



인사평가는 인적자원관리의 중요한 기능으로 강조되고 있으며, 각종 의사결정과정에서 근본이 되는 활동이다(Cardy & Dobbins, 1994). 최근 급변하는 환경 속에서 인적자원의 효율적인 관리는 기업의 성과를 좌우한다. 특히 인적자원관리의 핵심이 되는 인사평가의 효과적인 운용은 기업의 성과에 중요한 요소이다. 그래서 많은 조직은 효과적인 인사 평가시스템의 구축을 통해 기업의 생존과 성장을 동시에 도모하고자 노력하고 있으며(Ferris & Judge, 1991), 구성원의 성과와 능력을 정확하게 평가하여 개인이 조직의 효과성에 기여할 수 있도록 관리하는 인사평가의 중요성이 더욱 부각되고 있다.

인사평가의 중요성은 공감하더라도 운영이 제대로 이루어지지 않는 경우가 많은 것이 현실이다. 2007년 기업의 직장인 463명을 대상으로 ‘인사평가 제도에 대한 견해’에 대해 설문 조사를 진행한 결과, 과반수 정도가 인사평가를 단순히 형식적인 제도로 생각하고, 자신이 근무하고 있는 회사의 인사평가 프로세스에 대해서는 ‘전혀 신뢰할 수 없다’는 응답은 10.2%이었다. 특히 ‘자신이 공정한 인사평가를 하고 있는가?’란 질문에 무려 75.4%가 ‘그렇다’고 응답한 반면, ‘자신이 공정한 평가를 받고 있다고 생각하는가?’란 질문에 대해서는 58.5%의 직장인은 자신이 공정한 인사평가를 받고 있지 못하다고 생각하는 것으로 나타나 평가자와 피평가자의 입장에서 상반된 견해를 보였다 (한국경제, 2007. 08. 23).

인사평가 연구는 평가를 하는 평정 척도와 평가 오류에 대한 관심에서 출발하여 평가의 주체인 평가자를 훈련시키는 방안으로 이어졌다(Bernadin & Pence, 1980; Hauenstein, 1998; Landy & Farr, 1980). 다양한 평가자 훈련은

장·단점을 가지고 있지만, 각각의 훈련이 가지는 문제점을 보완하여 효율적으로 이루어진다면 평가자들은 정확한 평가를 할 수 있을 것이다. 그리고 훈련을 통해 개선된 정확한 평가는 피평가자들이 인사평가 제도를 공정하게 지각하여 수용하도록 도와주어, 결과적으로 인사평가를 형식적인 절차에서 인적자원관리시스템의 중요한 수단으로 인식할 수 있는 계기를 마련해 준다.

본 연구에서는 평가자 훈련의 유형에 따라 평가 정확성에 차이가 있는지에 대해서 실험을 통해 알아보았다. Woehr와 Huffcutt(1994)는 통합 분석을 통해 다양한 평가자 훈련들 중 참조틀 훈련(Frame-of-Reference Training)이 평가의 정확성 부분에서 가장 효과적인 훈련이라고 한 반면, Stamoulis와 Hauenstein(1993)은 참조틀 훈련은 피평가자들 간의 구별에 있어 다소 부족하며 이에 대안으로서 변산성 훈련(Rater Variability Training)을 제안했다. 본 연구에서는 참조틀 훈련과 변산성 훈련이 평가 정확성에 미치는 영향의 차이점을 알아보려고 한다. 평가 정확성은 Cronbach(1955)가 제시한 4가지 정확성 구성요소를 활용할 것이다. Cronbach의 4가지 정확성 구성요소는 크게 평가 차원간 정확성과 평가 대상간 정확성으로 구분되기 때문에 본 연구에서 알아보려고 하는 평가자 훈련의 효과성 차이를 차별적으로 알아볼 수 있을 것이다.

한편, 평가자에게 더욱 평가를 신중하게 할 수 있는 동기를 부여한다면 더욱 정확한 평가를 할 수 있다(Mero, Guidice, & Brownlee, 2007; Harris, 1994; Latham, 1986; Murphy & Cleveland, 1995). 실제 평가 장면에서는 평가의 결과에 대한 설명을 다른 사람들에게 하거나 평가에 대한 이유를 설명해야 하는 경우가 많다. 개

인이 평가에 대한 자신의 관점을 다른 사람들에게 정당화해야 할 필요성과 압력을 책임성(accountability)이라고 한다. 기존 연구에 의하면 책임성이 부여된 집단이 부여되지 않은 집단에 비해 문제 해결을 하기 위해서 더욱 적극적인 시도를 하고 더 복잡한 인지적 노력을 들이는 정보처리 전략을 사용한다(McAllister, Mitchel, & Beach, 1979). 따라서 본 연구에서는 평가자들에게 책임성을 부여하였을 때 더욱 정확한 평가를 한다는 연구들을 바탕으로 책임성이 평가의 정확성에 대한 평가자 훈련의 영향에 어떠한 변화를 만드는지 알아보하고자 한다.

또한, 평가 당시의 평가자 기분에 따라 평가자의 인사평가의 결과가 달라지는지에 대하여 연구하였다. 사회인지이론(social cognition theory) 관점에서 볼 때 개인이 판단을 할 당시의 기분(mood)에 따라 판단의 결과가 달라진다(Srull & Wyer, 1989; Izard, 1991). 긍정(positive) 기분의 사람들은 더 호의적이며 긍정적인 판단을 하며 편법적(heuristic) 정보처리과정을 거치고, 부정(negative) 기분의 사람들은 더 부정적인 판단을 하며 체계적, 분석적인 정보처리과정을 거친다(Forgas, 1987). 인사평가 역시 평가자의 판단이 개입되기 때문에 정확한 평가를 위해서는 평가자의 기분이 평가에 미치는 영향을 고려할 필요가 있을 것이다.

종합하면, 본 연구의 목적은 평가자 훈련의 유형에 따른 평가 정확성 효과성의 차이를 검증하고, 평가 책임성 그리고 평가자 기분에 따라 평가 정확성의 효과성에 차이가 있는지를 알아보는 것이다. 평가정확성은 직무수행의 정답에 해당하는 진점수(true score)가 필요한데, 조사연구를 통해서 진점수를 확인하거나 설정할 수 없기 때문에 실험연구를 실시

하였다. 각 평가자 훈련의 유형에 맞는 훈련을 실시하고 실험 참가자들은 강의 평가 영상을 시청한 후 평가의 정확성을 통해 훈련의 효과성 차이를 검증하였다. 또한, 실험의 중간에 실험대상자의 평가 책임성과 기분을 조작하는 처치를 통해 이러한 동기적, 정서적 요소들이 평가정확성에 상호작용 효과를 일으키는지 알아보았다.

### 인사평가의 정확성

직무수행의 평가 정보는 인사 결정, 급여 결정의 근거가 될 뿐 아니라 개인에게 피드백을 제공하고 보완할 부분에 대한 훈련의 필요성을 규명해주는 등 조직 구성원 개발에 필요한 훈련의 근거가 된다(Carroll & Scneier, 1982). 이는 종업원들의 육성과 관리를 위한 기준이 된다는 측면에서 인사평가의 목적과 일치하는 것이다. 이러한 인사평가의 기능들을 효율적으로 운용하기 위해서는 먼저 인사평가의 정확성이 전제가 되어야 한다(Cardy & Doddins, 1994).

인사평가의 정확성은 크게 두 가지의 측정 방법이 있다. 간접적인 측정과 직접적인 측정이다(Sinclair, 2000). 간접적인 측정은 평가결과의 심리측정적 속성(psychometric properties; 예, 신뢰도, 타당도)과 평가자 오류(rater error)를 측정하는 것이다. 하지만 평가 오류나 심리측정 속성은 피평가자의 직무수행의 진점수를 반영하는 것은 아니며 평가의 정확성을 대표한다기보다는 부정확성에 대한 준거라는 비판이 있다.

보다 최근의 연구자들은 인사평가의 효과성에 관하여 직접 준거를 활용한다(Hauenstein, Facticeau, & Schmidt, 1999; Kemter & Moser,

2007; Murphy & Cleveland, 1995; Sinclair, 2000; Stamoulis & Hauenstein, 1993). 직접적인 측정 방법은 통제된 상황에서 직무수행의 수준에 대한 진점수를 알고 있을 때 평가점수와 진점수의 차이에 대한 정확성을 알아보는 것이다. 실제 직무상황에서는 평가자들의 다양한 배경과 맥락에 의해서 진점수를 구하는 것이 용이하지 않고 다면평가를 통한 평가자 간의 상관도 낮기 때문에 평가점수의 평균을 진점수로 간주하기 어렵다. 또, 정확성과 공정성에 이의를 제기할 수도 있는 인사평가 점수를 직무수행의 진점수로 간주하는 경우 인사평가의 정확성을 파악하는 연구에서 순환논리에 빠질 수 있다. 따라서 객관적 평가를 조작할 수 있는 실험연구를 통해서 직무수행의 진점수를 얻는 것이 일반적이며 평가자 행동을 보다 정확히 파악할 수 있는 방법이라고 할 수 있다.

직접적 측정의 평가정확성은 Cronbach(1955)가 제시한 방법으로 파악할 수 있다. Cronbach는 대인지각 및 판단과 관련하여 간접적인 정확성 측정의 문제점을 지적하며, 실제적 현상과 지각적 판단에서 차원별, 피평가자별로 나타나는 변산성을 통해 파악할 수 있는 4가지 구성요소를 구분하였다. 본질적으로 평가 정확성은 평가자의 관찰점수와 실제점수의 차이를 비교함으로써 파악할 수 있다. 이러한 차이는 정확성의 계산방식을 어떻게 조작화하는가에 따라 달라진다. Cronbach는 기존의 단일 지수로 정확성을 계산하는 조작화의 단순성을 비판하면서 정확성(즉, 차이)을 다양한 측면으로 구분하였다. 이렇게 함으로써 하나로 합한 점수로 정확성을 사용할 때 하위 차원에서 발생할 수 있는 과다/과소 추정이 상쇄되거나 무시되는 현상을 막을 수 있는 것이다. 예를 들어, 같은 정확성을 보이는 평가자라도 피평

가자를 정확하게 파악하는지, 또는 평가차원을 정확하게 파악하는 지를 변별할 수 있는 것이다. Cronbach의 정확성 개념은 계산 방식에서 그 동안 약간의 변화가 있었으나 기본개념은 현재까지 인사평가 연구에서 대체로 유지되고 있다.

Cronbach의 요소는 전반적 정확성(elevation; EL), 대상 정확성(differential elevation; DE), 차원 정확성(stereotype accuracy; SA), 대상내 차원 정확성(differential accuracy; DA)이다. 전반적 정확성(EL)은 평가자의 평가점수에서 전반적인 평균의 차이 즉, 모든 피평가자에 대해서 모든 차원에서 평정치의 평균이 진점수와 가까운지를 판단하는 정확성이다. 대상 정확성(DE)은 주로 피평가자들의 효과로 인한 변산성 즉, 피평가자들을 잘 변별하여 서로 다르게 평가를 내렸는지를 판단하는 정확성이다. 차원 정확성(SA)은 평가 차원의 효과로 인한 변산성을 주로 반영하는 것으로 평가 차원을 잘 변별하는지에 대한 정확성이다. 마지막으로 대상내 차원 정확성(DA)은 피평가자와 평가 차원의 상호작용으로 인한 변산성을 반영하는 것으로, 피평가자들에 따른 평가 차원들의 변별에 대한 정확성이다.<sup>1)</sup>

1) 평가정확성은 다음의 식으로 계산되었다.

$$EL^2 = (\bar{x}_{..} - \bar{t}_{..})^2,$$

$$DE^2 = \frac{1}{n} \sum_i [(\bar{x}_i - \bar{x}_{..}) - (\bar{t}_i - \bar{t}_{..})]^2,$$

$$SA^2 = \frac{1}{k} \sum_j [(\bar{x}_{.j} - \bar{x}_{..}) - (\bar{t}_{.j} - \bar{t}_{..})]^2,$$

$$DA^2 = \frac{1}{kn} \sum_i \sum_j [(x_{ij} - \bar{x}_i - \bar{x}_{.j} + \bar{x}_{..}) - (t_{ij} - \bar{t}_i - \bar{t}_{.j} + \bar{t}_{..})]^2$$

,  $n$  = 피평가자 수,  $k$  = 평가차원 수,  $\bar{x}_{..}$  = 모든 평가차원과 피평가자들에 대한 평가자들의 관찰된 점수 평균,  $\bar{t}_{..}$  = 모든 평가차원과 피평가자들에 대한 평가자들의 관찰된 진점수 평균,  $\bar{x}_i$  = 피평가자별 평가자들의 관찰된 점수 평균,

### 평가 정확성 향상을 위한 평가자 훈련

평가의 정확한 수준과 차원을 알게 되면 평가자들에게 어떤 처방을 해야 하는지에 대해 논할 수 있다. 즉, 평가의 정확성을 높이는 방법을 모색하여 인사평가의 질(quality)을 향상시킬 수 있다. 인사평가 연구의 초기에는 평가자 오류 훈련(Rater Error Training; RET)과 같은 훈련을 통해 평가의 오류를 줄이거나 평가의 양식이나 도구의 개발과 같은 심리측정학적(psychometric) 속성을 개선시키고자 하였다. 그러나 Bernardin과 Pence(1980)의 연구에 의하면, 평가자 오류 훈련 집단, 참조틀 훈련 집단, 통제 집단의 평가 오류와 정확성을 비교하면 평가자 오류 훈련 집단의 참가자들이 참조틀 훈련 집단과 통제 집단의 참가자들 보다 평가오류는 줄었으나 평가의 정확성도 줄었다. 이는 후속 연구에서도 유사하게 나타난다(Pulakos, 1984). 이후 Landy와 Farr(1980), Feldman(1981)은 평가의 주체인 평가자의 역할에 더욱 관심을 가져야하며 그들의 인지적 과정(cognitive process)에 초점을 두어야한다고 주장하였고, 평가자들에게 직접적인 개입을 하여 평가자의 정확성을 향상시키는 방안들이 발전되어 왔다.

#### 참조틀 훈련(FOR)과 변산성 훈련(RVT)

평가자 오류 훈련과 같이 평가오류를 줄이는 방법은 효과적인 인사평가를 위한 평가의 질, 정확성이 떨어지는 부정적인 결과로 인해

서 평가 과정에서 간과되었던 평가자의 역할에 관심을 가지게 되었다(Landy & Farr, 1980; Feldman 1981). 참조틀 훈련은 평가차원에서 정확성을 증가시키기 위해서 직무수행을 평가할 때 공통적으로 참조하는 기준(즉, 틀)을 평가자에게 제공한다(Sulsky & Day, 1992). 평가자들에게 우수한, 평균적인, 부족한 수행을 나타내는 사람들을 보여주고 그들에 대한 평가자들의 평가가 얼마나 정확했는지에 관하여 피드백을 준다. 훈련의 목적은 각 수행차원에서 어떤 수행이 효과적인지에 대한 판단이 일치되도록 하기 위하여 평가자들을 훈련시키는 데 있다.

참조틀 훈련에서 우선 기본적으로 평가자는 평가 차원을 훈련 받는다. 평가 차원 훈련은 효율적인 수행의 범위를 나타내는 행동의 예시에 초점을 둔 직무평가 차원의 논의를 포함한다. 다음의 단계는 좋은, 평균, 나쁜 직무수행을 나타내는 중심 사건을 포함하는 짧은 글이나 비디오테이프를 보게 된다. 평가자는 이러한 중심사건을 평가하게 되고, 평가자들은 자신의 평가의 근거를 기록해둔다. 훈련자는 평가대상이 되는 속성(target) 점수에 대한 정보를 제공한다. 마지막으로, 평가자들은 이전에 시행한 평가에 대한 양적(quantitative), 질적(qualitative) 피드백을 받게 되고, 평가자들의 평가점수와 타겟 점수와의 차이에 대해 논의가 이루어진다(Bernardin & Buckley, 1981).

참조틀 훈련은 평가 차원과 관련된 측면에서 뛰어난 정확성을 나타내었지만, 피평가자들 간의 변별에 있어 문제가 있는 것으로 나타났다(Stamoulis & Hauenstein, 1993). 참조틀 훈련은 Cronbach(1955)의 4가지 구성요소 중에서 평가 차원에 기반을 둔 평가 정확성 요소(SA, DA)의 경우에는 평가의 정확성이 가장 증가하

$t_i$  = 피평가자별 평가자들의 진점수 평균,  $x_j$  = 평가 차원별 평가자들의 관찰된 점수 평균,  $t_j$  = 평가 차원별 평가자들의 진점수 평균,  $x_{ij}$  = 평가 차원 j 에서 피평가자 i에 대한 평가자들의 관찰된 점수,  $t_{ij}$  = 평가 차원 j에서 피평가자 i에 대한 평가자들의 진점수 평균

는 것으로 나타났다. 하지만 피평가자에 초점을 둔 평가 정확성(EL, DE)에서는 정확성을 증진시키는 효과가 적었다. 다시 말해서, 참조틀 훈련은 차원들 간의 변별 정확성은 증가하지 않지만, 피평가자들간의 변별 정확성은 증가시키지 못한다는 것이다. 그 이유는 참조틀 훈련의 초점은 피평가자들 간(between)의 차이가 아닌 평가 차원을 아우르는 각각의 피평가자들 내(within) 평가에 대한 논의에 집중하도록 하기 때문이다(Hauenstein, Facticeu & Schmidt, 1999).

이를 보완하는 훈련으로서 변산성 훈련을 제안하였는데 변산성 훈련의 목적은 평가자들의 평가 결과가 피평가자들의 직무수행의 진변량과 일치되도록 평가자들의 변산성을 증가시키는 것이다. 즉, 피평가자들 간의 변별 가능성을 증가시키는 것이다(Hauenstein, 1998). 이러한 목적을 달성하기 위해 변산성 훈련은 전통적인 평가자 오류 훈련과 참조틀 훈련의 요소들을 결합하였다. 변산성 훈련은 평가자 오류 훈련처럼 다른 피평가자들간의 구별을 강조하는 것을 포함하지만 평가자 오류 훈련과는 달리 평가 오류에 초점을 두지 않는다(Hauenstein, 1998). 평가 오류는 정확성과 관련이 없기 때문이다(Murphy & Balzer, 1986). 피평가자들 간의 상대적인 변별을 위해 변산성 훈련은 다음의 단계로 이루어진다. 피평가자들의 직무수행에 대한 평가에 있어 변산성을 증가시키도록 훈련을 받게 되고, 실제로 평가를 해 봄으로써 피평가자들을 상대적으로 변별하는 연습을 한다. 마지막으로 평가자의 평가에 대한 피드백을 제공한다. 참조틀 훈련의 피드백은 실제 수행(진점수)과 평가자의 평정 사이의 차이점에 집중하는 반면, 변산성 훈련의 피드백은 피평가자들 사이의 상대적인 차이점에 관한 논의에 초점을 두는 것이 주요하게

다르다(Sinclair, 2000).

본 연구에서는 먼저 Sinclair(2000)의 선행 연구들을 바탕으로 참조틀 훈련과 변산성 훈련의 훈련 유형에 따라 어떤 효과성을 나타내는지에 대해 알아보려고 한다(가설 1, 2). Cronbach (1955)의 평가 정확성 구성요소를 종속변인으로 하여, 참조틀 훈련은 평가 차원과 차원별 직무 수행수준에 대해 학습하고, 진점수와 차이를 확인하고 줄이는 것을 강조하기 때문에 주로 평가차원에 기반을 둔 정확성 구성요소인 차원 정확성(SA)과 대상내 차원 정확성(DA)에서 변산성 훈련보다 더욱 정확한 평가를 할 것이며, 변산성 훈련은 진변량과 일치도록 학습을 하고, 개별 피평가자들 간의 직무수행 차이를 확인하는 것을 강조하기 때문에 평가 차원들을 아우르는 정확성 구성요소인 전반적 정확성(EL)과 대상 정확성(DE)에서 참조틀 훈련보다 더욱 정확한 평가를 할 것이라고 예상된다. 이러한 결과를 바탕으로 다음과 같은 가설을 도출하였다.

**가설 1.** 전반적 정확성(EL)과 대상 정확성(DE)에는 변산성 훈련이 참조틀 훈련보다 더 효과적일 것이다.

**가설 2.** 차원 정확성(SA)과 대상내 차원 정확성(DA)에는 참조틀 훈련이 변산성 훈련보다 더 효과적일 것이다.

#### 평가 정확성에 대한 평가 책임성의 영향

Murphy와 Cleveland(1995)는 인사평가를 순수한 인지적인 요인만으로 이루어지는 과정이라고 볼 수 없으며, 인지적 요인 이외의 평가자-피평가자간의 상호작용에서 존재하는 다양한 사회적 요인과 인사평가 맥락의 상황적 요인

들이 인사평가에 영향을 미친다고 하였다. 이러한 사회적, 상황적 요인들은 정확한 평가를 하고자 하는 평가자의 동기와 관련된다. Harris (1994)는 평가자는 정확한 정보를 가지고 있다고 가정하고 동기화되어지지 않은 평가자들은 동기화 되어진 평가자들에 비해 신중한 정보 처리 기술을 사용하지 않는다고 하였다.

평가자의 동기를 향상시키는 방법들 중 다른 사람들에게 자신의 평가에 대한 설명이나 이유를 전해야 되는 방법을 책임성(accountability)이라고 한다. 책임성은 자신의 관점을 다른 사람에게 정당화해야 할 필요성 또는 압력으로 정의된다(Tetlock, 1983). 인사평가 맥락에서 책임성은 평가를 더 정확하게 하도록 하고, 책임성이 낮은 조건 참가자들은 관대한 평가를 하는 경향이 있다(Mero, Guidice, & Brownlee, 2007). Hauenstein(1992)은 평가를 하는 상황에서 책임성의 부여 여부에 따른 평가의 정확성의 차이에 대한 연구에서도 실험을 통해 피평가자의 직무수행을 관찰하기 전 후에 피평가자와 대면상황에서 평가에 대한 피드백을 주어야 하는 경우 평가의 관대화 경향이 있는 것으로 나타났다. 이는 피평가자와의 면담을 통해 피평가자의 저수행자에 대한 피드백을 주어서 불편한 관계가 되는 것을 꺼리기 때문에 정확한 평가를 하지 못한다는 것이다. 또한 책임성이 있으면 고수행자의 대비 오류(contrast error)를 줄이며 고수행자 행동을 잘 회상하며, 저수행자에 대한 관대화 오류가 줄고 평가의 정확성이 높아지는 것으로 나타났다(Palme & Feldman, 2005). 대체적으로 기존 연구는 평가자들에게 인사평가 시스템 상에서 책임성에 대한 지각을 높여준다면 더 정확한 평가가 가능하다는 것을 보여준다.

### 평가자 훈련과 책임성의 상호작용

본 연구에서는 동기적 압박인 책임성을 부가하는 것이 훈련 유형에 따른 평가 정확성의 효과성에 영향을 주는지 알아보고자 한다. 기존 연구를 볼 때 책임성은 평가자들에게 더욱 적극적으로 문제에 대해 이해하고 보다 복잡한 인지적 노력을 가하여 정보처리를 하도록 만든다는 것을 추론할 수 있다. 본 연구에서는 인사평가의 맥락에서 평가자들이 받은 평가자 훈련 유형에 따라 참조틀 훈련 조건은 평가 차원과 차원에 따른 직무수행의 수준을 알게 되고 진점수와의 차이를 줄이려는 인지적 학습과정을 거치고, 변산성 훈련은 피평가자들의 수행의 진변량과 정확하게 일치되어지도록 평가자들의 변량을 증가시켜 피평가자들 간의 구별가능성을 증가시키려는 인지적 학습을 거치게 된다. 이에 책임성을 부가하는 경우 평가자들은 각각의 훈련에서 중요시 여기는 인지적 과정을 더욱 정교하게 거치고, 신중하게 정보처리를 할 것으로 예상된다. 따라서 다음과 같은 가설을 도출 하였다.

**가설 3-1.** 전반적 정확성(EL)과 대상 정확성(DE)의 경우, 변산성 훈련이 참조틀 훈련보다 효과적인 결과는 책임성이 부가될 경우 더 증가할 것이다.

**가설 3-2.** 차원 정확성(SA)과 대상내 차원 정확성(DA)의 경우, 참조틀 훈련이 변산성 훈련보다 효과적인 결과는 책임성이 부가될 경우 더 증가할 것이다.

### 평가정확성에 대한 평가자 기분(mood)의 영향

Forgas(1987)는 두 가지 주된 방법으로 정서가 사회적 판단 정확성에 영향을 미친다고 하

었다. 첫째는 정보 효과(information effect)로서 긍정기분의 사람들은 다른 사람들을 위해서 더 호의적으로 보고하고 더 긍정적으로 평가하는 경향이 있고, 부정기분에 있는 사람들은 더 부정적인 판단을 하게 한다는 것이다. 두 번째는 프로세스 효과(processing effects)로서 긍정기분은 기존의 일반 지식구조에 의존하게 만들고 편법적(heuristic) 정보처리를 하고 세부 사항에 대한 주의를 적게 기울이게 만드는 반면 부정기분은 체계적, 분석적 정보처리를 하게 만들며 세부 사항에 주의를 기울이게 만든다. Schwarz(2002)도 부정기분은 체계적인 판단을 증가시키고, 대조적으로 긍정기분은 체계적인 판단과정을 감소시키고 휴리스틱을 나타나게 한다고 주장한다.

Sinclair(1998)는 인사평가에서도 긍정기분 상태의 평가자는 긍정적인 정보를 기억에서 회상하고, 긍정적으로 직무수행을 평가(즉, 관대한 평가)하려는 경향이 있다고 하고, Lefkowitz(2000)도 문헌들을 리뷰한 결과 평가자들이 긍정기분일 때 더 관대한 평정을 하고, 덜 엄격한 처벌을 하는 경향이 있으며, 더 부정확한 평가를 하게 된다는 것을 밝혔다. 이것은 긍정기분일 때 전략과 문제 해결에 대한 만족의 역치가 낮기 때문에 세부사항에 대한 주의를 덜 기울이고 질적 수행을 요구하는 전략에서도 깊이 생각하지 않는다. 하지만 부정기분조건에서는 현재 상황이 위협적이고, 해롭다는 신호로 받아들여 현재 상황을 변화시키려고 동기화 된다(Forgas, 1987).

평가자 훈련을 통해 평가자의 평가 기술을 향상시키는 효과성이 평가자의 기분에 따라 평가가 달라진다는 것은 평가자 훈련의 비용과 시간을 낭비하는 것이라고 할 수 있다. 따라서 인사평가에 있어서 평가자 기분이 미치

는 영향을 파악하여 효과적인 평가자 훈련에 참고할 필요가 있다. 본 연구에서는 기분이 인지적 과정에 영향을 미친다는 연구 결과를 고려하여 긍정기분 조건의 참가자들은 관대하며 평가에 주의를 기울이지 않고 체계적인 판단과정이 감소하고 휴리스틱을 사용하게 될 것으로 예상되며, 부정기분 조건의 참가자들은 더 많은 주의를 기울여 평가를 할 것이라는 기대를 하고 있다. 따라서 다음과 같은 가설을 도출하였다.

**가설 4-1.** 전반적 정확성(EL)과 대상 정확성(DE)의 경우, 참조틀 훈련집단에 비해 변산성 훈련집단에서 부정기분 집단이 긍정기분 집단보다 더 정확하게 평가할 것이다.

**가설 4-2.** 차원 정확성(SA)과 대상내 차원 정확성(DA)의 경우, 변산성 훈련집단에 비해 참조틀 훈련집단에서 부정기분 집단이 긍정기분 집단보다 더 정확하게 평가할 것이다.

#### 평가자 기분과 책임성의 상호작용

Bodenhausen, Kramer,과 Susser(1994)는 책임성과 기분의 효과를 함께 고려한 연구는 많지 않다고 소개를 하며 긍정기분의 참가자들이 고정관념적(stereo type) 편파를 사용하여 판단을 한다는 결과에서 책임성이라는 인지적 동기를 부여한다면 참가자들은 고정관념적 단서의 의존성이 줄어들 것이라고 가정을 하였다. 연구 결과는 긍정기분의 참가자들에게 책임성이 부여되었을 때, 통제된 기분 조건에 속한 참가자들의 고정관념적 편파 수준과 차이가 나지 않을 정도로 낮아지는 것을 발견하였다. 또한 Bless, Bohner, Schwarz,와 Strack(1990)은 설득적 메시지에 대해 실험 참가자들의 일치(설득)되어지는 경향을 연구하면서 책임성 부여



와 기분 조작을 통해 어떤 변화가 나타나지는지를 알아보았다. 책임성은 집중 토론과 가벼운 토론을 통해 판단자의 의견을 표출하도록 조작하고, 기분은 긍정기분과 부정기분으로 구분할 때, 설득되는 경향은 기분이 책임성의 영향을 조절하는 것으로 나타났다. 즉, 부정기분 조건의 참가자들은 집중 토론 조건에서 더욱 일치되지 않는 경향을 나타냈고, 긍정기분의 참가자들에게서는 토론의 형태에 따른 변화가 거의 없는 것으로 나타났다.

본 연구에서는 책임성이 부여된 참가자들은 책임성이 부여되지 않은 조건의 참가자들에 비해 더욱 신중한 정보처리 과정을 거치더라도, 긍정기분 조건에서는 인지적 노력의 영향이 약화되어 현재 느끼는 긍정 기분에 보다 의존하게 되어 부정확한 평가를 할 것으로 예측한다(Schwarz & Clore, 1983, 1988). 반면 부정기분의 조건에서, 책임성이 부여된 참가자들은 책임성에 의해 더 복잡한 인지적 노력을 들이는 정보처리 전략을 선택하고 기분의 영향으로 더욱 신중한 정보처리과정을 거치면서 더욱 정확하게 평가할 것이다. 따라서 다음과 같은 가설을 도출하였다.

**가설 5.** 책임성과 평가자 기분의 상호작용 효과가 평가정확성에 영향을 줄 것이며, 구체적으로, 부정기분 집단이 긍정기분 집단보다 더 정확한 정도는 책임성이 부가될 때 더 증가할 것이다.

#### 평가자 훈련과 책임성, 평가자 기분(mood)의 삼원상호작용

이상의 논의에서 평가자의 인지적 정보처리 과정을 도와주고 정확한 평가를 할 수 있

도록 하기 위해 평가자 훈련이 필요하며(DeNisi, 1996), 책임성은 복잡한 인지적 노력을 들이도록 정보처리를 유도하여 정확한 평가를 하도록 하며(Schlenker, 1985), 기분에 따라 체계적이고 분석적인 정보처리 과정을 거친다고 하였다(Forgas, 1987). 각 요인들끼리 상호작용을 한다는 선행연구를 종합해 보면 평가자 훈련이 기여하는 정확성에는 이런 요인들의 상호작용이 더 크게 작용할 것으로 예상할 수 있다. 즉, 평가자 훈련 참가자들은 훈련에 맞는 핵심부분을 학습하고 이를 인지적 정보처리 기준으로 여기게 된다. 이때 책임성이 부여된 참가자들은 책임성이 부여되지 않은 참가자들에 비해 자신이 학습한 훈련의 평가 기준에 맞도록 신중한 정보처리 과정을 거치면서 더욱 정확한 평가를 하고자 할 것이다. 하지만 참가자들이 느끼는 기분에 의해 긍정기분의 참가자들은 인지적 노력을 덜 들이고 현재의 기분에 의존하여 보다 덜 정확한 평가를 할 것으로 예상이 되고, 부정기분의 참가자들은 더욱 복잡한 인지적 노력을 들이는 정보처리 전략을 사용함으로써 더욱 정확한 평가를 할 것으로 예상이 된다. 따라서 다음의 가설을 도출하였다.

**가설 6-1.** 전반적 정확성(EL)과 대상 정확성(DE)의 경우, 책임성이 부여되고 부정적인 기분 상태이면서 변산성 훈련을 받는 조건에서 가장 정확하게 평가할 것이다.

**가설 6-2.** 차원 정확성(SA)과 대상내 차원 정확성(DA)의 경우, 책임성이 부여되고 부정적인 기분 상태이면서 참조틀 훈련을 받는 조건에서 가장 정확하게 평가할 것이다.

## 연구 방법

### 연구 참가자

본 연구는 서울 및 부산 소재 4년제 대학에 재학하는 157명의 대학생을 대상으로 실험연구를 진행하였다. 실험 참가자들 중 불성실한 참가자를 제외한 146명의 참가자들을 대상으로 자료를 분석에 이용하였다. 분석에 포함된 전체 참가자들의 인구통계학적 변인들을 살펴보면 성별은 남자(78명, 53.4%)가 여자(68명, 46.6%)보다 조금 많았다. 참가자들의 나이의 범위는 20~27세였고 평균은 22.1세였다(SD=1.98세).

### 실험 설계

본 연구의 실험설계는 Stamoulis와 Hauenstein (1993), Sinclair(2000)의 실험절차를 주로 참고하여 구성하였다. 평가자 훈련은 참조틀 훈련, 변산성 훈련으로 나누어 실시하며, 평가자 기분에 따른 집단 구분은 긍정적 기분 조건과 부정적 기분 조건으로 조작을 하였고, 책임성도 책임성을 부여하는 집단과 부여하지 않는 집단으로 구분되는 조작을 하였다. 즉, 2(훈련 조건) × 2(기분) × 2(책임성) 등 8가지 조건으로 설계하였다. 실험 집단의 인원은 실험도중 일부 불참자가 발생하여 셀마다 약간 차이가 있었다(셀 크기=17~20 명).

실험 1단계에서 피험자의 사전평가 점수를 바탕으로 초기 평가 정확성을 측정하여, 가설 검증 시에 공변인으로서 사용하여 피험자 개인의 초기 평가정확성 수준을 통제하고자 하였다. 사전 평가 정확성은 대상 정확성(DE)에서 두 훈련 조건의 차이가 있었고, 책임성 부여

조건과 기분 조작 조건에서 유의미한 차이는 없었다. 그러나 절대 점수의 차이는 다소 있었기 때문에 초기 평가 정확성을 공변인으로서 통제하였다.

### 진점수 측정

Cronbach의 4가지 평가 정확성 구성요소를 구하기 위해서는 피평가자들의 강의 수행 수준에 대한 진점수(target score)가 필요하다. 이를 위해 먼저 8명의 대학원생을 강사로 선정하여 강의하는 내용을 동영상으로 녹화 및 편집하였다. 강의 평가 영상물은 대학원생이 “설문지 개발 과정”이라는 주제를 프레젠테이션하는 모습을 촬영하여 제작을 하였다. 피평가자들의 강의 수준은 사전에 미리 설정해 두었다. A~C 등 3 명의 피평가자는 사전 평가 대상자이고, D~H 등 5명의 피평가자는 사후 평가 대상자이다. 7점 척도에서 1~2점은 하, 3~4점은 중, 5~7점은 상의 수준으로 정하였다. 사전 평가 대상자의 3명의 강의 수준과 사후 평가 대상자 5명의 강의수준은 무선적으로 섞어서 제시하였다.

강의평가에 대한 진점수를 구하기 위해서 인사평가와 관련된 전공의 대학원생 11명에게 실험의 목적을 설명하고 두 평가자 훈련(FOR과 RVT)을 실시한 뒤, 강의평가에 사용될 강의 영상물에서 강의자 역할을 하는 8명에 대해 평가를 실시하였다. 이들의 평가 점수의 동의도는  $r_{wg}$  지수가 .83~.98로 높은 수치를 나타내었다. 이것은 대학원생들이 8명의 피평가자들의 강의 수준에 대해 공통적인 시각으로 바라보고 있음을 나타내며, 진점수로 사용하는데 적절하다고 판단할 수 있다. 평가 차원별 평균 점수와 총점의 평균 점수를 보면 사전에 피평가자들에게 설정하였던 수준과 같게 나타났다.

본 연구의 가설을 검증하기 위한 4가지 평가 정확성을 계산하기 위해서 대학원생들에게서 구한 이 점수를 진점수로 사용하였다.

### 실험 절차

실험은 다음과 같은 시간과 절차를 거쳐 진행되었다. 1) 참가자들에게 실험에 대한 개략적인 소개를 한 후, 일정과 평가의 목적, 평가자 작성 요령에 대한 설명을 하고 그 시점에서 실험참가자의 기분을 측정하는 설문지를 실시하였다. 2) 실험참가자가 3명의 피평가자들이 강의를 하는 사전 평가 영상물을 시청한 뒤 평가를 실시하였다. 3) 실험참가자들을 8가지 조건 중 한 집단에 무선적으로 할당된 후, 두 가지 평가자 훈련 중 하나의 평가자 훈련을 실시하였다. 4) 5분 정도의 휴식을 가졌다. 5) 사전에 설계된 대로 참가자들이 속한 집단에 맞는 기분 조작 영상물을 시청하도록 하였다. 6) 사후 평가를 하기 전에 책임성 부여 집단에게 책임성을 부여하는 언어적 조작을 하고 통제 집단에게는 언급하지 않았다. 7) 5명의 피평가자들이 강의를 하는 사후 평가 영상물을 시청한 뒤 평가를 실시하였다. 8) 실험의 목적을 설명하고 참가자들의 질문에 응답을 하면서 실험을 종료하였다. 이상의 실험은 1시간 30분 동안 진행되었고, 세부적인 절차는 다음과 같다.

### 실험 1단계

#### 강의 평가 대상 영상물 제시

사전 평가를 실시하기 전에 참가자들에게 실험의 목적과 실험 일정에 대한 소개를 실시하였다. 참가자들이 평가를 하게 될 영상물에

대한 소개와 평가자 작성요령에 대해 설명하였다. 강의 영상물의 평가 차원별(구성력, 전달력, 강의내용) 구성은 피평가자 별로 A는 상/상/중, B는 중/중/하, C는 하/하/중으로 구분하였고, 전반적인 수준은 A는 상, B는 중, C는 하로 구분하였다. 영상물 시청에 앞서 현재 참가자들의 기분을 측정하였으며, 3명의 피평가자들의 강의 영상물은 순서효과를 고려하여 무선적으로 제시하였다.

#### 사전 평가

이 단계의 평가는 피평가자의 평가정확성에 관한 기준점수(baseline)를 측정하기 위한 목적이었다. 강의 평가 척도는 Sinclair(2000)의 연구에서 사용된 평가 척도를 수정하여 사용하였다. 평가 차원은 구성력(강의 자료의 논리성, 강의 전개와 자연스러움, 강의 장비 사용 능력), 전달력(대중적 말하기 능력, 강의 자료 구성과 내용에 관한 것이 아닌 강의 전달 방식 자체), 강의내용(강의 내용과 관련된 강의 내에서 제시된 정보의 폭, 강사가 강의를 이해하고 있는 정도) 등 3가지 차원이며, 전반적 평가를 더해 총 4가지를 측정하였다.

참가자들은 각기 다른 3명의 피평가자들의 강의 수준을 3가지 평가 차원별로 평가를 하고, 관찰된 평가 결과는 평가자 훈련을 실시할 때 피드백의 자료로 사용하였다. 평가자 작성은 강의 영상물을 시청하면서 각각의 피평가자들에 대한 메모를 하고, 마지막 강의 영상물 시청이 끝난 후 점수를 부여 하도록 하였다.

### 실험 2단계: 평가자 훈련

참조틀 훈련, 변산성 훈련의 설계는 Stamoulis

와 Hauenstein(1993), Sinclair(2000)에서 사용되어진 것을 기초로 하였다. 평가자 훈련 시작 시점에 훈련 자료를 나누어 주고, 기본적인 훈련은 영상물을 제작하여 모든 참가자들에게 표준화된 상태로 실시하였다. 참가자들의 질문에 응답하거나 추가적인 설명은 실험자가 훈련자(trainer)로서 현장에서 직접 설명을 하였다. 공통적으로 실시된 훈련 내용은 강의평가의 의의, 목적, 중요성에 대한 것과 평가를 할 때 관찰/메모의 중요성에 대한 것이었다. 그 다음으로 평가자 훈련 유형에 맞는 핵심 내용에 대해 설명을 해주었다.

참조틀 훈련은 다음과 같은 내용으로 구성되었다. 첫째, 참가자들에게 강의 평가의 목적이 피평가자들의 자기 개발을 위해 평가 차원별로 우수하거나 부족한 평가 차원을 구분하여 피드백을 주는 것임을 알려주었다. 둘째, 강의 평가의 평가 차원에 대한 상세한 설명을 실시하고, 참가자들의 질문에 응답을 하였다. 셋째, 참가자들에게 강의 중 피평가자가 하는 행동들을 정리한 훈련 자료를 제공해 주었고, 각각의 행동 지시문이 어떤 평가 차원에 속하게 되는지를 체크하도록 하였다. 넷째, 평가 차원과 강의 행동을 연결하여 참가자들에게 설명해 주었고 참가자들의 질문에 응답해 주었다. 다섯째, 평가차원별 강의 행동의 수준에 대해 설명을 해주었다. 각 차원에서 강의 행동의 수준은 상, 중, 하로 구분 하였다. 여섯째, 참가자들이 사전 평가하였던 평가 결과에 대해 피드백을 주었다. 피드백을 제공할 때에는 미리 설정해 두었던 진점수와 참가자들의 관찰된 점수를 비교하여 얼마나 비슷하고 차이가 나는지를 알려주었다. 피드백은 피평가자별로 평가 차원에 관하여 제공되었고 참가자들이 작성한 메모가 정확하게 되었는지를

검토해주었다. 마지막으로, 참조틀 훈련을 요약한 훈련 자료를 바탕으로 핵심 요약정리를 해주었고 참가자들의 질문에 응답해 주었다.

변산성 훈련의 핵심 내용은 다음과 같다. 첫째, 참가자들에게 강의 평가의 목적이 동영상 강의의 강사를 선정하기 위해 피평가자들 간 정확한 변별을 하는 것임을 알려주었다. 둘째, 강의 평가의 평가 차원에 대해 설명하고, 참가자들의 질문에 응답을 하였다. 셋째, 참가자들에게 강의 중 피평가자가 하는 행동들을 정리한 훈련 자료를 제공해 주었고, 각각의 행동 지시문의 행동 수준(상, 중, 하)에 대해 체크 하도록 하였다. 넷째, 평가 차원과 강의 행동 수준을 연결하여 참가자들에게 설명해 주었고 참가자들의 질문에 대해 응답해 주었다. 다섯째, 참가자들에게 평가 정확성을 향상시키기 위해서는 진변량과 동일하게 하는 것임을 알려주고, 피평가자들의 상대적인 차이에 중점을 두고 점수의 차이가 확연하게 나도록 평가를 해야 함을 설명해 주었다. 여섯째, 참가자들이 사전 평가하였던 평가 결과에 대해 피드백을 주었다. 피드백을 제공할 때에는 미리 설정해 두었던 진변량과 참가자들의 관찰된 점수를 비교하여 얼마나 비슷하고 차이가 나는지를 알려주었다. 피드백은 평가 차원별로 제공되었고, 참가자들이 작성한 메모가 정확하게 되었는지를 검토 해 주었다. 마지막으로, 변산성 훈련을 요약한 훈련 자료를 바탕으로 핵심 요약정리를 해주었고 참가자들의 질문에 응답해 주었다.

훈련자는 훈련이 끝난 뒤 참가자들에게 그들의 사전 평가했던 평가지를 나누어 주고, 훈련 유형에 맞게 피드백을 제공하고, 평가자들이 어떠한 것에 초점을 두어야 하는지에 대해 논의를 하였다. 그리고 참가자들의 질문에

대해 응답을 해주었다. 마지막으로, 각 훈련이 포함하는 내용을 간단히 요약하여 제시를 하였다.

### 실험 3단계

#### 기분 조작

평가자 훈련 후 참가자들에게 다음 절차를 위한 준비시간이 필요하다고 하면서 그 동안 영상물을 보여주겠다고 안내하고 기분을 조작하였다. 긍정기분 조건은 유쾌한 기분을 느낄 수 있도록 코메디 영상물을 편집하여 제공하였다. 부정기분 조건은 실제 가족의 가장이 죽음을 맞는 과정에 대한 현장 다큐멘터리 영상물을 편집하여 제시하였다. 두 조건에서 영상물을 제시한 시간은 6 분으로 같도록 하였다.

#### 책임성 조작

책임성 부여 조건의 참가자들에게 평가자의 역할이 중요하다는 것을 강조하면서 참가자의 평가 결과는 실제 평가의 자료로 사용되어지며 피평가자에게 전달되어 진다는 것을 인식시켰다. 그리고 참가자들에게 나누어준 강의 평가지에 이름과 학번을 쓰도록 하였다. 책임성을 부여하지 않는 통제 집단에게는 위와 같은 과정을 거치지 않았다.

#### 강의 평가 대상 영상물 제시

사후 평가 영상물은 5가지 수준(최상, 상, 중, 하, 최하)으로 구성되었다. 강의 평가 영상물 시청에 앞서 기분 조작을 확인하기 위해서 기분 조작 영상물을 시청한 후 현재 참가자들이 느끼는 기분을 측정하였다. 5명의 피평가자들의 강의 영상물은 순서효과를 고려하여

무선적으로 보여주었다.

#### 강의 평가

사전 평가와 동일한 방법으로 실시되었다. 단지 3명의 피평가자가 아닌 5명의 피평가자를 평가한다는 것이 차이점이다.

#### 분석

연구에 사용된 기분 측정 척도의 신뢰도를 확보하고 평가자 훈련의 유형이 평가 정확성에 미치는 영향과 책임성과 평가자 기분의 상호작용 효과를 검증할 위해 신뢰도 분석, 상관분석, t-검증, 공변량 분석(ANCOVA)을 실시하였고 SPSS 12.0 for windows 프로그램을 사용하였다.

## 결 과

### 평가 정확성 측정

참가자들에게 매긴 강의 평가 점수를 전점수와 비교하였다. 피평가자들의 강의 능력을 평가하는 강의 평가지는 총 8장으로서 사전 평가시 3명, 사후 평가시 5명을 평가할 수 있도록 구성되었다. 강의 평가는 Likert 7점 척도(1점: 매우 비효과적인 강의, 7점: 매우 효과적인 강의)를 사용하였다. 평가 정확성의 측정은 Cronbach(1955)의 4가지 정확성 구성요소인 전반적인 정확성(EL), 대상 정확성(DE), 차원 정확성(SA), 대상내 차원 정확성(DA)을 통해 알아보았다. 4가지 정확성 구성요소의 값은 적을수록 정확한 평가를 하였다고 해석이 된다. 8가지 실험 조건의 평가정확성은 표 1과 같이 도출되었다.

표 1. 실험 조건별 사후 평가 정확성 평균 점수

정확성	책임성	기분	FOR			RVT		
			M	SD	N	M	SD	N
EL	무	긍정	.38	.41	20	.21	.24	19
		부정	.15	.19	17	.21	.27	17
	유	긍정	.21	.28	17	.15	.17	20
		부정	.24	.57	18	.36	.60	18
DE	무	긍정	.74	.41	20	.54	.35	19
		부정	.72	.34	17	.69	.36	17
	유	긍정	.85	.58	17	.55	.39	20
		부정	.77	.59	18	.62	.38	18
SA	무	긍정	.15	.13	20	.14	.14	19
		부정	.30	.35	17	.20	.24	17
	유	긍정	.19	.19	17	.12	.08	20
		부정	.16	.23	18	.17	.18	18
DA	무	긍정	.90	.53	20	.85	.34	19
		부정	.82	.52	17	.74	.37	17
	유	긍정	.75	.28	17	.82	.41	20
		부정	.82	.26	18	.88	.34	18

주) 평균값은 낮은 점수가 정확한 지표임.

기분 조작 점검

기분 평가는 사전 평가를 실시하기 전과 기분 조작 영상물을 시청한 뒤(사후 평가를 실시 전)에 측정하였다. 측정도구는 정서측정 관련 기존 연구(이주일 & 김명언, 1999; 한태영 & 이세란, 2005)를 바탕으로 방향성(긍정, 부정)과 각성 수준의 차원(높은 각성, 중간 각성, 낮은 각성)으로 나누어 구성하였고, 부정 정서를 나타내는 3문항(화남, 짜증남, 우울함)과 긍정 정서 3문항(행복함, 유쾌함, 편안함)을 사용하여 5점 척도로 측정했다. 기분에 대한 측정

은 사전, 사후 평가를 실시하기 전에 각각 실시하였다. 두 정서측정 도구의 신뢰도( $\alpha$ )는 사전(긍정 .73, 부 .84), 사후 (긍정=.75, 부정=.82) 모두 양호했다.

사후에 대응표본 t-검증을 통해 조작을 확인한 결과, 긍정기분 집단은 사전에 가지고 있던 부정적인 기분이 사후에 측정하였을 때에 유의미하게 감소하였다( $t=4.26 p<.01$ ). 긍정적인 기분도 사전보다 사후에 증가 하였지만 유의미하지는 않았다. 부정기분 집단은 유의미하게 긍정적 기분은 감소하고, 부정적인 기분은 증가하였다( $t= 3.18 p<.01, t= -8.35$

$p < .01$ ). 이와 같은 결과를 바탕으로 본 연구에서 의도한 기분 조작은 성공적으로 조작되었다고 볼 수 있다.

가설 검증

먼저 실험 전에 피험자의 평가정확성 수준이 유사한지 확인하기 위하여 사전평가의 참조틀 훈련과 변산성 훈련 조건의 비교, 책임성 유무 조건의 비교, 및 부정-긍정 기분 조건의 비교를  $t$ -검증으로 실시하였다. 두 훈련의 사전 평가는 DE에서만 유의미한 차이가 있었다( $t = -3.37, p < .01$ ). 반면 책임성 부여 조건과 기분 조작 조건에서의 사전 평가는 모두 유의미한 차이가 나타나지 않았다. 그러나 피험자의 초기 상태를 통제하기 위해서 평가자 훈련 유형과 평가 정확성 향상의 관계에 대한 검증은 공변량 분석(ANCOVA)을 실시하였다. 이때 공변량은 피험자들에게 실시한 사전 강의평가의 점수를 투입하여 피험자 개인의 평가정확성 수준을 통제하고자 하였고, 사전 점수와 사후 점수의 상관은 DA를 제외하면  $-.05 \sim .06$  사이로 유의미하지 않은 상관을 보였다. DA는 사전, 사후의 상관이  $.25$ 로 유의미한 상관을 보였다.

표 2의 결과와 같이 평가자 훈련과 책임성, 평가자 기분의 삼원상호작용은 유의미하지 않은 것으로 나타났다(EL은  $F(1,137) = .05, ns$ ; DE는  $F(1,137) = .00, ns$ ; SA는  $F(1,137) = 1.75, ns$ ; DA는  $F(1,137) = .01, ns$ ). 세 조작 조건의 상호작용은 책임성과 평가자 기분의 상호작용이 전반적 정확성(EL) 측면에서만 유의미하게 나타났다(EL;  $F(1,137) = 4.36, p < .05$ ). 하지만 본 연구에서 제시한 가설 5에서 예측한 것과 다른 방향이었다. 즉, 책임성이 부여된 집단의 참가

표 2. 평가 정확성에 대한 공변량 분석 결과

	EL	DE	SA	DA
	F (1, 137)			
사전평가	1.25	3.25	.79	12.76
훈련(A)	.08	7.57**	1.76	.18
책임성(B)	.00	.15	1.19	.54
기분(C)	.03	.44	2.96	.05
A * B	.37	1.09	.16	.95
A * C	2.69	1.62	.01	.01
B * C	4.36*	.10	1.97	.90
A * B * C	.05	.00	1.75	.01

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

자들 중 긍정기분의 참가자들은 정확성이 향상된 반면, 부정기분의 참가자들은 정확성이 감소하였다(그림 1). 책임성이 부여되지 않은 집단의 참가자들은 긍정기분일 때 정확성이 감소하는 반면, 부정기분에서 더 정확한 평가를 하는 것으로 나타났다. 마지막으로 따라서 가설 3, 4, 5, 6은 지지되지 않았다.

또, 표 2에 제시하였듯이, 평가 정확성에 대한 훈련의 주효과는 대상 정확성(DE)에서만 유의미하였다( $F(1,137) = 7.57, p < .01$ ). 대상 정확성(DE)은 훈련의 유형에 따른 차이가 유의미하게 나타남으로써 참조틀 훈련을 받은 경우와 변산성 훈련을 받은 경우, 평가 정확성이 유의미하게 달라진다는 사실이 검증되었다. 반면, 다른 평가 정확성 요소들에서는 참조틀 훈련과 변산성 훈련의 효과 차이는 유의미하지 않았다(EL은  $F(1,137) = .08, ns$ ; SA는  $F(1,137) = 1.76, ns$ ; DA는  $F(1,137) = .18, ns$ ). 따라서 가설 1은 부분적으로 지지되었고, 2는 지지되지 않았다. 이러한 연구 결과는 참조틀 훈련이

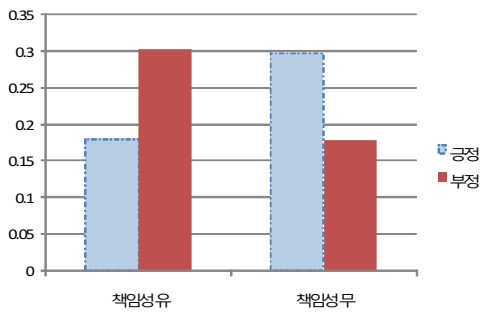


그림 1. 전반적 정확성(EL)에 대한 책임성과 평가자 기분의 상호작용

차원 정확성(SA), 대상내 차원 정확성(DA)에서 효과적이고, 변산성 훈련은 전반적 정확성(EL), 대상 정확성(DE)에서 효과적이라는 Hauenstein (1998)의 연구와 다른 결과이다. 평가자 훈련의 주효과 외에 다른 변인의 주효과는 유의미하지 않은 것으로 나타났다.

이상에서 상호작용 효과가 대부분 유의미하지 않았기 때문에, 본 연구에서는 추가적으로 세 주효과 변인이 네 가지 평가 정확성 구성 요소별로 얼마나 영향을 주었는지 알아보았다.

먼저, 표 3에 제시된 바와 같이 사전평가에 비해서 훈련 후 사후평가에서 평가 정확성이 얼마나 향상되었는지 평가자 훈련의 효과를 세부적으로 알아보았다. 참조틀 훈련 집단은

대상내 차원 정확성(DA)에서 유의미하게 평가 정확성이 개선되었다( $t=-4.19, p<.01$ ). 이는 기존의 연구와 일치하는 방향이다. 반면, 차원 정확성(SA)은 유의미한 변화가 없었고, 대상 정확성(DE)는 정확성이 오히려 감소되는 방향으로 유의미하였다( $t= 3.51, p<.01, t=.80, p>.01$ ). 변산성 훈련 집단은 차원 정확성(SA)을 제외한 3가지 평가 정확성이 모두 향상되는 경향을 나타내었고, 모두 유의미 하였다(EL;  $t=-2.65, p<.01$ , DE;  $t=-3.01, p<.01$ , DA;  $t=-5.29, p<.01$ ). 이 결과는 변산성 훈련이 전반적 정확성(EL), 대상 정확성(DE)를 개선시킨다는 기존의 주장과 유사하나, 추가적으로 대상내 차원 정확성(DA)도 개선되는 결과가 나타난 것이다.

표 4는 책임성 효과에 대한 추가분석으로 네 가지 평가 정확성 요소별로 책임성 부여에 따라 평가 정확성이 얼마나 향상되는지를 나타낸 것이다. 책임성이 없는 조건에는 대상내 차원 정확성(DA)측면에서만 평가 정확성이 유의미하게 향상되었다( $t=-3.97, p<.01$ ). 반면, 책임성이 부여된 집단의 경우에는 EL, DA가 유의미하게 평가 정확성 지수가 감소하여 개선되는 결과를 나타내었다(EL;  $t=-1.96, p<.05$ , DA;  $t=-5.49, p<.01$ ). 즉, 책임성 여부를 분리하여 볼 때는 책임성 부가 집단의 평가정확성이

표 3. 훈련 유형별 평가 정확성 평균 점수

	FOR 훈련 (n=72)			RVT 훈련 (n=74)		
	사전	사후	t값	사전	사후	t값
EL	.31(.46)	.25(.40)	-.95	.46(.68)	.23(.36)	-2.65**
DE	.49(.48)	.76(.48)	3.51**	1.06(1.36)	.60(.37)	-3.01**
SA	.17(.20)	.20(.24)	.80	.15(.20)	.16(.17)	.16
DA	1.14(.62)	.83(.42)	-4.19**	1.36(.92)	.83(.37)	-5.29**

주) 낮은 점수일수록 정확한 평가임, 괄호 안의 값은 표준편차. \* $p<.05$ , \*\* $p<.01$



표 4. 책임성 조작여부별 평가 정확성 평균 점수

	책임성 부가하지 않음 (n=73)			책임성 부가 (n=73)		
	사전	사후	t값	사전	사후	t값
EL	.36(.48)	.24(.30)	-1.76	.41(.67)	.24(.44)	-1.96*
DE	.76(.69)	.67(.37)	-1.10	.81(1.34)	.69(.50)	-.70
SA	.17(.23)	.19(.23)	.69	.15(.18)	.16(.17)	.29
DA	1.14(.73)	.83(.44)	-3.97**	1.37(.84)	.82(.33)	-5.49**

괄호 안의 값은 표준편차임. \* $p < .05$ , \*\* $p < .01$

표 5. 기분 조작별 평가 정확성 평균 점수

	긍정기분 조건 (n=76)			부정기분 조건 (n=70)		
	사전	사후	t값	사전	사후	t값
EL	.45(.70)	.24(.30)	-2.35*	.32(.42)	.24(.48)	-1.22
DE	.92(1.30)	.66(.49)	-1.61	.64(.69)	.70(.43)	.83
SA	.14(.20)	.15(.14)	.17	.17(.21)	.21(.26)	.78
DA	1.36(.79)	.84(.40)	-5.43**	1.14(.79)	.82(.38)	-4.00**

괄호 안의 값은 표준편차임. \* $p < .05$ , \*\* $p < .01$

좀 더 많은 측면에서 개선되는 것으로 나타났지만, 두 집단을 직접 비교할 때는 가설을 지지하는 정도의 차이는 없었다.

표 5에 제시하였듯이 4가지 평가 정확성 구성요소별로 긍정기분 집단과 부정기분 집단의 평가 정확성 향상 차이를 추가적으로 알아본 결과, 긍정기분 집단은 전반적 정확성(EL), 대상내 차원 정확성(DA)에서 유의미하게 평가 정확성 지수가 감소하였다(EL;  $t = -2.35$   $p < .05$ , DA;  $t = -3.97$   $p < .01$ ). 반면, 부정기분 집단은 대상내 차원 정확성(DA)에서만 평가 정확성 지수가 유의미하게 감소하였다( $t = -4.00$   $p < .01$ ). 즉, 기분조작 조건을 분리하여 살펴본 결과도 본 연구의 예측과 달리 긍정기분 조작 조건에

서 평가정확성이 좀 더 많이 개선되는 결과가 나타났다.

## 논 의

연구의 결과를 요약하면, Cronbach의 4가지 평가 정확성 구성요소들 중 대상 정확성(DE)은 두 유형의 평가자 훈련의 효과성에서 차이가 명확하게 나타났다. 즉, 피평가자들의 변별을 핵심 내용으로 하는 변산성 훈련이 그 목적에 맞게 DE를 향상시키는 면에서 참조들 훈련보다 우수한 것으로 나타났으며, 세부적으로 사전 사후 평가 정확성으로 비교하면 변

산성 훈련은 DE 측면의 정확성을 높이는 효과를 나타내지만, 평가차원들의 변별을 핵심 내용으로 하는 참조틀 훈련은 피평가자별 변상성을 오히려 적게 만들어서 DE 측면의 정확성을 오히려 감소시켰다(RVT;  $t=-3.01, p<.01$ , FOR;  $t=3.51, p<.01$ ). 즉, 피평가자들간의 변별을 위해서는 변산성 훈련이 더 효과적이라는 것을 의미한다.

그러나 대상내 차원 정확성(DA), 전반적 정확성(EL)에서는 두 평가자 훈련 모두 정확성 향상에는 도움이 되었지만 훈련 간 차이가 나타나지 않았다. 이러한 결과는 두 훈련 모두 참가자들에게 관대화 오류(halo effect)를 방지하는데 차이 없이 효과적이라는 결과로 Sinclair (2000)의 연구 결과와 동일한 결과이다.

그리고 차원 정확성(SA)은 통계적으로 유의미하지는 않았으며, 사전-사후 평균차이를 비교해 보면 두 훈련 모두 정확성이 약간 감소하는 결과가 나타났다. 평가자 훈련은 훈련을 통해 평가자들이 가지고 있는 평가에 대한 인지적 범주화(cognitive categories)를 변경을 하는 것인데, 본 연구의 결과와 사전-사후 평균 변화를 볼 때 평가자들이 인지적 범주화 변경에 실패하여 기존에 가지고 있던 인지적 범주 내에서 평가를 한 것으로 보인다. 특히 참조틀 훈련은 평가차원에 관한 인지적 범주화를 변경하는 정도가 크다는 것이 기존의 연구결과인데(Stamoulis & Hauenstein, 1993), 본 연구에서는 참조틀 훈련이 변산성 훈련보다 상대적으로 더 많은 훈련내용으로 구성되어 피험자들에게 더 어려웠을 수 있었다는 측면 때문에 예상한 결과가 나타나지 않았을 수 있다. 만약 이러한 추론이 맞다면 참조틀 훈련을 통해 인지적 범주화를 변경하는 것을 촉진하기 위해서는 훈련 자체가 인지적 부담이 과도하

지 않도록 설계하는 것이 중요할 것이다. 그러나 본 연구의 결과를 다른 측면에서 생각할 수도 있다. 그것은 평가 영상물 제작에 있어서 평가 차원에 대한 조작의 문제로, 본 연구는 기존 연구들(Hauenstein, Facticeau, & Schmidt, 1999; Sinclair, 2000; Stamoulis & Hauenstein, 1993)에 비해 피평가자 별로 평가 차원들의 난이도 구분이 명확하지 않았다. 즉, 피험자가 피평가자들 간의 수준 차이를 통해 피평가자 내(즉, 차원)의 수준을 쉽게 일반화하지 못하도록 평가 차원의 난이도를 다양하게 배열하면서 전체적인 수준(즉, 피평가자 간)이 비슷하도록 조작했는데 이 측면이 영향을 주었을 가능성이 있다. 이 점 또한 평가자들에게 인지적 부하를 많이 주어서 평가차원별 정확성에 강점이 있는 참조틀 훈련의 효과가 나타나지 못하도록 만든 요인이 되었을 수도 있다.

평가 책임성의 경우 주효과나 평가자 훈련과 상호작용이 나타나지 않았다. 이러한 결과의 이유로는 책임성 부여에 대한 참가자들의 지각 강도에서 충분한 차이를 만들지 못했을 가능성을 생각할 수 있다. 본 연구에서는 책임성을 부여하는 조건이 참가자들에게 평가자에 학번과 이름을 쓰도록 하고, 참가자의 평가 점수는 연구자에게 전달되어 실제 평가의 평가 자료로 사용이 되며 피평가자들에게 전달이 된다는 것이었다. 이러한 조건이 피험자에게 주의를 정확한 평가를 해야 한다는 동기를 자극하는 면에서 어느 정도 효과가 있었다(표 4 참조). 그러나 실험 자체가 참가자들에게 상당한 몰입을 요구하였기 때문에 책임성을 부여하지 않은 집단도 동기수준이 높을 수 있다. 이는 표 4에서 정확성이 향상된 정도가 두 집단에서 큰 차이가 나지 않는 점으로 추론할 수 있다. 결과적으로 두 집단의 차이가

훈련의 효과를 달라지게 만들기에는 충분하지 않은 차이였을 수 있다. 추후 연구에서는 책임성 부여의 차이가 실험 과정 내에서 유지가 되는 상황으로 조작을 하거나, 직접 평가 근거에 대해 서술을 하도록 하거나 발언을 하는 조건으로 구성하는 것이 바람직할 것이다.

평가자 기분은 실험전후를 비교해 볼 때 적절히 조작된 것으로 추론되나, 평가에는 큰 영향을 주지 않는 것으로 나타났다. 이러한 결과의 이유로 기분(mood)의 특성상 조작 시점에 일시적으로 존재하고 평가 시점에는 효과가 약화되었을 가능성을 생각할 수 있다. 그래서 기분 조작 측정의 결과는 유의미하게 나타났지만, 사후 평가 영상물을 시청하고 평가를 하는 과정에까지는 기분 조작의 효과가 유지되지 않은 것으로 추론할 수 있다. 또 다른 이유는 기분 조작 영상물의 내용에서 부정적 기분으로 조작하기 위한 영상물(한 가정의 가장이 생을 마감하는 현장다큐멘터리)을 시청한 후 부정적 기분이 사회정의에 대한 태도를 형성했을 가능성도 있다. 또한, 본 연구의 기분 조작이 실험 참가자들에게 인지과정과 상호작용하기 보다는 실험(평가)에 대한 동기나 몰입의 저하와 같은 태도를 형성하였을 가능성이 있다.

책임성과 평가자 기분의 상호작용 효과는 전반적 정확성(EL)에서만 책임성이 있는 경우 기분이 긍정적인 집단이 더 정확하게 평가를 하는 반면, 책임성이 없는 경우에는 부정기분 집단이 긍정기분 집단보다 더 정확한 평가를 하는 것으로 나타났다. 이러한 결과는 Bodenhausen, Kramer,과 Susser(1994)의 연구 결과와 유사하다. 즉, 긍정기분의 개인은 정보를 처리하는 데 있어서 인식론적 동기가 부족하기 때문에 편파가 잘 일어나는데, 이에 책임

성을 부여한다면 판단의 편파가 줄어드는 것이다. 반면에 부정기분 집단의 경우에는 심리적으로 더 부담이 되는 책임성까지 부여함으로써 앞서 평가자 기분에서 설명한 것과 같이 부정정서가 정교한 판단을 하도록 하는 효과를 넘어서서 부작용으로 작용하는 것으로 추론된다.

그러나 피평가자의 전체 점수부여에 해당하는 전반적 정확성(EL)을 제외한 다른 평가 정확성 요소들에서는 책임성과 평가자 기분의 상호작용이 나타나지 않은 점을 주목할 필요도 있다. 다른 정확성 구성요소들(즉, DE, SA, DA)은 피평가자별로 변별을 하거나, 평가차원별로 변별을 해야 하는 정보처리과정을 포함하는 평가 정확성 구성요소로서 전반적 정확성(EL)에 비해 정교한 판단을 요구한다. 따라서 EL과 같이 포괄적인 판단에는 기분이 포괄적으로 작용하지만, 보다 정교한 정보처리에서는 그 영향이 약하거나 기분의 일시적 효과가 유지되지 못했을 가능성이 있다.

이상의 연구 결과가 실제 인사평가에 주는 시사점은 다음과 같다. 첫째, 본 연구에서는 평가 정확성을 제고하기 위해 평가자 훈련이라는 방법이 어느 정도 효과적인지를 파악하여 조직의 평가제도 구축 요소에 반영할 수 있는 근거를 제시하고자 하였다. 본 연구에서의 결과는 이전 연구와 대체로 유사하게 참조틀 훈련은 평가자들에게 평가 차원에 대해 정확한 변별을 하는 데 도움을 주어 피평가자의 장단점을 파악하여 피드백을 제공하는 인사평가의 목적을 달성하는 데 효과적일 수 있다는 점을 보여준다. 그와 더불어 변산성 훈련은 참조틀 훈련이 피평가간 변별을 통한 인사결정의 목적에 취약한 점을 보강하는 역할을 담당할 수 있음을 시사한다. 즉, 평가자간 변별

에 대한 정확성을 높이고자 한 기존의 평가 오류 훈련(rater error training)이 정확성보다는 오류에 치중하여 실제로 정확성을 높이지 못하는 취약점을 극복할 수 있는 대안으로 변산성 훈련을 고려할 수 있음을 보여준다. 특히, 변산성 훈련은 대상 정확성(DE)에서 평가정확성 향상도가 증가함으로서 효과를 발휘할 수 있는 정확성 차원을 실증적으로 보여주었다. 이러한 연구결과를 바탕으로 조직의 평가제도 구축 시 평가 목적에 맞는 평가자 훈련을 실시하는 것을 제안한다.

둘째, 평가 정확성 향상을 도모하는 훈련 중 보편적으로 활용되는 참조틀 훈련을 실무로 적용할 때 평가자의 인지적 부담을 고려할 필요가 있다. 참조틀 훈련은 필요 시 평가자(주로 상사)가 인사평가에 대해 가지고 있는 잘못된 인지적 범주를 바꾸거나 적절한 범주를 형성하는 목적이 있다. 하지만 본 연구에서 가설이 지지되지 못한 바를 인지적 과부하로 추론할 수 있다면 평가자가 인지적 범주의 변경에 어려움을 겪게 될 수 있는 인지 과부하를 반영할 필요가 있다. 현실적으로 참조틀 훈련을 실시할 때에는 평가자는 평가를 부가적인 임무로 여긴다는 점(DeNisi, 1996)을 고려하여 주기적이고 시간적 여유를 가지고 실시되어야 평가자 훈련에 투자된 시간과 비용의 효과를 얻을 수 있을 것이다.

셋째, 현장의 인사평가에서는 평가자의 관대화를 많이 우려하는데, 관대화를 주로 반영하는 전반적 정확성(EL)은 평가자의 기분과 책임성의 상호작용의 영향을 받는다. 따라서 조직이 평가자의 책임성을 부가한 제도를 운용할 때에도 평가자들에게 기분의 영향력이 있음을 주지시켜서 되도록 긍정적인 정서상태일 때 평가를 하도록 유도하여 평가 정확성 향상

을 도모할 수 있을 것이다.

한편, 본 연구는 이러한 결과를 실제 인사평가 장면에서 적용하고자할 때 고려해야 할 몇 가지 제한점을 지니고 있다. 첫째, 실제 조직에서 이루어지는 평가자와 피평가자간의 상호작용은 통제된 실험 연구에서 이루어지는 것보다 더 복잡하다는 것이다. 예를 들어, 실험은 1시간 30분 정도 진행되었지만 실제 조직에서 이루어지는 직무수행 평가는 6개월, 혹은 1년 간격으로 시행되므로 본 연구의 변인들이 작용하는 시간프레임이 다를 수 있다. 또한, 대학생을 대상으로 하여 영상물 내의 피평가자를 평가하도록 하였다는 점도 다르다. 그러나 본 연구의 약점인 외적타당도는 심리적 현실성(psychological reality) 측면에서는 심각한 약점이 아닐 수도 있다. 피험자인 대학생들은 일정기간의 훈련과정을 거치면서 강의자의 행동을 관찰해왔으며 공식적으로나 비공식적으로 평가할 기회가 많다는 점에서 조직현장의 실제 평가자와 유사하다고 볼 수 있다. 둘째, 실험에 사용된 영상물들은 제한점을 가지고 있다. 강의 평가를 위해 사용된 영상물들은 평가차원들의 난이도를 섞어서 피평가자 별로 쉽게 구분하지 못하도록 조작하였다. 하지만 실험을 진행하면서 피평가자의 차원별 난이도가 많이 다르게 구성된 경우 피험자들이 평가를 하는데 어려움을 겪는 것으로 보였다. 이 점은 평가자 훈련의 효과성이 크게 나타나지 않은 원인 중 하나일 것이다. 또한, 부정기분을 유도하기 위해 Robbins과 DeNisi (1998)는 슬픈 영화를 편집하여 조작 영상물로 사용한 반면, 본 연구에서는 실제 사건을 담은 다큐멘터리를 편집하여 제작하였다. 이러한 부정기분 조작 영상물이 부정정서 조작 이상의 인지과정에 영향을 주었을 가능성도 있

다. 셋째, 본 연구에서 평가 동기의 효과를 살펴보기 위해 실험조작을 통해 책임성을 부여하고자 한 설계방법이 정확히 얼마나 효과가 있었는지에 대한 조작확인이 결여되었다. 실험상황에서 조작 후 실험참가자들의 반응을 관찰하였고 책임성과 관련된 분석에서 책임성을 부가한 집단의 평가정확성이 체계적으로 높은 결과는 확인되었으나 이런 점이 조작확인으로 충분하지 못한 면이 있다. 또 본 연구의 결과에서 책임성 차원에서 유의미한 효과가 나타나지 않은 것은 조작의 문제일 가능성도 있으나 이를 정확히 확인하기 힘들다는 점이 제한점 중에 해당하며 본 연구에서 관련 부분의 결과해석에 주의가 필요하다.

요약하자면, 이상의 제한점에도 불구하고 본 연구는 인사평가 연구의 정확성에 영향을 주는 요소를 파악하기 위한 가장 이상적인 방법으로 실험설계 방법을 채택하여 내적타당도를 확보한 점과 평가오류가 아닌 평가정확성을 직접적으로 연구했다는 점에 의의를 둘 수 있다. 본 연구의 결과를 바탕으로 실제 조직의 인사평가 장면에서 평가자의 인지, 정서, 및 동기가 어떻게 작용하는지에 대하여 시사점을 제공해 줄 수 있으며, 보다 정교한 미래 연구에 도움을 줄 수 있을 것이다.

### 참고문헌

이주일, 김명언 (1999). 조직 장면에서 체험된 정서와 표현된 정서간 불일치의 효과: 경쟁가설의 검증. 한국심리학회지: 산업 및 조직, 12(2), 171-193.  
한국경제 (2007). 인사고과 시즌, 흔히 볼 수 있는 평가자 유형 1위 '내 사람 챙기기

형'. 8월 23일자.

한태영, 이세란 (2005). 청소년의 정서관련 변인과 학업 스트레스의 관계: 체험정서, 표현정서, 및 정서지능의 영향. 한국심리학회지: 학교, 2(1), 37-58.  
Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: New response sets and decreasing accuracy. *Journal of Applied Psychology*, 65, 60-66.  
Bernardin, H. J., & Buckley, M. R. (1981). Strategies in rater training. *Academy of Management Review*, 6, 205-212.  
Bless, H., Bohner, G., Schwarz, N., & Strack, F. (1990). Mood and persuasion: A cognitive response analysis. *Personality and Social Psychology Bulletin*, 16, 331-345.  
Bodenhausen, G. V., Kramer, G. P., & Susser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of Personality and Social Psychology*, 66, 621-632.  
Cardy, R. L., & Dobbins, G. H. (1994). *Performance Appraisal: Alternate Perspectives*. Cincinnati: South-Western Publishing.  
Carroll, S. J., & Scneier, C. E. (1982). *Performance appraisals and revel systems: The identification, measurement, and development of performance in organizations*. Glenview, IL: Scott. Foresman.  
Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychology Bulletin*, 52, 177-193.  
DeNisi, A. (1996). *Implications for theory and practice*. In *Cognitive approach to performance appraisal* 151-172. Routledg New York: NY.  
Feldman, J. M. (1981). Beyond attribution theory: a perceptual -cognitive viewpoint. In K. R.

- Rowland & G. R. Ferris (Eds.), *Personal and Human Resource Management*, 4, (pp. 45-99). Greenwich, CT: JAI Press.
- Ferris, G. R., & Judge, T. A. (1991). Personnel /human resources management: A political influence perspective. *Journal of Management*, 17, 447-488.
- Forgas, J. P., (1987). Mood Effects on Person-Perceptron Judgments. *Journal of Personality and Social Psychology*, 53, 53-60.
- Harris, M. M. (1994). Rater motivation in the performance appraisal context: A theoretical framework, *Journal of Management*, 20, 737-756.
- Hauenstein, N. M. A. (1998). Training raters to increase the accuracy of appraisals and the usefulness of feedback. In J. W. Smither (Ed.) *Performance Appraisal* (pp.404-442). Jossey Bass San Francisco.
- Hauenstein, N. M. A. (1992). An information-processing approach to leniency in performance judgments. *Journal of Applied Psychology*, 77, 485-493.
- Hauenstein, N. M. A., Facticeau, & Schmidt, J. A. (1999). *Rater variability training: An alternative procedure to rater error training and frame-of-reference training*. Unpublished Manuscript, Virginia Polytechnic Institute and State University, Blacksburg.
- Izard, C. E. (1991). *Psychology of emotions*. New York: Plenum.
- Kemter, v., & Moser, K. (2007). Frame-of-reference training: Evaluation of rating accuracy within a double-pretest-one- posttest-design. *LASER Discussion Papers-paper*, 1, 1-31.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87, 72-107.
- Latham, G. P. (1986). Job performance and appraisal In C. L. Cooper & I. Robertson (Eds). *International Review of Industrial and Organizational Psychology*, New York: Wiley
- Lazarus, R. S. (1991), *Emotion and Adaptation*. New York: Oxford University Press.
- Lerkowitz, J. (2000). The role of interpersonal affective regard in supervisory performance ratings: A literature review and proposed causal model. *Journal of Occupational & Organizational psychology*, 73, 67-85.
- McAllister, D. W., Mitchell, T. R., & Beach, L. R. (1979). The contingency model for the selection of decision strategies: An empirical test of the effects of significance, accountability, and reversibility. *Organizational Behavior and Human Performance*, 24, 228-244.
- Mero, N. P., Guidice, R. M. & Brownlee, A. L. (2007). Accountability in a Performance Appraisal Context: The Effect of Audience and Form of Accounting on Rater Response and Behavior. *Journal of Management*, 33, 223-252
- Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organization, and goal-based perspectives*. Sage publications.
- Murphy, K. R., & Balzer, W. K. (1986). Rater errors and rating accuracy. *Journal of Applied Psychology*, 71, 39-44.
- Palmer. J. K., & Feldman. J. M. (2005).

- Accountability and Need for Cognition Effects on Contrast, Halo, and Accuracy in Performance Ratings. *The Journal of Psychology: Interdisciplinary and Applied*, 139, 119-138
- Pulakos, E. D. (1984). A comparison of training programs: Error training and accuracy training. *Journal of Applied Psychology*, 69, 581-588.
- Robbins, T. L., & DeNisi A. S. (1998). Mood VS. Interpersonal Affect: Identifying Process and Rating Distortions in Performance Appraisal. *Journal of Business and Psychology*, 12, 313-325.
- Schlenker, B. R. (1985). *The self and social life*. New York: McGraw-Hill.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of wellbeing: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45, 513-523.
- Schwarz, N., & Clore, G. L. (1988). How do I feel about it? The informative function of affective states. In K. Fiedler & J. Forgas (Eds.). *Affect, cognition, and social behavior* (pp.44-62). Toronto, Ontario, Canada: Hogrefe.
- Schwarz, N. (2002) Feelings as information: Moods influence judgments and processing strategies. In T. Gilovich, D. Griffin & D. Kahneman (Eds.). *Heuristics and biases: The psychology of intuitive judgment* (pp.534-547). New York: Cambridge University Press.
- Sinclair, A. L. (2000). *Differentiating Rater Accuracy Training Programs*. Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirement for the degree of Master of Science in Industrial /Organizational Psychology.
- Sinclair, R. C. (1998). Mood, categorization breadth, and performance appraisal: The effects of information acquisition and acquisition and affective state on halo, accuracy, information retrieval, and evaluations. *Organizational Behavior and Human Decision Processes*, 42, 22-26.
- Srull, T. K., & Wyer, R. S. Jr. (1989). Person memory and judgment. *Psychological Review*, 96, 58-83.
- Stamoulis, D. T., Hauenstein N. M. A. (1993). Rater Training and Rating Accuracy: Training for Dimensional Accuracy Versus Training for Ratee Differentiation. *Journal of Applied Psychology*, 78, 994-1003.
- Sulsky, L. M., & Day, D. V. (1992). Frame-of-reference training and cognitive categories: An empirical investigation of rater memory issue. *Journal of Applied Psychology*, 77, 501-510.
- Woehr, D. J., & Huffcutt, A. I. (1994). Rater Training for Performance Appraisal: A Quantitative Review. *Journal of Occupational and Organizational Psychology*, 67, 189-205.
- 1차 원고접수 : 2008. 07. 5  
2차 원고접수 : 2008. 10. 7  
3차 원고접수 : 2008. 11. 15  
최종게재결정 : 2008. 11. 23

## **The Effect of Rater Training Type on Rating Accuracy: Focusing on Interaction Effect of Accountability and Rater Mood**

**Jae Han Cho**

**Tae Young Han**

Kwangwoon University

The purpose of this study was to compare the effects of two types of rater training on rating accuracy in performance appraisal context. It also investigated whether accountability and rater mood interact with the effects. Specifically, the study examined the effects of frame-of-reference training (FOR) and rater variability training (RVT) in terms of elevation (EL), differential elevation (DE), stereotype accuracy (SA), and DA (differential accuracy), which are conceptualized by Cronbach (1955). The study also studied 2-way interactions, comparing the differences of accuracy according to the conditions whether accountability is required or not, and positive or negative mood is imposed. In addition, it tested a 3-way interaction effect among the three conditions. The study conducted an experiment with 2 x 2 x 2 design in which 146 undergraduate students were assigned for lecture evaluation. The results revealed that accuracy in DE was improved more in RVT than FOR, while the other types of accuracy were not significantly different. The effects of accountability and mood were not found significant, either. The interaction effect of accountability and rater mood was significant in EL in the way that the accountability-required group rated more accurately when it was in positive mood whereas the accountability-not-required group was more accurate in negative mood. The other interaction effects were not significant, neither was the 3-way interaction. Theoretical and practical implications including limitations were discussed in performance appraisal context.

*Key words* : Performance Appraisal, Frame-of-Reference Training, Rater Variability Training, Accountability, Rater Mood, Cronbach's Accuracy Components,