

텔레마케터 선발용 역량검사의 구성형식별 신뢰도와 타당도: Ipsative와 Normative 측정간 비교*

이 종 구[†]

대구대학교 심리학과

한 영 석

중앙공무원교육원

본 연구에서는 자기평정척도의 응답왜곡을 방지하기 위해 역량검사를 3가지 형식(normative, random ipsative, fixed ipsative)으로 제작하여 각 검사형식별 신뢰도와 타당도를 비교하였다. 조사대상자는 국내 L기업에 근무하는 여성 텔레마케터 272명이다. 3가지 검사 형식과 긍정응답왜곡(faking good)간의 상관분석 결과, 표준적 형식에서는 모든 역량이 응답왜곡과 유의한 정적 상관을 보였으나 두 가지 선택형 형식에서는 일부를 제외하고 모두 상관이 .10이하로 낮게 나타났다. 두 가지 선택형 형식의 신뢰도는 표준적 형식보다 다소 낮게 나타났으나 하나의 역량을 제외하고는 수용할만한 신뢰도를 보였고 타당도는 표준적 형식(.18)보다 다소 높게(.23, .31) 나타났다. 선택형 형식에서 모든 역량은 응답왜곡 경향이 사라짐과 동시에 신뢰도와 타당도는 수용할 만하게 나타나 현장에서 확대 적용가능성이 있을 것으로 보인다.

주요어 : 텔레마케터, 표준적 측정, 선택형 측정, 긍정응답왜곡, 역량

* 익명의 세 분 심사위원의 세심한 조언에 진심으로 감사드립니다.

† 교신저자 : 이종구, 대구대학교 심리학과, 경북 경산시 진량읍 내리리 15번지, gooya@daegu.ac.kr

인사선발 도구로서 심리검사는 그 내용과 형식이 어떠한 직무수행과 연관된 역량을 측정하기 위한 목적으로 제작되고 활용된다. 통상 인지적인 역량은 적성검사를 통해, 필요 지식은 지원서를, 경험은 전기자료(biographical data)를, 성격과 동기부분과 관련된 역량은 성격(또는 인성)검사의 형식을 빌어 역량검사와 면접을 통해 측정한다. 본 연구에서 주로 다룰 영역은 성격과 관련된 역량 부분이다. 성격검사의 타당도에 대한 Barrick과 Mount(1991)의 메타분석 결과, 성격검사의 수행에 대한 예측력이 높은 것으로 나타났다. 선발장면에서 성격검사의 도입은 급격한 증가 추세이나 긍정응답왜곡(faking good) 가능성이 크기 때문에 검사개발자나 기업 모두 선발결정에서 성격검사의 비중을 어느 정도 두어야 할지에 대해 확신하지 못하는 경우가 많다. 반대로 징집용 성격검사의 경우 부정응답왜곡(faking bad)이 매우 심한 편이다.

선발이나 평가 장면에서 가장 많이 이용되는 자기보고형(self-report) 검사의 형식은 문항별 평정형으로 규준적(normative) 형식이다. 이 형식의 강점은 상대 척도로서 규준(norm)만 제공되면 개인차를 쉽게 파악할 수 있다는데 있다. 그러나 이러한 형식은 응답왜곡에 대한 방지책의 미비로 인해 가장 보편적인 형식이며 다루기 쉬움은 물론 개인간 상대적 비교가 가능한 척도임에도 인사 담당자가 가장 불신하는 형식이기도 하다. 이러한 이유로 역량과 성격검사를 제작할 때 응답왜곡에 심하게 노출되어 있는 규준적 형식의 문제점을 해결하려는 다양한 방법들이 시도되고 있다.

긍정응답왜곡(faking good)을 줄이는 방법 중 대표적인 것이 검사 형식을 선택형(ipsative)¹⁾

1) Ipsative는 그 의미에 따라 강제선택형 또는 선택

형식으로 바꾸는 것이다(Bowen, Martin, & Hunt, 2002; Hicks, 1970). 성격검사에서 선택형 측정방식을 사용할 경우 응답왜곡이 줄어드는 결과는 여러 연구에서 일관되게 나타난다(김명소, 이현주, 2006; 이종구, 한영석, 신강현, 허창구, 박지훈, 최우성, 2007; Saville & Willson, 1991).

그러나 응답 왜곡의 방지가 성격검사가 갖추어야 할 중요한 요건 중 신뢰도나 타당도 이상으로 중요한 요건은 아니다. 응답왜곡을 방지하기 위해 사용하는 방법이 검사의 신뢰도나 타당도를 손상시킨다면 그런 형식을 도입하기 보다는 다소의 응답왜곡이 있더라도 신뢰도와 타당도가 보장되는 검사를 사용함이 옳다. 다만 검사 형식을 바꾸더라도 신뢰도와 타당도가 최소한 규준적 형식의 검사와 비슷하게 유지된다면 당연히 검사 형식의 변경을 고려해 보아야 할 것이다. 본 연구에서는 일차적으로 규준적 형식의 응답 왜곡 정도를 확인하고 두 가지 선택형 형식의 검사가 긍정응답 왜곡을 방지하는지의 여부를 확인하고자 한다. 그런 연후에 검사 형식별 신뢰도와 타당도를 비교하고자 한다.

응답왜곡과 통제방법

인지능력 검사를 제외한 대부분의 선발도구(지원서, 인성검사, Biodata, 면접 등)에서 정도

형, 개인내 비교형, 상대 참조형 등의 용어로 사용되기도 한다. 그러나 어느 번역어 모두 Ipsative의 의미를 충분히 담고 있지는 않다. 입사티브(ip-sah-tiv)라는 원어 발음 그대로 사용하는 것도 방법이지만 본 연구에서는 가장 근접하다고 판단되는 '선택형'이라는 번역어를 사용하기로 하였다.

의 차이는 있으나 응답왜곡 경향이 나타난다. 긍정적 응답왜곡은 특정 직무를 얻으려 할 때, 즉 인사선발상황에서 보편적으로 일어나며, 부정적 응답왜곡은 특정 직무를 회피하려고 할 때(예, 입대상황) 일어난다. 긍정적 응답왜곡은 사회적으로 바람직한 특성을 채는 문항에 실제보다 좋게 응답하고, 사회적으로 바람직하지 않은 특성은 실제보다 그렇지 않은 것으로 응답하는 것을 말한다. 부정적 응답 왜곡은 사회적으로 바람직한 특성에는 실제보다 더 그렇지 않은 것으로, 바람직하지 않은 특성은 실제보다 더 그런 것으로 응답하는 경우이다. 이러한 응답왜곡이 모든 사람에게 같은 방향으로 일정한 크기로 나타난다면 문제될 소지가 크지 않다. 그러나 응답왜곡이 일부 지원자에게 편중되게 나타나고 이것이 선발 결정에 영향을 준다면 적절한 통제 방법을 찾아야 할 필요가 있다.

Schmitt와 Oswald(2006)의 연구결과 직업을 얻으려 동기화된 지원자가 현직자에 비해 긍정적 응답왜곡 경향이 더 큰 것으로 나타났다. 또 Jackson, Wroblewski 및 Ashton(2000)의 연구 결과에서도 솔직한 응답을 부탁한 조건에 비해 지원자를 가정한 조건에서 응답왜곡이 더 높은 것으로 나타났다. 이러한 응답왜곡은 선발결정에 심각한 영향을 줄 수 있는데 선발비율이 적을수록(경쟁률이 높을수록) 소수의 응답왜곡에 더 심각하게 영향을 받는다(Griffith, Chmielowski, Snell, Frei, & McDaniel, 2000).

자기보고식 검사의 응답왜곡으로 인해 평균이 달라지는 점(Zickar, Rossé, Levin, & Hulin, 1996)에 대해 대부분 동의하지만 응답왜곡이 타당도에 미치는 영향에 관해서는 상반된 연구결과가 많다. Barrick과 Mount(1996)는 응답

왜곡이 준거관련 타당도에 영향을 주지 않는다고 보고한 반면 Worthington과 Schlottmann(1986)은 응답왜곡이 준거관련 타당도 감소에 영향을 준다고 보고하였다. Rossé, Stecher, Miller 및 Levin(1998)은 응답왜곡을 하더라도 타당도 변화가 크지 않는 것은 분포의 양극단 점수의 순위가 달라지더라도 상관계수의 크기 변화가 크지 않기 때문이라고 보았다. 그러나 응답왜곡이 설령 타당도에 영향을 주지 않더라도 선발결정시 상위에 있는 개인들의 순위를 바꾸어 놓을 수 있다는 점에서 심각히 고려되어야 할 것이다. 정도에서 차이가 있으나 비교적 효과적인 것으로 알려진 응답왜곡 통제 방법들은 다음과 같다(이은정, 박동건, 2003).

사전경고

검사 실시 전에 응답자들에게 왜곡된 응답을 하면 불이익이 있으니 정직하게 응답하라는 지시를 할 경우 응답왜곡이 다소 줄어들 수 있다. Doll(1971), Schrader와 Osburn(1977)의 전기자료(biodata) 연구에서 거짓말 척도(lie scale)에 대한 사전 경고시 응답왜곡이 줄어드는 것으로 나타났다. 그러나 사전경고에도 불구하고 응답왜곡을 한 응답자의 점수를 통제할 수 없다는 단점을 지닌다.

응답왜곡 탐지 척도의 사용

응답왜곡의 정도를 파악하기 위해 응답왜곡 척도를 이용할 수 있다. 긍정 또는 부정 왜곡 척도는 정상범위에 있는 사람들이 행동하기 어려운 극단적으로 긍정적이거나 부정적인 문항으로 구성되어 있다(Rossé et al., 1998). 신뢰도와 타당도가 확보된 응답왜곡 탐지 척도를 사용할 경우 응답을 극단적으로 왜곡시킨 지

원자에게 불이익을 줄 수 있다. 그러나 적당히(극단적이지 않게) 응답을 왜곡한 경우 이에 대한 제재가 어렵다는 문제가 있다. 선발 상황이 아닌 현직자를 대상으로 할 경우(예, 자기 평가) 응답왜곡자에 대한 처리가 애매할 수 있다. 극단적인 응답왜곡자에 대해 재검사를 실시할 수 있으나 반복해서 응답왜곡을 할 경우 적절한 조정방법을 찾기 어렵다.

점수 교정

응답왜곡 통제를 위한 하나의 방편으로 역량 검사 점수를 응답왜곡의 정도에 따라 깎거나 높여주는, 즉 점수 교정 방법을 생각해 볼 수 있다. 점수의 교정(correction)은 회귀 가중치를 이용하는 것이 가장 일반적이다(Ellingson, Sackett, & Hough, 1999). 이 방법을 이용하려면 역량 검사를 실시할 때 응답왜곡 척도를 동시에 실시해야 한다. 역량점수의 교정절차는 다음과 같다.

첫째, 역량점수(Y)를 종속변수로, 응답왜곡 점수(X)를 독립변수로 하여 회귀가중치(b)를 계산한다. 둘째, 회귀가중치를 이용하여 역량 점수를 조정하여 교정점수를 산출한다. 교정 점수(Yc)의 산출 공식은 다음과 같다.

$$\text{교정점수 } Y_c = Y - b(X - \bar{X})$$

만약 역량점수와 응답왜곡 점수간의 상관관계가 높고 역량점수(Y)가 50점인 A, B, C 세 사람의 응답왜곡 점수(X)가 각각 10, 20, 30점이고 응답왜곡 점수의 평균이 20점이라면 세 사람의 교정 전 역량점수는 같지만 교정된 점수(Yc)는 A > B > C 순이 될 것이다. 교정점수를 이용할 경우 응답왜곡을 한 사람의 점수가 하향 조정될 수 있다는 장점이 있지만 응답왜

곡을 하지 않고 솔직하게 응답한(응답왜곡 점수가 낮은) 사람의 점수도 상향 조정된다는 문제를 지닌다.

점수 교정에서 제기되는 또 다른 문제는 응답왜곡을 사회적 바람직성과 같은 개념으로 볼 때 나타난다. 보통 응답왜곡을 측정하기 위해 거짓말 척도(lie scale)나 사회적 바람직성(social desirability) 척도를 사용하는데 Hurd, Barrett, Miguel, Tan 및 Lueke(2001)의 연구 결과에서 사회적 바람직성은 적응성, 성실성, 정서적 안정성, 정직성 등 가치 있는 성격특성과 관련이 있는 것으로 보고된 바 있다. 따라서 이 같은 경우 사회적 바람직성 점수를 토대로 성격 요인의 점수를 교정한다면 수행과 관련된 성격특성 분산을 줄여 오히려 준거관련 타당도를 낮추는 결과를 초래할 수 있다.

우수수행자와의 거리(distance)

이 방법에서 개인의 역량점수는 우수수행자들의 역량점수 평균과 개인의 역량 점수간의 차이값이다. 여러 개의 역량을 동시에 측정하였을 경우 역량별로 차이값을 계산한 후 절대값의 합산점수를 이용할 수 있다. 극단적인 차이에 가중치를 두고자 할 경우에는 차이의 제곱합을 이용할 수 있다. 어떤 경우이건 차이값이 작을수록 좋은 점수를 받는다. 이 방법은 응답자들이 응답왜곡을 하더라도 좋은 점수를 보장하기 어렵기 때문에 응답왜곡 방지책으로는 비교적 양호한 방식에 속한다. 전기자료(biographical data)의 경험적 점수화(empirical keying)에서 주로 이 방식을 사용하며(이종구, 2001; Guion, 1965; Hogan, 1994), 전기자료가 성격검사에 비해 응답왜곡에 덜 영향 받는 것은 문항의 사실확인 가능성으로 인해 솔직한 응답가능성이 높다는 점과 함께 경

험적 점수화 방법도 한 가지 이유이다. 그러나 우수수행자 집단을 사전에 구성하기 어렵고 설령 구성한다고 하더라도 현재의 우수수행자(과거 수행에 의한 평가)이기 때문에 미래에 필요한 역량을 지닌 사람을 선별하기 어렵다는 문제를 지니고 있다.

선택형 방식의 측정

선택형 측정에서는 사회적 바람직성이 비슷한 내용들로 선택지 집합을 구성하고 이 중에서 자신에게 가장 부합되는 것과 부합되지 않는 것을 선택하게 하여 부합되는 것은 양수(+)값을, 부합되지 않는 것은 음수(-)의 값을, 선택되지 않은 것은 0점을 부여한다. 한 집합의 채점결과의 합이 0이 될 때 이를 선택형(ipsative) 측정이라고 한다(Baron, 1996).

선택형 측정방식(표 1 참조)에서는 선택지 집합 내 문항들의 사회적 바람직성이 비슷한 가운데 부합(best fit)과 비부합(worst fit) 선택지

를 강제로 선택해야 하기 때문에 응답왜곡 가능성이 거의 없다는 강점을 지닌다. 응답왜곡 방지가 유일한 목적이라면 이 방법이 가장 권장된다. 그러나 이 방법은 규준적 척도보다 하위척도의 신뢰도가 대체로 낮아지는 것으로 보고되고 있고, 이론상 개인간 비교가 불가능하다는 문제점이 있으며(Johnson, Wood, & Blinkhorn, 1988), 모든 사람의 총점이 동일하기 때문에 개인차가 주 관심인 선발장면에 이용하기가 대단히 어렵다는 문제를 지닌다.

표 1에서 규준적 측정 방식에서는 평정형 4개의 문항이 있으며, 선택형에서는 4개가 하나의 선택 집합이 된다. 만약 규준적 측정 방식에서 표 1과 같이 응답한(4, 3, 3, 2) 사람이 나 3, 2, 2, 1 또는 5, 3, 3, 2와 같이 1번을 가장 높게 평가하고 2와 3번은 중간, 1번을 가장 낮게 응답한 모든 사람은 선택형 측정 방식에서 동일한 점수를 가지게 된다(①은 가장 일치, ④는 가장 불일치).

표 1. 규준적 및 선택형 측정방식의 예

규준적 측정					
	전혀 아니다		매우 그렇다		
1) 나는 꾸준한 거북이형이다 -----	①	②	③	④	⑤
2) 나는 사람 많은 것을 좋아한다 -----	①	②	③	④	⑤
3) 나는 새로운 변화를 좋아한다 -----	①	②	③	④	⑤
4) 나는 리더 경험이 많다 -----	①	②	③	④	⑤
선택형 측정					
	가장 일치	가장 불일치			
1) ① 나는 꾸준한 거북이형이다	☑	○			
② 나는 사람 많은 것을 좋아한다	○	○			
③ 나는 새로운 변화를 좋아한다	○	○			
④ 나는 리더 경험이 많다	○	☑			

선택형 측정의 문제점에 대한 해결책

선택형 측정이 가지는 주요문제는 신뢰도 감소, 모든 개인의 총점이 동일 또는 유사하다는 점, 개인간 비교가 어렵다는 점 들이다 (Baron, 1996). 표준적 검사형식에 비해 신뢰도 (여기서는 내적 일관성)가 낮아지는 문제는 애초에 변별도(문항-총점 상관)가 높은 문항들로 하위척도를 구성하여 신뢰도가 다소 낮아 지더라도 수용할만한 선까지 낮아지게 하는 방법 외에 별다른 해결책이 없다.

개인들의 총점이 비슷하다는 것과 개인간 비교가 어렵다는 것은 유사한 문제이기도 하고 전혀 별개의 문제이기도 하다. 가령 두 역량을 측정하는데 한 역량당 10문항씩 있고 이를 양자택일형으로 구성하였다고 하자. 선택된 문항에는 +1점, 선택되지 않은 문항에는 -1점을 부여할 경우, 한 역량에 속한 10문항이 모두 선택된다면 이 역량의 점수는 10점이 될 것이고 나머지 한 역량의 점수는 -10점이 될 것이다. 어떤 식으로 선택되건 두 역량 점수의 평균은 0점이 된다. 만약 3개의 역량을 측정하고자 할 때 3지선다형으로 구성하여 자신과 가장 부합되는 것과 가장 부합되지 않는 것을 하나씩 선택하게 하여 부합되는 것은 1점, 선택되지 않은 것은 0점, 부합되지 않는 것에는 -1점을 부여해도 세 역량점수의 평균은 0점이 될 것이며 개인간 총점의 분산 (variance)도 0이 될 것이다.

인사선발 과정에서 핵심역량들의 총점을 이용하는 경우가 많은데 위와 같은 선택형 형식에서는 모든 사람의 역량 총점은 0점이기 때문에 총점 이용이 불가능해진다. 총점의 분산이 0이 되는 문제를 해결할 한 가지 방법이 부분적 선택형 측정(partial ipsative measure) 방

법이다(Hicks, 1970). 부분적 선택형 측정이란 다수의 역량을 측정하여 그 중 일부분만을 이용하거나 다수의 선택지 집합에서 일부만을 선택하게 하여 채점하는 방식을 말한다. 예컨대 30가지 역량을 측정하여 그 중 핵심이 되는 10개의 역량만을 채점하고 이들의 총점을 이용한다면 총점의 분산이 커지고 개인차가 드러날 수 있다(Hicks, 1970).

문항에서 선택지의 수에서 강제로 선택해야 하는 선택 수를 n 값이 2 이상이 되게 선택지를 구성하면 이 또한 총점의 분산이 0보다는 커지게 된다. 만약 선택지 수에서 선택수를 n 값이 1일 경우, 예컨대 3개 역량을 측정하고자 하여 3지선다형으로 구성하고 자신과 가장 부합되는 것 1개와 부합되지 않는 것 1개를 선택하게 했다고 하자. 이 경우 부합되는 것은 +1점, 부합되지 않는 것은 -1점, 선택되지 않은 것은 0점이 부여되어 세 역량 점수의 평균이나 총점은 0이 된다. 이와 다르게 4지선다형(선택지의 수=4)이나 5지선다형(선택지의 수=5)으로 제시하고 이 중 자신과 부합되는 것과 부합되지 않는 것 각각 1개씩을 선택하게 하면(선택 수=2) 총점이 개인마다 동일하지 않게 된다. 선택수에 비해 선택지수가 커질수록 총점의 분산은 더 커질 가능성이 있다. 그러나 총점의 분산을 높이기 위해 선택지의 수를 늘리게 되면 지나치게 일부만을 채점하게 되어 문항의 낭비가 심하게 된다.

총점의 분산이 영에 가까울 경우 개인간 비교는 가능하지 않다. 그렇다고 총점의 분산이 영보다 아주 크게 나오더라도 표준적이 아닌 선택형 측정에서 나온 결과라면 이 결과 자체가 개인간 비교가 가능하다는 것을 보장해 주지는 못한다. 이 문제에 대한 실질적인 해결책은 사실상 없다고 볼 수 있다. 가령 5개 역

량을 측정하여 2개 역량의 총점을 구하는 식의 부분적 선택형 방식으로 측정하더라도 실제로는 5개 역량이 모두 낮은데 그중 하나의 역량은 상대적으로 더 높고 다른 하나의 역량은 상대적으로 더 낮은 사람과 실제로는 5개 역량이 모두 높은 가운데 2개의 역량이 상대적으로 높거나 낮은 두 사람의 총점은 같을 수 있다. 물론 자신이 가진 역량 중 어떤 부분이 상대적으로 높고 어떤 부분이 상대적으로 낮은지, 또는 소수의 핵심역량의 총점이 개인내에서 상대적으로 높은지 여부에 관심이 있다면, 즉 개인간 비교에 관심을 두지 않고 한 개인의 이해에만 관심이 있다면 선택형 측정 결과를 사용하는 것이 아무런 문제를 야기하지 않는다. 문제는 선택형 측정결과를 개인간 비교를 위해 사용하고자 할 때 발생한다. 부분적 선택형 측정이 개인간 비교가 가능하다는 것을 보장해 주지는 못하지만 총점의 분산을 발생하게 하고 이 총점과 준거간의 관련성을 보인다면 개인간 비교의 근거로 삼을 수는 있을 것이다(Hicks, 1970).

연구과제

본 연구에서 확인하고자 하는 것은 다음과 같다. 첫째, 입사지원 상황에서 응답왜곡이 어느 정도 일어나는가의 문제이다. 이는 규준적 역량척도와 응답왜곡 탐지척도간의 상관을 통해 일부 확인할 수 있으며, 각 역량에 대해 자신의 있는 그대로에 대한 보고 결과와 지원 상황을 가정했을 때의 응답결과의 차이로 왜곡의 정도를 확인할 수 있다. 둘째, 선택형 형식의 척도를 구성하되 측정된 역량 중 일부만을 점수화 할 경우(부분적 선택형 측정 방식 도입) 선택형 형식의 가장 큰 문제로 지적되

어 온 총점분산의 동일성 문제를 해결할 수 있는가를 확인하고자 한다. 셋째, 선택형 척도의 신뢰도 감소가 이 척도를 이용할 수 없을 만큼 심각한가의 정도를 확인하고자 한다. 만약 최소한의 신뢰도도 확보되지 않는다면 응답왜곡을 방지한다고 하더라도 선택형 측정방식의 도입을 심각히 재고해 보아야 할 것이다. 마지막으로 선택형 형식에서 이상의 문제가 어느 정도 해결될 경우 규준적 척도와 선택형 척도의 타당도를 비교하고자 한다.

연구방법

예비문항자료수집

긍정응답왜곡 척도

긍정응답왜곡 척도의 제작과정은 다음과 같다. 현직 텔레마케터 35명과 대학생 86명(남 30명, 여 56명)이 조사에 참여하였다. 텔레마케터에게는 회사생활을 할 때를 가정하게 하였고 대학생들에게는 학교생활 상황을 가정하게 하였다. 조사대상자들에게 매우 바람직한 행동이지만 사람인 이상 실제로 행하기는 매우 힘든 행동을 10가지씩 쓰게 하였다. 의미상 유사한 행동들을 제외하고 남은 257개의 행동을 문항화하였다. 모든 문항에서 회사의 일과 관련되는 단어는 일반적인 단어로 수정하였다(예, 상사->윗사람). 리더직 텔레마케터 3명과 연구진 5명이 지나치게 현실성이 없는 문항을 제외하고 50문항을 예비척도에 포함시켰다. 100명의 텔레마케터에게 각각의 행동을 어느 정도 하는지에 대해 5점 척도에 응답하게 하여 그렇다(5점 또는 4점)에 응답한 비율이 15% 미만인 30개 문항을 최종 선정하였다.

Megargee(2004)는 MMPI-2의 희귀반응 척도(infrequency scale) 연구에서 응답비율이 0.0%-5.4%인 문항은 매우 좋고(excellent), 5.4%-10.4%인 문항은 좋고(good), 10.4%-15.4%인 문항은 쓸 만하다고(fair)고 주장하였다. 그는 빈번한 발생의 기준을 15.5%이상으로 보았다. 물론 희귀반응 응답비율에 대한 절대적인 기준은 없다.

핵심역량 선정

텔레마케터에게 필요한 역량자료 수집은 우수텔레마케터(64명), 리더직 텔레마케터(16명)들을 대상으로 한 워크샵과 그룹인터뷰를 통해 이루어 졌다. 우수수행자들의 과업별 성공행동 사례분석에서 시초 52개의 역량이 추출되었다. 52개 역량 각각이 텔레마케팅 업무를 잘 수행하기 위해 어느 정도 관련되는가를 평정하게 하여 3.5점(5점 척도) 이상인 28개 역량을 선정하였다. 집단토의를 통해 28개 역량 중 역량내 행동들이 유사하다고 판단되는 역량들을 통합하여 최종적으로 22개 역량이 도출되었다. 22개 각각의 역량명과 역량 정의 및 대표적인 행동 기술문을 리더직 텔레마케터 34명에게 제시하고 평가와 선발상황에서 반드시 고려해야할 역량을 선택하게 하였다. 조사대상자의 90% 이상(31명)에게서 선택된 6개 역량을 핵심역량으로 선정하였다. 고객지향, 성취지향, 문제해결 역량은 34명 모두, 긍정적 사고와 팀지향성은 32명(94.1%), 자기관리역량은 31명(91.2%)에게서 선택되었다.

역량검사의 예비문항 선정

문항개발은 현직자 인터뷰에서 나온 내용을 주로 이용하였다. 현직자 인터뷰에는 우수텔레마케터 64명(지난 6개월간 수행평가에서 상

위 20%이내)과 리더직 텔레마케터 16명이 참여하였다. 성공행동사례는 입사 후 직무수행과 입사 전 학창시절 상황 2가지로 나누어 수집하였다. 입사 후에 대해서는 직무분석 결과를 토대로 주요 과업들(tasks) 및 맥락수행 상황 각각에서 5가지씩의 성공행동 사례를 기술하게 하였다. 입사 전에 대해서는 현재의 과업수행과 관련된 성공경험 또는 현 과업을 수행하는데 도움이 되었다고 판단되는 경험들을 가능한 한 많이 기술하게 하였다(약 3시간 소요). 성공사례를 내용분석 하여 얻은 역량관련 행동은 2,070개였다.

현직자 인터뷰 외에 추가로 참고한 내용은 이종구(2001)과 Owens(1976)의 전기자료 연구에서 사용된 문항들이다. 본 연구에서 개발하고자 한 역량검사는 대졸 선발용으로 활용하고자 하였기 때문에 검사 실시 대상자의 다수는 직장생활경험이 없는 대졸자이며 소수의 경력자도 포함된다. 이런 이유로 최종 문항은 직무와 무관한 용어로 작성하고자 하였다. 이종구(2001)의 연구에서 사용한 자료의 원천 중 하나는 MBC 다큐멘터리 성공시대의 줄거리를 내용분석 하여 얻은 것이다. 이 프로그램에서 조명한 사람들의 다수는 기업에 종사한 사람이며, 내용 속에는 이들의 성장과정과 좌절극복 과정이 잘 드러나 있으며, 직장생활이나 기업을 일으키는데 지대한 역할을 한 성공행동사례를 담고 있다. 따라서 대상인물들이 보인 행동들에는 직장생활에서의 우수수행 관련 역량행동들이 다수 포함되어 있다. 이 내용분석에서 얻은 문항은 400개였다. Owens(1976)의 전기자료 질문지(114문항)는 대학생의 수행을 예측하기 위해 제작된 것이다. 이 질문지에는 역량행동과 비교적 무관한 성장배경이나 흥미 정보도 포함되어 있으나 대학생활 중에 능력

발휘와 연관된 다양한 활동들이 포함되어 있다. 시초 문항풀에 속한 2,584개 문항 중 중복되는 내용과 텔레마케터 역량과 관련성이 적거나 22개 역량에 포함되지 않은 문항들을 일차로 제외시켰다. 남은 1,256개의 문항(역량당 40-60개)을 실무자 4명과 연구진 8명이 토의를 통해 역량당 25문항 씩 총 550문항을 예비 검사 문항으로 선정하였다.

예비검사

최종문항 선정과 선택형 형식에서 선택집합(option set)을 구성하기 위해 텔레마케터 100명이 550문항의 역량검사를 두 가지 상황에서 평정하였다. 두가지 검사를 수행하는데 약 2시간 정도 소요되었다.

한 번은 자신이 회사에 잘 보여야 하는 지원자라고 생각하고 응답하게 하였다. 표 2는 6개 핵심역량에 속한 문항들을 2개씩 예로 제

시한 것이다.

이 응답결과는 각 문항이 회사에 입사하는데 얼마나 바람직한 특성인지를 파악하여 선택형 척도를 구성할 때 선택 집합내 4개 문항의 바람직성 정도를 유사하게 맞추는데 활용하였다. 이 응답결과를 각 문항의 사회적 바람직성을 의미하는가에 대해서는 논란이 있을 수 있다. 그러나 본 연구에서는 지원자라고 가정된 문항별 응답결과에 회사에서 요구하는 특성이 반영될 수 있다고 판단하였다. 이 결과를 각 문항에 대해 일종의 사회적 바람직성을 평가한 것으로 간주하였다.

다른 한 번은 자기 평가 상황으로 자신의 있는 그대로 솔직하게 응답하라고 주문하였다. 모두 Likert 식 5점 척도를 사용하였다. 예비검사 결과를 토대로 사회적 바람직성이 지나치게 높거나(4.3점 초과) 낮음(3.7점 미만) 문항을 제외시키고 문항-총점 상관의 양호한

표 2. 역량별 문항 예와 사회적 바람직성 평균

역량	문항	M(SD)
고객	다른 사람들이 좋아하는 스타일을 빨리 캐치해내는 편이다	4.10(.79)
고객	복잡하거나 어려운 내용을 타인의 눈높이에 맞춰서 쉽게 설명	4.07(.71)
성취	내가 맡은 일에 대해서 확실하게 매듭을 짓는다	4.12(.63)
성취	친구나 동료들이 해결하지 못하는 문제에 도전하는 것을 좋아한다	4.07(.68)
문제	문제의 원인을 파악하기 위해서 다양한 정보를 활용한다	4.05(.49)
문제	기존의 문제점을 보완할 수 있는 방법들을 찾아내는 일을 잘 한다	4.10(.66)
긍정	매사 잘될 것이라는 확신을 가지고 일을 한다.	4.10(.62)
긍정	사소한 일에서도 일의 의미를 찾고자 열정적으로 일을 한다	4.05(.49)
팀	개인보다는 내가 속한 조직/집단을 우선시 한다	4.07(.64)
팀	함께 일하는 동료가 어려운 일을 당하면 그냥 지나치지 않는다	4.05(.54)
자기	긴장상황에서 마음을 다스려 긴장을 풀 수 있다	4.07(.68)
자기	마음을 잘 조절하여 하기 싫은 일이라도 계속해 나가는 편이다	4.07(.71)

주) 고객-고객지향, 성취-성취지향, 문제-문제해결, 긍정-긍정적사고, 팀-팀지향성, 자기-자기관리

(40 이상) 문항을 중심으로 역량 당 14문항이 최종 선정되었다. 정해진 문항 제외기준은 없으나 4.3점을 초과하는 문항을 제외시킨 이유는 역량별 문항들의 평정치 평균이 다소 차이가 있어 지나치게 점수가 높은 문항이 있으면 비슷한 점수대의 선택집합을 구성하기 어려웠기 때문이다.

검사대상자

국내 L기업의 고객센터에 근무하는 여성 텔레마케터 272명이다. 근무기간의 평균은 1년 8개월이다.

측정도구 및 절차

역량검사

본 검사는 22개 역량 당 14문항으로 총 308개 문항에 대해 3가지 형식으로 검사를 구성하였다. 표준적 형식은 4점 Likert 방식의 척도이며 22개 역량을 고루 섞어 제시하였다. 선택형 형식은 2가지로 구성하였다. 하나의 선택집합에는 4개 문항(모두 다른 역량)이 포함되어 자신과 가장 부합되는 내용과 가장 부합되지 않는 내용 2개를 선택하게 되어 있다. 첫 번째 구성방식은 무작위 선택형 형식(random ipsative measure)으로 77개의 선택 집합 구성을 무작위로 하였다. 선택집합의 두 번째 구성방식은 하나의 선택집합에 재고자 하는 핵심역량 1개 문항과 다른 역량 3개 문항을 포함시키는 방식이다. 따라서 선택 집합내에서 6개의 핵심역량 간에는 경쟁이 없고 핵심역량과 다른 역량간의 경쟁만 있어 핵심역량을 고정시키고 다른 역량 문항을 무선적으로 배치한다는 의미에서 이 방식을 본 연구에서

는 고정 선택형 측정(fixed ipsative measure)이라고 명명하였다. 예컨대 고정 선택형 측정에서는 고객지향과 성취지향이 핵심역량이면 하나의 선택 집합에 이 두 가지 역량이 동시에 포함되지 않는다. 단 모든 선택집합에서 핵심역량 1개와 다른 역량 3개를 경쟁시키려면 핵심역량이 6개이면 비핵심역량은 18개가 있어야 한다. 본 연구에서 사용된 역량은 모두 22개로 2개가 모자라게 된다. 따라서 고정선택형이라고 하더라도 각 역량당 10개 문항은 완전히 고정선택형으로 구성하였고 4개 문항은 선택집합내에 다른 핵심역량이 포함되어 있다. 하나의 선택 집합에 포함된 4개의 문항의 사회적 바람직성 정도는 최대한 유사하게 되도록 문항을 배정하였다. 각 선택 집합에 포함된 4개 문항의 사회적 바람직성(지원 상황을 고려한 응답 점수) 정도가 정확하게 같지는 않다. 여기서 최대한이라는 용어를 쓴 이유는 각 역량별로 14개 문항의 사회적 바람직성 정도를 1위부터 14위까지 나열한 후 최대한 같은 순위끼리 경쟁하는 선택 집합이 많게 구성하였기 때문이다. 같은 순위끼리의 선택집합을 구성하기 어려운 상황에서는 차 순위의 문항을 포함시켰다. 표준적 검사에는 308문항과 긍정왜곡 척도 30문항이 포함되어 실시되었다. 피검사자가 한 가지 형식의 검사를 완성하는데는 약 40분 정도 소요되었다.

긍정응답왜곡 척도

예비연구에서 확정된 30문항으로 이루어진 척도를 사용하였다.

준거

최근 6개월간의 평균 수행평가점수이다. 5등급(S, A, B, C, D)의 수행평가는 1개월 단위

로 산출되는데 이를 5점에서 1점 사이의 점수로 바꾸어 적용하였다. 수행평가 항목으로는 전화응대의 수, 응대 품질, 고객평가, 업무숙지도 및 상사의 역량평가(5단계의 BARS 척도) 등이 있다.

연구결과

표 3은 표준적 척도에서 긍정왜곡점수와 2가지 역량점수(입사 지원자를 가상한 역량평가와 자기평가) 간의 상관과 2가지 상황에서 평가한 역량점수를 비교한 결과이다(100명의 예비검사 결과). 상관분석 결과, 역량점수와 긍정왜곡 점수간 상관은 자기평가보다 지원상황을 가상했을 때 더 높게 나타났다. 가상평가(지원상황)와 자기평가간의 평균차이에 대한 t 검증 결과 가상평가가 자기평가보다 모든 역량에서 유의하게 높게 나타났다. 이러한 결과는 표준적 척도의 경우 일반적인 자기평가 상황에서도 응답왜곡이 일어나지만 입사를

가정했을 때 응답왜곡 경향성이 현저하게 더 높아질 가능성이 있음을 보여주는 결과이다.

표 4는 부분적 선택형 형식을 사용했을 때 개인차가 충분히 드러나는 지를 확인하기 위해 표준적 형식과 2가지 부분적 선택형 형식간 변이계수(coefficient of variability)를 이용하여 표준편차를 비교하기 위한 결과이다. 변이계수(coefficient of variability: CV)는 $100 \times (\text{표준편차} / \text{평균})$ 으로 평균대비 표준편차의 비율이다. 이 수치는 측정치의 단위가 다른 변수들의 표준편차를 비교하기 위해 사용된다. 단, 척도의 수준이 절대영점이 있는 비율척도일 때 가능하다. 물론 본 연구에서 사용한 척도가 비율척도는 아니다. 변산의 존재여부나 크기가 중요한 이슈이지만 비교할만한 마땅한 지수가 없고, 척도의 값이 영점에서 시작하여 이 지수를 이용하기로 하였다.

표준적 형식의 경우 4점 척도(0점에서 3점 사이)이고 2가지 선택형 측정방식의 경우 3점 척도(0점에서 2점 사이)에 해당한다. 변이계수를 통한 표준편차의 비교 결과, 6개의 개별 역량의 경우 2가지 부분 선택형 형식은 표준적 형식보다 변이계수가 다소 낮게 나타났다. 무선 선택형의 경우 표준적 형식과 큰 차이를 보이지 않았으나 고정선택형의 경우 고객지향 역량을 제외하고는 대부분 표준적 형식과 무선 선택형에 비해 다소 낮게 나타났다. 변이계수의 최대 폭을 0에서 표준적 측정치 사이로 본다면 2가지 선택형은 0보다는 표준적 측정치에 가깝기 때문에 하위 역량 점수를 이용하여 개인차를 파악하고자 하는데는 큰 문제가 없을 것으로 보인다. 총점에 대한 변이계수 또한 표준적 형식에 비해 무작위 부분 선택형과 고정 선택형 모두 다소 낮게 나타났으나 이 또한 0보다는 표준적 측정치에 가깝기

표 3. 가상평가와 자기평가 비교

	긍정응답왜곡		가상 평가 r1	자기 평가 r2	가상 평가 M (SD)	자기 평가 M (SD)	t
	가상 평가	자기 평가					
고객지향	.65*	.33*	4.09(.41)	3.65(.45)	5.89*		
성취지향	.64*	.41*	3.94(.39)	3.51(.44)	6.13*		
문제해결	.58*	.39*	4.03(.39)	3.49(.45)	7.32*		
긍정적사고	.72*	.48*	3.93(.37)	3.52(.46)	6.16*		
팀지향성	.60*	.55*	4.00(.41)	3.70(.45)	4.47*		
자기관리	.63*	.39*	4.06(.36)	3.46(.42)	8.45*		
전 체	.65*	.44*	4.02(.36)	3.57(.41)	6.91*		

* p<.001

표 4. 검사형식별 평균, 표준편차 및 변이계수

	변이계수(100*(표준편차/평균))					
	규준적	무선선택형	고정선택형	규준적	무선선택형	고정선택형
고객지향	1.61(.49)	0.99(.29)	0.96(.27)	30.19	29.29	28.14
성취지향	1.64(.48)	1.12(.26)	1.18(.25)	29.33	23.21	21.20
문제해결	1.44(.52)	0.88(.26)	0.99(.22)	35.92	29.55	22.22
긍정적사고	1.60(.53)	0.97(.29)	1.19(.24)	33.28	29.90	20.18
팀지향	1.85(.47)	1.23(.31)	1.18(.28)	25.62	25.20	23.73
자기관리	1.49(.55)	0.93(.30)	0.97(.29)	36.87	32.26	29.90
역량총점	1.60(.44)	1.02(.24)	1.08(.18)	27.54	23.53	16.67

때문에 총점의 활용가능성이 있다고 판단할 수 있다. 측정형식에 관계없이 총점에 대한 변이계수가 개별 역량에 대한 계수보다 다소 낮게 나타났다.

표 5는 규준적 형식, 2가지의 부분 선택형 형식과 긍정응답왜곡간의 상관분석 결과이다. 그 결과, 규준적 형식에서 측정된 모든 역량은 응답왜곡과 유의한 정적 상관을 보인 반면, 2가지 부분 선택형 형식으로 측정된 모든 역량은 응답왜곡과 유의한 상관을 보이지 않았다.

본 연구에 참여한 현직자들의 경우 매월 역

량평가를 받아 자사의 핵심역량을 파악하고 있음에도 불구하고 선택형 형식에서 응답왜곡이 나타나지 않아 지원자가 지원 회사의 핵심역량을 파악하고 응답한다고 하더라도 선택형 형식에서는 응답왜곡이 쉽지 않을 것으로 보인다. 만약 하나의 선택집합 속에 지원하는 회사의 핵심역량 하나와 다른 역량 3개가 들어 있으며 응답자가 지원하는 회사의 핵심역량을 알고 있다면 의도적 응답왜곡이 가능할 수 있다. 그러나 본 연구에서 고정 선택형으로 측정해도 핵심역량점수들이 긍정응답왜곡 점수와 유의한 상관을 보이지 않아 선택집합을 어떤 식으로 구성해도 선택형 형식에서는

표 5. 검사형식별 응답왜곡척도와의 상관

	규준적	무선 선택형	고정 선택형
고객지향	.28*	.01	-.04
성취지향	.28*	.04	.10
문제해결	.25*	-.03	-.00
긍정적 사고	.24*	.06	.02
팀지향성	.34*	.10	-.05
자기관리	.26*	-.00	-.08

* p<.01

표 6. 검사형식별 내적 신뢰도

	규준적	무선 선택형	고정 선택형
고객지향	.79	.64	.64
성취지향	.86	.60	.66
문제해결	.90	.49	.43
긍정적 사고	.84	.55	.57
팀지향성	.81	.76	.58
자기관리	.85	.57	.70

표 7. 준거와의 상관

	고객 지향	성취 지향	문제 해결	긍정 사고	팀 지향	자기 관리	역량 평균
규준적	.07	-.01	-.01	.04	.07	.11	.18**
무선 선택형	.11	.01	.16*	.04	.09	-.03	.23***
고정 선택형	.20**	.04	.06	-.04	.17*	.12†	.31***

+ p<.10, * p<.05, ** p<.01, *** p<.001

응답왜곡 가능성이 거의 없다고 간주할 수 있을 것이다.

표 6은 각 검사 형식별 하위 역량의 내적일관성 계수(Cronbach's alpha)이다. 검사형식별 신뢰도 분석 결과 2가지 선택형 형식의 신뢰도가 다소 낮게 나타났으나 문제해결을 제외한 나머지 역량의 신뢰도는 수용할 만한 것으로 나타났다.

표 7은 검사형식별 6개의 역량과 준거와의 상관분석 결과이다. 전체적으로 준거와의 상관성이 높지는 않지만 선택집합을 구성할 때 핵심역량간 경쟁을 배제시킨 고정선택형이 다른 형식에 비해 더 높은 타당도를 보였다.

논 의

응답왜곡을 줄임과 동시에 개인간 변별 및 신뢰도와 타당도가 확보된 역량검사를 제작하기 위해 선택형 측정방식을 도입한 본 연구의 결과를 요약하고 논의하면 다음과 같다.

첫째, 규준적 척도를 사용하여 입사지원을 가정하고 응답했을 때가 솔직한 응답을 요청한 자기평가보다 모든 역량에서 높은 점수를 보였다. 또 입사지원 상황은 물론 솔직한 응답을 부탁한 자기평가도 긍정응답왜곡 척도와

모두 유의한 상관을 보였다. 이 결과는 직업을 얻으려 동기화된 지원자의 긍정적 응답왜곡이 재직자보다 더 크다는 Schmitt와 Oswald (2006) 그리고 Jackson 등(2006)의 연구결과와 일관된다. 현직자의 자기평가가 긍정응답왜곡 척도와 정적 상관을 보인 결과는 규준적 척도일 경우 사회적 바람직성에 의한 긍정왜곡 경향을 제거하기는 어렵다는 점을 시사한다.

둘째, 규준적 측정 방식의 경우 긍정적 응답왜곡과 유의한 상관을 보인 반면, 2가지 부분적 선택형 측정 결과는 긍정적 응답왜곡과 유의한 상관을 보이지 않는 것으로 나타났다. 이 결과는 무선적인 선택형이 아닌 고정 선택형 측정방식을 사용하더라도 응답왜곡만큼은 충분히 방지할 수 있음을 의미한다. 22개 역량 중 6개의 역량 점수만을 이용하고 또 평가되는 핵심역량이 무엇인지를 잘 알고 있는 현직자에게서 나온 이 결과는 선택집합을 구성할 때 어떤 식으로 조합하더라도 응답왜곡이 쉽지 않음을 시사한다.

셋째, 검사의 구성형식별 개인차 측정이 가능한가의 문제는 변이계수를 토대로 검토하였다. 완전 선택형의 경우 총점의 분산(variance)이 영이어서 개인차를 전혀 측정할 수 없다. 본 연구결과 역량별 변이계수는 규준적 형식이 2가지 선택형 총점보다 다소 더 크게 나타

났으며 총점의 경우도 마찬가지로 나타났다. 비록 규준적 측정치에는 다소 못 미치지만 선택형 측정의 변이 계수가 영보다는 규준적 측정치에 가깝게 나타나 총점활용도 가능한 것으로 판단된다. 물론 큰 변이계수가 선택형 형식의 개인차 측정의 문제를 해결해주는 것은 아니다. 그러나 부분적 선택형이 총점 분산이 영으로 되는 문제를 어느 정도 해결해 주고 있음은 물론 보다 높고 안정적인 타당도가 확보된다면 사용을 권고해볼만 한 것으로 판단된다.

넷째, 신뢰도와 타당도에 대한 검토 결과 선택형 측정은 모든 하위역량에서 규준적 측정정보보다 신뢰도가 떨어지는 것으로 나타났다. 단, 낮아진 신뢰도가 선택형 검사형식의 도입을 완전히 백지화 할 만큼은 아니다. 물론 신뢰도가 아주 낮은 일부역량의 문항은 앞으로 수정되어야 할 필요가 있다.

다섯째, 타당도에 대한 검토 결과, 고정 선택형 측정방식이 규준적 측정정보보다 더 높은 것으로 나타났다. 전반적인 타당도는 유의했으나 유의하지 않은 개별역량이 많아 어느 측정방식도 만족할 만한 수준에 이르지 못한 것으로 보인다. 그 이유는 역량이나 성격척도의 직접적인 수행예측력이 낮은 탓일 수도 있으며, 역량 측정에 다소 부적절한 문항이 포함되었기 때문일 수도 있으며, 본 연구의 수행 준거가 매우 복합적인데 기인했을 수도 있다. 본 연구에서의 수행준거는 회사에서 제공한 고과등급으로 다양한 항목들이 종합된 것이다. 추후 연구에서 세분화된 수행준거별로 역량들의 예측력을 검토해볼 필요가 있다. 다소 부적절한 문항의 개입가능성에 대해 추후에 문항총점상관과 준거와의 관계를 고려하여 역량 당 14문항에서 10문항 또는 7문항으로 줄

인 후 준거와의 상관을 구했을 때 타당도가 상당히 증가되었다(모든 역량이 준거와 유의). 그러나 이러한 결과는 추가자료 수집을 통해 교차타당화가 이루어져야 받아들여 질 것으로 보인다.

본 연구를 통해 부분적 선택형 형식의 도입이 응답왜곡을 방지함은 물론 규준적 척도보다 다소 더 높은 타당도를 얻어내 응답왜곡에 대한 하나의 대안적 방법일 수 있음을 확인하였다. 그러나 총 77집합 308문항 중 채점되는 문항은 6개 역량(역량 당 14문항)의 84문항으로 224문항은 때움질(filler) 문항으로 사용한 셈이 된다. 또한 역량 내에서 소수의 문항만을 사용한 경우 채점되는 문항의 비율은 더 낮아지게 된다. 부분적 선택형 형식의 연구에서 채점되는 문항수가 전체문항에서 어느 정도이어야 하는가에 대한 가이드라인은 없다. 향후 다양한 시뮬레이션을 통해 사용되는 역량 또는 문항 비율과 총점 변량간의 관계 또는 타당도에 관한 가이드라인이 제시되어야 할 것이다. 또 전체 역량 중 일부 역량을 사용하는 부분선택형이 아닌 선택집합의 수를 본 연구(4개)보다 더 늘리는 방안도 강구해볼 필요가 있다.

본 연구에서 응답왜곡에 대한 인사담당자의 불신을 해소하기 위한 대안으로 선택형 척도의 다양한 측면을 고찰하였으나 그럼에도 불구하고 규준적 척도가 응답왜곡에 그다지 민감하지 않다는 연구결과도 꾸준히 보고되고 있다. Hogan, Barrett 및 Hogan(2007)의 연구에서 동일한 직무에 재지원하고 같은 성격검사에 반복 응답한 5,266명에 대한 분석 결과 5.2%이하만이 점수가 이전보다 향상된 것으로 나타났으며 두 번의 검사에 걸쳐 극단적인 점수변화를 보인 사람은 극소수(3명, .06%)인

것으로 나타났다. 이러한 결과를 토대로 이들은 성격검사에서의 응답왜곡이 실제선발 상황에서 별 문제가 되지 않는 것으로 결론지었다. 본 연구에서도 예비검사결과에서 긍정응답왜곡과 역량점수간의 상관성이 매우 높게 나타났다(표 3 참조) 본 검사에서는 이보다 낮은 상관성이 관찰되었다(표 5 참조). 현직자에게 지원상황을 가장하고 응답할 때 다소 과도한 과장이 나타난다면 응답왜곡과의 상관 또한 과장되었을 가능성이 있다.

물론 Hogan 등(2007)의 연구에서 반복측정시 검사점수의 변화가 적다는 것 자체를 응답왜곡이 일어나지 않았다고 해석한 것에 대해 반론을 제기할 수 있다. 지원자들이 시초 응답에서 상당한 긍정왜곡을 하였다면 두 번째 응답에서 긍정왜곡을 하였다더라도 별다른 점수 변화가 관찰되지 않을 수도 있다. 이와는 달리 첫 번째 지원에서 탈락한 지원자가 두 번째 지원에서 동기화된다면 검사-재검사간 상당한 점수차이를 보일 가능성이 있다. 만약 두 번째 지원에서 동기화된 지원자임에도 점수차이를 보이지 않았다면 규준적 척도가 응답왜곡에 민감하지 않다는 결론을 수용할 수 있을 것이다. Hogan 등(2007)의 결론에 대한 수용여부에 관계없이 이들의 연구에서와 같이 실제 지원자에 대한 자료를 통해 규준적 척도의 응답왜곡 문제를 검토해야 할 필요가 있을 것이다.

요약하면, 역량별 신뢰도는 규준적 형식이 두 가지 선택형 형식보다 우수한 것으로 나타난 반면, 타당도는 선택형 형식이 규준적 보다 다소 더 높은 것으로 나타났다. 선택형 형식에서 약간의 신뢰도 감소가 있었으나 어느 정도 수용할 만하기 때문에 보다 타당도가 높게 나타난 선택형 형식의 도입을 고려해 볼만

하다. 문항 수를 줄였을 때의 타당도 증가에 대한 교차타당화가 이루어진다면 사용된 모든 문항을 이용하기 보다는 변산이 다소 작더라도 주요 문항을 선정해 선발에 이용하는 것이 보다 합리적일 것으로 보인다.

참고문헌

- 김명소, 이현주 (2006). 성격검사의 형식이 응답왜곡(faking)에 미치는 효과: Normative 형식과 Ipsative 형식의 비교. *한국심리학회지: 산업 및 조직*, 19(3), 371-393.
- 이은정, 박동건 (2003). 성격검사의 응답왜곡 탐지: 직무바람직성의 개념화 및 선발 결정에의 영향. *한국심리학회지: 산업 및 조직*, 16(2), 121-151.
- 이종구 (2001). 전기자료(Biodata)의 타당화 연구: 보험 판매사원의 수행예측. *한국심리학회지: 산업 및 조직*, 14(3), 83-107.
- 이종구, 한영석, 신강현, 허창구, 박지훈, 최우성 (2007). 인성검사의 Ipsative와 Normative 형식과 Faking간의 관계. *한국 산업 및 조직 심리학회 춘계 학술대회 발표논문집*, 108-109.
- Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology*, 69(1), 49-56.
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Barrick, M. R., & Mount, M. K. (1996). Effect of impression management and self-deception on the predictive validity of personality

- constructs. *Journal of Applied Psychology*, 81(3), 261-272.
- Bowen, C., Martin, B. A., & Hunt, S. T. (2002). A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires. *The International Journal of Organizational Analysis*, 10(3), 240-259.
- Doll, R. E. (1971). Item susceptibility to attempted faking as related to item characteristic and adopted fake set. *Journal of Psychology*, 77, 9-16.
- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, 84(2), 155-166.
- Griffith, R. L., Chmielowski, T., Snell, A. F., Frei, R. L. & McDaniel, M. A. (2000). Does faking matter? An examination of rank order changes in applicant. *Paper presented at the 15th annual meetings of the Society at Industrial and Organizational Psychologists*, New Orleans, LA.
- Guion, R. M. (1965). *Personnel testing*(ch. 13). McGraw-Hill.
- Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, 74, 167-184.
- Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92(5), 1270-1285.
- 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.
- Hurd, J. M., Barrett, G. V., Miguel, R. F., Tan, J. A., & Lueke, S. B. (2001). When do response distortion scales reflect faking? A meta-analysis. *Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology*, San Diego, CA.
- Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced-choice offer a solution? *Human Performance*, 13(4), 371-388
- Johnson, C. E., Wood, R., & Blinkhorn, S. F. (1988). Spuriouser and spuriouser: The use of ipsative personality tests. *Journal of Occupational Psychology*, 61, 153-162.
- Megargee, R. A. (2004, May). *Developmental and initial validation of an MMPI-2 infrequency scale (Fc) for use with criminal offenders*. Paper presented at the 39th Annual Symposium on Recent Developments on the MMPI-2 /MMPI-A, Minneapolis, MN.
- Moffett, R. G. (1996). *Relationship between the big five personality factors and biodata factors*. Doctoral Dissertation. University of Auburn, Alabama, USA.
- Owens, W. A. (1976). Background data. In M. D. Dunnette(Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally.
- Rossé, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of

- response distortion of preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83(4), 634-644.
- Saville, P., & Willson, E. (1991). The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology*, 64, 219-238.
- Schrader, A. D. & Osburn, H. G. (1977). Biodata faking: Effects of induce subtlety and position specificity. *Personnel Psychology*, 30, 395-404.
- Vinchur, A. J., Schippmann, J. S., Switzer, III, F. S., & Roth, P. C. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology*, 83(4), 586-597.
- Worthington, D. L., & Schlottmann, R. S. (1986). The predictive validity of subtle and obvious empirically derived psychology test items under faking conditions. *Journal of Personality Assessment*, 50, 171-181.
- Zickar, M. J., Rosse, J. G., Levin, R. A., & Hulin, C. L. (1996). Modeling the effects of faking on personality tests. *Paper presented at the 11th annual meeting of the Society for Industrial and Organizational Psychology*, San Diego, CA.
- 1차 원고접수 : 2008. 3. 28
2차 원고접수 : 2008. 7. 2
3차 원고접수 : 2008. 8. 16
최종게재결정 : 2008. 8. 18

**The reliability and validity of tele-marketer selecting
competency tests in different formats:
A comparison between Ipsative and Normative Measures**

Lee, Jong Goo

Department of Psychology
Daegu University

Han, Young Seok

Competency Development Center
Central Official Training Institute

The purpose of this study was to make three types of competency tests - Normative, Random Ipsative, and Fixed Ipsative - to prevent response distortion in self-rating scales and then compare their reliability and validity. Participants were 272 female tele-marketers working at L company in Korea. As a result of correlation analysis for three types of test formats with a faking scale, coefficient of correlation between normative format and the faking scale was positively significant in every subscale, while those between two types of ipsative formats and the faking scale were lower than .10 except just some subscales. The degree of reliability in both ipsative formats was a little lower than that in normative format, but both ipsative formats had acceptable reliability except one subscale. The degree of validity for ipsative formats ($r = .23$, $r = .31$) was higher than that of normative format ($r = .18$). From these results in which the subscales in ipsative formats had no faking effects and reasonable reliability and validity, competency tests in ipsative format seem to be applied more broadly in the human resource management field.

Key words : tele-marketer, normative measure, ipsative measure, faking good, competency