개월이었다.

최종분석 자료는 업적 평가를 기준으로 자료를 3등분하여 중간집단을 제외한 상하집단이다. 업적평가는 L사에서 제공한 인사고과 점수인데 개인별 수주, 클레임 수, 고객불만처리 등의 점수를 통합한 점수이다. 우수집단의 최하위 및 부진 집단의 최상위 동점자료는 모두 우수 또는 부진집

표 1. 조사대상자

시기	남		여		연령 M (SD)	재직기간 M (SD)
	n	%	n	%		
1차	15	(3.4)	417	(96.5)	27.15(3.39)	1.08(.79)
2차	6	(1.9)	311	(98.1)	27.82(3.42)	1.46(.77)
3차	3	(1.3)	226	(98.7)	27.50(3.05)	1.29(.65)

표 2. 검사시기별 우수-부진 대상자수

시기	우수		부진		전체 n
	n	%	n	%	
1차	140	(49.65)	142	(50.35)	282
2차	106	(50.24)	105	(49.76)	211
3차	76	(49.35)	78	(50.65)	154

표 3. 1, 2, 3차 및 교차타당화 분석 자료의 수

시기	우수		부진		전체 n
	n	%	n	%	
1차	85	(49.13)	88	(50.87)	173
2차	85	(49.13)	88	(50.87)	173
3차	48	(52.17)	44	(47.83)	92
교차	94	(48.21)	101	(51.79)	195

단에 포함시켰다. 검사시기별 및 우수-부진 집단별 분석에 포함된 자료의 수는 표 2에 제시되어 있다. 자료의 동질성을 확보하기 위해 남자자료는 분석에서 모두 제외하였다.

표 3은 실제 분석에 이용된 자료의 수이다. 표 2의 검사시기별 우수-부진 집단 중에서 1차와 2차 검사에 모두 응답한 자료의 수는 173명이며, 3차 응답자중 1차와 2차에 응답한 자료의 수는 92명이었다. 1, 2, 3차 자료에 포함되지 않은 우수-부진 대상자들은 교차타당화에 이용하였다.

결 과

문항선별

단일자료의 타당도 검증을 위해서는 219문항 중 1차 개발집단 자료에 대한 분석결과를 토대로 하였다. 219문항 각각에 대한 우수-부진 집단간 개별 문항에 대한 t-검증과 우수-부진집단×문항 선택지의 빈도자료를 대상으로 한 χ^2 검증 결과를 토대로 하였다. 매우 많은 문항들을 이미 여러 단계를 거쳐 제외하였기 때문에 최종단계에서 가급적이면 많은 문항을 보유하고자 t-검증과 χ^2 검증 결과 중 어느 한 곳에서 .10 수준에서 유의한 88문항을 최종적으로 선별하였다. 1, 2차 통합 자료에 대한 문항 선별은 1, 2차에 응답한 자료를 개별자료로 간주하고 t-검증과 우수-부진집단×문항선택지의 빈도자료를 대상으로 한 χ^2 검증 결과를 토대로 하였는데 총 75문항이 선별되었다.

검사-재검사 신뢰도

표 4의 검사-재검사 신뢰도는 표 3의 1차 자료

표 4. 검사-재검사 신뢰도(단일자료 가중치)

자료	1차자료	2차자료
2차자료	.70*	
3차자료	.65*	.76*

* p<.001

표 5. 검사-재검사 신뢰도(통합자료 가중치)

자료	1차자료	2차자료
2차자료	.61*	
3차자료	.64*	.74*

* p<.001

를 대상으로 88문항의 각 선택지에 대한 WAB 가중치를 도출한 후(Guion, 1965) 이 가중치를 2차(재검사) 자료와 3차 자료에 적용하여 개인별로 검사 총점을 구하여 검사 시기간 상관을 구한 것이다. 그 결과 1, 2차간 검사-재검사 신뢰도는 .70으로 나타났다.

표 5의 검사-재검사 신뢰도는 1차와 2차의 응답 결과를 통합하여 75문항에 대해 WAB 가중방식을 적용하여 가중치를 도출한 수 1, 2, 3차 자료에 이 가중치를 부여하여 검사 시기간의 상관을 구한 것이다. 그 결과 1차와 2차 자료간의 검사-재검사 신뢰도는 .61로 나타났다. Hogan (1994)은 이전 연구에서 전기자료의 검사-재검사의 신뢰도의 범위가 .60에서 .90 사이로 보고한 바 있는데 선행연구들에 비하면 본 연구에서 나온 검사-재검사 신뢰도는 다소 낮은 편으로 볼 수 있다.

타당도 비교

표 6은 2가지 자료를 대상으로 가중치를 도출한 후 타당도를 계산한 결과이다. 단일자료에 대

표 6. 타당도

자료	1차	2차	3차	교차타당화
1차 예측	.64	.27	.25	.24
통합예측	.65	.56	.55	.53

한 결과는 1차 자료의 개발 집단을 대상으로 가중치를 유도해 낸 후 이 가중치를 모든 집단에 적용하여 업적평가와의 상관관을 구한 것이다. 그 결과 개발집단(1차)에 대한 타당도는 .64로 매우 높게 나타났으나 교차타당화의 타당도가 크게 감소되었으며(.64→.24) 동일한 대상자에게 실시한 2차와 3차 자료에서도 타당도 감소가 큰 것으로 나타났다. 따라서 소표본에서 유도된 경험적 점수화의 타당도가 매우 불안정적임을 보여주었다.

통합자료에 대한 결과는 1차와 2차 자료를 단순히 통합한 자료를 대상으로 가중치를 유도해낸 후 이 가중치를 1, 2, 3차 및 교차타당화 자료에 적용하여 준거와의 상관관을 구한 것이다. 단순 통합은 173명의 1, 2차 자료를 모두 개별 자료로 간주하여(346명으로 간주) 통합가중치를 계산하였다. 그 결과 가중치를 계산하는데 포함된 1, 2차 자료에 대한 타당도가 높음은 물론 교차타당화 자료와 3차 자료에 대한 타당도 감소가 거의 일어나지 않는 것으로 나타났다.

논 의

전기자료의 타당도를 높이기 위해 다양한 정보 원천으로부터 수행과 관련된 문항목록을 구성하고 보다 안정적인 타당도 확보방안을 모색하고자 시도된 본 연구의 결과를 요약·논의하면 다음과 같다.

지금까지 전기자료의 타당화 연구에서 가장 많

이 사용되어온 문항점수 부여방법은 경험적 점수화(empirical keying) 방법이다. 이 방법은 사용하기가 쉽고 높은 타당도를 얻을 수 있다는 장점이 있으나 타당도가 매우 불안정하고, 타당도에 대한 설명적 근거, 즉 문항-준거간 이론적 연결고리를 제공해주기 어렵다는 문제가 계속 지적되어 왔다. 또 타당도를 높이기 위해 WAB 기법을 비롯하여 다양한 가중치부여 방법이 제안되어 타당도의 개선이라는 목표는 어느 정도 달성되었으나 그 어느 기법도 표본의 변화나 시간경과에 따른 타당도의 불안정성을 해결하고 있지는 못하다.

본 연구에서는 전기자료 문항-수행준거간 이론적 연결고리를 찾는데 일조하고 보다 안정된 타당도를 얻기 위해 다양한 원천으로부터 597문항(성장기 경험 397문항, 입사전 최근경험 200문항)을 수집하였다. 이들 문항들 중에서 타당화 연구에 사용한 문항들에 대한 용이한 해석을 위해, 그리고 타당도를 높이기 위해 선별기준으로 수행에 영향을 주는 것으로 알려진 심리적 구성개념들을 이용하였다. 이 과정을 통해 성장기 경험 112문항, 입사전 최근 경험 56문항이 선정되었고, 우수 텔레마케터 면접을 통해 46문항이 추가되어 최종 219문항의 전기자료 검사 목록이 구성되었다. 그 결과 개발집단의 타당도가 .64와 .65(표 6 참조)였으며 자료를 통합했을 때 검증집단의 타당도가 .53으로 나타나 Hunter와 Hunter(1984)가 보고한 전기가료의 타당도 .30~.40보다 높게 나타났다.

연구 2에서 검사-재검사 신뢰도는 .61에서 .76 사이로 나타나 선행 연구 결과에서 보인 신뢰도 범위내에 포함되었으나 높은 수준의 신뢰도를 보이지는 않았다. 연구 2의 주 목적인 가중치 개발 자료에 따른 타당도 비교 결과 개발 집단의 타당도에서는 차이가 없었으나 재검사 자료와 교차타당화 자료에서의 타당도는 상당한 차이를 보였

다. 단일 자료로 가중치를 유도했을 경우 재검사와 교차타당화에서 상당한 타당도 감소를 보였으나 1, 2차 자료를 통합하여 유도된 가중치를 적용했을 경우 타당도 감소가 거의 나타나지 않았다.

전자자료의 검사-재검사 신뢰도와 타당도 분석 결과를 연관지어 볼 때 1개월 간격으로 실시된 1, 2, 3차 검사간의 신뢰도가 크기 면에서는 비교적 안정되고 일관되게 나타났으나 그다지 높게 나타나지는 않았다. 따라서 검사-재검사 신뢰도가 .61 ~ .76 정도라면 교차타당화 집단은 물론 동일 집단에 반복실시된다고 하더라도 어느 정도의 타당도 감소는 충분히 관찰 가능하리라고 예상할 수 있다. 본 연구에서 단일자료를 가중치 개발 자료로 이용했을 때 .37~.40 정도의 타당도 감소를 보였는데 이 결과는 검사-재검사 신뢰도의 크기를 고려할 때 가능한 결과로 보여진다.

반면, 2차례에 걸친 자료를 통합하여 가중치 개발 집단을 구성함으로써 가중치 개발에 이용된 사례수의 증가는 물론 개인의 반응편차가 제거되었을 때 개발집단의 타당도는 단일자료를 이용했을 때와 비슷하게 나타났으나 타당도 감소 현상은 크게 줄어들었다. 따라서 이론적 측면보다 실제적 측면에 보다 초점을 둔다면 실시상의 어려움이 따르지만 반복 실시한 자료를 통합하여 가중치를 개발하는 방법을 타당도 감소 문제의 해결 대안으로 고려해 볼 수 있을 것이다.

그러나 본 연구의 결과가 폭넓게 수용되려면 몇 가지 해석상의 주의와 이론적인 검토 및 체계적인 보완 연구를 필요로 한다. 선행 연구들에서 대표본을 이용했을 때의 타당도는 일관되게 안정적으로 나타난 바 있다. 본 연구에서는 회사의 규모상 대표본을 이용하기 어려워 2차례에 걸친 응답결과를 통합하여 표본의 크기를 늘린 후에도 가중치를 유도한 결과 비교적 안정된 타당도를

유지할 수 있었다. 그러나 이 방법은 개인의 반응 평균이 단일 반응의 결과에 비해 예측오차를 줄여줄 것이라는 추론에 의거하여 시도한 것이지만 명확한 이론적 근거에 의거한 것은 아니다. 이에 더하여 2차례에 걸친 응답결과가 동일하지는 않지만 동일인의 2회에 걸친 반응을 단순 통합하여 2배수의 사례로 간주하여 분석자료로 이용하는데 대한 비판이 따를 수 있다. 그러나 본 연구에서 검토하고자 하는 주된 부분이 통계적 유의도 검증에 있지 않다. 통계적 검증이 결과물의 핵심부면이라면 반복측정에 의한 사례수 늘이는 당연히 심각한 문제를 안게 된다. 그러나 주 관심이 상관의 크기라면 동일자료를 몇 배로 늘이더라도 그 크기는 동일하게 유지되기 때문에 타당도의 안정성(상관의 크기)에 초점이 두어진 본 연구에서는 이 문제가 분석상의 심각한 문제를 불러 일으키지는 않을 것으로 판단된다.

둘째, 본 연구에서 가중치 개발에 사용된 문항의 수는 단일자료의 경우 88개, 통합자료의 경우는 75개이다. 가중치의 산출은 option-keying 방법을 사용했으므로 문항들에 포함된 선택지의 수를 문항수로 간주한다면 300개 내외인 셈이 된다. 따라서 Nunnally(1978)가 권고한 최소한의 사례수는 1500명 정도가 되는데 본 연구에서 자료를 통합하여 자료의 수가 늘어나 다소 안정된 타당도가 확보되었다고는 하지만 경험적 점수화 중 WAB 방법을 이용하기에는 자료의 수가 매우 부족한 편이다. 향후 검사 대상자의 수가 1000명 이내의 단일 회사에 대해 전자자료 검사가 개발될 경우 새로운 자료의 지속적인 추가와 가중치 조정작업이 뒤따라야 할 것이다. 이와 함께 유사 직종의 다른 회사에서 실시한 자료를 통합한 가중치 조정작업이 필요하리라고 본다.

참고문헌

- 강혜련 (1996). 경험적 척도화에 의한 전기자료의 타당화 연구. *한국심리학회지: 산업 및 조직*, 9(2), 1-16.
- 박동건, 전인식 (2001). 전기자료(Biodata) 문항의 가중치 부여체계간의 타당도 연구: 분석집 단크기에 따른 비교연구. *한국심리학회지: 산업 및 조직*, 14(1), 101-114.
- 박영석 (1995). 이직자의 생활사. *한국심리학회지: 산업 및 조직*, 8(1), 75-102.
- 이종구 (2001). 전기자료(Biodata)의 타당화 연구: 보험판매사원의 수행 예측. *한국심리학회지: 산업 및 조직*, 14(3), 83-107.
- 이종구, 윤창영, 임대열 (1999). 영업수행의 예측에서 전기자료의 이용. *한국심리학회지: 산업 및 조직*, 12(1), 141-153.
- Carlson, K. P., Scullen, S. E., Schmidt, F. L., Rothstein, H., & Erwin, F. (1999). Generalizable biographical data validity can be achieved without multi-organizational development and keying. *Personnel Psychology*, 52(3).
- Dean, M. A. (1999). *On biodata construct validity criterion-related validity, and adverse effect*. Unpublished doctoral dissertation, Louisiana State University.
- Dunnette, M. D., Kirchner, W. K., Erickson, J., & Banas, P. (1960). Predicting turnover among female office workers. *Personnel Administration*, 23, 45-50.
- Guion, R. M. (1965). *Personnel testing(ch.13)*. McGraw-Hill.
- Hogan, J. B. (1994). Empirical keying of background data measures. In Stokes, G. S., Mumford, M. D., & Owens, W. A.(Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction*. Palo Alto, CA: CPP books.
- Hunter, J. E., & Hunter, K. F.(1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96. 72-98
- Hurt, H. T. & Joseph, K. (1977). Scales for the measurement of innovativeness. *Human Communication Research*, 4, 58-65.
- Karas, M. (1997). *Combining biodata keying strategies to meet the goals of science and practice*. Symposium paper presented at the Second Australian Industrial and Organizational Psychology Conference, Melbourne, Australia.
- Kavanagh, M. J., & York, D. R. (1972). Biographical correlates of middle managers' performance. *Personnel Psychology*, 25, 319-332.
- Kirton, M. J. (1976). Adaptors and innovators: A description and measure. *Journal of Applied Psychology*, 6, 622-629.
- 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.
- 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.
- Mitchell, T. W., & Klimoski, R. J. (1986). Estimating the validity of cross-validity estimation. *Journal of Applied Psychology*, 71. 311-317.
- Nickels, B. J. (1994a). Generation procedures for biodata items. In Stokes, G. S., Mumford, M. D., & Owens, W. A.(Eds.), *Biodata handbook:*

- Theory, research, and use of biographical information in selection and performance prediction.* Palo Alto, CA: CPP books.
- Nickels, B. J. (1994b). The nature of biodata. In Stokes, G. S., Mumford, M. D., & Owens, W. A.(Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction.* Palo Alto, CA: CPP books.
- Nunnally, J. (1978). *Psychometric theory.* New York: McGraw-Hill.
- Owens, W. A. (1976). Background data. In M.D. Dunnette(Ed.), *Handbook of industrial and organizational psychology.* Chicago: Rand McNally.
- Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology, 35,* 1-62.
- Vinclair, A.J., Schippmann, J.S., Switzer, III, F.S., & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology, Vol. 83,* No. 4, 586-597.]
- Wernimont, P. F., & Campbell, J. P. (1968). Signs sample, and criteria. *Journal of Applied Psychology, 52,* 372-376.
- Whitney, D. J. & Schmitt, N.(1997). Relationship between culture and responses to biodata employment items. *Journal of Applied Psychology, 82,* 113-129.
- 1 차원고접수 : 2002. 6. 26.
수정원고접수 : 2002. 11. 19.
최종게재결정 : 2002. 11. 23.

A Possible Alternative Method to Resolve Unstable Validity of Biographical Data in Small Samples

Jong-Goo Lee

Daegu University

This study was conducted to examine a possible alternative method to resolve unstable validity of biographical data in small samples. In this study, 219 biographical items were derived from life history with incumbents about early and recent life experience respectively. The participants of this study were tele-marketers in L company and the survey was repeatedly conducted for three times: 432 cases, 316 cases, 229 cases. Based upon the results of categorical data analysis (χ^2 test), 88 items among 219 biographical items was selected. Test-retest reliability test showed that the reliability coefficient was .70. But the significant validity shrinkage was observed in the cross-validation and retest data. To establish stable validity of biographical data, the combined weights with first and second data were recalculated. The results showed relatively stable validities in test, retest and cross-validation data.

key words : biographical data, validity, WAB, small samples