한국심리학회지: 산업 및 조직
Korean Journal of Industrial and Organizational Psychology
2003. Vol. 16, No. 2, 121-151

© 한국 산업 및 조직심리학회 2003

성격 검사의 응답왜곡 탐지: 직무 바람직성의 개념화 및 선발 결정에의 영향

이 은 정 박 동 건†

고려대학교 심리학과

유능한 인재를 선발하기 위한 방법으로 성격검사를 사용하여 얻을 수 있는 이점들이 많다. 그러나 인사 선발과 같이 사람들이 긍정적인 인상을 보이기 위해 노력하는 장면에서, 응답왜곡의 지표로서 사회적 바람직성은 중요한 논쟁거리가 되어 왔으며 이에 대해서는 여러 가지 주장들이 존재한다. 이러한 맥락에서, 본 연구는 사회적 바람직성이 응답왜곡의 타당한 지표가 될 수 있는가라는 근본적인 물음에서 출발한다. 이에 따라, 본 연구는 첫째, 선발장면에서 나타나는 성격검사에 대한 응답왜곡의 지표로서 사회적 바람직성 측정 도구의 적절성을 검토하고, 둘째, 서로 다른 직무(운항 승무원, 생활 설계사)에서 동일한 성격검사에 대한 응답왜곡이 다르게 나타나는지를 확인하고, 이러한 직무 바람직성이 사회적 바람직성과 구별되는 개념인지를 조사하는 것을 목적으로 한다. 이에 추가적으로, 선발장면에서의 의도적인 응답왜곡이 성격검사의 준거관련 타당도에는 그다지 큰 영향을 미치지 않으나, 채용 결정에 변화를 야기한다는 최근의 연구 결과들과 맥을 같이하여, 본 연구에서 구성 타당도가 확인된 직무바람직성 개념을 근거로 직무별로 다르게 나타나는 응답왜곡이 선발장면에서의 채용 결정에 미치는 영향을 살펴보았다. 결과에 따르면, 첫째, 기존에 응답왜곡의 지표로 흔하게 사용하고 있는 사회적 바람직성 척도들에는 응답왜곡을 적절히 탐지하지 못하는 문항들이 상당수 포함되어 있음이 나타났다. 둘째, 성격검사 척도 자체 내에 응답왜곡에 민감하게 반응하는 문항들이 포함되어 있음이 나타났다. 셋째, 응답왜곡에 민감하게 반응하는 문항들은 직무별(운항승무원 직무와 생활설계사 직무)로 어느 정도 차이를 보임으로써 특정 직무에 따라 바람직한 방향으로 응답왜곡이 일어나는 직무 바람직성이 존재하는 것으로 보인다. 넷째, 이러한 직무 바람직성은 기존의 사회적 바람직성과 구별되는 구성개념이다. 다섯 째, 직무바람직성에 영향 받는 문항들의 효과를 제거한 검사점수를 사용할 경우, 피검자들 사이의 서열에 변화가 나타나서 누가 채용되는 지의 채용 결정이 바뀌는 것으로 나타났다.

주요어 : 성격검사, 응답왜곡, 직무바람직성, 선발결정

† 교신저자 : 박 동 건, 고려대학교 심리학과, 02-3290-2066, sykhpark@korea.ac.kr

1990년 대 초 이래로 성격 측정 도구들과 정직성 검사(integrity testing)는 산업/조직 장면에서 많이 사용되어 왔으며(Hough & Schneider, 1996), 조직 장면에서 성격 척도의 사용에 관한 문헌 및 연구들도 상당하다. Hough, Eaton, Dunnette, Kamp 및 McCloy(1990)는 성격 척도들이 다양한 직무 관련 행동을 예측하는 데 타당하며, 성격 척도들의 준거 관련 타당도는 바람직한 방향으로 응답을 왜곡하고자 노력한 사람들에 의해서도 절하되지 않음을 대규모 자료를 사용하여 밝혔다. 그리고 여러 연구들은 성격 측정 도구들이 상급자 평가를 포함한 직무상의 주관적/객관적 성과 지표들 이외에, 맥락 수행, 약물 중독, 작업장에서의 폭력과 같은 다양한 준거를 예측한다는 사실을 밝혔다(Ones, Viswesvaran & Schmidt, 1993; Hough et al., 1990; Ones, Viswesvaran & Reiss, 1996 등). 또한, Barrick과 Mount(1996)는 통합 분석을 통해 성실성 및 외향성과 다양한 직무 관련 수행 준거간의 작지만 일관적인 관계를 발견했다. 비록 이러한 관계가 전형적인 인지 능력 검사에 비해 작지만, 성격은 인지 능력과는 매우 다른 인간 특성을 측정한다는 점에서 인지 능력 측정치와 함께 사용할 때 증분 타당도를 이룰 수 있을 것이다(Zickar & Drasgow, 1996).

이와 같이 성격 검사들을 인사 선발에서 사용하는 것의 유용성에 관한 많은 연구 결과들에도 불구하고, 여전히 인사 선발에서 성격 측정치를 사용하는 데에 대한 우려의 목소리가 있다. 특히, 직무 지원자들이 사회적으로 바람직한 방향으로 자신의 응답을 왜곡한다는 것이 성격 검사와 정직성 검사를 실제 장면에서 사용하는 데 문제가 된다는 주장이 많다(Gatewood & Feild, 1994; Gomez-Mejia, Balkin, & Cardy, 1995; H. G. Heneman & R. L. Heneman, 1994; Scarpello, Ledvinka, & Bergmann, 1995). 즉, 지원자들이 성격 검사의 문항에 답할 때 의도적으로 진실을 과장하거나 거짓 응답을 함으로써 진짜 자신보다 더 긍정적으로 보이려고 할 수 있다는 것이다. 그 결과, 심리 측정 척도의 예측 타당도를 떨어뜨리고 선발 결정을 바꿀 수 있으며, 그로 인해 조직에서 가장 적격의 사람들을 고용할 수 있는 가능성이 감소할 수 있다(Dunnette, McCartney, Carlson & Kirchner, 1962; Ellingson, Sackett & Hough, 1999).

인사 선발과 같이 사람들이 긍정적인 인상을 보이기 위해 노력하는 장면에서, 지원자들의 의도적인 응답 왜곡과 관련된 이러한 문제들을 극복하기 위한 노력으로 연구자 및 실무자들은 다음과 같이 여러 전략을 사용해왔다: (1) 응답 왜곡에 대해 경고하는 것, (2) 바람직성이나 구성개념이 미묘한 것으로 밝혀진 문항들을 사용하는 것(예를 들어, 취미나 좋아하는 것 등을 질문), (3) Minnesota Multiphasic Personality Inventory(MMPI; Dahlstrom, Welsh & Dahlstrom, 1972)의 L척도와 같이 타당화 척도를 사용한 왜곡에 대한 교정, (4) 특정한 문항 형식(예를 들어, forced choice 방식)을 사용하는 것. 1950년 대 이래로, 사회적 바람직성 척도와 같은 타당화 척도들을 사용하여 응답 왜곡을 탐지하는 방법은 이러한 전략 들 중에서도 대표적으로 사용되어 왔다(Ones et al., 1996). 실제로 많은 성격 검사들은 응답 타당도 척도를 포함한다. 예를 들어, Eysenck Personality Inventory(Eysenck & Eysenck, 1968), 형용사 체크리스트(Adjective Checklist; Gough & Heilbrun, 1980), California Personality Inventory(CPI; Gough, 1975), MMPI(Dahlstrom, Welsh, & Dahlstrom, 1972, 1975), Personality Research Form(Jackson, 1967), Assessment of Background and Life Experiences (Hough, et al., 1990) 모두 응답 타당도 척도를 포함하고 있다.

이러한 응답 타당화 척도들은 많은 성격 검사들에 포함되어 의도적인 응답왜곡의 지표로 사용되어 왔으며, 자기보고식 성격검사에서의 응답왜곡을 다룬 선행 연구들은 전형적으로 이러한 성격 검사 타당화 척도(예를 들어, Crowne & Marlowe, 1960; Meehl & Hathaway, 1946; Paulhus, 1984)를 사용하여 성격검사의 내적 혹은 외적 타당도 위협을 다뤘다. 그러나 그러한 타당화 척도가 실제로 응답왜곡을 적절히 탐지해 낼 수 있는 지에 대한 측정 적절성에 관심을 둔 연구는 그리 많지 않다.

또한 이러한 응답 타당화 척도들의 대부분은 사회에서 일반적으로 바람직하다고 여겨지는 것들, 혹은 여러 직무들에 걸쳐 전반적으로 바람직하다고 여겨지는 경향성들에 초점을 두었다. 즉, 지원자들이 자신의 응답을 지원하는 직무에 따라 다르게 왜곡할 수 있다는 가능성을 간과하는 경향이 있었던 것이다. 성격검사의 준거 관련 타당도 및 구성타당도에 대해 연구한 무수한 기존 연구들 또한, 성격검사에 대한 응답왜곡이 직무별로 다르게 나타날 수 있으며, 따라서 응답왜곡이 성격검사 준거관련 타당도 및 구성타당도에 미치는 영향 또한 직무별로 다를 수 있음을 간과하였다. 그러나 비록 성격 검사가 아닌 생활사 검사를 통해서이지만, 응답편의를 예측하는 데 사회적 바람직성이 아닌 직무 바람직성이 보다 우월하다는 결론을 내린 Kluger와 Colellar(1993)의 연구를 볼 때, 직무에 따라 다르게 나타나는 응답왜곡을 살피는 것은 가치 있는 작업일 것이다.

응답 왜곡의 문제를 극복하고자 하는 노력들과 그에 대한 논의가 진행되고 있는 한편, Hough와 Schneider(1996)는 그들의 저서에서, 그동안의 연구들을 살펴볼 때 사회적 바람직성은 성격과 수행간의 관계에 있어서 신화적인 조절변인(mythical moderators)이며, 사람들은 지시받았을 경우 응답왜곡을 할 수 있으나 실제 지원 장면에서의 왜곡의 정도는 지시받았을 경우의 왜곡의 정도만큼 크지 않다고 결론내릴 수 있다고 했다. 그러나 보다 최근에 Ross, Stecher, Miller 및 Levin (1998)은 성격 검사의 예측 타당도에는 의도적인 응답 왜곡이 그다지 큰 영향을 주지 않을 수 있으나, 실제 채용 상황에서 누가 고용될 지의 채용 결정에 영향을 줄 수 있다는 그들의 연구 결과를 통해, 응답왜곡이 실제 채용 상황에서 간과할 수 없는 문제임을 보였다. 성격검사에 대한 응답왜곡의 효과를 조사한 대부분의 연구들이 응답왜곡이 성격검사의 예측 타당도 및 구성타당도에 미치는 영향을 조사한 것이나, 실제 인사선발 상황에서의 관심은 응답왜곡이 채용결정에 미치는 영향일 것이다. 따라서 Ross 등(1998)의 이러한 연구 결과들을 보다 심도있게 살펴보는 것이 필요할 것이다.

이러한 연구 맥락에서, 본 연구는 (1) 선발장면에서 나타나는 성격검사에 대한 응답왜곡의 지표 및 타당화 척도로서 사회적 바람직성 척도의 측정 적절성을 검토하고, (2) 서로 다른 직무에서 동일한 성격검사에 대한 응답왜곡이 다르게 나타나는 지를 확인하고, (3) 이러한 직무 바람직성이 사회적 바람직성과 구별되는 개념인지를 조사하며, Ross 등(1998)의 연구와 맥을 같이하여 (4) 응답왜곡의 영향을 제거했을 때 선발결정에서 차이가 나타나는 지의 유무와 그 차이의 정도를 직무 바람직성에 근거하여 살펴볼 것이다.

응답왜곡의 탐지와 극복을 위한 전략: 사회적 바람직성 척도

선발상황에서 바람직한 방향으로 자신을 보이고자 하는 응답 왜곡(fake good)을 탐지해내고자 하

는 노력은 대부분 인상관리(impression management)와 자기-기만(self-deception)을 측정하는 사회적 바람직성 척도들을 통해 이루어져왔다(Pauhlus & John, 1998). 예를 들어, Lanning(1989)은 바람직한 방향으로(fake good) 응답왜곡 하도록 지시한 대학생 집단과 일반적인 응답자들을 구별하기 위해, "good impression" 척도의 점수와 CPI 척도의 점수를 사용하여 회귀 방정식을 계산했다. 회귀 방정식을 통해 도출된 점수는 67%의 정확률(hit rate)[1]을 보였으며, 실제로 정직한 응답자들이나 동일한 임계 점수를 사용했을 때 거짓 응답자로 잘못 분류된 비율(false positive rate)이 1%였다. 그러나 두 표본의 조합이 다를 때 회귀 방정식에 근거한 이러한 결과의 일반화는 어렵다. 실제로, 이 연구에서 인상관리 탐지 척도 점수는 대학생 집단과 현직자 집단간의 구분에 있어서도 유사한 정확률을 보였다. 또한, Dwight와 Alliger(1997), 그리고 Kroger와 Turnbull(1975)은 바람직한 방향으로 응답왜곡(fake good)을 하는 방법을 응답자들에게 교육할 경우 정직한 응답자 집단과 보다 유사한 사회적 바람직성 점수를 가짐을 발견했다. 즉, 비록 사회적 바람직성 척도가 응답왜곡 탐지를 위해 설계되었지만 그러한 척도들 자체도 응답왜곡에 민감하게 반응하지 못할 수 있는 것이다. 이러한 연구들을 고려해볼 때, 조직 장면에서 응답왜곡을 탐지해내는 사회적 바람직성 척도들의 효과성은 여전히 명확하지 않다. 그럼에도 현재까지의 연구들은 응답왜곡을 탐지하는데 사회적 바람직성 척도에 과도하게 의존하는 경향이 있다.

1) 응답왜곡 상황에서, 정확률은 탐지 척도에서 특정 임계 점수로 옳게 분류된 응답 왜곡자들의 비율을 의미.

응답왜곡의 탐지와 극복을 위한 대안 전략: 직무 바람직성

고용상황에서의 의도적인 응답왜곡에 대한 대부분의 선행 연구들에서 사용한 사회적 바람직성 척도들은 일반적인 통념이나 가치에서 바람직한 경향성이나 여러 직무 전반에 걸친 바람직성에 초점을 두었으며 특정한 직무들에 따라 바람직성이 다르게 나타날 수 있음은 간과하였다. 그러나 지원자들은 그들 자신이 특정 직무에서 중요하다고 생각하는 특성을 지니고 있다고 보이고 싶을 것이며 이러한 특성은 일반적인 의미에서 사회적으로 바람직하지 않을 수도 있다. 예를 들어 경호 직무에 대한 지원자들은 "나는 내 갈 길을 막는 사람은 밀쳐낸다"라는 문항에 대해 '그렇다'고 답하는 경향이 있을 수 있으나, 그 응답은 사회적으로 바람직하지 않을 수 있다. 이와 유사하게, "나는 토하는 사람을 보면 참기 힘들게 불쾌하다"라는 문항에 대해서 간호 보조 직무 지원자들은 '전혀 그렇지 않다'라고 답하는 것이 그 직무에 가장 적합할 것이며, 실제로 불쾌한 기분이 들더라도 그렇지 않다고 응답하는 경향이 있을 것이다.

이러한 직무 바람직성 응답의 가능성을 Kluger와 Colella(1993)는 학생들을 대상으로 실험 상황에서 생활사 검사를 사용하여 증명했다. 또한 그들은 생활사 문항의 평균과 변량 모두에서 응답편의를 예측하는 데 사회적 바람직성보다 직무 바람직성이 보다 우월하다는 결과를 얻었으며, 그 결과를 통해 지난 연구들이 실제 선발 장면에서 응답을 왜곡하려는 동기에 대한 상관 개념만(즉, 직무 바람직성이 아닌 사회적 바람직성)을 측정했기 때문에 응답 왜곡의 진정한 효과를 어느 정도 과소추정 했을 것이라는 것이라 제안했다. 또한, Furnham(1990)은 지원자들이 도서관 직

원(librarian), 은행원(banker), 광고직 간부(advertising executive)의 각각의 직위에 따라 다른 패턴으로 응답왜곡을 함을 보였다.

이러한 연구들을 통해서 볼 때, 실제 고용 장면에서 지원자들은 그들이 지원하는 직업에 따라 다양한 양상으로 응답을 왜곡할 것이라 추정해 볼 수 있다. 그러나 응답왜곡에 대한 기존의 연구들은 직무별로 필요시 되는 요건들이나, 바람직한 종업원 상(像)이 다를 수 있음을 고려하지 않고 전반적으로 바람직한 경향성만을 고려하는 것으로 보이는 응답왜곡탐지 척도들을 사용한 것의 거의 대부분이었다. 이러한 측면에서 볼 때, 실제로 직무 바람직성이 성격 검사에 대한 응답에서 나타나는 지, 이러한 특정 직무별 바람직성이 기존의 사회적 바람직성 척도에 대한 응답에서도 직업별 차이를 야기하는 지를 고찰해 보는 것은 기존 연구들에서 명확히 밝히지 못했던 응답 왜곡의 전략들에 대한 통찰을 할 수 있도록 하며, 성격검사에서의 응답왜곡을 보다 적절히 이해할 수 있는 기반을 마련할 수 있을 것이다.

응답왜곡과 성격검사의 예측 타당도

여러 성격 검사에서의 응답왜곡 가능성을 조사한 많은 선행 연구들은 사회적 바람직성 척도나 거짓말 척도 같은 타당화 척도를 응답왜곡의 지표로 사용하였으며(예를 들어, Crowne & Marlowe, 1960; Meehl & Hathaway, 1946; Paulhus, 1984), 상당수 연구들의 관심은 성격 검사의 외적 타당도에 응답왜곡이 미치는 영향을 살펴보는 것에 있었다. 특히, 많은 연구자들이 성격검사의 준거 관련 타당도에 대한 의도적인 응답왜곡의 효과를 조사하였는데 그 결과는 혼재되어 있다.

많은 연구자들이(Barrick & Mount, 1996; Christiansen, Goffin, Johnston & Rothstein, 1994; Hough, 1998; Hough et al., 1990; McCrae & Costa, 1983; Ones & Viswesvaran, 1998 등) 의도적인 응답 왜곡이 성격 척도의 준거 관련 타당도에 영향을 미치지 않는다는 것을 발견했다. 한편, 다른 많은 연구자들(Douglas, McDaniel & Snell, 1996; Dunnette et al., 1962; Ironson & Davis, 1979; Norman, 1963; Schmit & Ryan, 1992; Schmit, Ryan, Stierwalt & Powell, 1995; Zickar & Drasgow, 1996 등)은 의도적인 왜곡이 타당도에 심각한 영향을 미친다는 것을 발견했다. 또 다른 연구자들은(예를 들어, Dunnette, Paullin & Motowidlo, 1989; Kamp, 1996; Paajanen, 1988) 왜곡이 성격검사 타당도에 영향을 미치는 것은 사실이나 그 정도는 그다지 크지 않다고 보고했다.

Hough(1998)는 이러한 혼재되는 결과들을 연구가 시행된 장면에 따라 구분함으로써 명확히 이해할 수 있다고 하며, 응답 왜곡을 하고자 하는 동기가 다르게 나타날 수 있는 장면들에 따라 연구를 범주화하였다. 이러함 범주화를 통한 Hough(1998)의 연구와 그 이후의 Jackson, Wroblewski 및 Achton(2000)의 연구는 성격검사의 구성 타당도는 실험실 장면에서 직접적으로 응답왜곡을 하도록 한 연구에서는 부적인 영향을 받았으나, 실제 지원 장면에서는 그다지 심각하지 않았다는 것을 보였다. 최근에, Robie와 Zickar(2001)도 직무 지원자와 현직자들 간에 성격 검사 도구가 측정 동등성을 갖는다는 것을 보였다. 실제 지원 장면에서는 응답 왜곡이 그다지 심각하지 않다는 주장을 지지하는 설득력 있는 증거는 대규모 통합 분석 연구를 통해서도 제시되었다. Ones, Viswesvaran과 Schmidt(1993)와 그 이후에 Ones, Viswesvaran 과 Reiss(1996)의 통합 분석 연구를 통해서도 사회적 바람직성이 성격

검사와 수행 준거간의 관계에서 예측변인이나 억제변인, 매개변인으로써 역할을 하지 않는다는 것이 나타났다.

이러한 많은 연구들이 실제 지원 장면에서의 의도적인 응답왜곡이 성격검사의 예측타당도에 미치는 영향이 문제가 되지 않을 정도라는 것을 제안하지만, 보다 최근의 연구들은 이러한 응답 왜곡이 채용결정에는 영향을 줄 수 있다고 제안한다(Christiansen, Goffin, Johnston, & Rothstein, 1994; Hough, 1998; Rosse, Stecher, Miller, & Levin, 1998; Zickar, Rosse, & Levin, 1996 등).

채용결정에 대한 응답왜곡의 영향

앞에서 언급한 바와 같이, 현재는 응답 왜곡이 일어날 지라도, 그것은 성격 검사의 예측 타당도에 영향을 미치지 않는다(R. Hogan, Hogan, & Roberts, 1996; Hough & Schneider, 1996)는 주장이 우세한 상황이다. 그러나 보다 최근에 응답 왜곡이 성격 검사에 중요한 문제는 아니라는 Ones 등(1995, 1996)의 결론에 대해 의문을 제기하고, 이전 연구에서 포착되지 않았던 응답 왜곡의 효과가 있을지 모른다고 주장하는 연구들이 나타나고 있다.

특히 Ross 등(1998)은 이전 연구에서 예측 타당도에 대한 응답왜곡의 효과가 나타나지 않은 것에 대한 방법론적 방법론적인 이유를 제시했다. Ross 등(1998)은 상관 계수는 변인들간의 선형적 관계에 대한 가정에 영향을 덜 받는(즉, robust한) 추정치라는 것을 보인 Drasgow와 Kang(1984)의 연구결과에 근거하여, 상관계수의 이러한 경향성은 이변량 분포의 특정 범위에서의 서열 순위 변화에 민감하지 못한 결과를 낳는다(지원자들의 상위 점수자들 간의 변화와 같은)고 했다. 즉, 상관 분석은 응답 왜곡에서의 차이 때문에 생길 수 있는 지원자들의 서열 순위 변동에 민감하게 반응하지 못한다는 것이다. 예를 들어, 점수 표본에서 최상위 집단에 속한 사람들이 응답왜곡을 한 거라면, 비록 검사의 관찰된 (동시) 타당도가 전체 표본에서 바뀌지 않을지라도, 예측변인 분포에서 최상단에 속하는 지원자들(즉, 채용 상황에서 가장 고용될 가능성이 높은 지원자들)에게 있어서의 그 타당도는 거의 0에 가까울 수도 있을 것이다(Douglas et al., 1996; Levin, 1995; Zickar, Rosse, & Levin, 1996). 이에 덧붙여 Ross 등(1998)은 왜곡을 효과를 잡아내는 데 상관 분석의 민감성을 낮추는 방법론적 요인들을 몇 가지 더 제안하였다(Ross 등, 1998을 참조).

여기서 중요하게 살펴보아야 할 것은, Ross 등(1998)이 제안한 통계적 문제들이나 그 밖의 영향으로 인해 예측 타당도에 미치는 효과를 탐지하지 못한 상황일지라도, 지원자들의 의도적인 응답 왜곡은 누가 채용될 지에 매우 큰 효과를 미칠 수 있다는 것이다. 앞에서 기술한 바와 같이 성격 척도가 응답왜곡에 영향을 받기 쉬운 점을 고려해 보면, 선발 장면에서 시행하는 성격 검사에서 단지 몇 명의 지원자들만이 극단적인 응답왜곡을 했다고 가정해 보더라도, 이러한 지원자들이 실제로는 그 성격 차원에서 그들과 동일한 진점수를 가지고 있는 솔직한 지원자들보다 더 높은 점수를 받을 것이라는 것을 쉽게 예상해 볼 수 있다. 즉, 검사에서 높은 점수를 받아 선발 비율이 낮을 때 채용되게 되는 지원자들 중 상당수가 그들의 검사 점수를 의식적으로 왜곡했을 뿐, 실제로 그러한 높은 진점수를 갖지는 못한 사람들일 수 있다는 것이다. Ross 등(1998)은 NEO-PI-R에서 직무 지원자들에서 나타나는 응답 왜곡이 직무 현직자들에게서 나타나는 응답왜곡 보다 훨씬 크고, 응답 왜곡에서 유의한 개인차가 존재하며, 직무 지원자들에게서

나타나는 응답왜곡이 채용 결정에 유의한 효과를 미친 다는 것을 보임으로써 이러한 그들의 주장을 입증했다.

Christiansen 등(1994)[2]은 경찰학교(police academy) 신입생들을 대상으로, 전체 표본에서 나타났던 성격 점수와 수행간의 정적 관계가(n=442, ρ=0.18), 성격 점수 분포의 상위 50%만을 대상으로 하여 살펴봤을 때 거의 0에 가까워진다는 것을 발견했다(n=214, ρ=0.02). 뿐만 아니라, 성격 점수에서 상위 53번 째 까지의 사람들만을 대상으로 했을 때에는 이 관계가 심지어 부적으로 나타났다(n=53, ρ=-0.80). 한편, 성격 점수와 사회적 바람직성간의 관계는 전체 표본에서(n=442, ρ=0.31) 보다 상위 50%에서 더욱 강했으며 (n=214, ρ=0.37), 상위 53번 째 까지의 사람들만을 대상으로 했을 때는 이보다 더 강한 관계가 나타났다(n=53, ρ=0.77). 또한, Becker와 Colquitt (1992)는 생활사 검사에 대한 응답왜곡이 누가 채용될 지에 작게나마 영향을 미친다고 결론내렸다.

그러나 이 연구들을 제외하고 현재까지 실제 채용 결정에 대한 응답왜곡의 효과를 고려한 연구들은 거의 찾아보기 힘들다. 그럼에도 실제 인사선발 상황에서 검사 수검자들을 그들의 점수에 따라 서열화 하여 가장 높은 점수를 받은 사람부터 선발 비율에 따라 선발하는 하향식 (top-down) 선발 방법이 대부분 조직에서의 일반적인 관행이기 때문에 이에 대한 연구 결과들은 중요한 함의를 갖는다.

이러한 점을 고려해 볼 때, 지원자들의 응답왜곡이 실제 채용 결정에 미치는 영향을 살펴보는 것도 매우 의미 있는 일이라 판단되며, 특히 실제 장면에서 성격검사의 타당화 척도나 그 밖의 대안적인 방법들을 통해 응답왜곡으로 인한 효과를 적절히 제거하여 채용결정이 정확히 이루어질 수 있도록 하는 노력이 필수적일 것이다.

연구 목적 및 가설

이상의 논의를 바탕으로, 본 연구는 네 가지 목적으로 설계되었다. (1) 선발장면에서 나타나는 성격검사에 대한 응답왜곡의 지표 및 타당화 척도로서 사회적 바람직성 척도의 측정 적절성을 검토하고, (2) 서로 다른 직무에서 동일한 성격검사에 대한 응답왜곡이 다르게 나타나는 지를 확인하고 직무 바람직성 측정 도구를 개발하고, (3) 이러한 직무 바람직성이 사회적 바람직성과 구별되는 개념인지를 조사하며, (4) 의도적인 응답왜곡이 선발 결정에 변화를 야기하는 지의 여부와 그 정도를 직무 바람직성에 근거하여 살펴보는 것이다.

이를 위해서는, 실제 고용 상황에서 실제 직무 지원자들의 응답 왜곡에 대한 이해가 선행되어야 할 것이다. 앞서 언급했듯이, 실제 직무 지원 장면에 초점을 둔 이전 연구들은 적으나, 인상관리에 관한 문헌들은 상황적 요구(situational demands)가 인상 관리 행동에 영향을 미치는 중요한 요인이라는 것을 지적한다(Leary & Kowalski, 1990 등). 이에 근거하여 본 연구에서는 직무에 지원하는 데 고유하게 내재되어 있는 상황적인 요구가 직무 지원자들 표본에서 현직자들 표본에서 보다 더 높은 수준의 응답 왜곡을 낳을 것이라 기대한다. 즉, 직무 지원자들이 직무에 지원하는 과정의 일부로 성격 검사를 시행하여, 가장 긍정적인 방식으로 자신을 보이고자 동기화될 것이라 가정하는 것이다. 반면 정직한 조건의 실험 참가자들에게는 성격 검사의 결과가 연구

2) 여기서 제시한 Christiansen 등(1994)의 연구에서의 상관치들은 범위 제한에 대해 수정한 것.

목적으로만 사용되며 자신의 성격에 대한 결과를 피드백으로 제공하여 자신의 성격에 대해 되돌아 볼 기회가 될 것이라 이야기를 하였다. 이 상황에서는 스스로를 긍정적인 모습으로 보이고자 하는 동기보다는 솔직하게 자신의 모습을 보이고자 하는 동기가 더 강하게 나타날 수 있다.

이러한 측면에서 볼 때, 지원자 표본에서의 성격검사에 대한 응답의 평균이 정직한 집단에서보다 더 높고 표준편차는 더 적으며 부적으로 편포되어 있을 것이라 예상할 수 있다. 이에 덧붙여, 실험실 장면에서 응답왜곡의 효과가 실제 인사선발 상황시 지원자들에게서의 응답왜곡의 효과보다 크다는 기존 연구들(Hough, 1998; Jackson 등, 2000 등)에 기초하여, 본 연구에서는 각 실험조건에서의 성격 검사 척도의 응답에 대해 다음과 같은 가설을 세웠다.

가설 1: 성격 검사 척도에 대한 응답의 평균은 대학생들에게 선발상황을 가정하도록 한 실험 조건에서 가장 높을 것이며, 실제 직무 지원자 표본에서의 평균이 그 다음일 것이며, 통제집단에서 가장 낮을 것이다.

가설 2: 성격 검사 척도에 대한 응답의 분포는 정직한 집단과 비교하였을 때 대학생들에게 선발상황을 가정하도록 한 실험 조건에서 상대적으로 더 부적으로 편포되어 있을 것이며, 더 적은 표준편차를 보일 것이다.

다음으로, 앞서 언급한 바와 같이 사회적 바람직성 척도의 측정 적절성에 대한 의문이 제기되고 있는 시점에서 응답 왜곡의 지표로 사용되는 대표적인 타당화 척도중 하나인 Marlowe-Crowne (Crowne & Marlowe, 1960) 척도의 측정 적절성을 살펴볼 것이다. 만일 이러한 사회적 바람직성 척도에 응답왜곡을 적절히 변별하지 못하는 문항들이 상당수 포함되어 있으면 이 척도를 통한 점수를 응답왜곡의 지표로 사용하는 것이 부적절함을 나타내는 것이다. 한편, 성격 척도에 속한 문항들 중에는 응답왜곡을 적절히 변별할 수 있는 문항들이 있다면 타당화 척도를 검사에 추가적으로 포함시킬 필요성이 약화될 것이다. 이는 성격검사의 문항수준 분석을 통해 성격검사 자체에 내재되어 있는 응답왜곡에 영향을 받는 문항들을 타당화 문항들로 사용하는 것이, 추가적인 타당화 척도를 검사 시행시 포함시키는 것보다 검사 시간을 단축할 수 있는 등의 현실적인 측면에서도 이점을 가져올 수 있기 때문이다.

이러한 측면에서 본 연구에서는, 성격검사에 대한 응답왜곡의 지표로 흔히 사용되는 타당화 척도에 속한 문항들과 성격 검사 자체의 문항들의 응답 왜곡에 대한 민감성을 차별문항기능 분석을 통해 살펴볼 것이다. 단, 응답왜곡 탐지 척도의 측정적절성 및 성격검사 문항의 응답왜곡에 대한 민감성, 그리고 후에 언급할 직무 바람직성을 확인하는 작업은 탐색적인 수준의 연구이기에 이들에 대해서는 연구문제를 상정하였다.

연구문제 1: 차별문항기능 분석 방법을 사용하였을 때, 성격검사에서의 응답왜곡의 지표로 사용되는 타당화 척도의 문항들 중에는 응답왜곡을 적절히 변별하지 못하는 문항들이 상당수 있을 것이다.

연구문제 2: 차별문항기능 분석 방법을 사용하였을 때, 성격검사 문항들 중에는 응답왜곡을 적절히 변별할 수 있는 문항들이 있을 것이다.

다음으로 본 연구에서는 성격검사에 대한 응답 왜곡을 두 가지 직무(운항 승무원 직무와 생

활 설계사 직무)의 비교를 통해서 살펴볼 것이다. 즉, 각각의 직무에 대해서 특정 직무에 국한된 바람직한 응답 패턴이 있는지 혹은 직무에 대한 바람직한 응답은 사회적 바람직성 응답과 구분되지 않는, 직무간 일반화 할 수 있는 구성개념인지의 여부를 이 두 직무의 비교를 통해 조사해볼 것이다. 운항 승무원과 생활 설계사 직무는 각각 서비스 직종과 영업 직종에 포함되는 직무들이다. 따라서 본 연구에서는 선행연구들에 기초하여, 이 두 직무에서 나타나는 응답왜곡이 서로 다른 패턴을 보일 것이라 가정한다. 즉, 두 직무 모두에 대한 지원자들은 인사선발 상황에서(혹은 실험실 장면에서) 응답 왜곡을 하려고 동기화되며, 자신들이 지원하는 직무에서 바람직한 경향성이 어떤 것인지를 각 문항들에 대해 분별있는 방식으로 충분히 통찰을 할 것이라 예상할 수 있기에, 본 연구에서는 다음과 같은 연구문제를 제기하였다.

연구문제 3: 성격검사 문항들에 대한 응답왜곡은 운항승무원 직무와 생활설계사 직무에서 서로 다른 패턴으로 나타날 것이다. 즉 차별문항기능 분석 방법을 사용했을 때, 운항승무원 직무에 지원하는 집단과 생활설계사 직무에 지원하는 집단에서 차별적으로 기능한다고 밝혀진 문항들은 서로 다를 것이다.

다음으로, 본 연구에서 직무 간에 다른 패턴으로 나타난 응답 경향성을 직무 바람직성이라 명명한다면, 이러한 직무 바람직성이 기존의 사회적 바람직성과 구별되는 개념인지, 직무 바람직성이 기존의 사회적 바람직성보다 응답왜곡을 보다 적절히 변별해 낼 수 있는 지를 조사하는 과정이 필수적일 수 있다. 본 연구에서는 이를 확인적 요인 분석을 통해 검증할 것이다.

연구문제 4: 성격 척도 자체의 문항을 통해 밝혀낸 직무 바람직성은 사회적 바람직성과 구별되는 구성개념일 것이다.

끝으로, 본 연구의 주요한 목적 중 하나는 응답 왜곡이 채용 결정에 미치는 효과를 조사하는 것이다. Zickar 등(1996)과 Douglas 등(1996)은 모두 응답 왜곡이 성격 점수 분포의 상위에 속하는 지원자들의 서열 순위를 변화시킨다는 것을 보이는 시뮬레이션을 제시했다. 이러한 시뮬레이션은 응답 왜곡이 선발 도구로 성격 검사를 사용하는 하향식(top-down) 채용 체계에서 누가 채용될지에 영향을 준다는 것을 나타낸다. 이것은 선발 비율이 적을 때 특히 그러할 것이다(Levin, 1995; Zickar et al., 1996). 성격 검사 점수에서 응답 왜곡의 효과를 수정하면 채용 결정이 바뀔 것이고, 높은 응답 왜곡 점수를 가진 지원자들이 잘못 채용되는 것이 줄어들 것이다. 그러나 기존의 사회적 바람직성을 측정하는 타당화 척도 자체가 응답왜곡을 적절히 측정하지 못한다면, 이에 기반한 점수 수정은 오히려 역효과를 낳을 수도 있다. 따라서 본 연구에서는 기존의 사회적 바람직성 척도가 아닌 본 연구에서 밝혀진 직무 바람직성 측정 문항들을 통해, 응답자들의 성격 검사 점수를 수정하고 이러한 점수 수정이 채용 결정을 변화시키는 지 즉 피검자들의 서열 순위가 변하는 지의 여부를 살펴볼 것이다.

가설 3: 성격검사 자체를 통해 도출한 직무 바람직성 문항들의 효과를 제거한 검사점수를 사용할 경우, 원 성격검사 점수를 사용하였을 때와 피검자들의 서열 순위가 달라질 것이다.

방 법

조사 대상 및 절차

자기보고식 성격검사에 대한 응답왜곡을 다룬 선행 연구들은 내적 혹은 외적 타당도 위협을 주제로, 전형적으로 실험실 장면에서 수행되었다. 그러나 실험 연구에서 응답 왜곡의 정도가 과대 추정되며, 현실에 일반화 하는 데 어려움이 있다. 따라서 본 연구에서는 실험 연구와 더불어 여러 직무의 현직자들을 대상으로 한 현장 조사 연구도 병행하여 성격 검사의 응답왜곡에 대한 경험적인 연구를 수행할 것이다.

연구 대상은 크게 학생과 현직자들의 두 집단으로 구분해 볼 수 있으며, 특히 실제 직무 지원자들에게 있어서는 피험자내 설계를 사용하는 것이 현실적인 어려움이 있기에, 서로 다른 지시를 받은 다른 집단의 사람들의 응답을 비교하는 피험자간 설계를 사용한다. 또한 전반적인 직무에 대한 바람직성이 아닌 직무별로 특정한 응답왜곡 경향이 나타나는 지를 살펴보기 위한 비교연구를 진행하기 위해, 본 연구에는 크게 5가지 유형의 표본이 포함된다. 각 실험 조건의 참가자들에 대한 보다 자세한 사항을 표 1에 제시하였다.

전체 응답자의 성별 및 평균 연령 등의 인구통계학적 특성은 표 2에 제시하였다. 직무지원자들의 경우, 각 직무의 특성상 운항승무원 직무에는 남자가 여자에 비해 월등히 많은 반면, 생활설계사 직무에 있어서는 남자 지원자는 한명도 없는 것으로 나타났다.

각 실험 집단에 대한 지시 사항은 다음과 같다.

1. 정직한 조건(대학생): "성격 검사의 타당성을 조사하기 위한 연구입니다. 자신의 성격 그대로 솔직하게 응답해 주십시오."

2. 운항승무원직 지원 상황 가정 조건(대학생): "항공기 운항승무원들 선발을 위한 검사의 타당성을 조사하기 위한 연구입니다. 자신이 항공기 운항승무원이 되고자 하여 그 직무에 지원하였는데 선발과정의 일부로 검사에 응하게 되었다고 가정하고 응답하여 주십시오."

표 1. 조사 대상 및 실험 조건

구 분	조사대상	실험 조건	참여 인원	조사방식
통제집단	대학생	정직하게 응답	337명	실험 일정표를 배포하여 참가희망자들은 원하는 시간에 와서 실험에 참여.
실험집단	대학생	1. 항공기 운항승무원 직무에 지원하는 입사상황 가정	215명	
		2. 생활 설계사 직무에 지원하는 입사상황 가정	193명	
	직무 지원자	3. 항공기 운항 승무원 직무에 대한 실제 지원자	216명	입사시험 과정의 일부로 검사 시행.
		4. 생활 설계사 직무의 실제 지원자	245명	
계			1206명	-

표 2. 전체 응답자의 인구통계학적 특성

구 분	조사대상		성 별		연 령	
통제집단	대학생		남	196명	평균	21.44
			여	137명	범위	19세~29세
실험집단	대학생	운항승무원	남	168명	평균	24.47
			여	47명	범위	19세~34세
		생활설계사	남	87명	평균	22.16
			여	106명	범위	19세~27세
	직무 지원자	운항승무원	남	188명	평균	25.18
			여	27명	범위	20세~34세
		생활설계사	남	0명	평균	28.80
			여	245명	범위	21세~33세

3. 생활설계사직무 지원상황 가정 조건(대학생): "생활설계사 선발을 위한 검사의 타당성을 조사하기 위한 연구입니다. 자신이 생활설계사 직무에 지원하였는데 그 선발과정의 일부로 검사에 응하게 되었다고 생각하고 응답하여 주십시오."

4. 실제 직무 지원자들: "본 검사는 여러분의 선발과정의 일부로 시행되는 것이며, 검사 결과는 선발결정에 영향을 미칠 것입니다. 본 검사에는 거짓말을 탐지하는 척도가 포함되어 있으니 정직하게 응답하여 주십시오."

측정

본 연구에서는 성격 5요인에 기반한 성격 검사 중 Hogan 성격 검사(HPI; Hogan & Hogan, 1992)를 사용하여 자기보고식 성격검사에 대한 의도적 응답왜곡의 효과를 조사하였다. 특히 본 연구의 관심이, 조직 장면에서 사용되는 성격검사에서의 응답왜곡에 있기에 Hogan 성격검사의 직업 척도(HPI Occupational Scales)를 사용하였으며, 6개의 내용 척도 중에서, 검사 길이, 단일차원성, 본 연구에서 상정한 직무에서의 인사 선발과의 관련성을 고려하여 서비스 지향성(service orientation)[3] 15문항, 신뢰성(reliability)[4] 15문항, 스트레스에 대한 인내(Stress Tolerance)[5] 15문항을 선정하여 사용하였다. 성격검사 타당화 척도로는 기존에 가장 널리 사용되어온 Marlowe-Crowne 척도(Crowne & Marlowe, 1960) 33문항을 사용하였다. 그러나 생활설계사 직무의 실제 지원자 집단에 있어서는 현실적인 여건상 사회적 바람직성 척도를 실험에 포함시킬 수 없었다.

자료 분석

기존에 성격검사에서의 응답왜곡을 조사한 연구들의 대부분은 응답왜곡 집단과 정직한 집단간, 혹은 직무 지원자 집단과 현직자 집단간의 성격검사 점수에서의 평균차이를 검증하거나, 응답왜곡이 성격검사에서 높은 점수를 받는 것과

3) 상냥하고 남에게 도움을 주는 능력.
4) 규칙과 절차를 따르고자 하는 의지.
5) 압력과 역경을 극복하고 효율적으로 대처하는 능력.

관련있다는 결과를 제시하였다(예를 들어, Hough, 1998; Hough, Eaton, Dunnette, Kamp & McCloy, 1990). 즉, 이러한 연구들은 개별 문항이 아닌 척도 수준 즉 문항들의 합성 수준(item composite level)에서 응답왜곡의 효과를 조사한 것이다.

그러나 응답왜곡은 천성적으로 문항 수준의 현상이다: 사람들은 척도가 아니라 개별 문항에 응답하며 왜곡을 한다. 따라서 문항 수준에서 초점을 두고 응답왜곡을 연구하는 것이 중요하다(Zickar et al., 1999). 이에 추가로, Zickar와 Robie (1999)는 성격검사에 대한 응답왜곡을 개별 문항 수준에서 연구하면 응답 과정에 대해 통찰해 볼 수 있다고 주장하며, 응답 왜곡에 대한 문항 수준 연구를 통해 응답왜곡을 탐지해낼 수 있는 도구를 개발할 수 있을 것이라 제안했다.

이러한 측면에서 본 연구에서는 문항반응이론에 기반을 둔 차별문항 기능 탐지 절차를 사용하여 응답을 왜곡하는 사람들과 정직하게 응답하는 사람들 간의 차이를 문항 수준에서 살펴볼 것이다. 즉 성격척도 자체의 문항들 중 응답왜곡에 민감하게 반응하는 문항들이 있는지, 그리고 그러한 문항들이 직무에 따라 다르게 나타나는지를 파악하여, 특정 직무별 바람직성을 측정할 수 있는 문항들을 성격 척도 자체 내에서 밝혀낼 수 있을 것이라는 논리이다.

이러한 논리에 따라, 본 연구는 크게 5가지 과정으로 분석되었다. 먼저 연구에 사용된 척도들의 신뢰성과 기술통계량을 분석하였다. 다음으로 MULTILOG 5.1 (Thissen, 1988)을 사용하여 각 척도의 문항 모수치들을 추정하였다. 이 과정을 통해 추정된 문항 모수치들을 사용하여 자료가 본 연구에서 상정한 문항반응모형에 적합한지와, 각 척도의 일차원성을 web version MODFIT 1.1(Levine & Drasgow, 2001)을 사용하여 검증하였다. 모형 적합성과 일차원성이 검증된 차원들에서 차별문항기능(DIF)은 Thissen(2001)이 개발한 IRTLRDIF v.2.0b를 사용하여 파악하였다. 끝으로, DIF를 보인다고 파악된 문항들이 성격검사 척도의 내적 구조에 미치는 영향을 살펴보기 위한 확인적 요인분석은 AMOS 4.0을 사용하여 시행하였다.

이상의 분석 과정 중, 모형 추정과 적합성 검증 및 차별문항기능 파악 과정에 관한 보다 상세한 사항을 다음에 기술하였다.

모형 추정과 적합도(Model estimation and fit)

2모수 로지스틱 모형(two-parameter logistic model; 2PLM)은 그 간결성과 성격과 태도 척도를 사용한 자료의 유형에 매우 적합하다는 이유로 이러한 척도에서 광범위하게 사용되어 왔다. 성격 평정에서 2모수 모형의 효용성을 기술한 많은 연구자들(Embretson & Reise, 2000 참조)과 맥을 같이하여, 본 연구에서도 각 실험 조건에서 각 척도별로 2모수 모형의 모수치들을 추정하였다. 추정된 모형들의 적합도는 Drasgow, Levine, Thissen, Williams 및 Mead(1995)가 제안한 카이 자승(chi-square; χ^2) 통계량을 사용하여 살펴보았다. 카이 자승 통계량은 모형을 통한 응답 기댓값과 실제 자료에서 관찰된 빈도간의 차이를 보는 지표이다. 단일 문항(single items)들에 대해서, 그리고 두 문항씩의 쌍(doubles)과 세 문항씩의 쌍(triples) 각각에 대한 카이 자승 통계치를 계산한다(Reiss & Waller, 2002). 일반적으로 카이자승 통계량을 자유도로 나눈 것을 모델 부합도를 판단하는 지표로 사용되며, Zickar 와 Drasgow(1996)는 이 지표가 3보다 작으면 수용할만 하다고 했다. 카이 자승 통계량이 큰 것은 가정한 IRT 모형이 자료에 적합하지 않다는 것을 나타낸다. 모형이 적합하지 않을 경우에는 집단간 능력차이가 차별적 문항기능으로 오인될 수 있으며, 이렇게 오인된 차별적 문항기능은 적용되는 모형의 부적절함을

반영하는 것이지 문항의 편파성을 반영하는 것이라 보기는 어렵기에(김신영, 2001), 모형의 적합성 검증은 차별문항기능을 밝히기에 앞서 필수적으로 이루어져야 하는 절차이다.

차별문항기능 분석(Differential item functioning analyses)

차별문항기능(differential item functioning; DIF)은 동일한 문항이 어느 하위 집단에서 다른 집단들과 다르게 측정하거나 다르게 기능한다는 것을 의미한다. 즉, 적어도 그 하위 집단에 있어서 그 문항은 덜 타당한 것이다(Thissen, 2001). 이 연구에서, 직무 지원 상황에 놓였다고 생각하라고 지시받은 개인이나 실제 직무 지원자들이, 정직하라고 지시받은 동일한 잠재능력을 지닌 개인과 실제로 그 특성에서 동일한 능력을 지니고 있더라도 특정 문항에 긍정하는 반응 혹은 부정하는 반응을 선택할 가능성이 다르다면, 그 문항은 응답 왜곡 지시에 따른 DIF 증거일 것이다.

DIF의 가장 직접적인 측정방법은 두 집단간의 문항 모수에서의 차이를 밝히는 것이다(Thissen, 2001). 저자들은 실험적인 응답왜곡 조건에 따른 문항 반응 함수에서의 차이를 조사하기 위해 IRT 분석을 했으며, 두 집단의 문항에 대한 응답을 평가하는 데에는 Thissen, Steinberg와 Wainer (1988, 1993)가 제안한 우도비(Likelihood Ratio; LR) 검증방법을 사용했다. LR 검증방법은, 간결(compact) 모델의 우도를 확장(augmented) 모델의 우도와 비교한다. 간결 모델은 어느 문항도 유의한 DIF를 갖지 않는 모델이며, 확장 모델은 하나나 그 이상의 문항들이 유의한 DIF를 갖는 것이다. Thissen 등(1993)에 따르면, LR 검증은 다음과 같이 표현할 수 있다.

$$G^2(df) = 2log[\frac{Likelihood[A]}{Lilelihood[C]}]$$ 6)

G^2는 df와 동일한 자유도의 카이자승 분포를 지닌다고 가정된다. LR-DIF 검증은 매우 일반적이며 비교집단과 준거집단[7] 문항 모수들에 관한 여러 다양한 가설들을 검증할 수 있으며, LR 검증의 장점은 DIF 분석에 앞서 비교집단과 준거집단 문항 모수들의 동등화 과정을 추가적으로 시행하지 않는다는 것이다(Raju & Ellis, 2002). 95%의 유의수준에서 자유도 1일 때 G^2값 3.84를 임계치로 사용하여 이 이상의 G^2값을 보이는 문항들은 차별문항기능을 보인다고 말할 수 있다(Thissen, 2001).

연구 결과

기술통계량 및 신뢰도

표 3에 각 실험 조건에서 관측된 척도별 기술통계량을 제시하였다. 각 조건에서 척도별 신뢰도(내적 일관성 계수: Cronbach's α)는 .69에서 .88로 수용할만하다고 판단되었다. 보다 구체적으로, 서비스 지향성을 제외한 모든 척도에서 통제집

6) A와 C는 확장(augmented) 모델과 간결(compact) 모델을 의미하며, 확작모델은 간결모델의 모든 모수들을 포함하며, 그 외에 추가적인 모수들을 더 지니고 있다. 즉 종종 간결 모델은 확장 모델에 위계적으로 nested되어 있다고 말하곤 한다. 이 절차의 목적은 확장모델에서의 추가적인 모수들이 0과 유의하게 다른지를 검증하는 것이다.

7) IRT 기반 DIF 분석을 시행할 때, 응답자 표본을 준거(참조) 집단과 하나나 그 이상의 비교(초점) 집단으로 나눈다. 비교 집단은 일반적으로 연구자가 관심을 두고 있는 하위집단이며, 준거집단은 비교의 기준이 되는 집단이다.

표 3. 척도별 기술 통계량

척도	Cronbach's α	평균	표준편차	왜도	첨도
서비스지향성(13문)					
통제집단	.70	9.68	2.59	-0.62	-0.27
운항승무원 학생	.78	11.80	1.94	-2.20	4.46
운항승무원 지원자	.70	11.47	1.93	-1.70	2.62
생활설계사 학생	.75	11.64	1.95	-1.94	3.95
생활설계사 지원자	.69	10.06	2.35	-0.98	0.38
신뢰성(15문)					
통제집단	.71	7.99	3.02	-0.04	-0.40
운항승무원 학생	.76	11.39	2.73	-1.12	1.20
운항승무원 지원자	.74	11.51	2.63	-1.42	2.64
생활설계사 학생	.77	10.58	3.11	-0.80	0.58
생활설계사 지원자	.72	8.73	3.04	-0.14	-0.60
스트레스 인내력(15문)					
통제집단	.73	7.92	3.21	-0.18	-0.52
운항승무원 학생	.79	12.33	2.59	-1.39	1.85
운항승무원 지원자	.73	12.18	2.31	-1.37	2.25
생활설계사 학생	.86	12.29	3.22	-1.55	2.00
생활설계사 지원자	.75	8.34	3.28	0.42	-0.76
사회적 바람직성(33문)					
통제집단	.72	14.88	4.99	0.23	0.15
운항승무원 학생	.88	21.38	6.78	-0.51	-0.54
운항승무원 지원자	.85	20.75	6.15	-0.20	-0.70
생활설계사 학생	.85	22.18	6.09	-0.78	0.26
생활설계사 지원자	-	-	-	-	-

단에서 가장 낮은 내적 일관성 계수가 나타났는데 이는 이전 연구들과도 일치하는 결과이다(예를 들어, Zickar 등, 1999).

각 척도에서, 의도적인 응답왜곡이 나타날 것이라 가정해 볼 수 있는 실험집단들에서의 평균값이 정직한 조건에 속한 집단에서의 평균값보다 일관적으로 더 높게 나타났다. 또한 기존 연구들에서 나타났던 응답 왜곡 표본에서 성격 점수 분포의 부적 편포 현상은 본 연구에서도 지지되었다(모든 척도에서 응답 왜곡 집단들이 가장 부적으로 편포되어 있었다). 그러나 본 연구에서는 서비스 지향성 척도를 제외하고는, 기존의 성격검사에 대한 응답왜곡에 관한 연구들에서 제시한 범위의 축소(range restriction) 현상은 나타나지 않았다. 각 척도에서의 조건 간 차이를 일원 변량분석(One-way ANOVA)을 통해 살펴본

표 4. 변량분석 결과

		자승합	자유도	평균자승	F
서비스지향성	집단간	1004.71	4	251.18	50.82***
	집단내	5935.75	1201	4.94	
	전체	6940.46	1205		
신뢰성	집단간	2673.83	4	668.46	78.297***
	집단내	10253.52	1201	8.54	
	전체	12927.34	1205		
스트레스인내력	집단간	5251.30	4	1312.83	148.05***
	집단내	10649.57	1201	8.87	
	전체	15900.87	1205		
사회적 바람직성	집단간	9523.88	3	3174.63	90.76***
	집단내	33474.12	957	34.98	
	전체	42997.99	960		

* $p<.05$, ** $p<.01$, *** $p<.001$

결과 모두 p<.001수준에서 유의하였다: 서비스지향성 $F(4, 1201) = 50.82$, $\eta^2 = .15$; 신뢰성 $F(4, 1201) = 78.30$, $\eta^2 = .21$; 스트레스 인내성 $F(4, 1201) = 148.05$, $\eta^2 = .33$. 결과는 표 4에 제시하였다.

Tukey 방식을 사용한 사후 분석에서는 각 척도별로 다양한 패턴으로 차이가 나타났다. 우선 신뢰성 척도에서는 통제집단은 모든 지원자 집단(학생 및 실제 지원자)보다 더 낮은 평균으로 유의미한 차이를 보였다(p<.001).

또한, 운항승무원 지원자 집단 간에는 즉, 운항승무원 지원자를 가장한 학생 집단과 실제 지원자 집단 간에는 유의한 차이가 보이지 않았으나(p>1.0), 생활설계사 집단 간에는 즉, 생활설계사 직무 지원자를 가장한 학생 집단과 실제 지원자 집단 간에는 유의한 차이가 나타났다. 덧붙여, 운항승무원 직무의 지원자 집단(학생 및 실제 지원자)와 생활설계사 직무의 지원자 집단(학생 및 실제 지원자)간에도 유의한 차이를 보여(p<.05), 본 연구에서 가정한 직무에 특수한 바람직성 응답의 존재를 지지하는 결과가 나타났다.

다음으로, 서비스 척도 및 스트레스 척도에 있어서도 신뢰성 척도와 마찬가지로 통제집단은 생활설계사 직무의 실제 지원자 집단을 제외한 모든 지원자 집단(학생 및 실제 지원자) 보다 더 낮은 평균 점수를 나타내며 유의미한 차이를 보였다(p<.001). 생활설계사 직무의 실제 지원자 집단은 통제집단보다 낮은 평균 점수를 보였으나 그 차이가 유의미하진 않았다(p>.10). 스트레스 인내력 척도와 서비스 척도에 있어서는 항공승무원 직무의 지원자들과 생활설계사 직무의 지원자들 간에는, 항공승무원 직무의 학생 표본과 생활설계사 직무의 실제 지원자 표본간의 대비를 제외하곤(p<.001), 유의한 차이가 나타나지 않았다. 또한, 생활설계사 직무의 실제 지원자집단은 통제집단을 제외한 모든 비교 집단보다 유의하게 낮은 평균 점수를 보인 예상치 못한 결과가 나타났다(p<.001). 종합하면, 성격검사 척도 점수에 대한 가설 1은 부분적으로만 지지되었으

표 5. 연구에 포함된 척도들간의 상관계수

	1	2	3
1. 서비스지향성	1.00		
2. 신뢰성	.47**	1.00	
3. 스트레스인내력	.52**	.67**	1.00
4. 사회적 바람직성	.59**	.69**	.68**

** $p<.01$

주. 생활설계사 직무의 실제 지원자 집단에서는 사회적 바람직성 척도에 응답하도록 하지 않았기에, 제시된 상관은 전체 표본에서 생활설계사 직무의 실제 지원자 집단을 제외한 961명 표본에서 산출한 것임.

며, 성격검사 척도 점수 분포의 부적 편포 및 범위제한에 대한 가설2도 부분적으로 지지되었다.

다음으로 표 5는 연구에 포함된 성격 척도들간 상관계수를 나타낸 것이다. HPI의 직업 척도가 HPI의 원척도에 속한 HIC들을 분해하여 다시 새로운 요인들로 결합한 것이기에, 척도간 상관이 높게 나타났다. 뿐만 아니라, 성격검사 타당화 척도인 Marlowe-Crown 척도도 HPI의 다른 성격 척도들과의 상관이 높게 나타났다. 이러한 유의한 상관은 사회적 바람직성 척도가 중요한 성격 특성들을 측정하고 있으며, 따라서 의도하지 않았던(nuisance) 변량의 지표로 간주해서는 안된다는 기존 연구들에서의 주장(예를 들어, Kriedt & Dawson, 1961)과도 일치하는 결과이다.

직무에 특수한 바람직성 응답의 존재를 파악하는 데 보다 구체적인 정보를 얻기 위해 본 연구에서는 각 특정 직무에 따른 바람직성의 경향을 개별 문항 수준에서 차별문항기능 분석 방법을 사용하여 살펴보고자 한다.

일차원성 검증 및 모형-자료 적합성 검증

이상적으로는 IRT 모형의 부합도는 독립적인 교차 타당화 표본을 사용하여 평가해야 하지만, 본 연구에서는 독립적인 두 표본을 구성하기에 표본 크기가 너무 작아서 문항 모수치들이 잘못 추정될 우려가 있기에 단일 표본을 사용하여 모형 부합도를 판단할 수밖에 없었다. Drasgow, Levine, Thissen, Williams와 Mead(1995)가 제안한 카이 자승(chi-square; χ^2) 통계량을 통한 모형 적합성 검증 결과는 표 6에 제시하였다.

카이 자승 통계치는 2모수 문항 반응 모형이 정직한 표본 뿐 아니라 지원자 조건의 표본들 모두에서 자료에 잘 부합한다는 것을 보였다(스트레스 인내력 척도의 생활설계사 지원상황인 학생표본을 제외하고). Zickar와 Drasgow(1996)이 그들의 연구에서 단일 문항들에 대한 카이자승 통계치가 두 문항 쌍이나 세 문항 쌍들에 대한 카이자승 통계치보다 큰 것을 발견하였으며, 이와 대조적으로 Drasgow 등(1995)은 두 문항 쌍이나 세 문항 쌍들에 대한 카이자승 통계치가 단일 문항에 대한 것보다 큰 것을 발견하였다. 이러한 기존 연구들의 명확하지 않은 결과는 본 연구에서도 동일하게 나타났다. 즉 특정 척도의 특정 표본에서는 Zickar 등(1996)과 동일한 패턴이 나타났으나, 동일 척도라 할지라도 다른 표본에서는 Drasgow 등(1995)과 동일한 패턴이 나타났다.

그럼에도 본 연구에서도 두 문항 쌍이나 세 문항 쌍들에 대한 카이자승 단일 문항에 대한 것 뿐 아니라, 두 문항쌍이나 세 문항쌍들에 대한 카이 자승 검증 결과도 수용할 만 하였다(스트레스 인내력 척도의 생활설계사 지원상황인 학생표본을 제외하고). 이는 문항간의 상호작용이 2모수 모형으로 잘 모델링 되었다는 것을 의미하며, 척도의 일차원성가정이 위배되지 않았음을 나타낸다. 비록 세 척도 각각에 적합하지 않은 문항들이 하나 이상씩은 있었으나, 이 문항들

표 6. 카이자승 적합도 통계량

척 도	부적합문항 수[a]			χ^2/df 평균		
	단일문항	두 문항 쌍	세 문항 쌍	단일문항	두 문항 쌍	세 문항 쌍
서비스지향성(13문)						
통제집단	0	2	0	0.01	0.84	1.10
운항승무원 학생	0	0	0	1.79	1.12	0.91
운항승무원 지원자	0	0	0	0.30	0.32	0.24
생활설계사 학생	0	0	0	0.47	0.39	0.38
생활설계사 지원자	0	0	0	0.00	0.40	0.59
스트레스인내력(15문)						
통제집단	0	1	0	0.01	0.68	0.85
운항승무원 학생	0	0	0	0.01	0.31	0.89
운항승무원 지원자	0	0	0	0.18	0.58	1.20
생활설계사 학생	14	10	0	5.31	3.06	2.07
생활설계사 지원자	0	1	0	0.03	0.67	1.05
신뢰성(15문)						
통제집단	0	0	0	0.01	0.59	0.97
운항승무원 학생	0	0	0	0.01	0.27	0.55
운항승무원 지원자	0	0	0	0.07	0.46	0.28
생활설계사 학생	0	2	0	0.24	0.92	1.66
생활설계사 지원자	0	1	1	0.00	0.67	1.03
Marlowe-Crown 척도						
통제집단	0	1	0	0.01	0.56	0.92
운항승무원 학생	0	2	0	0.04	0.45	0.68
운항승무원 지원자	0	0	0	0.02	0.41	0.71
생활설계사 학생	0	2	1	0.01	0.52	0.83
생활설계사 지원자	-	-	-	-	-	-

주. a: chisquare/df>3

을 부적합성을 보인 문항들이라 간주하여 연구에서 아예 제외시키지는 않았다. 이는, 세 표본 모두에서 동일하게 적합하지 않다고 일관되게 검증된 문항들은 하나도 존재하지 않았기 때문이며, 대신 그 문항들이 특정한 결과를 야기하는 데 영향을 주는지를 계속 추적하여 살펴보기로 하였다. 이에 추가로, 스트레스 인내력 척도의 일차원성 및 모형 적합도가 생활설계사 지원 상황인 학생 표본에서만 부적합하게 나왔는데, 다른 조건 및 다른 표본에서는 이 척도의 일차원성 및 모형 적합도가 수용할만 하였기에, 일단 두고 추후 분석에서 계속 살펴보기로 하였다.

차별문항 기능 분석

검사 타당화 척도의 차별문항 기능 분석

지금까지 많은 연구들에서 성격검사의 타당화 척도로 사용되어왔으며, 본 연구에서도 사용한 Marlowe-Crowne 척도는 사회적 바람직성을 탐지하고자 개발된 척도이므로 대부분의 문항들이 차별문항기능에 민감하게 반응하는 것이 이상적일 것이다. 즉, 이 척도에 포함된 문항들은 정직한 집단과 지원자 집단에서 차별적으로 기능할 것으로 기대할 수 있다. 차별문항기능은 IRTLRDIF를 이용하여, Thissen(2001)이 제안한 95%의 유의수준에서 자유도 1일 때 G^2값 3.84를 임계치로 사용하여 파악하였다. 결과를 다음의 표 7에 제시하였다.

사회적 바람직성 척도 전체 33문항 중 정직한 집단을 운항승무원 지원자 중 학생표본의 응답과 비교했을 때 응답왜곡에 민감하게 반응하지 못한 문항(차별문항기능을 보이지 않은 문항)이 17문항이나 되었으며, 운항승무원의 실제 지원자와 비교했을 때도 13문항이 차별기능을 보이지 않았으며, 생활설계사 지원자 중 학생 표본과 비교했을 때에도 전체 문항 중 반 이상인 17문항이 응답왜곡에 민감하게 반응하지 못하는 것으로 나타났다. 즉 응답왜곡에 민감하게 반응해야 하는 타당화 척도에 포함된 반 정도의 문항들이 실제로는 지원자들의 응답과 정직한 집단의 응답간의 차이에 민감하게 반응하지 못하는 것으로 나타났으며, 이는 타당화 척도의 측정 부적절성을 반영하는 것으로 볼 수 있다. 즉, 성격검사에서의 응답왜곡의 지표로 사용되는 타당화 척도의 문항들 중에는 응답왜곡을 적절히 변별하지 못하는 문항들이 상당 수 있을 것이라는 연구문제 1을 지지하는 결과가 나타났다.

성격 척도의 차별문항 기능 분석

이상의 결과를 통해 알 수 있듯, 기존의 사회적 바람직성 척도도 응답왜곡에 민감하게 반응하지 못하는, 즉 차별문항기능을 보이지 않는 상당수 포함되어 있음을 알 수 있다. 반면에, 성격척도 자체에 응답왜곡에 민감하게 반응하는 문항들이 내재되어 있다면 이러한 문항들을 사용하여 응답왜곡을 탐지하는 것이 보다 절약적일 수 있다. 표 8에 모든 세 척도에서 정직 조건과 지원자 조건간의 비교를 위해 차별문항기능(DIF) 분석을 한 결과를 제시하였다.

각 척도에 따라, 모든 비교조건에서 동일하게 DIF를 보인다고 나타난 문항들은 전반적인 직무 바람직성 혹은 사회적 바람직성을 나타내는 문항들이라 해석할 수 있을 것이다. 이러한 전반적인 직무 바람직성 혹은 사회적 바람직성을 나타내는 문항들은 서비스 지향성 척도에서 단 한 문항, 신뢰성 척도에서 두 문항만 나타났다(이러한 문항들은 표에 밑줄을 쳐서 제시). 즉 이러한 결과는, 실험실 조건과 현장 조건 모두에서 지원자들은 척도 수준이 아닌 각 문항 수준에서 바

표 7. 타당화 척도 중 DIF를 보이지 않은 문항

검사 타당화 척도	DIF를 보이지 않은 문항
Marlowe-Crown(33문)	
정직집단 vs. 운항승무원 학생	s1, s6, s7, s8, s9, s10, s11, s12, s14, s15, s16, s17, s18, s20, s27, s31, s33
정직집단 vs. 운항승무원 지원자	s1, s6, s8, s10, s12, s14, s15, s16, s17, s18, s20, s27, s33
정직집단 vs. 생활설계사 학생	s2, s3, s7, s8, s11, s12, s13, s15, s16, s17, s18, s20, s26, s28, s31, s32, s33

표 8. 척도별 DIF를 보인 문항의 수

척 도	DIF를 보인 문항
서비스지향성(13문)	
①정직집단 vs. 운항승무원 학생	서1, 서2, 서3, 서4, 서5, 서8
②정직집단 vs. 운항승무원 지원자	서1, 서2, 서3, 서5, 서9
③정직집단 vs. 생활설계사 학생	서1, 서4, 서5, 서7
④정직집단 vs. 생활설계사 지원자	서2, 서4, 서5, 서7, 서10
신뢰성(15문)	
①정직집단 vs. 운항승무원 학생	신2, 신4, 신6, 신8, 신12
②정직집단 vs. 운항승무원 지원자	신2, 신5, 신6, 신7, 신8, 신9, 신12, 신13
③정직집단 vs. 생활설계사 학생	신2, 신4, 신5, 신9, 신11, 신12, 신13
④정직집단 vs. 생활설계사 지원자	신1, 신2, 신4, 신5, 신6, 신10, 신11, 신12, 신13
스트레스 인내력(15문)	
①정직집단 vs. 운항승무원 학생	스1, 스2, 스3, 스10, 스13
②정직집단 vs. 운항승무원 지원자	스2, 스3, 스10, 스13
③정직집단 vs. 생활설계사 학생	스5, 스6, 스12, 스14
④정직집단 vs. 생활설계사 지원자	스1, 스3, 스6, 스8, 스9, 스10, 스12, 스13

람직한 방향으로 응답왜곡을 하며, 자신이 지원하는 상황의 요구에 따라 지원자들이 판단하는 바람직성의 기준이 상당히 달라질 수 있음을 나타내는 것이다. 따라서 성격 검사 문항들 중에는 응답왜곡을 적절히 변별할 수 있는 문항들이 있을 것이라는 연구문제 2도 지지하는 결과가 나타났다.

본 연구에서 지원자들의 응답에 영향을 줄 수 있는 상황적 요건은 크게 두 가지로 살펴볼 수 있을 것이다. 첫째는 응답자들이 실제 지원자인지 학생들인지 즉, 실험실 상황인지 현장 상황인지의 상황적 요건이며, 둘째는 그들이 지원하고자 하는 직무가 운항승무원 직무인지 생활설계사 직무인지의 상황적 요건이다. 따라서 본 연구에서 가정한 바와 같이 지원자들이 판단하는 바람직성의 기준이 그들이 지원하고자 하는 직무에 따라 달라지는 것인지를 파악하기 위해 비교집단들 간의 중첩을 조사하였다. 즉 각 척도에서, 운항승무원 직무의 지원자들간의 중첩비율이나 생활설계사 직무의 지원자들간의 중첩비율이, 학생표본간의 중첩비율이나 지원자 표본간의 중첩비율보다 더 커야 한다는 논리이다. 다시 말해, 정직집단과 운항승무원 학생들을 비교하였을 때 DIF를 보인 문항들과 정직집단과 운항승무원 실제 지원자들을 비교했을 때 DIF를 보인 문항들 전체 중 이 두 비교조건에서 공통적으로 DIF라고 나타난 문항들의 비율이, 정직집단과 운항승무원 학생을 비교했을 때와 정직집단과 생활설계사 학생들을 비교했을 때 DIF를 보인 문항들 전체 중 이 두 비교조건에서 공통적으로 DIF라고 나타난 문항들의 비율보다 더 큰 지를 조사한 것이다. 이에 대한 결과를 다음의 표 9에 제시하였다.

우선, 서비스 지향성 척도를 보면 ①과 ②의

비교에서 DIF를 보인 전체 문항은 7문항이다(서1, 서2, 서3, 서4, 서5, 서8, 서9). 그 7문항 중 ①과 ②의 비교에서 공통적으로 DIF를 보인 문항들은 4문항이다(서1, 서2, 서3, 서5). 따라서 진정으로 정직한 집단과 운항승무원 직무의 응답자들간의 DIF를 보이는 문항들은 4/7*100= 57.14%라 할 수 있다. 그리고, 동일 척도에서 ③과 ④의 비교에서 DIF를 보인 문항들은 총 6문항(서1, 서2, 서4, 서5, 서7, 서10)이며, 이들 중 ③과 ④의 비교에서 공통적으로 DIF라 나타난 문항들은 3문항(서4, 서5, 서7)이다. 따라서 정직한 집단과 생활설계사 직무의 응답자들간에 DIF를 보인 문항들은 3/6*100=50%라 할 수 있다.

다음으로, 동일한 서비스 지향성 척도에서 ①과 ③의 비교에서 DIF를 보인 전체 문항은 7문항이다(서1, 서2, 서3, 서4, 서5, 서7, 서8). 그 7문항 중 ①과 ③의 비교에서 공통적으로 DIF를 보인 문항들은 4문항이다(서1, 서4, 서5). 따라서 정직한 집단과 학생 응답자들간에 DIF를 보이는 문항들은 3/7*100=42.86%라 할 수 있다. 끝으로, ②와 ④의 비교에서 DIF를 보인 전체 문항은 8문항이며(서1, 서2, 서4, 서3, 서5, 서7, 서9, 서10), 그 문항들 중 공통인 문항들은 2문항이다(서2, 서5). 따라서 정직한 집단과 실제 지원자 응답자들간에 DIF를 보인 문항들은 2/8*100=25%라 할 수 있다.

종합하여 보면, 정직한 집단과 운항승무원 직무의 응답자들간에 DIF를 보인 문항들의 비율 57.14%과 정직한 집단과 생활설계사 직무의 응답자들간에 DIF를 보인 문항들의 비율 50%는 모두, 정직한 집단과 학생 응답자들간에 DIF를 보인 문항들의 비율인 42.86% 및 정직한 집단과 실제 지원자 응답자들간에 DIF를 보인 문항들의 비율인 25%보다 더 높게 나타났다. 따라서 정직한 집단과 직무에 지원하는 상황에 처한 응답자들간의 응답에 차이가 나타나게 한 것은 그들이 학생인지 실제 지원자인지에 의해 좌우되었다기보다는, 그들이 어느 직무에 지원했는지에 의해 더 큰 영향을 받았다고 잠정적으로 결론내릴 수 있다.

이와 동일한 방식으로 각 비교 집단별 DIF 문항들의 중첩비율을 계산한 것이 표 9에 제시되어 있다. 표 9를 통해 알 수 있듯, 서비스 지향성과 신뢰성 척도 모두에서 운항승무원 직무의 지원자들 간의 중첩비율이나 생활설계사 직무의 지원자들 간의 중첩비율이, 학생표본 간의 중첩비율이나 지원자 표본 간의 중첩비율보다 더 크게 나타나, 차별문항기능이 직무간의 차이에 의해 나타난 것임을 지지해주고 있다. 그러나 생활설계사 직무에 지원하는 실험상황에서 응답한 학생 표본에서, 스트레스 인내성 척도가 일차원성도 지지되지 못했으며 2모수 모형에도 적합하

표 9. 비교집단별 DIF 문항들의 중첩비율

척도			
서비스지향성	①② 4/7*100=57.14%	①③ 3/7*100=42.86%	①②>③④>①③>②④
	③④ 3/6*100=50%	②④ 2/8*100=25%	
신뢰성	①② 4/9*100=44.44%	①③ 3/9*100=33.33%	③④>①②>②④>①③
	③④ 6/10*100=60%	②④ 5/12*100=41.67%	
스트레스인내성	①② 4/5*100=80%	*①③ 0/9*100=0%*	①②>②④
	*③④ 2/10*100=20%*	②④ 3/9*100=33.33%	

지 않은 것으로 나타났기에, 문항 반응이론의 기본 가정이 충족되지 못한 상태에서 차별문항기능의 검증이 이루어졌다. 따라서 스트레스 인내성 척도에 있어서는 ④의 비교결과를 신뢰할 수 없기에, 이 척도에 한해서는 직무간 차이에 의한 DIF 문항들인지에 관해 결론을 지을 수 없었다. 그러나 종합하여 볼 때, 성격검사 문항들에 대한 응답왜곡은 운항승무원 직무와 생활설계사 직무에서 서로 다른 패턴으로 나타날 것이라는 연구문제 3도 지지하는 결과가 나타났다.

확인적 요인 분석

문항반응이론(Item Response Theory)만을 이용해서는 다변량 요인구조가 표본에 따라 불변하는지의 여부를 검증할 수 없다(Zickar & Robie, 1999). 따라서 본 연구에서는 요인 구조 차이를 검증하기 하고, 각 조건간의 비교를 통해 도출해 낸 특정 직무별 바람직성 요인의 측정 적절성을 확인하고, 동일 직무별 측정 동등성을 검증하기 위해 확인적 요인분석(confirmatory factor analysis; CFA) 방법을 사용하였다. 이를 통해 차별문항기능(DIF) 분석을 통해 밝혀진 결과를 보다 심층적으로 이해할 수 있을 것이다.

모형의 적합도를 판단하는 기준으로는 일반부합지수(GFI: Goodness of fit Index), 수정부합지수(AGFI: Adjusted Goodness of fit Index), 원소간 근접오차(RMSEA: Root Mean Square Error of Approximation), 원소간 평균차이(RMR : Root Mean square), 비교부합지수(CFI: Comparative Fit Index), 표준부합지수(NFI: Normed Fit Index) 등을 사용하였고 모수 추정(parameter estimation) 방식은 가장 널리 사용되는 최대우도 추정(Maximum Likelihood estimation)을 사용하였다.

직무바람직성 요인의 구성 타당도

본 연구자는 앞 절에서 성격 검사 및 사회적 바람직성 척도에서 추출한 직무 바람직성 탐지 문항들이 타당한 요인구조를 가지는지를 파악하기 위해 측정모형을 설정하고 부합도 지수를 산출하였다. 잠재변인인 직무 바람직성은 3가지 측정변인을 지표로 한다. 모형에서 각 문항들을 측정 변인으로 할 경우 지표의 수가 많아지므로 Williams와 Anderson(1994)의 방법과 같이 각각 3~4문항씩을 무선적으로 선택하여 3개의 측정변인으로 구성하였고, 이 때 지표를 구성하는 문항들은 각 직무마다 탐색된 직무 바람직성 탐지 문항들이다. 또한 잠재변인의 척도 단위를 확정하고 식별성(identification)의 문제를 해결하기 위해 각 잠재요인의 λ중 하나를 1에 고정 시켰다(박광배, 2000). 이렇게 구성된 직무바람직성 요인은 크게 2가지로, 생활설계사 직무 바람직성은

표 10. 직무 바람직성 요인의 측정모형 적합성

	χ^2	df	χ^2/df	RMR	GFI	AGFI	NFI	CFI	RMSEA
운항승무원 직무바람직성									
운항승무원(학생)	10.58	2	5.29	0.035	0.977	0.886	0.897	0.912	0.142
운항승무원(지원자)	1.12	2	0.56	0.011	0.997	0.987	0.984	1.000	0.000
생활설계사 직무 바람직성									
생활설계사(학생)	0.34	2	0.17	0.005	0.999	0.996	0.998	1.000	0.000

생활 설계사 표본에서 도출된 문항들로 구성되었고, 운항승무원 직무 바람직성은 운항 승무원 표본에서 도출된 문항들이다.

이러한 모형에 대해 각각 2가지 직무 표본에서 모두 부합도를 산출하였으며 그 결과는 표 10과 같다. 표 10에서 보는 바와 같이 직무바람직성 요인의 측정모형은 운항승무원(지원자) 표본과 생활설계사(학생) 표본에서 모두 각 부합도 지수의 수용기준을 충족시키는 것으로 드러났다. 하지만 운항승무원(학생) 표본에서는 RMR이 .05 미만, GFI와 CFI는 수용 기준인 .90 이상이었으나, AGFI와 NFI가 각각 .886과 .897로 수용 기준에 미치지 못하는 것으로 나타났다.

이와 같은 결과를 바탕으로 본 연구자가 가정한 직무바람직성 개념은 3표본 중 2표본에서 매우 높은 부합도를 보였으며 이에 어느 정도 만족할 만한 타당성을 가지고 있다고 할 수 있을 것이다. 또한 운항 승무원(학생) 표본에서의 부합도가 일부 수용 기준을 충족시키지 못한 것에 대해서는 논의 부분에 몇 가지로 언급하도록 하겠다.

직무바람직성 요인과 사회적 바람직성 요인의 변별

본 연구자의 2번째 관심은 직무 바람직성 요인이 사회적 바람직성과 구별되는 개념인가 하는 문제이다. 이를 파악하기 위해 본 연구자는 2가지 경쟁모형을 설정하고 비교하였다. 먼저 1요인 모형은 직무 바람직성의 문항들과 사회적 바람직성을 측정하는 문항들이 1요인을 구성한다는 모형이고, 2요인 모형은 사회적 바람직성과 직무 바람직성이 각각 구별되는 2개의 요인을 구성하고 있는 것이다. 본 연구자는 만일 직무 바람직성과 사회적 바람직성이 구별된다면 2가지 모형의 적합도 지수에 있어 유의한 차이가 있을 것이라고 가정한다. 이에 대한 분석 결과는 표 11에 제시되어 있는데, 모형 1과 모형 2 같이 위계적인 모형의 차이 검증은 보통 χ^2의 차이값을 기준으로 한다. 이 방법은 표본 크기에 민감하다는 문제가 있어 표본이 조금만 커지더라도 모형을 부적합하게 판단할 가능성이 있지만 2가지 모형을 비교할 수 있다는 장점이 있다. 2가지 모형의 χ^2의 차이는 또 하나의 χ^2이 되며 이때의 자유도는 2모형의 자유도의 차이와 같다. 표 11에서 보는 바와 같이 2요인 모형은 운항

표 11. 직무 바람직성과 사회적 바람직성의 변별타당도

		χ^2	df	$\Delta\chi^2$	GFI	AGFI	NFI	CFI	RMSEA
운항승무원 (학생)	2요인a	12.15	4	3.22	0.98	0.92	0.95	0.96	0.098
	1요인	15.37	5		0.97	0.92	0.94	0.95	0.098
운항승무원 (지원자)	2요인	1.53	4	3.98*	1.00	0.99	0.99	1.00	0.000
	1요인	5.41	5		0.99	0.98	0.98	1.00	0.000
생활설계사 (학생)	2요인b	5.46	4	3.92*	0.99	0.96	0.98	1.00	0.04
	1요인	9.38	5		0.98	0.95	0.98	0.99	0.06

주> a: 운항승무원 직무 바람직성 vs. 사회적 바람직성(Marlowe-Crowne)
b: 생활설계사 직무 바람직성 vs. 사회적 바람직성(Marlowe-Crowne)

* $p<.05$

승무원(지원자) 표본과 생활 설계사(학생) 표본에서 1요인 모형에 비해 더 나은 부합도를 보였다 ($\Delta\chi^2>3.84$, df=1, $p<.05$).

이와 같은 결과를 바탕으로 본 연구자가 가정한 직무바람직성 개념은 3표본 중 2표본에서 사회적 바람직성과 변별되는 개념이라는 증거를 보였으며, 운항 승무원(학생) 표본에서는 사회적 바람직성과 직무 바람직성이 구분되지 못하고 1개의 바람직성 개념으로 설명되었다. 이와 같은 결과는 운항 승무원(학생) 표본에서 직무 바람직성의 요인구조가 만족할만한 부합도를 보이지 못했던 사실과도 연관이 있는 것으로 보인다. 이러한 결과를 종합하여 볼 때, 성격 척도 자체의 문항을 통해 밝혀낸 직무 바람직성은 사회적 바람직성과 구별되는 구성개념일 것이라는 연구문제 4를 부분적으로 지지하는 결과가 나타난 것으로 볼 수 있다.

채용 결정에서의 변화

끝으로, 앞선 분석 결과를 통해 도출한 직무 바람직성 측정 문항들을 사용하여 피검자들의 능력을 다시 추정하고 선발 비율을 적용하였을 때 채용되는 사람들과, 직무 바람직성 측정 문항들과 상관없이 피검자들의 능력 추정치를 사용했을 때 채용되는 사람들이 얼마나 일치하는 지의 정도를 살펴보았다.

우선, 전체 성격 문항에서 각 직무별로 특정하게 바람직성 경향이 나타나는 문항들을 제거하고 문항 모수치를 추정한 후, 다시 피검자들 전체의 능력을 추정하였다. 그리고 그 능력 추정치를 이용하여 피검자들을 서열화하였다. 다음으로, 직무 바람직성을 고려하지 않은 원 성격척도를 이용해 피검자들 능력을 추정한 후 그를 또 서열화하였다. 그리고 10%와 20%의 선발 비율을 각각 적용했을 때(하향식 서발 방식), 각 서열 범주에서 선발되는 사람들이 얼마나 일치하는 지의 일치율을 계산하였다. 그 결과를 다음의 표 12에 제시하였다.

표 12를 보면 알 수 있듯, 운항 승무원 실험조건인 학생 표본만 제외하고 선발 비율이 낮을수록 두 가지 방식(직무 바람직성을 고려 vs. 직무 바람직성을 고려하지 않음)에서의 일치율이 상대적으로 더 낮게 나타났다. 이는 선발 비율이 낮을수록 응답 왜곡이 누가 채용되는 지의 채용 결정에 더 큰 효과를 미친다는 선행 연구(Rosse

표 12. 선발비율에 따른 선발 일치율

	전체표본수	선발비율(%)	선발인원(명)	일치율(%)
운항승무원(학생)	215명	10	23	73.9
		20	53	71.7
운항승무원(지원자)	216명	10	22	54.5
		20	43	76.7
생활설계사(학생)	193명	10	19	42.1
		20	40	72.5
생활설계사(지원자)	245명	10	25	64.0
		20	49	77.6

et al., 1998)와도 일치하는 현상이다. 따라서 성격검사 자체를 통해 도출한 직무 바람직성 문항들을 통해 검사점수를 수정할 경우, 원 성격검사 점수를 사용하였을 때와 피검자들의 서열 순위가 달라지게 되어 채용 결정에 변화가 나타날 것이라는 세 번째 가설도 지지되었다.

기존에 사용되었던 사회적 바람직성 척도가 성격검사 이외에 추가적으로 포함된 검사인 반면, 본 연구에서의 직무 바람직성 측정 문항들은 성격 검사 자체에 내재해 있는 문항들이라는 차이로 인해, 성격검사에서 사회적 바람직성을 고려하였을 때 채용결정에서의 변화는 살펴보지는 못했다. 그러나 조직에서 성격검사를 선발과정의 일부로 사용할 때, 일반적인 관행은 개개인의 점수에서 사회적 바람직성을 통제하는 것이 아니며, 일정수준 이상의 사회적 바람직성 점수를 받은 사람들은 선별해 내고, 나머지 사람들에 대해서는 원 점수를 그대로 사용하는 것이다. 이 점을 고려한 다면, 본 연구에서 직무 바람직성을 고려했을 때의 서열과, 전체 성격 척도에서의 점수를 사용했을 때의 서열에서 차이가 나타난 것도 중요한 함의를 갖는다고 할 수 있을 것이다.

논 의

특히 오늘날과 같은 무한 경쟁 환경에서, 이윤추구를 목적으로 하는 기업에서는 조직의 역량을 극대화하고, 최대한의 영리를 달성하기 위해 조직 운용의 주체인 "사람"을 잘 선발해야 한다. 따라서 어떤 경우라도 선발도구는 그 타당성의 검증이 필수적인 것이다. 본 연구는 타당한 인사선발을 확립하기 위한 기초 연구로써 성격 검사의 응답과 성격 변인의 예측력에 관한 근본적인 이해를 목적으로 한다.

본 연구는 사회적 바람직성이 응답왜곡의 타당한 지표가 될 수 있는가라는 근본적인 물음에서 출발하였다. 이에 따라, 본 연구는 첫째, 선발장면에서 나타나는 성격검사에 대한 응답왜곡의 지표로서 사회적 바람직성 측정 도구의 적절성을 검토하고, 둘째, 서로 다른 직무(운항 승무원, 생활 설계사)에서 동일한 성격검사에 대한 응답왜곡이 다르게 나타나는지를 확인하고, 이러한 직무 바람직성이 사회적 바람직성과 구별되는 개념인지를 조사하는 것을 목적으로 한다. 이에 추가적으로, 선발장면에서의 의도적인 응답왜곡이 성격검사의 준거관련 타당도에는 그다지 큰 영향을 미치지 않으나, 채용 결정에 변화를 야기한다는 최근의 연구 결과들과 맥을 같이하여, 본 연구에서 구성 타당도가 확인된 직무바람직성 개념을 근거로 직무별로 다르게 나타나는 응답왜곡이 선발장면에서의 채용 결정에 미치는 영향을 살펴보았다.

본 연구에서 도출된 결과는 다음과 같이 요약해 볼 수 있다. 첫째, 동일한 척도에 대한 실제 지원자 집단과 지원 상황을 가정한 실험 집단에서(즉, 응답왜곡 집단)의 척도별 응답 평균이 정직하게 응답한 통제 집단의 척도별 응답 평균보다 높게 나타났다. 둘째, 차별문항기능 분석을 통해, 기존에 응답왜곡의 지표로 흔하게 사용하고 있는 사회적 바람직성 척도들에는 응답왜곡을 적절히 탐지하지 못하는 문항들이 상당수 포함되어 있음이 나타났다. 셋째, 반면에 성격검사 척도 자체 내에는 응답왜곡에 민감하게 반응하는 문항들이 포함되어 있음이 나타났다. 넷째, 응답왜곡에 민감하게 반응하는 문항들은 직무별(운항승무원 직무와 생활설계사 직무)로 어느 정도 차이를 보임으로써 특정 직무에 따라 바람직한 방향으로 응답왜곡이 일어나는 직무 바람직성이 존재하는 것으로 보인다. 다섯 째, 이러한

직무 바람직성의 기존의 사회적 바람직성과 구별되는 구성개념이다. 여섯 째, 특정 직무별 응답왜곡을 탐지할 수 있는 문항들을 사용하여 피검자들의 능력 추정치를 수정하였을 때, 피검자들 사이의 서열에 변화가 나타나서 누가 채용되는 지의 채용 결정이 바뀌는 것으로 나타났다.

시사점 및 한계

본 연구의 학문적 시사점은 기존에 응답왜곡의 지표로써 사용되고 있는 사회적 바람직성과의 비교를 통해 응답왜곡을 보다 더 효율적으로 탐지할 수 있는 척도를 개발하고 성격 및 태도 변인들과의 관계성을 파악함으로써 검사에서 나타날 수 있는 개인차를 더욱 정교하게 이해하는 틀을 제공하였다는 것이다. 응답왜곡에 관한 논쟁은 공공연히 있었음에도 불구하고 아직 응답왜곡이 나타나는 과정에 대한 실체를 파악하려는 학자는 매우 적었으며 그에 대한 연구는 미흡한 실정이다. 즉, "성격검사에서 응답자의 진점수는 무엇인가?", "지원자들의 응답왜곡에 의해 성격검사의 예측력이 어느 정도 영향을 받는가?", 그리고 "각 직무마다 고유한 응답왜곡이 존재하는가?"와 같은 물음에 대답할 수 있는 구체적인 근거를 마련해 줄 것이다.

또한, 국내에서도 유태용과 민병모(2001)가 10개 연구의 통합분석을 통해 다양한 장면에서의 수행을 예측하기 위한 5요인 성격모델의 사용가능성을 살펴본 바 있지만, 실제 선발 상황에서 성격검사의 효용 및 응답왜곡에 관한 국내 연구는 찾기 힘든 형편이다. 이러한 상황에서 본 연구는 국내에서 인사선발 상황에서 성격검사를 사용할 때 나타나는 응답왜곡 현상을 이해할 수 있는 밑거름이 될 수 있을 것이다.

다음으로 실무적인 시사점으로는, 응답 왜곡 탐지 척도의 개발을 통해 인성검사에서 수검자의 응답 왜곡을 추정하여 절감하고 인사선발의 효율성을 증가시켜 직무수행을 더욱 정확하게 예측할 수 있다는 것이다. 이러한 노력은 능력검사의 사용과 더불어 인성검사를 사용함으로써 예측 타당성이 증가하게 되어 전체적인 선발의 타당도를 높일 수 있는 방안을 마련하는 셈이 된다. 본 연구의 결과는 인사 선발 장면에 있어 지원자들의 성격을 정확하게 측정하여 미래의 업무수행능력을 가늠할 수 있는 검사도구 개발에 활용될 수 있을 것이다. 또한 직무마다 응답왜곡의 패턴을 파악하여 결과를 제시할 수 있으므로, 경영자의 입장에서 선발을 원하는 직무에 최적한 인재를 가려낼 수 있을 것이다.

다른 한편으로는 이러한 시사점에도 불구하고 본 연구는 다음과 같은 한계를 가진다. 첫째, 각 실험 조건 별 조사 대상이 적어 연구 결과가 안정적이지 못할 수 있다. 특히 문항반응이론을 통해 모형 적합도를 검증하기 위해서는 교차 타당화 표본이 필요함에도 이러한 조건을 충족시키지 못했다.

또한, 이러한 표본의 문제로 인해 생활설계사 직무에 조건에서 대학생 집단과 실제 직무 지원자 집단간의 연령 차이가 상대적으로 크게 나타나, 결과 해석에 있어 유의해야 할 것으로 여겨진다.

둘째, 본 연구에서 사용한 하위척도들 간의 상관이 상당하여 본 연구를 통한 결과 해석에 주의를 기울이 필요가 있다. 이러한 상관은 실험 참가자들이나 직무 지원자들의 바람직성 응답 경향에 기인한 것일 수도 있으며, 그 밖의 여러 이유를 알수 없는 원인들에 의한 것일 수 있다. 이러한 높은 상관은 기존에 실제 지원자들이나 실험 참가자들을 대상으로 응답왜곡을 연구한 결과들과도 동일한 것이다. 그러나 본 연구의 분

석 과정의 일부로 모형-자료 부합도 검증을 하였을 때 각 척도는 다른 척도의 문항들이 합해졌을 때는 1차원성을 만족시키지 못하는 것으로 나타났다. 따라서 여러 요인들에 기인하여 하위 척도들간의 상관이 상당함에도 불구하고 본 연구에서 사용한 하위 척도들은 변별이 되는 것으로 여겨진다.

셋째, 실제 업무 수행을 비롯한 준거자료를 얻지 못하여, 성격검사의 외적 타당도에 대한 증거가 부족하다. 따라서 성격검사의 응답왜곡이 타당도에 미치는 영향을 체계적으로 살펴보는 데 한계를 갖고 있다. 또한, 선발 결정에 대한 성격검사의 응답왜곡의 효과를 조사하는 데 있어서도, 실제 조직에서는 성격검사 결과만을 근거로 선발이 이루어지는 경우는 극히 드묾에도 실제 채용된 사람들이 누구인지의 자료를 얻지 못하여 선발 비율에 따른 채용 결정에서의 변화를 가상적으로 살펴볼 수밖에 없었다.

넷째, 학생들을 대상으로 한 실험 결과에 대해서는 연구에서 나타난 차이를 일반화하여 이해하는 데 한계가 있다. 다섯 째, 본 연구에서는 차별문항기능 분석만을 시행하였는데 이와 동시에 차별검사기능(Differential Test Functioning) 분석을 시행한다면, 응답왜곡에 민감하게 반응하는 문항들 중 어떤 문항의 영향이 커서 검사 수준에의 결과에 영향을 미칠지를 고려할 수 있으므로, 기존의 척도 수준 연구들과의 연계점을 찾는 데 용이할 수 있었을 것이다. 끝으로, 본 연구에서는 직무에 따른 바람직성 응답의 차이를 단지 통계적인 분석기법을 통해서만 살펴보았으나, 직무별로 차이가 나타난 각 문항들에 대한 심층적인 내용분석이 이루어진다면 직무바람직성에 대한 보다 깊은 이해를 할 수 있을 것이다.

참고문헌

김신영 (2001). **차별적 문항기능의 추출방법**. 교육과학사.

박광배 (2000). **다변량분석**. 학지사.

유태용, 민병모 (2001). 다양한 장면에서 수행을 예측하기 위한 5요인 성격모델의 사용가능성과 한계: 국내 연구결과의 통합분석. **한국심리학회지: 산업 및 조직**. 14(2). 115-134.

Alliger, G. M., Lilienfeld, S. O., & Mitchell, K. E. (1995). The susceptibility of overt and convert integrity tests to coaching and faking. *Psychological Science, 7,* 32-39.

Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81,* 262-272.

Bass, B. M. (1957/). Faking by scales applicants on a forced choice personality inventory. *Journal of Applied Psychology, 41,* 403-404.

Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking of a biodata form: An analysis along several dimensions of item type. *Personnel Psychology, 45,* 389-406.

Christiansen, N. D., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology, 47,* 847-860.

Conger, A. J., & Jancson, D. N. (1972). Suppressor variables, prediction, and the interpretation of psychological relationships. *Educational and sychological Measurement, 32,* 579-599.

Costa, P. T., & McCrae, R. R. (1985). *The NEO*

Personality inventory manual. Odessa, FL: sychological Assessment Resources.

Cowles, M., Darling, M., & Skanes, A. (1992). Some characteristics of the simulated self. *Personality and Individual Differences, 13,* 501-510.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of Psychopathplogy. *Journal of Consulting Psychology, 24,* 349-354.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook. Vol 1.* Clinical interpretations.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1975). *An MMPI handbook. Vol 2.* Research applications.

Dicken, C. (1963). Good impression, social desirability, and acquiescence as suppressor variables. *Educational and Psychological measurement, 23,* 699-720.

Douglas, E. F., McDaniel M. A., & Snell, A. F. (1996). The validity of non-cognitive measures decays when applicants fake. *Academy of Management Proceedings.*

Drasgow, F. & Kang, T. (1984). statistical power of differential validity and differential prediction analysis for detecting measurement nonequivalence. *Journal of Applied Psychology, 69,* 498-508.

Drasgow, F., Levine, M. V., Tsien, S., Williams, B., & Mead, A. D. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement, 19,* 143-165.

Dunnette, M. D., McCartney, J., Carlson, H. C., & Kirchner, W. K. (1962). A study of faking behavior on a forced-choice self-description checklist. *Personnel Psychology, 15,* 13-24.

Dunnette, M. D., Paullin, C., & Motowidlo, S. J. (1989). Development and validation of the Kelly Applicant Profile for use in screening candidates for positions with Kelly assisted living services. MN.

Dwight, S. A., & Alliger, G. M. (1997, April). Using response latencies to identify overt integrity test dissimulation. In M. McDaniel (Chair), *Faking on non-cognitive measures: The Extent, impact, and identification of assimilation.* Symposium conducted at the meeting of the Society for Industrical and Organizational Psychology, St. Louis, MO.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden.

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,* 155-166.

Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists.* Hillsdale, NJ: Erlbaum

Eysenck, H. J., & Eysenck, S. B. G. (1968). *Manual for the Eysenck Personality Inventory.* San Diego, CA: Educational and Industrial Testing Service.

Fiske, S. T., & Taylor, S. E. (1991). Social cognition (2nd ed.). New York: McGraw-Hill.

Furnham, A. (1990). Response bias, social desirability and dissimulation. *Personality and individual differences, 7,* 385-400.

Gatewood, R. D., & Feild, H. S. (1994). *Human resource selection.* Orlano, FL: Dryden Press and

Harcourt Brace College Publichers.

Gomez-Mejia, L R., Balkin, D. B., & Cardy, R. L. (1995). *Managing human resources.* Englewood Cliffs, NJ: Prentice Hall.

Gough, H. G.(1975). *Manual for the Personnel Reacction Blank.* Palo Alto, CA: Consulting Psychologist Press.

Gough, H. G., & Heilbrun, A. B., Jr. (1980). *The Adjective Checklist manual.* Palo Alto, CA: Consulting Psychologists Press.

Heneman, H. G., III, & Heneman, R. L. (1994). *Staffing organizations.* Madison, WI: Mendota House.

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measuremnt and employment decisions: Questions and answers. *American Psychologist, 51,* 469-477.

Hough, L. M.(1998). Effects of Intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance, 11,* 209-244.

Hough, L. M., & Schneider, R. J. (1996). *Personality traits, taxonomies, and applications in organizations.* In K. R. Murphy(Ed.), Individual Differences and Behavior in Organizations(pp. 31-88). San Francisco: Jossey-Bass.

Hough, L. M., & Schneider, R. J. (1996). *The frontiers of I/O Personality Research.* In K. R. Murphy (Ed.), Individual differences & behavior in organizations (pp. 31-88). San Francisco: Jossey-Bass.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validity of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75,* 581-595.

Ironson, G. H., & Davis, G. A.(1979). Faking high or low creativity scores on the Adjective Check List. *Journal of Creative Behavior, 13,* 139-145.

Jackson, D. N. (1967). *Personality Research Form manual.* Goshen, NY: Research Psychologists Press.

Jackson, D. N., Wroblewski, V. R., & Achton, M. C.(2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance, 13,* 371-388.

Kamp, J. D. (1996). Personal communication, May 30, 1996.

Kluger, A. M., & Colella, A. (1993). Beyond the mean bias: The effect of warning against faking on biodata item variances. *Personnel Psychology, 46,* 763-780.

Kriedt, P. H., & Dawson, R. I. (1961). Response set and the prediction of clerical job performance. *Journal of Applied Psychology, 45,* 175-178.

Kroger, R. O., & Turnbull, W. (1975). Invalidity of validity scales: The case of the MMPI. *Journal of Consulting and Clinical Psychology, 43,* 48-55.

Lanning, K. (1989). Faking personality profiles on a standard personality inventory. *Personality and Individual Differences, 10,* 437-443.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107,* 34-47.

Levin, R. A. (1995). Self-presentation, lies and bullshit: The impact of impression management on emplpyee selection. *Presented at the annual meeting of Society of Industrial and Organizational*

Psychology, Orlando, FL.

Levine, M. V., & Drasgow, F. (2001). web version MODFIT 1.1. IRT Modeling Lab. http://work.psych.uiuc.edu/irt/

McCrae, R. R., & Costa, P. T. Jr. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51,* 882-888.

Meehl, P., & Hathaway, S. (1946). The K Factor as a supressor in the MMPI. *Journal of Applied Psychology, 45,* 76-79.

N. S. Raju & B. B. Eliis(2001). Differential item and test functioning. *Measuring and analyzing behavior in organizations.* 156-187. SF: Jossey-Bass.

Norman, W. T. (1963). Personality measurement, faking, and detection: An assessment method for use in personnel selection. *Journal of Applied Psychology, 47,* 225-241.

Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance, 11,* 245-269.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81,* 660-679.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L., (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance [Monograph]. *Journal of Applied Psychology, 78,* 679-703.

Paajanen, G. E. (1988). The perdiction of counterproductive behavior by individual and organizational variables. Unpublished doctoral dissertation, University of Minnesota.

Pauhlus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66,* 1025-1060.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598-609.

Reiss, S. P., & Waller, N. G. (2001). *Item response theory for dichotomous assessment data.* Measuring and analysing behavior in organization. 88-122. SF: Jossey-Bass.

Ross, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83(4),* 634-644.

Scarpello, V. G., Ledvinka, J., & Bergmann, T. J. (1995). *Human resource management.* Cincinnati, OH: South-Western College Publishing.

Schmit, M. J., & Ryan, A. M. (1992). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology, 78,* 966-974.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80,* 607-620.

Thissen, D. (2001). *IRTLRDIF v.2.0b User's guide.* Chicago, IL: Scientific Software, Inc.

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group

differences in trace line. In H. Wainer & H. Braun(Eds.), *Test Validity*. Hillsdale, NJ: Erlbaum, 147-169.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response model. In P. W. Holland & H. Wainer(Eds.), *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates, 67-113.

Williams, L. J., & Anderson, S. E. (1994). An alternative approach to the estimation of method effects by using latent variable models: Applications in organizational behavior research. *Journal of Applied Psychology, 79,* 323-331.

Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrumnent using approproateness measurement. *Applied psychological Measurement, 20,* 71-87.

Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology, 84,* 551-563.

Zickar, M. J., Rosse, J. G., Levin, R. A., & Hulin, C. L. (1996). Modelling the effecto of faking on personality tests. Unpublished manuscript.

1 차원고접수 : 2003. 5. 26
수정원고접수 : 2003. 8. 12
최종게재결정 : 2003. 8. 24

한국심리학회지: 산업 및 조직
Korean Journal of Industrial and Organizational Psychology
2003. Vol. 16, No. 2, 121-151

© Korean Society of Industrial and Organizational Psychology 2003

Job desirability as indicator of faking in personality test: Conceptualization and Impact on selection decision

Eun-jung, Lee Dong-Gun Park
Korea University

Although personality tests have many advantages over personnel selection, there are disputes about using social desirability scales as indicators of faking in selection context. This point is a critical issue to effective selection and performance prediction in trends that personality tests are widely used in selection. This study is based on the question whether social desirability can be the indicators of faking, purposed to understand faking in personality test, systematically. To do so, I examined the usefulness of social desirability scale in the faking research, found faking-prone items in Hogan Personality Inventory, investigated faking-prone items to be differed according to jobs, conceptualized Job Desirability constructs and explored whether it could be differentiated from Social desirability. In addition, I examined the effects of faking according to job desirability on the selection decision. This is based on the researches to reveal that faking has little impact on criterion validity but some impact on selection decision. This study shows four results. First, social desirability scales have many items disable to detect faking, Second, there are faking-prone items in personality test itself. Third, there are different faking-prone items in different jobs, so faking is occurred according to desirability of specific job. Fourth, Job desirability can be differentiated from Social desirability. Finally, when faking-prone items in personality test were used to estimate the abilities(θ) of subjects, there are change in rankings of person is occurred.

key words : Entrepreneur, Competency, Biodata, personality test