

평가형식 및 척도화 방법이 다면평가에서 관대화 오류 및 후광효과에 미치는 영향

김 명 소[†]

호서대학교 산업심리학과

이 광 진

ORP 연구소

본 연구는 다면평가 상황에서 평가자에게 제공되는 평가형식(리커트 척도 형식, 행동기술평정척도 형식) 및 척도화 방법(균형적 척도, 긍정적 어조의 불균형적 척도)이 평가의 관대화 오류와 후광효과에 어떠한 영향을 미치는지를 살펴보고자 하였다. 이를 위해, 현재 역량 다면평가를 실시하고 있는 A기업의 다양한 직군과 직위의 현직자 133명을 대상으로 자기·상사·부하·동료로 구성된 440명의 평가자들로 하여금 피평가자의 역량 수준을 측정토록 하였으며, 각 독립변인의 효과성을 검증하고자 각 수준에 따라 2×2 피험자간 실험설계를 실시하였다. 관대화 오류 분석방법으로 2×2 ANOVA와 왜도(degree of skewness)분석을 실시한 결과, 평가형식은 피평가자의 점수에 유의미한 영향을 미치지 않는 것으로 나타난 반면 척도화 방법의 경우 유의미한 차이를 보였다. 구체적으로, 긍정적 어조의 불균형적 척도는 각각 상대적으로 낮은 평가점수와 작은 음의 왜도지수를 지닌 것으로 나타나 관대화 오류가 더 적은 것으로 해석된다. 한편, 후광효과 분석방법으로 표준편차 비교 분석법과 상관비교 분석을 실시한 결과, 평가 형식에서 행동기술평정 척도는 상관비교분석을 사용한 경우에 리커트 척도와 비교해 유의미한 차이를 보이며 후광효과가 더 적은 것으로 나타났으나, 척도화 방법은 두 분석결과에서 서로 상반되는 결과를 보이며 그 효과를 입증하지 못했다. 또한 이 같은 결과를 토대로 본 연구의 의의 및 제한점과 함께 앞으로의 추후 연구과제에 대해 논의하였다.

주요어 : 다면평가, 평가형식, 척도화 방법, 관대화 오류, 후광효과, 평가오류 측정

[†] 교신저자 : 김명소, 호서대학교 산업심리학과, mskim@hoseo.edu

오늘날 조직들은 급변하는 조직 내·외부 환경변화에 발맞추어 인재 확보 및 육성을 조직의 비전과 목표 달성의 주요 과제로 설정하고, 선발, 평가, 보상, 육성 등 인사전반에 걸쳐 강도 높은 혁신을 실시하고 있다. 그러나 인적자원관리의 이름과 내용이 그 어떠한 것으로 바뀐다 하여도 인적자원관리의 성패는 정확하고 수용성이 높은 평가제도의 존재를 전제로 한다. 만일 피평가자가 평가시스템에 불만족을 표시하거나 또는 불공정하다고 인식한다면 평가결과를 수용하기 어려울 것이며, 이는 궁극적으로 근무의욕 상실, 직무 및 조직 만족도 저하, 팀웍 해체, 이직 등을 낳게 하는 주요 원인이 될 것이다. 또한 피평가자들이 평가가 불공정하고 부정확하다고 느끼는 정도만큼 평가결과를 보상 및 육성 등과 연계시키기란 그만큼 더 어려워 질 것이다. 이와 같은 상황을 배경으로 평가의 정확도와 피평가자인 구성원의 수용도를 높이기 위해 평가 내용의 변화, 평가자 훈련, 평가양식과 절차의 변화, 평가대상의 확산 등 여러 측면에서 기업들의 노력이 계속되고 있다.

특히, 이런 노력의 일환으로 평가주체에 따른 평가방식의 다양화가 추구되어 전통적으로 이루어지던 상사에 의한 평가에 제한되는 것이 아니라 다양한 평가주체를 포함하는 다면평가¹⁾제도의 활용이 급증하고 있다. 다면평가 제도는 수행평가, 승진, 임금 결정 등과 같은 인적자원관리(HRM)의 용도로만 사용되는 것이 아니라, 인적자원개발(HRD)의 측면에서도

소중한 자료가 될 수 있으며 최근 들어 교육 장면에서 진단의 목적으로 그 활용도가 높아지고 있다. 구체적으로, 다면평가제도는 평가자료가 전방위적으로 수집된다는 점에서 상급자에 의한 일방적인 평가와 구별이 된다(김판석, 오성호, 이선우, 1999). 기존의 상사에 의한 평가뿐만 아니라 자기평가 및 동료와 부하, 나아가 이해관계에 있는 기관이나 고객과 같은 외부인들의 평가까지도 포함하는 포괄적인 평가체계를 의미한다(Milkovich & Boudreau, 1994; 임창희, 1999). 즉 한 명의 평가대상자를 두고 평가대상자 자신을 포함하여 다수의 평가자가 다양한 수준에서 다양한 측면으로 평가하여 그 결과를 결합하는 평가방법을 의미한다. 일반적으로, 구체적인 평가절차는 조직마다 다르지만 공통적인 것은 대체로 피평가자를 잘 알고, 또 그 평가결과가 피평가자에 의해 가치 있는 것으로 받아들여질 수 있는 사람들에 의해 평가가 이루어진다는 것이다. 이와 같은 다수의 평가자가 평가에 참여할 경우에 상사가 혼자 평가할 때보다, 평가자의 개인적 선호경향, 편견, 고정관념과 같은 주관적 오류가 개입될 가능성이 적어질 것으로 기대된다.

그러나, 다면평가제도의 장점과 이를 수용하는 조직구성원의 긍정적인 반응과 함께 다면평가를 맹신해서는 안 된다는 목소리 또한 높아지고 있다. 이는 어느 평가방법들과 마찬가지로 평가의 본질상 평가의 대상도 사람이고, 평가의 주체도 사람이라는 측면에서 평가과정상 오류는 여전히 발생할 수 있고(Bernardin & Beatty, 1984), 또한 이러한 평가오류는 평가의 속성상 완전히 제거하기 어려운 것이 현실이다(Bretz, Milkovich, & Read, 1992). 따라서 다면평가를 통해 다수의 평가주체자로부터 평가

1) 국내의 경우, 다면평가 또는 360° 평가로 불리지만 외국에서는 360 degree feedback, multi-source assessment, multi-rater feedback, multi-perspective ratings, full-circle feedback 등으로 다양하게 표현되고 있다.

정보를 얻더라도 그 수집 과정이 잘못된다면 오류가 발생할 수 있고, 나아가 평가자의 조직 내 위치에 따라 달라지는 평가기준과 평가조망으로 인해서 평가의 신뢰성이나 일관성이 낮아 질 수 있다(박영석, 2003; Viswesvaran, Ones & Schmidt, 1996). 또한, 타인평가의 경우 일반적으로 상사보다는 동료, 동료보다는 부하 직원이 더 관대한 평가를 하는 것으로 나타난다(한영석, 김명소, 2007). 자기평가의 경우에는 다면평가의 용도에 따라서도 관대화 수준이 달라진다. 즉, 다면평가가 현재의 수준을 진단해보는 목적으로 사용될 경우에 자기평가에서 관대화 오류가 덜 나타나지만, 실제 수행평가 상황이나 선발 장면에서는 자기평가가 가장 관대화 경향이 높다. Harris와 Schaubroek(1988)는 자기평가의 관대화 오류에 대한 연구들에 대한 메타분석을 통해 자기평가의 경우 평균적으로 상사평가에 비교해서 표준편차 .50 정도 높고, 동료평가보다는 표준편차 .25정도 높다고 보고한 바 있다. 평가주체자의 직급과 평가의 용도에 따라서 평가결과가 달라지는 현상들은 측정에서 다수(multiple)의 의미가 항상 긍정적인 측면만을 내포하는 것이 아니며 다면평가의 신뢰도나 타당도가 위협을 받을 수 있다는 점을 시사해 주고 있다.

이러한 이유로 전통적인 수행평가를 대체해서 다면평가의 도입을 주장한 학자들마저도 관대화 오류와 같은 평가오류 문제를 지적한 바 있다(Thornton, 1980). 이 같은 문제를 반영하듯 최근 행정안전부²⁾는 ‘성과평가의 관대화 지수’를 개발하여 도입키로 결정하였다(문화일보, 2007. 7. 28). 고위공무원단 제도가 시행된

2) 행정안전부: 새정부 출범일인 2008년 2월 25일 이후 중앙인사위원회가 행정안전부로 통합되었음.

이후 처음으로 실시된 직무평가에서 83.5%가 ‘탁월’ 및 ‘우수’로 평정되고 ‘보통’은 15.4%였으며, ‘미흡’은 1.1%, ‘불량’은 0%로 평정됨으로써 그 동안 문제시 되었던 평가오류가 인사제도의 실효성을 위협하고 있는 것에 대한 제도적 장치를 마련한 것이다. 정부 조직 외에도 360도 다면평가를 실시하는 대부분의 기업에서 강제배분 척도(forced-distribution scale), 강제선택형 문항, 불균형 척도 등과 같은 다양한 방법을 활용하여 관대화 오류를 포함한 평가오류를 감소시키는데 관심을 기울이고 있다. 그 동안 많은 연구자들이 평가오류가 다면평가의 실효성 및 타당성에 미치는 영향에 대해 연구를 수행해왔는데, 이는 크게 ‘누가’ 평가하는지에 따라 발생하는 평가오류 즉, 평가주체(자기, 상사, 동료, 부하)간에 발생하는 평가오류와 어떠한 기준이나 측정방법(평가형식 및 척도)을 가지고 평가를 하느냐에 따라 발생하는 평가오류로 나누어 볼 수 있다. 평가주체 간에 발생하는 평가오류의 경우, 평가수준내 평가오류(예: Cooper, 1981; Lance & Woehr, 1986; Murphy & Anhalt, 1992)나 평가자간 일치도(예: Harris & Schaubroek, 1998; 장재윤, 1999)를 검토하는데 주로 초점을 두고 연구들이 진행되었다. 이들 연구에서는 공통적으로 특정 피평가자에 대한 여러 다면평가자들의 평가가 서로 일치하지 않는다는 결과가 보고된 바 있다(Levy, Cawley, & Foti, 1998; 한영석, 김명소, 2007).

평가기준이나 측정방법에 따른 평가오류의 경우, 서로 다른 평가형식 및 척도화 방법이 평가오류에 미치는 영향을 중심으로 연구되었다(Schwab, 1975; Finley, 1977; Mount, 1984; Fox, 1994; Ronen & Zisu, 1999; Jones & Fletcher, 2002). 평가주체에 따른 평가오류에 관한 연구

들은 국·내외 안팎에서 다양하고 활발하게 진행된 반면, 평가형식 및 측정척도가 평가자 오류에 미치는 영향에 관한 연구들은 상대적으로 적을 뿐 아니라 대다수가 다면평가 보다는 자기평가를 중심으로 연구되었다. 그 결과 또한 연구마다 서로 다르게 보고되어서 일관성 있는 연구결과가 도출되지 못한 실정이다. 예를 들면, 평가자의 책임을 강조하는 증거 기반 평가형식이 관대화와 후광오류를 줄이는데 기여한다는 연구결과도 있는 반면(Harris, 1994), 리커트 형식이 더 효과적이라는 결과도 제시되었다(Jones & Fletcher, 2002). 게다가 대다수가 국외에서 진행된 연구 결과이며, 국내에서는 아직까지 이 분야의 연구가 매우 미흡한 실정이다. 특히 국내에서는 평가오류 자체를 어떻게 측정, 평가할 것인지에 대한 방법론에 대해서도 아직까지 명확히 알려지지 않았다.

따라서 본 연구는 다면평가 상황에서 평가자에게 제공되는 평가형식과 척도화 방법이 평가오류에 어떠한 영향을 미치는지 실제 국내 A 기업의 현직자 역량에 대한 다면평가 자료를 바탕으로 살펴보고자 한다. 이를 위해 먼저 평가오류의 종류와 측정방법들을 검토하고, 평가형식과 척도의 종류와 관련된 연구결과들을 살펴보고자 한다.

평가오류

관대화 오류와 후광효과

평가오류는 한 개인이 다른 사람을 관찰하고 평정할 경우 체계적으로 발생하는 오류로, 편견이나 주관성 혹은 외생적인 영향을 받지 않은 객관적인 평가결과와 실제 인간의 부정확한 판단과정에 의한 평가결과와의 차이를

의미한다(Feldman, 1981). 현재까지 알려진 평가오류로는 대비효과(contrast effect), 첫인상 효과(first impression effect), 후광효과(halo effect), 유사성 오류(similar-to-me error), 관대화 오류(leniency error), 엄격화 오류(severity error), 중심화 경향(central tendency), 최신오류(recency error), 논리적 오류(logical error) 등이 있다. 이와 같은 오류들 중 가장 대표적인 것은 관대화 오류와 후광효과인데, 이는 보통 고정오류(constant-errors)로 불리며, 평가되는 차원들간에 잘못된 유사성이 반영될 경우 발생된다.

특히, 관대화 오류는 평가오류에 관한 기존 연구들에서 가장 많이 다루어졌으며, 여러 학자들에 의해서 개념화되었다. Motowidlo와 Borman(1977)은 관대화 오류(leniency error) 또는 관대화 경향(leniency tendency)이란 피평가자의 능력이나 성과를 실제보다 더 높게 평가하는 것으로 엄격화 경향(severity tendency)과 반대되는 개념으로 정의하였으며, Saal, Downey, 및 Lahey(1980)는 평가시 척도의 중앙치 또는 진점수로부터 일관되게 높게 평가하는 것을 관대화 오류로 정의하였다. 또한, DeCotiis(1977)에 의하면 관대화 오류란 진정한 성과수준으로 측정된 객관적인 기준보다 일관적으로 높게 평가하는 평가자에 기인하는 반응세트(response set)이다. Bernardin, Alvares, 및 Cranny(1976)는 평가자의 평가점수 평균이 평가 척도상에서 중간지점에서 양적인 방향으로 이동하는 경향이라고 정의한 바 있다. 이밖에도, 관대화 오류는 다른 평가자들이 부여한 점수 혹은 다른 측정방법에 의해서 도출된 점수보다 높게 평가된 평정치를 의미한다(Jones & Fletcher, 2002). 이 같은 관대화 오류는 평가 상황에서 가장 흔히 나타나는 오류로서 평가 결과를 왜곡시키는 위험을 안고 있으며(Bretz,

1992), 실제로 다면평가가 확산되면서 기업의 입장에서는 평가의 변별력(즉, 피평가자간 개인차 측정 능력) 저하로 인해 가장 큰 문제점으로 지적되어 왔다.

관대화 오류의 발생원인은 크게 두 가지 측면에서 살펴볼 수 있다. 첫 번째는 자기평가시 발생하는 관대화 오류이며, 두 번째는 타인평가시 발생하는 관대화이다. 자기평가시 발생하는 관대화의 경우, 가장 대표적인 설명은 평가시 자기방어심리가 발동하여 자신에 대해 보다 더 높은 점수를 부여한다는 것이다(Steel & Ovalle, 1984). 자기개념 연구자들에 의하면 인간은 스스로 가치 있게 여기는(self-worth) 느낌을 유지하거나 증진하려고 한다. 그 결과 성격검사와 같은 정답이 없는 도구의 경우 대체적으로 사람들은 타인이 자신을 평가하는 것보다 자신을 더 좋게 평가하려는 경향을 보인다(Campbell & Fehr, 1990). 한편, 타인평가시 발생하는 관대화의 경우 자기평가시 발생하는 관대화보다 의도적인 경향이 있다. 피평가자에게 낮은 점수를 부여할 경우 발생할 수 있는 충돌 즉, 인간관계의 저해와 팀워크의 상실, 타인의 처우와 사기 등에 미치는 영향을 고려하여 후하게 평가한다고 볼 수 있다(Longenecker, Gioia, & Sims, 1987). 또한 박경규(2001)는 타인(부하)평가 상황에서, 첫째 부하를 낮게 평가하여 갈등이나 대립을 초래할 필요가 없는 것, 둘째 자기부하를 타부서의 구성원과 비교해 승진에 유리하게 하기 위한 것, 셋째 낮은 평가가 곧 평가자 자신의 책임으로 간주될 수 있다는 점을 관대화 오류의 주된 원인으로 제시하였다.

또 하나의 대표적 평가오류인 후광효과(halo effect)도 다양한 평가주체나 평가도구들을 대상으로 연구되었으며, 평가자가 피평가자의

단편적인 특성에 대한 지식을 기준으로 모든 특성에 대해 전반적으로 좋거나 또는 나쁘게 평가하는 경향을 의미한다. 하지만 후광효과에 대한 개념적 정의는 학자들마다 각기 다른 입장을 보여 왔다. Borman(1975)은 후광효과에 대해 평가자가 피평가자를 다양한 성과차원의 수준에 따라 구별하지 않고 전반적인 인상을 통해 평가하는 경향이라고 정의하였으며, Bernadin(1974)은 서로 다른 차원에 대해 피평가자를 동일한 수준으로 평정하는 경향이라 하였다. 또한, Latham과 Wexley(1981)는 피평가자 직무성과의 단편적인 측면을 다른 모든 직무성과차원으로 부적절하게 일반화하는 경향으로 설명하였다.

일반적으로 후광효과는 평가자가 피평가자의 인적특성이나 부정적인 요소에 비중을 두고 평가할 때 과도하게 발생하며(Fiscaro, 1988), 피평가자들이 역으로 인상관리를 통해 평가자에게 긍정적이거나 부정적인 인상을 만들으로써 의도적으로 후광효과를 조장할 수도 있다(Wayne & Ferris, 1990). 이러한 후광효과가 발생하면 평가도구의 하위척도간 변별력을 떨어뜨리는 결과를 낳고, 그 결과 한두 가지의 하위척도만을 사용할 경우보다 심리측정적 특성(즉, 신뢰도나 타당도)이 개선되지 않는다. 다시 말해서, 하위척도들이 중복적인 정보만을 제공해줌으로써 다수의 문항이나 척도를 사용하는 것에 대한 타당성을 확보할 수 없게 된다. 실무에서는 다면평가시 개별 평가자가 평가해야 할 대상자가 증가하게 되므로 평가도구의 길이가 중요한 이슈이다. 또한, 직무 혹은 직군 배치나 육성의 측면에서 개별 구성원의 특성(혹은 역량)의 프로파일과 장/단점을 정확하게 진단하는 것이 매우 중요한데, 후광효과가 높으면 이러한 분석이 어렵게 된다.

따라서 후광효과는 관대화 오류와 함께 간과할 수 없는 오류에 해당한다.

측정 방법

관대화 오류를 측정하는 방법은 크게 세 가지 방법, 즉, 평균, 변량분석, 왜도(skewness)로 나뉜다. 이들 방법 중 가장 자주 사용되는 것은 측정요소의 평균값을 평가척도의 값과 비교하는 것으로(Bernardin et al., 1976), 측정요소의 평균점수가 평가척도의 중간점(예: 5점 척도에서 중간점은 3점)을 넘으면 관대화로 간주하고 중간점보다 낮으면 엄격화로 간주하는 것이다. 그러나 이 같은 방법은 중간점이 실제 측정요소의 중간수준을 잘 대표하는지에 대한 문제점을 지니고 있다. 만일 측정요소의 중간 수준이 평가척도의 중간점보다 높다면, 평가점수가 평가척도의 중간점보다 높다하더라도 관대화가 아닐 수 있기 때문이다.

두 번째 방법은 평가자(또는 다른 독립변인) × 피평가자 × 평가차원의 삼원 변량분석 방법에 기초한다. 여기서 평가자(또는 다른 독립변인)의 주효과가 통계적으로 유의할 경우, 특히 종속변인인 평가점수 분산의 많은 부분을 설명한다면 이는 관대한 경향이 있는 것으로 간주된다(Friedman & Comelius, 1976). 예를 들어, 평가주체(자기, 상사, 부하, 동료)에 따른 관대화 오류를 비교·분석하는데 있어 처치변인인 평가주체에 따라 동일 피평가자의 평가 점수가 유의한 차이를 보일 경우, 특히 처치변인인 평가주체가 피평가자의 점수차이(분산)에 대해 큰 설명량을 지니면 평가주체에 따라 상대적인 관대화 경향이 존재하는 것으로 간주된다. 물론, 평가형식이나 평가훈련과 같이 관대화에 영향을 미칠 수 있는 변인은 모두 독립변인으로 포함될 수 있다.

세 번째 방법은 차원 평가점수의 분포에 대한 왜도를 평가하는 방법이다(Landy, Farr, Saal, & Freytag, 1976). 왜도는 분포가 편포된 정도에 대한 측정치이며, 정적편포일 경우 그 끝이 왼쪽으로 치우친 것이고 부적편포일 경우 오른쪽으로 치우친 것이다. 여기서 유의한 부적편포는 관대화 오류로 간주되며, 정적편포는 엄격화 오류로 간주된다. 그러나 왜도의 경우 통계적 유의미성을 측정하는 방법이 없다는 제한점이 있어서, 해석상 주의가 기울여야 한다.

한편, 후광효과를 측정하는 방법은 측정치에 대한 측정 방법에 따라 상관관계분석 및 요인분석, 분산(표준편차)분석, 변량분석(ANOVA)으로 나뉜다. 첫 번째 상관분석 방법은 구체적으로는 다속성-다방법(Multitrait-Multimethod: MTMM)이라 불리는데, 피평가자가 받은 점수를 각 평가차원의 점수로 구분하여 상관관계를 검토하는 것이다(Keaveny & McGann, 1975). 여기서 동일방법-이질속성(Monomethod-Heterotrait)에서의 상관관계가 높을 경우 다른 차원의 행동을 차별화하지 못한 것을 나타내며, 따라서 후광효과가 존재한다고 해석한다. 그러나 이는 진정한 후광효과가 존재하는 경우와 실제로 후광효과가 아닌 경우를 명확히 구별해내지 못하는 단점이 있다. 예를 들어, 실제로 두 평가 차원에서 모두 높은 수행수준을 보이는 경우 후광효과로 간주될 수 있다. 또한 상관관계는 두 변인이 같은 방향으로 변하는 정도, 즉 선형결합의 정도만을 나타내기에 한 차원에서 평가 점수가 일관성 있게 낮고 다른 차원에서는 평가점수가 일관되게 높아도 후광효과와 오류로 간주될 수도 있다. 그러나 이와 같은 문제점에도 불구하고 아직까지 상관분석 방법은 후광효과를

측정하는데 있어 가장 널리 쓰이며 방법 중에 하나이다.

후광효과를 측정하는 두 번째 방법은 차원의 상관행렬에 대한 요인분석이다(Kraut, 1975). 구체적으로, 요인의 수가 적게 나타날수록 후광효과가 더 많이 나타난다고 간주한다. 그러나 이 방법은 첫 번째 방법인 상관관계에 대한 행렬을 기본적인 분석 자료로 이용하기 때문에 앞서 제시된 상관관계의 문제점을 그대로 안고 있다.

세 번째 방법은 피평가자가 받은 각 차원의 분산(표준편차)에 초점을 맞추고 있다(Bernardin & Walter, 1977). 즉, 평가자(또는 다른 독립변인)를 통해 얻어진 피평가자의 각 수행차원 점수들의 분산을 산출하고, 각 수행차원 점수간에 분산이 작으면 후광효과가 존재한다고 간주한다. 그러나 이 같은 방법은 수행의 실제수준(평가점수)을 이용하지 않고 이들의 분산을 이용함으로써 실제 수행수준을 고려하지 못하는 문제점을 지니고 있다.

네 번째 방법은 평가자 × 피평가자 × 평가차원의 삼원변량 분석에 기초한 것으로(Dickinson & Tice, 1977), 평가자 × 피평가자의 상호작용이 통계적으로 유의할 경우 후광효과가 있는 것으로 간주된다. 특히, 평가점수 분산의 많은 부분을 설명하는 경우 후광효과가 큰 것으로 해석한다. 이는 피평가자들의 평가 점수 차이가 평가차원에 따라 구분지어지는 것이 아니고, 평가자와 피평가자의 관계에 따라 전반적으로 측정차원을 동일하게 평가하는 지표(측정차원간 분산 및 상관)에서 유의한 차이를 보일 경우 상대적인 후광효과가 존재하는 것을 의미한다.

평가형식 및 척도화 방법

평가형식

리커트 척도 형식은 응답자들을 직접 척도화하는 응답자 중심(subject-centered)의 접근 방법이며, 일차원성 또는 일개념성의 속성을 띠고 각 문항은 측정하려는 속성에 대해 긍정이나 부정적의 태도를 나타내는 문항들로 구성되어있다(김명소, 1992). 리커트 척도 형식은 전형적 수행 검사에서 가장 많이 활용되는 방법으로, 일반적으로 하나의 구성개념에 대해 단편적이며 단일 차원적 속성을 띠는 세부적인 문항들로 구성되고 최종적으로 이를 합산해 구성개념을 측정한다.

Jones와 Fletcher(2002)는 평가형식에 따른 평가오류를 분석하는 연구에서 영국 공무원 선발 시 측정하는 역량(competency)을 단일 차원적 속성을 대표하는 요인들로 보다 세부적으로 분류시켜 측정하고 이를 합산하는 리커트 척도 형식을 이용한 바 있다. 이들 연구자들은 또한 리커트 척도 형식과의 비교 목적으로 ‘증거 기반 역량측정(Competency/evidence)’ 방법을 함께 사용하였다. 이는 피평가자의 역량수준에 대한 평가를 할 때 특정 점수에 대한 증거(즉, 해당 역량과 관련된 구체적 행동을 관찰하고 이를 기록함)를 제시하게 하는 방법으로, 평가자의 평가에 대한 책임감을 증진시켜 주는데 초점을 두고 있다. 본 연구에서는 평가 장면에서 평가자가 관찰한 행동을 기록하고 이와 같은 증거를 바탕으로 피평가자를 평가한다고 하여, 이를 행동기술평정척도(Behavioral description rating scale)라 명명하였다. 현재 국내의 경우 고위공무원단 역량평가를 위해서 이와 같은 방법이 활용되고 있다.

Harris(1994)는 평가자 동기모델에서 평가 상황에서 평가자의 태도는 평가결과에 영향을 미칠 수 있기 때문에 구체적으로 평가에 대해

책임감을 느끼고 진지하게 평가에 임하는 동기화된 평가자의 경우 피평가자를 좀 더 분석적이며 체계적으로 바라보고(deliberate processing), 또 평가되는 각각의 차원(능력 및 역량)에 대해서도 좀 더 면밀히 세부적으로 살펴볼 것이라고 주장하였다. 반면에 동기가 낮은 평가자의 경우 성급하게 휴리스틱(heuristic)에 기반한 접근방법을 사용하여 단순한 의사결정을 하는 경향이 있다. 이와 관련하여 Cooper(1981)는 평가자의 평가에 대한 책임감과 주의집중은 서로 다른 평가차원들에 대한 구별을 용이하게 하며, 그 결과 후광효과를 줄일 수 있다고 주장한 바 있다.

Jones와 Fletcher(2002)는 위에서 언급한 바와 같이 선발상황에서 Harris의 평가자 동기모델을 이용하여, 자기평가시 평가점수에 대한 근거를 제공하는 행동기술평정척도가 평가자에게 평가에 대한 책임감을 부여해 리커트 형식을 사용할 경우 보다 평가오류를 감소시킬 것이라는 가설을 세웠다. 그러나 결과는 오히려 반대로 나타나 Harris의 평가자 동기모델이 지지받지 못했다. 즉, 평가자에 평가역량에 대해 세부적인 행동기술문들이 제시되고 이에 대한 평정치를 합산하는 방식을 취하는 리커트 형식이 오히려 관대화 오류를 감소시키는데 효과적인 것으로 나타났으며, 후광효과는 평가형식 간에 유의미한 차이를 보이지 않음에 따라 그들이 제시한 가설은 모두 지지되지 못했다.

척도화 방법

일반적으로 제공되는 대다수 평가척도는 호의적인 응답과 비호의적인 응답의 수가 동일하게 제시되며 척도의 중간점수를 중심으로 양측이 대칭되는 균형적인 형태로 구성된다. 반면에 불균형적 척도는 중간점수를 중심으로

균형을 이루기보다는 한 측이 좀 더 강한 혹은 약한 어조의 응답으로 구성된다. 5점 척도를 예로 들 때, 세 개의 호의적인 응답과 한 개의 중립적인 응답, 한 개의 비호의적인 응답을 지니는 형태로 구성된다(Friedman & Amoo, 1999). Kasten과 Weintraub(1999)는 평가결과가 부정편포를 구성할 것으로 예측될 경우(즉, 피평가자들의 평가점수가 대부분 높을 경우), 긍정적 어조의 불균형적인 척도를 사용하면 피평가자간 개인차를 증가시킬 수 있다고 제안한 바 있다. 즉, 대부분의 점수가 몰려있는 중간 혹은 중상 지점의 점수들을 좀 더 낮은 점수로 이동시킬 수 있는 기능을 함으로써 관대화 경향이나 후광효과를 감소시키는 효과가 기대된다.

몇몇 연구자들은 피험자 응답에 영향을 미치는 척도 기술문의 내용(anchor)과 위치(position)에 대해서 연구를 수행했는데, Friedman과 Leefter(1981)의 연구에서 피험자들은 척도에 대한 기술문이 어디에 위치하는지(예를 들면, 1, 2, 3, 4, 5점 점수 자체) 보다는 실제로 이용되는 기술문의 내용을 보고 응답한다는 것이 확인되었다. 구체적으로 동일한 문항에 각각 척도의 위치와 이에 해당하는 기술문을 바꿔 설문을 실시하고 비교분석한 결과, 척도의 위치가 바뀌었더라도 피험자들은 자신이 동의하는 기술문의 내용을 바탕으로 설문에 응답하는 모습을 보였다. 이러한 결과는 평가체계와 형태가 수행평가 오류에 유의한 영향을 미칠 수 있다는 가능성을 시사하고 있다.

Fox, Caspy 및 Reisler(1994)는 이스라엘 경찰관을 대상으로 평가센터를 통해 관리직 역량을 평가하는 상황에서 긍정적 어조의 불균형적 척도와 균형적 척도가 평가오류에 미치는 영향에 대해 경험적인 검증을 시도하였다. 구

체적으로, 긍정적인 어조의 불균형적인 척도(1=다른 사람들에 비해 다소 부족하다; 2=다른 사람들과 같다; 3=다른 사람들 보다 다소 낫다; 4=다른 사람들에 비해 월등히 낫다; 5=어느 누구보다도 우수하다)를 사용했을 때 균형적 척도(1=다른 사람에 비해 크게 부족하다; 2=다른 사람들에 비해 다소 부족하다; 3=다른 사람들과 같다; 4=다른 사람들보다 다소 낫다; 5=다른 사람들에 비해 월등히 낫다)를 사용한 경우보다 일관적으로 관대화 오류뿐만 아니라 후광효과가 낮게 나타났다.

또한, Fox와 동료들의 연구(1994)를 바탕으로 유사한 연구를 수행한 Jones와 Fletcher(2002)도 이와 동일한 결과를 보고한 바 있다. 이들 연구자에 의하면 공무원 선발과정으로 균형적 척도와 불균형적인 척도를 이용한 상황에서 피험자들의 자기평가 점수를 분석한 결과, 척도의 중간점을 보다 하향화시킨 긍정적 어조의 불균형 척도의 경우 균형적 척도와 비해서 상대적으로 보다 낮은 평정점수를 보임에 따라 관대화 오류가 적은 것으로 해석했다. 후광효과의 경우, 여성 평가자 집단에서만 불균형 척도를 사용할 때 평가차원들 좀 더 독립적으로 평가하는 경향을 보였다.

연구 목적 및 가설

위에서 기술한 바와 같이, 그 동안 평가형식과 척도화 방법이 평가오류에 미치는 영향에 관련된 연구들은 대체로 선발장면에서 자기평가를 중심으로 진행되었다. 또한, 평가형식과 척도화 방법이 각각 별도로 연구되어 왔다. 이에 본 연구는 이들 연구결과들을 실제 기업의 역량평가 용도로 실시되고 있는 다면평가 상황에 적용하여(해당 기업의 역량평가

는 전체 업적평가의 30%를 차지함), 평가형식과 척도가 관대화 오류와 후광효과에 어떠한 영향을 미치는 살펴보는데 그 목적을 두고 있다. 대부분의 기존 연구들에서는 한 연구에서 관대화 오류나 후광효과를 측정하는 방법으로 한 가지 방법만 사용하였으나 본 연구에서는 각각의 평가오류를 측정하는데 두 가지 측정 방법들을 사용하여 각 오류 결과를 교차적으로 확인하고자 한다.

구체적으로, Jones와 Fletcher(2002)는 Harris(1994)가 제안한 평가자 동기모델을 바탕으로 증거기반 측정 혹은 행동기술평정척도(Behavior description rating scale)가 평가자의 평가에 대한 책임감을 향상시켜 평가오류를 감소시키는 역할을 할 것이라는 가설을 세웠으나 이를 입증하는 데는 실패하였다. 이에 본 연구는 동기모델의 효과를 재차 규명하고자 다음과 같은 가설을 세웠다.

가설 1. 관대화 오류는 다면평가에서 평가자에게 제공되는 평가형식에 따라서 달라질 것이다. 구체적으로 리커트 형식이 제공되는 상황보다 행동기술평정척도 형식이 제공되는 상황에서 관대화 오류가 적게 나타날 것이다.

가설 2. 후광효과는 다면평가에서 평가자에게 제공되는 평가형식에 따라서 달라질 것이다. 구체적으로 리커트 형식이 제공되는 상황보다 행동기술평정척도 형식이 제공되는 상황에서 후광효과가 적게 나타날 것이다.

또한, Jones와 Fletcher(2002), Fox와 동료들(1994)은 자기평가 상황에서 긍정적인 어조의 불균형 척도가 관대화와 후광효과를 감소시킨다는 결과를 보고하였다. 이에 본 연구는 앞서 자기평가에서 나타난 결과를 다면평가 상

황에 적용하여 이를 확인코자 하며, 구체적인 가설은 다음과 같다.

가설 3. 관대화 오류는 다면평가에서 평가자에게 균형적인 척도가 제공되는 상황보다 긍정적인 어조의 불균형인 척도가 제공되는 상황에서 보다 적게 나타날 것이다.

가설 4. 후광효과는 다면평가에서 평가자에게 균형적인 척도가 제공되는 상황보다 긍정적인 어조의 불균형인 척도가 제공되는 상황에서 보다 적게 나타날 것이다.

방 법

조사 대상자 및 연구절차

현재 역량 다면평가를 실시하여 수행평가 및 승진에 반영하고 있는 A기업의 다양한 직군과 직위의 현직자 133명을 대상으로 자기·상사·부하·동료로 구성된 440명의 평가자들로 하여금 30개 역량에 대한 다면평가를 실시하였다(표 1). HR 팀의 주관 하에 2007년도 5월에 일주일간 사내인트라넷을 이용해 평가질문지를 송부·수거하는 형식으로 평가가 진행되었다.

연구 설계

다면평가 상황에서 평가형식과 척도화 방법이 평가결과에 어떠한 영향을 미치는지를 살펴보기 위해 본 연구는 피평가자와 평가자로 구성된 440명의 피험자집단을 각각의 독립변인 수준에 따라 배정한 2(평가형식) × 2(평가척도) 피험자간 설계를 실시하였다(표 2).

도 구

본 연구는 A기업의 역량모델링 결과 도출된 5개 역량군의 30개 핵심역량을 측정하는 문항들을 개발한 후, 이를 불균형적 척도의 리커트 형식, 균형적 척도의 리커트 형식, 불균형적 척도의 행동기술평정척도 형식, 균형적 척도의 행동기술평정척도 형식으로 구성하였다. 각 평가형식 및 척도화 방법에 따른 문항들의 신뢰도(α)는 리커트 형식의 경우 역량요인별 신뢰도를 이용해 측정하였으며, 행동기술평정척도 형식은 문항(혹은 척도)이 각각 측정역량별로 한 개씩 제시되었기에 역량군별 신뢰도 분석을 실시하였다. 불균형적 척도의 리커트 형식은 .60~.87, 균형적 척도의 리커트 형식은 .54~.90, 불균형적 행동기술평정척도 형식은 .66~.87, 균형적 행동기술평정척도 형식은 .62~.79로 전반적으로 양호한 신뢰도 수준을 보였다.

독립변인의 조작

구체적으로 두 개의 독립변인에 따른 네 처지집단에 활용된 척도들은 다음과 같다. 첫째, ‘불균형적 척도의 리커트 형식’은 각 역량별로 구성된 5개의 행동 기술문(총 150 문항)에 대한 동의도를 묻는 방법으로 긍정적 어조의 불균형적 기술문(1=그렇지 않다; 2=보통이다; 3=그렇다; 4=많이 그렇다; 5=매우 그렇다)을 이용하여 평가하도록 구성되었다. 두 번째, ‘균형적 척도의 리커트 형식’ 역시 각 역량별로 구성된 5개의 행동 기술문에 대한 동의 정도를 평가하도록 하였지만 평가척도는 일반적인 균형적 기술문(1=전혀 그렇지 않다; 2=그렇지 않다; 3=보통이다; 4=그렇다; 5=매우 그렇다)을 이용하여 평가하도록 구성되었다.

표 1. 연구 참가자들의 인구통계학적 특성(결측치=미응답자)

피평가자	직급	사원	대리	과장	차장	연구원	주임 연구원	선임 연구원	결측치		
<i>n</i> =133		22	34	30	8	1	29	8	1		
자기	직급	사원	대리	과장	차장	연구원	주임 연구원	선임 연구원			
<i>n</i> =121		22	31	29	6	1	26	6			
평균 근무 기간: 4년 1개월											
상사	직급	사원	대리	과장	차장	부장	주임 연구원	선임 연구원	책임 연구원	결측치	
<i>n</i> =115		3	1	43	27	4	1	28	5	3	
평균 근무 기간: 7년 7개월											
피평가자와 근무한 기간: 2년 5개월											
평 가 자	동료	직급	사원	대리	과장	차장	연구원	주임 연구원	선임 연구원	결측치	
	<i>n</i> =110		11	30	28	4	1	28	5	3	
평균 근속 년수: 3년 11개월											
피평가자와 근무한 기간: 2년 1개월											
부하	직급	사원	대리	과장	차장	연구원	주임 연구원	선임 연구원	결측		
<i>n</i> =94		36	18	6	13	20	1				
평균 근속 년수: 2년 2개월											
피평가자와 근무한 기간: 1년 4개월											
전체	직급	사원	대리	과장	차장	부장	연구원	주임 연구원	선임 연구원	책임 연구원	결측치
<i>n</i> =440		72	80	106	37	4	15	75	39	5	7
평균 근속 년수: 4년 7개월											
피평가자와 근무한 기간: 1년 11개월											

우 그렇다)을 이용하였다.

세 번째, ‘불균형적 행동기술평정척도 형식’은 측정되는 역량에 대한 정의와 함께 긍정적 및 부정적 행동의 직무행동 예시문이 각각 5

개씩 주어지고 평가자로 하여금 해당 역량에 대한 관찰 직무행동을 작성하고 행동의 수와 질적 수준을 바탕으로 피평가자의 역량수준을 평가하도록 하였다. 척도화 방법은 긍정적 어

표 2. 독립변인 수준에 따른 실험참가자 인원수

평가형식	평가척도		전체
	불균형적 척도	균형적 척도	
리커트 척도	125	73	198
행동기술평정척도	136	106	242
전체	261	179	440

조의 불균형적 척도(1=긍정적 행동은 적고 부정적 행동이 많은 경우; 2=긍정적 행동과 부정적 행동의 수가 비슷한 경우; 3=부정적 행동은 적고 긍정적 행동이 많은 경우; 4=대부분 긍정적 행동인 경우; 5=모두 긍정적인 행동의 경우)를 이용하였다. 마지막으로 ‘균형적 행동기술평정척도 형식’ 역시 불균형적 척도의 행동기술평정척도와 같은 평가형식을 취하였으나, 평가척도는 3점을 기준으로 대칭을 이루고 있는 균형적 척도(1=거의 모두 부정적 행동인 경우; 2=긍정적 행동은 적고 부정적 행동이 많은 경우; 3=긍정적 행동과 부정적 행동의 수가 비슷한 경우; 4=부정적 행동은 적고 긍정적 행동이 많은 경우; 5=거의 모두 긍정적 행동인 경우)를 이용하였다.

종속변인의 측정

관대화 오류를 측정하는 방법으로 독립변인 수준별로 평정점수를 비교하는 변량분석법이 이용되었으며, 추가적으로 각 독립수준별 평정점수의 분포 형태를 활용하는 왜도(degree of skewness)분석법을 적용하여 변량분석에서 도출된 결과를 확인하고자 하였다. 구체적으로, 변량분석법의 경우 A기업의 30개 역량에 대한 피평가자의 역량수준에 대해 자기·상사·부하·동료로 구성된 다면평가자가 부여한 평정

점수의 평균값을 이용하였다. 여기서 평균점수를 사용한 이유는 A기업에서 실제로 평균점수를 활용해서 역량평가를 하고 있고, 각 독립변인 상황에서 동일 피평가자에 대한 평가주체별(자기, 상사, 부하, 동료) 평가점수를 비교해본 결과도 평균값을 이용했을 때와 평가오류 발생 증감정도에 있어 유사한 패턴을 보였기 때문이다. 또한, 본 연구의 주요 초점이 평가주체가 아닌 평가형식 및 척도화 방법인 점도 평균값을 사용한 이유 중에 하나이다. 본 연구에서는 Jones와 Fletcher(2002)의 관대화 오류 정의에 기반하여, 변량분석법에서 평가점수의 평균치가 클수록, 왜도분석에서는 부적적으로 편포되어 있을수록 상대적으로 관대화 오류가 나타난 것으로 간주한다.

후광효과를 측정하기 위해서 가장 널리 이용되고 있는 표준편차 분석과 다속성-다방법을 사용하였다. 표준편차 분석법은 독립변인 수준별로 표준편차를 산출하고 이에 대한 유의성 검증을 통해 표준편차가 상대적으로 크게 나타날 경우 후광효과가 적게 나타났음을 의미한다. 두 번째로 다속성-다방법 상관행렬 중 동일방법-이질속성에서의 내적상관(inter-correlation)을 산출하여 독립변인 수준별로 속성간 상관계수를 산출하고 각 처치집단간 평균상관의 유의성을 검증하였다. 상대적으로 상관이 낮은 경우, 이는 후광효과가 적게 나

타난 것으로 해석하였다. 위와 같은 모든 분석에 SAS 9.1(이종구, 2001)이 사용되었다.

결 과

평가 형식 및 척도화 방법에 따른 관대화 오류 분석

평가형식과 척도에 따른 관대화 오류를 측정하기 위해 각 독립변인 수준별로 A기업의 30개 역량에 대한 평균 평정점수를 비교하고자 변량분석을 실시한 결과, 표 3 및 4와 같이 나타났다. 평가형식은 피평가자의 점수에 유의미한 영향을 미치지 않는 것으로 나타난 반면, 척도화 방법의 경우 유의미한 차이를

보였다. 즉, 평가자들은 균형적 척도(M=3.91)를 사용할 때보다 불균형 척도(M=3.73)를 사용할 때 상대적으로 낮은 점수를 부여하는 것으로 나타났다. 이는 다면평가 상황에서 균형적인 척도보다 긍정적인 어조의 불균형 척도가 제공되는 상황에서 관대화 오류가 상대적으로 적게 나타날 것이라는 가설 3을 지지한다. 반면에 평가형식에 있어 리커트 형식이 행동기술평정척도 형식과 비교해 다소 낮은 점수를 보였으나 변량분석 결과 유의한 차이를 보이지 않음에 따라 관대화 오류에 별다른 영향을 미치지 않는 것으로 밝혀졌다. 물론 일반 기업에서 흔히 관찰되는 것처럼 모든 처지 상황에서 척도의 중간점수인 3점과 비교해 그 평균적인 수준(3.7이상)이 전반적으로 상향 평가되었음을 알 수 있다. 따라서, 본 연구에

표 3. 평가형식과 척도에 따른 평정점수의 평균과 표준편차

평가형식	평가척도				전체	
	불균형적 척도		균형적 척도			
	평균	표준편차	평균	표준편차	평균	표준편차
리커트 척도	3.76	.31	3.83	.29	3.79	.3
행동기술평정척도	3.71	.35	3.97	.26	3.83	.34
전체	3.73	.33	3.91	.28	3.81	.32

표 4. 평가형식과 평가척도에 대한 변량분석표

변량원	df	자승합	평균자승합	F	p	Scheffe
평가형식(A)	1	.07832	.07832	.840	.36	
평정척도(B)	1	.94510	.94510	10.1	<.01	Ub < B
A * B	1	.29590	.29590	3.16	.07	
오차	129	12.07590	.09361			
전체	132	13.48946				

* Ub: 불균형 척도, B: 균형척도

표 5. 평가형식과 평가척도에 따른 왜도 지수(degree of skewness)

평가형식	평가 척도		전체
	불균형적 척도	균형적 척도	
	왜도	왜도	
리커트 척도	-0.083	-0.097	-0.09
행동기술평정척도	-0.018	-0.022	-0.02
전체	-0.0505	-0.0595	

서 이용된 변량분석은 각 독립수준별로 측정된 평정치의 상대적인 관대화 차이를 중심으로 비교하는 것이 적절할 것이다.

참고로, 위와 같은 변량분석을 실시하기 전에 네 개 처지 집단간 기존의 역량 혹은 직무수행에서 차이가 있는지 살펴보기 위해서 A기업의 인사담당자가 제공해준 3개년 평균 업무고과 점수를 활용하여 집단간 차이분석을 실시한 결과 유의미한 차이가 나타나지 않았다. 그 결과 공변량 분석을 실시하지 않고 변량분석을 실시하였다.

또한, 변량분석 외에 왜도분석법을 이용하여 각 독립변인 수준별로 왜도지수를 살펴본 결과, 모두 음의 왜도지수를 지니며 부적편포의 형태를 띄고 있었다. 이는 각 조건에서 모두 어느 정도의 관대화 오류가 존재했음을 의미한다. 표 5와 같이, 각 독립 수준별로 산출된 왜도지수를 직접 비교한 결과, 행동기술평정척도 형식(-.02)은 리커트 척도형식(-.09)과 비교해 상당히 작은 음수의 왜도값을 지니며 대체로 '0'에 가까운 대칭분포의 형태를 보였으며, 불균형적 척도(-.0505) 역시 균형적 척도(-.0595)와 비교해 미미하지만 다소 작은 음수의 왜도값을 지니는 것으로 나타났다.

행동기술평정척도 형식과 불균형 척도법을 각각의 수준에서 살펴본 결과, 행동기술평정

척도 형식은 불균형 척도가 제시된 상황(-.018)과 균형적 척도(-.022)가 제시된 상황에서 모두 리커트 척도와 비교해 일관되게 부적편포 지수의 감소 패턴을 보이며 상대적으로 낮은 관대화 경향을 보였다. 또한, 불균형 척도 역시 리커트 척도형식(-.083)뿐 아니라 행동기술평정척도 형식(-.018)이 제시되는 상황에서 균형적 척도에 비해 부적편포 지수의 감소 패턴을 보여서 관대화 경향이 상대적으로 적게 나타났음을 확인할 수 있었다.

즉, 위 왜도분석 결과를 통해서 균형적인 척도보다 긍정적인 어조의 불균형척도가 제공되는 상황에서 관대화 오류가 적게 나타난 변량분석 결과를 확인할 수 있었고 가설 3이 다시 한 번 지지되었다. 그러나, 왜도지수 차이가 상당히 크게 나타났음에도 불구하고, 변량분석 결과가 유의미하지 않게 나타났다는 점과 통계적 유의도 검증을 실시하지 못하는 방법론적 한계 때문에 리커트 형식보다 행동기술평정척도 형식이 제공되는 상황에서 관대화 오류가 적게 나타날 것이라는 가설 1에 대해서는 기술적인 수준에서만 언급할 수 있다.

평가 형식 및 척도화 방법에 따른 후광효과 분석

평가 형식 및 척도화 방법에 따른 후광효과

표 6. 평가형식 및 평가척도별 역량간 표준편차

평가형식	평가척도		전체
	불균형적 척도	균형적 척도	
	표준편차	표준편차	
리커트 척도	.73	.67	.70
행동기술평정척도	.81	.69	.75
전체	.77	.68	

표 7. 평가형식 및 평가척도에 따른 표준편차에 대한 변량분석표

변량원	df	자승합	평균자승합	F	p	Scheffe
평가형식(A)	1	.00760	.00760	.25	.618	
평정척도(B)	1	.21103	.21103	6.95	<.01	Ub>b
A * B	1	.05825	.05825	1.92	.168	
오차	129	3.91433	.03034			
전체	132	4.21425				

* Ub: 불균형 척도, B: 균형척도

를 살펴보기 위한 첫 번째 분석방법으로 각 독립변인 수준별로 역량점수간의 표준편차에 대한 변량분석이 실시되었다(표 6과 7). 분석 결과, 평정척도에 따라 표준편차간에 유의미한 차이를 보였으며, 불균형 척도(M=.77)는 균형적 척도(M=.68)와 비교해 보다 큰 표준편차를 지니며, 상대적으로 역량간 구분이 좀 더 명확히 이루어졌음을 나타냈다. 이 같은 결과로 다면평가에서 균형적인 어조가 제공되는 상황보다 긍정적 어조의 불균형적 척도가 제공되는 상황에서 후광효과가 더 적을 것이라는 연구가설 4가 지지되었다. 반면 리커트 척도(M=.70)와 행동기술평정척도 형식(M=.75)간에 표준편차에 대한 효과의 차이가 유의하지 않음에 따라 평가형식은 후광효과에 영향을 미치지 않는 것으로 나타났다.

후광효과를 측정하는 두 번째 분석방법으로, 다속성 다방법 상관행렬 분석 중 동일방법-이질속성에서의 내적상관을 산출하였다(표 8). 이를 독립변인 수준별로 비교·분석한 결과, 평가형식 및 척도화 방법 모두 주효과가 통계적으로 유의미한 차이를 보였다(표 9). 또한 평가형식과 평가 척도화 방법간 상호작용 효과가 통계적으로 유의하였으나, 그림 5와 같이 서열 상호작용³⁾의 형태를 보임에 따라 추가적인 상호작용항 분석 없이 각 독립변인의 주효과를 설명할 수 있었다.

평가형식을 기준으로 살펴보면 행동기술평

3) 각기 다른 조건에서도 한 수준의 평균이 다른 수준의 평균보다 항상 동일하게 높을 경우 추가적인 상호작용항에 대한 분석없이 주효과 해석이 가능하다.

표 8. 평가형식 및 평가척도별 역량간 평균 상관

평가형식	평가척도		전체
	불균형적 척도	균형적 척도	
	평균 상관	평균 상관	
리커트 척도	.67	.65	.66
행동기술평정척도	.39	.29	.34
전체	.53	.47	

표 9. 평가형식 및 평가척도별 역량간 평균 상관에 대한 변량분석표

변량원	df	자승합	평균자승합	F	p
평가형식(A)	1	45.37233	45.37233	1977.69	<.0001
평정척도(B)	1	1.49266	1.49266	65.06	<.0001
A * B	1	.58755	.58755	25.61	<.0001
오차	1736	39.82736	.02294		
전체	1739	87.27989			

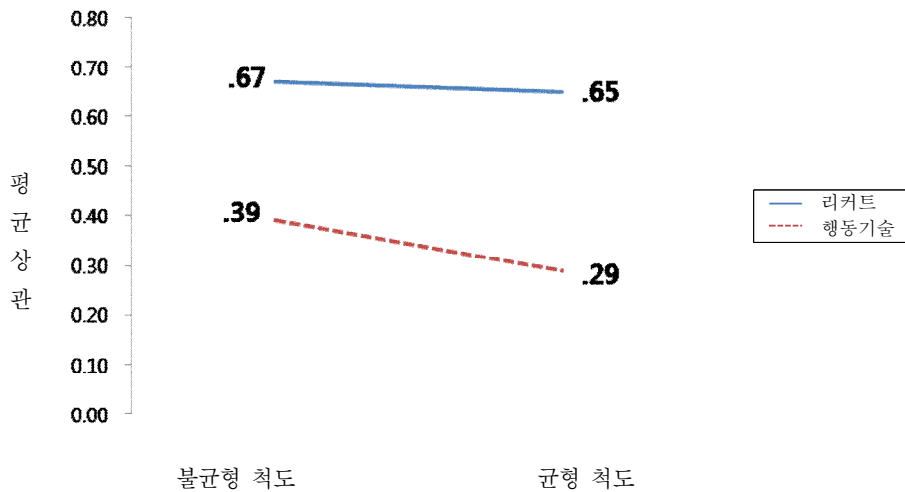


그림 1. 평가형식 및 척도화 방식이 후광효과에 미치는 영향

정척도 형식(평균상관=.34)은 리커트 척도 형식(평균상관=.66)과 비교해 상대적으로 낮은 역량간 상관을 보였으며, 이는 불균형 척도(평균상관=.39)와 균형적 척도(평균상관=.29)에서 모두 리커트 척도 형식보다 낮은 것으로 나타났다. 이 같은 결과는 행동기술평정척도 형식은 리커트 척도형식과 비교해 상대적으로 후광효과를 감소시키는데 효과적임을 의미하는 것으로 리커트 형식과 비교해서 행동기술평정척도 형식이 제공되는 상황에서 후광효과가 더 적을 것이라는 가설 2가 지지되었음을 의미한다.

또한, 척도화 방법의 경우 균형적 척도(평균상관=.47)는 불균형적 척도(평균상관=.53)와 비교해 상대적으로 낮은 평균상관을 보였으며, 리커트 형식(평균상관=.65)과 행동기술평정척도 형식(평균상관=.29)에서 모두 불균형적 척도보다 낮은 것으로 나타났다. 따라서 균형적인 척도가 제공되는 상황보다 긍정적 어조의 불균형적 척도가 제공되는 상황에서 후광효과가 보다 적게 나타날 것이라는 가설 4가 기각되었음을 알 수 있다.

논 의

본 연구는 최근 공사를 막론하고 많은 기업에서 지속적으로 확산되는 추세에 있는 다면평가 상황을 배경으로 평가도구의 형식과 척도화 방법이 평가의 신뢰성과 타당성을 저해시키는 평가오류에 어떠한 영향을 미치는지를 밝히는데 목적을 두었다. 이를 검증하기 위해 현재 역량 다면평가를 실시하고 있는 A기업의 현직자를 대상으로 자기, 타인, 동료, 부하로 구성된 평가자로 하여금 피평가자의 역량수준

을 측정하도록 하였다. 평가형식은 행동기술평정척도 형식과 리커트 척도 형식으로, 척도화 방법은 불균형 척도와 균형적 척도로 구성되었고, 그 결과 각각의 처치 수준별로 불균형적 리커트 형식, 균형적 리커트 형식, 불균형적 행동기술평정척도 형식, 균형적 행동기술평정척도 형식의 실험집단이 구성되었다. 주요 연구 결과를 요약하면 다음과 같다.

먼저, 변량분석을 통해서 평가형식과 척도화 방법이 관대화 오류에 미치는 영향을 분석한 결과, 척도화 방법 중 불균형 척도를 사용할 때 관대화가 더 적게 나타났으며, 평가형식에서는 유의미한 차이를 보이지 않았다. 한편, 부가적으로 실시한 왜도분석에서는 척도화 방법의 효과 외에 평가형식에서도 차이가 도출되어, 행동기술평정척도 형식에서 리커트 형식에 비해서 관대화가 감소되는 경향을 보였다. 그러나 왜도지수들을 비교하는 통계적 유의도 검증 절차가 부재하기 때문에 균형적 척도보다 긍정적 어조의 불균형 척도가 제시된 상황에서 관대화가 적게 나타날 것이라는 가설 3은 지지되었다고 결론내릴 수 있지만, 행동기술평정척도에서 관대화 오류가 적게 나타날 것이라는 가설 1에 대해서는 명확하게 결론내리기 어렵다. 따라서, 왜도분석의 경우 관대화를 측정하는 주 측정방법이라기 보다는 변량분석 결과에 대한 부가적인 확인 절차로 간주하는 것이 바람직할 것이다.

후광효과 검증을 위해 표준편차 분석법과 평가 차원간 상관분석법을 활용한 결과, 표준편차 분석에서는 불균형 척도가 균형적 척도보다 더 큰 표준편차를 지니며 후광효과가 적게 나타나는 결과를 보였으나, 평가형식의 경우 그 효과는 유의미한 차이를 보이지 않았다. 반면, 상관분석에서는 행동기술평정척도가 리

커트 척도와 비교해 상대적으로 후광효과가 적은 것으로 나타났다. 그러나 척도화 방법의 경우 표준편차 분석법과는 반대로 균형척도가 오히려 후광효과를 감소시키는데 효과적인 것으로 나타나 측정 방법간에 불일치한 결과도 출되었다. 이 같은 결과를 종합해볼 때, 평가형식에 따라 후광효과가 차이가 존재할 것이라는 연구가설 2는 표준편차 분석결과를 통해 부분적으로 지지되었으나, 균형적 척도가 제공되는 상황보다 긍정적 어조의 불균형적 척도가 제공되는 상황에서 후광효과가 보다 감소할 것이라는 연구가설 4는 분석결과의 비일관성으로 인해 입증되지 못하였다.

본 연구 결과, 기존 평가 상황에서 자주 이용되던 리커트 척도형식과 균형적 척도는 행동기술평정척도 형식 및 불균형적 척도와 비교해 평가오류에 좀 더 쉽게 노출되며 이로 인해 평가결과에 부정적 요소로 작용될 가능성이 상대적으로 높다는 것으로 밝혀졌다. 이에 반해 아직까지 활용 정도가 그다지 높지 않은 행동기술평정척도 형식과 불균형적 척도는 평가오류 중 가장 흔히 발생하는 관대화 오류 혹은 후광효과에 있어 상대적으로 안정적인 것으로 나타났다.

특히, 행동기술평정척도 형식은 Harris의 동기모델에 기초해 Jones와 Fletcher(2002)가 제시한 방법으로 평가시 그에 대한 증거를 제시토록 함으로써 평가에 대한 평가자의 책임감을 향상 시키고 그 결과 평가를 보다 객관적으로 수행할 것이라는 논리에 바탕을 두고 있다. 이는 또한 측정역량에 대한 역량정의와 더불어 이를 대표하는 긍정적 행동과 부정적 행동의 예시를 제공함에 따라 평가자가 측정되는 평가차원에 대해 보다 명확히 인지하고 평가차원들 간에 혼동을 최소화 할 수 있다는 장

점을 지니고 있다. 이와 더불어 평가시 제시하는 평가근거는 피드백 제공이나 교육훈련시 참고자료 등과 같이 향후 평가결과의 다양한 활용이 가능하다는 이점을 지닌다. 본 연구를 통해서 현재 고위공무원단 역량평가나 대기업 중심의 임원 평가에서 사용하는 행동기술 평정척도가 일반적인 리커트 척도에 비해서 평가오류가 적은 경향이 있다는 것이 경험적으로 밝혀짐으로써 이들 역량평가방법의 적절성이 지지되었다.

또한 불균형 척도는 Jones와 Fletcher(2002), Fox와 동료들의 연구(1994)에서 활용한 방법으로 균형적 척도에서 실제 평가시 이용 빈도가 매우 적은 부정적 기술문(예: 전혀 우수하지 않다)을 삭제하고 척도의 중앙점을 보다 하향화시켜 개인간 변별성을 극대화도록 한 척도화 방법이다. Friedman과 Leefer(1981), Kasten과 Weintraub(1999)가 제시한 것처럼, 평가자는 척도의 위치보다는 기술문의 내용을 바탕으로 평가하며, 따라서 긍정적 어조의 불균형적인 척도를 사용하면 대부분의 점수가 몰려있는 중간 혹은 중상 지점의 점수들을 좀 더 낮은 점수로 이동시킬 수 있어 관대화 경향이 감소되는 효과가 있다는 것이 본 연구에서 밝혀졌다. 최근 들어 수행(역량)평가가 확산되면서, 척도의 불균형성이 평가오류에 영향을 미치는지는 중요한 이슈가 될 수 있다. 왜냐하면, 특별히 새로운 도구나 측정방법을 고안하지 않아도 기존의 척도 기술문 내용만을 수정함으로써 평가오류를 감소시키는 효과가 있다면 경제적인 측면에서 매우 효율적이기 때문이다.

기존의 평가형식과 척도화 방법이 평가오류에 미치는 영향에 관련된 연구들은 대체로 선발장면이나 자기진단 상황에서 자기평가를 중심으로 진행되었으며, 대부분이 국외에서 수

행되었다. 또한, 기존의 다면평가에 대한 대부분의 연구들은 서로 다른 수준의 평가주체들(자기, 상사, 동료, 부하)간 일치도와 관련된 주제들을 주로 다루어왔다. 이와 더불어 평가의 정확성과 타당성을 저해하는 대표적인 평가오류인 관대화 오류와 후광효과에 대한 측정방법들이 국내 논문들에서 구체적으로 소개되지 않았다. 따라서 본 연구는 평가오류 측정방법들을 개관하고, 실제 기업의 수행평가 용도로 실시되고 있는 다면평가 결과를 바탕으로 평가형식과 척도가 관대화 오류와 후광효과에 어떠한 영향을 미치는 살펴보았다. 이 점에서 의의가 있다. 이를 통해 현재 다면평가를 실시하고 있는 기업 및 공공기관 인사담당자들에게 자신의 조직에서 시행하고 있는 다면평가에서 발생 가능한 평가오류에 대한 인식을 제고할 수 있는 기회를 제공하였다. 또한, 자사의 평가 시스템에서 나타나는 관대화 오류나 후광효과를 실증적으로 검증하여 평가에 반영하고 평가시스템을 수정, 보완할 수 있는 기틀을 마련하였다.

그러나 이 같은 학문적 의의나 실무에서의 시사점에도 불구하고 본 연구는 그 결과를 일반화 시키는데 있어 몇 가지 한계점을 지니고 있다. 먼저, 본 연구는 평가형식과 척도화 방법이 관대화 오류와 후광효과에 어떠한 영향을 미치는지를 살펴보기 위해 서로 다른 측정방법들을 함께 이용하였다. 원래의 의도는 연구 결과의 교차타당성을 확보하기 위한 것이었으나, 평가오류 측정 방법 간에 일부 비일관적인 결과가 도출되어 평가형식과 척도화 방법의 효과성을 입증하는데 있어 한계점으로 작용하였다. 사실 지난 수십 년간 관대화 오류와 후광효과에 대해서 연구가 수행되었고 측정방법들이 제시되었음에도 불구하고 각각

의 평가오류를 어떻게 개념화하고 어떻게 측정할 것인가에 대한 합의점은 도출하지 못한 상태이다. Saal, Downey, 및 Lahey(1980)의 연구에서도 두 평가형식(즉, 도식평정척도와 혼합표준평정척도)에 따른 평가오류들을 몇몇 측정방법을 통해 규명하였는데, 이들 연구자들은 비일관적인 결과에 대한 원인으로 각 측정방법마다 다르게 정의된 조작적 정의들을 꼽았다. 즉, 각각의 측정방법들은 평가오류에 대해서 각기 다른 조작적 정의를 바탕으로 제안되었으나, 아직까지 이들에 대한 상대적인 효과성을 비교하고 입증한 연구는 없는 실정이다. 따라서 후속연구에서는 평가오류 측정방법들에 대한 상대적 논리성 및 정확성을 비교하고, 그 우수성을 제시할 수 있는 연구가 이루어질 필요가 있다. 이 같은 결과가 정립될 때 보다 정확하게 평가오류에 대한 처치방법들의 효과성을 입증하는 것이 가능해질 것이다.

두 번째로, 본 연구에서 이용된 평가오류 측정방법들이 지니는 논리적 오류는 또 다른 한계점으로 작용하였다. Jeffrey, Bernardin, Peter, 및 Joseph (1995)은 자신의 연구에서 사용한 평균 및 왜도 비교를 통해 도출된 관대화 오류 결과를 ‘상대적 관대화’라고 지칭한 바 있다. 이와 마찬가지로 본 연구에서 이용된 변량분석과 왜도분석 역시 각 독립변인의 수준별로 측정된 평정치의 상대적인 차이만을 비교할 뿐 실제 진정한 평정점수를 고려하지 못한다는 문제점을 지니고 있다. 구체적으로 변량분석 및 왜도분석 결과 모든 처치수준에서 척도의 중간점수 3점 이상의 높은 평균값과 음의 왜도지수가 나타나, 전반적으로 관대화 경향이 발생한 가운데 상대적으로 어느 수준에서 관대화가 더 크게 나타났는지에 해석의 초점이 맞추어져 있다. 또한, 왜도분석의 경우 통

계적 유의도 검증 절차가 부재하기 때문에 좀 더 조심스러운 해석이 요구된다.

또한 후광효과 분석시 이용된 상관분석은 실제로 측정되는 두 차원 모두에서 피험자가 높은 성과를 보이는 경우, 이를 고려하지 못한다는 문제점을 지니고 있다. 예를 들어, 특정 피평가자의 경우 자료를 분석하는 능력도 뛰어나고 대인관계도 우수하다면 자연스럽게 두 차원의 평정 점수 간 상관관계가 높을 수밖에 없지만, 후광효과에 활용되는 상관분석 방법에 의하면 후광효과가 존재하는 것으로 해석된다. 뿐만 아니라, 행동기술평정척도에서는 측정역량에 대한 정의를 제시한 경우이고 리커트 척도방법인 경우 역량이 정확하게 제시 분류되어 있지 않기 때문에, 하나 혹은 두 개의 요인으로 피험자들이 받아들여서 반응한 경우 실제로는 하나 혹은 두 개의 요인이 되어 소위 동일방법-이질속성간의 상관이 자연스럽게 높아진다. 후광효과 분석에 이용된 표준편차 분석법에서 역시 분산 산출을 통해 후광효과와의 상대적 강도를 파악하는 방법으로 실제 성과수준을 고려하지 못한다는 면에서 상관계수와 유사한 문제점을 지닌다.

마지막으로 본 연구의 연구 설계방법에 있어 피험자 내 반복설계가 가장 바람직한 방법일 것이다. 그러나 현직자들을 대상으로 실제의 평가 상황에서 4 개의 처지변인(즉, 측정방법)을 반복해서 실시하는 것은 현실적으로 불가능하므로 피험자 간 설계를 실시하게 되었다. 추후 연구에서는 이 같은 한계점을 극복하여 피험자 내 반복설계를 실시한다면 보다 타당한 연구결과가 도출될 것으로 기대한다. 그 외에도 방법론적으로 후광효과의 표준편차에 대한 변량분석 부분은 표준편차가 정규분포를 따르지 않는 특성 때문에 변량분석결과

에 대한 해석에 제한이 따른다.

본 연구의 결과는 어떠한 평가형식 및 척도화 방법이 더 우수하다는 결론을 내리기에 앞서 평가의 오류를 최대한 줄이고 정확성을 향상시키는데 있어 과거 언급되지 않았던 평가형식 및 척도화 방법들을 고려할 것을 제안하고 있다. 따라서 학계는 평가오류 측정에 있어 보다 논리적이고 과학적인 방법을 개발하고, 다양한 측면에서 평가오류에 영향을 미칠 수 있는 다양한 변인들을 밝혀내고 평가의 타당성과 객관성을 높이는데 기여할 수 있는 연구들을 수행할 필요가 있다. 또한, 각 기업 및 조직의 인사담당자들은 이 같은 연구결과들을 바탕으로 구성원의 수행 혹은 역량 평가시 사전조사를 통해 보다 다양한 접근방법들을 비교, 적용하고 평가오류를 최소화하는 일에 노력을 기울여야 할 것이다. 이러한 상호 노력은 보다 타당한 평가시스템을 구축하고 노사 모두 수용할 수 있는 평가문화의 정착을 앞당기는 촉매제로 작용할 것임에 틀림없다.

참고문헌

- 김명소 (1992). 심리검사 제작의 이론과 실제: 척도화 방법. 1992년 한국심리학회 학술대회 발표집, 단행권, 단일호, 197-227.
- 김판석, 오성호, 이선우 (1999). 다면평가제 사례연구: 신화와 실상. 한국정책학회, 1999년도 하계학술대회 발표논문집(정부정책 및 정부개혁의 평가), 19-41.
- 문화일보 (2007). <사설>고위공무원단 실험 1년과 하나마나한 직무평가. 7월 28일자.
- 박경규 (2001). 부하의 체면중시가 상사-부하간 교환관계 및 조직행동에 미친 영향. 경

- 영학 연구, 30(4), 1093-1114.
- 박영석 (2003). 리더십 다면평가에서 나타난 평가자 조망. *한국심리학회지: 산업 및 조직*, 16(3), 75-87.
- 양진환 (1999). 인사고과의 조작적 정의에 대한 평정정확성. *강남대학교 논문집*, 34(3), 111-136.
- 이승계 (2005). 다면평가제도의 실효성에 관한 연구. *대한경영학회지*, 18(1), 297-329.
- 이종구 (2001). SAS와 통계자료분석. 학지사.
- 임창희 역 (1999). 360도 다면평가. 21세기 북스.
- 장재윤 (1999). 연구개발 관리자에 대한 다면평가의 일치도 및 수행과의 관계. *한국심리학회지: 산업 및 조직*, 12(1), 31-53.
- 중앙인사위원회 (2003). 다면평가 운영관련 유의사항.
- 한영석, 김명소 (2007). 고위공무원단 9개 리더십 역량이 성과평가에 미치는 영향. 2007년 한국 산업 및 조직 심리학회 추계학술대회 및 심포지움 발표집, 19-28. 12월 7일. 경기도 이천: SK Telecom 미래경영연구원.
- Bernardin, H. J., & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work*. Boston: Kent
- Bernardin, H. J., & Walter, C. S. (1977). Effect of rater training and diary-keeping on psychometric error in rating. *Journal of Applied Psychology*, 62, 64-69.
- Bernardin, H. J., Alvares, K. A., & Cranny, C. J. (1976). A recomparison of behavioral expectation scales to summated scales. *Journal of Applied Psychology*, 61, 564-570.
- Bernardin, H. J., Lashells, M. B, Smith, P. C., & Alvares, K. M. (1976). Behavioral expectation scales: Effects of development procedures and format. *Journal of Applied Psychology*, 61, 75-79.
- Bretz, Jr., R. D., Milkovich, G. T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions and implications. *Journal of Management*, 18(2), 321-322.
- Campbell, J. D., & Fehr, B. (1990). Self-esteem and perceptions of conveyed impression: Is negative affectivity associated with greater realism? *Journal of Personality and Social Psychology*, 58, 122-123.
- Chruch, A. H. (1995). First-rate multi-rater feedback. *Training and Development*, Aug, 42-43.
- Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, 90(2), 218-244.
- Decotiis, T. A. (1977). An analysis of the external validity and applied relevance of three rating formats. *Organizational Behavior and Human Performance*, 19, 247-266.
- Dickinson, T. L., & Tice, T. E. (1977.) The discriminant validity of scales developed by retranslation. *Personnel Psychology*, 30, 255-268.
- Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66, 127-148.
- Finley, D. M., Osburn, H. G., Dubin, J. A., & Jeanneret, P. R. (1977). Behaviorally based rating scales: Effects of specific anchors and disguised scale continua. *Personnel Psychology*, 30(4), 659 - 669.

- Friedman, B. A., & Comelius, E. T. (1976). Effect of rater participation in scale construction on the psychometric characteristics of two rating scale format. *Journal of Applied Psychology, 61*, 210-216.
- Friedman, H. H., & Leefer, J. R. (1981). Label versus position in rating scales. *Journal of the Academy of Marketing Science, 9*, 1-2.
- Fisicaro, S. A. (1988). A re-examination of the relation between halo error and accuracy. *Journal of Applied Psychology, 73*, 239-244.
- Fox, S., Caspy, T., & Reiser, A. (1994). Variables affecting leniency, halo and validity of self-appraisal. *Journal of Occupational and Organizational Psychology, 67*, 45-56.
- Friedman, H. H., & Amoo, T. (1999). Rating the rating scales. *The Journal of Marketing Management, 9*(3), 114-123.
- Harris, M. M. (1994). Rater motivation in the performance appraisal context: A theoretical framework. *Journal of Management, 94*, 20(4), 753-756.
- Harris, M. M., & Schaubroek, J. (1988). A meta-analysis of self-supervisor, self-peer and peer-supervisor ratings. *Personnel Psychology, 41*, 43-62.
- Jeffrey, S. K., Bernardin, H. J., Peter, V. & Joseph, P. (1995). Stability of Rater Leniency: Three Studies. *The Academy of Management Journal, 38*(4), 1036-1051.
- Jones, L., & Fletcher, C. (2002). Self-assessment in a selection situation: An evaluation of different measurement approaches. *Journal of Occupational and Organizational Psychology, 75*, 145-161.
- Kasten, R., & Weintraub, Z. (1999). Rating errors and rating accuracy: A field experiment. *Human Performance, 12*(2), 137-153.
- Keaveny, T. J., & McGann, A. F. (1975). A comparison of behavioral expectation scales and graphic rating scales. *Journal of Applied Psychology, 60*, 695-703.
- Landy, F. J., Farr, J. L, Saal, F. E., & Freytag, W. R. (1976). Behaviorally anchored scales for rating the performance of police officers. *Journal of Applied Psychology, 61*, 750-758.
- Latham, G. P., & Wexley, K. N. (1981). *Increasing productivity through performance appraisal*. Addison-Wesley Company.
- Levy, P. E., Cawley, B. D., & Foti, R. J. (1998). Reactions to appraisal discrepancies: Performance ratings and attributions. *Journal of Business and Psychology, 12*(4), 437-455.
- Longenecker, C. O., Gioia, D. A., & Sims, H. P. (1987). Behind the mask: The politics of employee appraisal. *Academy of Management Executive, 1*, 183-193.
- Mount, M. K. (1984). Psychometric properties of subordinate ratings of managerial performance. *Personnel Psychology, 37*(4), 687-702.
- Mikovich, G. T., & Boudreau, J. W. (1994). *Human Resource Management(7th ed.)*. Burr Ridge, IL: Irwin.
- Motowidlo, S. J., & Borman, W. C. (1977). Behaviorally anchored scales for measuring morale in military units. *Journal of Applied Psychology, 62*, 177-183.
- Murphy, K. R., & Blazer, W. K. (1989). Rater errors and rating accuracy. *Journal of Applied*

- Psychology*, 74, 619-624.
- Ronen, K. & Zisu, W. (1999). Rating errors and rating accuracy: A field experiment. *Human Performance*, 12(2), 137-153.
- Saal, F. E., Downey, R. G. & Lahey, M. A. (1980). Rating the ratings: Assessing the quality of rating data. *Psychological Bulletin*, 88, 413-428.
- Schwab, D. P., Heneman, H. G., & DeCottis, T. A. (1975). Behaviorally anchored rating scales: A review of the literature. *Personnel Psychology*, 28(4), 549 - 562.
- Steele, R. P. & Ovalle, N. K. (1984). Self-appraisal based upon supervisory feedback. *Personnel Psychology*, 37, 667-685.
- Thornton, III, G. C. (1980). Psychometric properties of self-appraisal and job performance. *Personnel Psychology*, 33, 263-271.
- Viswesvaran, C., Ones, D. S. & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81, 557-574.
- Wayne, S. J., & Ferris, G. R. (1990). Influence tactics, affect and exchange quality in supervisor-subordinate interactions: A laboratory experiment and field study. *Journal of Applied Psychology*, 75, 487-499.
- Wildt, A. R., & Michael, B. M. (1978). Determinants of scale response: Label versus position. *Journal of Marketing Research*, 15(2), 261-267.
- 1차 원고접수 : 2008. 1. 5
2차 원고접수 : 2008. 4. 2
3차 원고접수 : 2008. 4. 30
최종게재결정 : 2008. 5. 8

Format and scale effects on leniency error and halo effect in 360° feedback

Myoung So Kim

Hoseo University

Gwang Gene Yi

ORP Institute

The purpose of the present study was to investigate the effect of rating format(Likert, Behavioral description rating format) and scale type(balanced scale, positively unbalanced scale) on leniency error and halo effect in 360° feedback. The competency scale developed to appraise the job performance of employees in a large Korean company was administered to a sample of 440 raters who are composed of oneself, supervisor, subordinate, and peer. The raters of four groups rated the level of 133 ratees on thirty core competencies of the company through the 5-point scale. The results of 2×2 ANOVA and degree of skewness analysis indicated that positively unbalanced scale showed less leniency than the conventional measure of balanced scale, while there was no effect of rating format. Furthermore, the results of analyses of both standard deviation and correlation demonstrated that Behavioral description rating was more effective in reducing halo effect than Likert scale format. However, no scale effect was found since the two analyses of standard deviation and correlation produced inconsistent results. Finally, the limitations and future direction of the present study were discussed.

Key words : 360° feedback, rating format, scale type, leniency error, halo effect, measurement of rating errors