

인사고과의 양식의 효과성에 대한 이론적 고찰 - A Theoretical Study on The Effectiveness of Performance Appraisal formats -

이 광 회*
Lee, Kwang Hee
이 선 규**
Lee, Seon Kyu

Abstract

Appraisal format is the methods that is used to gather informations related to performance behaviors. Appraisal format has descriptive characteristics including evaluation dimensions, anchors, scales. The purpose of this paper is to study the effectiveness of a format characteristics of the appraisal results through literature review and to explore more appropriate appraisal format through comparing three types of formats, graphic scales, BARS, BOS.

According to the results, rating format's characteristics can affect the quality of appraisals in that appraisal format can provide raters with evaluation standards which help raters to evaluate effectively. Specific dimensions are better than general one dimension. The effectiveness of scale in some degrees depends on development procedure. In general anchors are important to reduce rating errors and behavioral anchors are better than numeric anchors. The systematically developed BARS and BOS can reduce rating errors but graphic scales has limitations

1. 서론

고과양식은 성과에 관련된 정보를 수집하기 위해 사용되는 수단이다. 고과양식은 어떤 표현적 특성, 즉 고과차원의 수, 고과항목들의 수와 유형, 고과척도의 형태 등을 가진다. 이러한 고과양식의 특징은 고과자들에게 성과행위를 고과하는 데 도움을 줄 수 있는 표준을 제공한다는 면에서 고과결과의 정확성과 효과성에 영향을 미칠 수 있다.

여기에서는 고과양식의 개발에 있어서의 일반적으로 제기 되어왔던 문제인 고과차원, 척도 기준(anchors) 등과 같은 특성이 고과양식의 효과성에 미치는 영향을 살펴보고자 한다. 그리고 전통적인 평가양식인 그림척도와 비교적 최근에 개발되어 관심의 대상이 되고 있는 BARS, BOS 등의 개발과정 및 효과성에 대한 검토를 통하여 보다 바람직한 고과양식에 대한 방향을 모색해보고자 한다.

2. 고과양식의 일반적 특성에 따른 효과성

2.1 고과차원

어떤 차원이 고과되어야 하는가라는 문제는 학자들의 논쟁거리가 되어 왔다. Kavanagh(1971)

* LG 경제연구원 선임 컨설턴트

** 금오공과대학교 산업공학과 부교수

은 고과척도 내용의 유형(즉, 성과결과 대 관찰가능한 직무행위 대 직무수행자의 개인특성)을 명확하게 선정할 수 없다고 주장했다.[1] 일부 학자들은 직무성과의 전반에 대한 고과는 보다 구체적인 고과에 비하여 외적인 요소에 영향을 더 받으며, 성과에 관련된 중요한 분산원천을 고려하지 못하는 것 같다고 주장했다.[2] Kavanagh(1973)는 적절한 고과차원을 선정하는 문제는 직무요건의 본질, 해당 직무에 대한 개인적 요소와 성과요소들의 관련성, 그리고 신뢰성, 고과 오류에 대한 저항, 그리고 구성타당성의 측면에서 각 유형이 갖는 실증적인 방어가능성에 의해서만 해결될 수 있다고 주장했다.[3]

고과차원의 수에 관하여서는 개인들은 이질적이며 다양한 정보를 동시에 다루는 데 한계를 갖는 것으로 보인다. 즉, Miller(1956)가 주장한, 이질적인 정보가 5 - 9 개를 넘을 경우 사람들은 정보의 처리에 한계를 갖는다는 원리(seven, plus or minus two dictum)가 고과행위에서도 적용되는 것으로 보인다.[4]

2.2 척도기준(Anchors)

고과척도는 전형적으로 세 가지 유형의 척도기준(anchors), 즉 숫자적, 형용사적, 행위적 척도기준 중의 하나를 사용한다. 이들 기준설정 시스템들의 상대적 효과성에 대한 몇 가지의 연구결과에 의하면 척도의 기준화가 구체적일수록 효과가 있다는 것을 보여주고 있다. 즉 Bendig(1952a, 1952b, 1953)에 의하면 척도에 대한 신뢰성은 척도기준 설정이 증가함에 따라 증진된다는 것을 발견했다.[5] Barrett 등(1958)은 기준이 없는 척도에 비하여, 기준이 설정된 척도가 효과성이 커졌다는 것을 보였다.[6]

여러 학자들은 단순한 숫자나 형용사적인 척도기준들에 비하여 행위적인 척도기준이 상대적으로 효과적이라고 주장했다.[7] BARS 양식은 척도기준의 행위적 특성에 크게 의존하기 때문에, BARS 양식에 대하여 긍정적인 태도를 갖는 거의 대부분의 연구들이 형용사적이거나 숫자적인 척도기준들보다는 행위적인 척도기준을 보다 긍정적인 것으로 본다. 그럼에도 불구하고 정교한 척도기준의 특성에 대하여, 의문을 제기하는 연구들이 있다. Finn(1972)는 척도기준들의 정의방법에 따른 고과의 평균이나 신뢰성에 차이가 없다는 것을 발견했다.[8] 따라서 척도기준이 수치적인가 혹은 기술적인가 하는 것은 중요하지 않은 것으로 보인다.

Kay(1959)는 고과척도로서 중요사건은 지나치게 상세하며, 상황에 따른 제약이 있다고 주장했다.[9] 척도기준의 수와 유형의 중요성은 성과차원정의(dimension definition)의 적절성과 관련이 있다.[10] 고과되는 차원에 대한 적절한 정의가 없으면, 고과자는 척도의 의미를 제공하는 척도기준들에 의존할 수 밖에 없다. Barrett 등(1958)의 연구에 의하면, 행위적 척도기준이 잘 정의되어 있고 차원정의가 없는(특성 명칭만 있음) 척도는 차원정의와 척도기준이 있는 척도나, 차원정의가 있고 척도기준이 없는 척도들보다, 높은 신뢰성, 낮은 HALO, 그리고 낮은 관대화를 가졌다는 것을 발견했다.[11]

일반적으로 볼 때, 척도기준은 중요한 것으로 보이며, 행위적 척도기준이, 숫자나 형용사적인 것보다 우수한 것으로 보인다.

2.3 척도(Scales)의 점수화

전통적으로 고과척도의 개발에 있어서 취약한 절차중의 하나는 태도기준에 척도의 값을 할당하는 과정이다. 숫자적이거나 형용사적인 기준을 가진 척도에 점수를 부여하는 방법에 대한 연구가 이루어져 왔지만,[12] 명확한 결론을 제시하지 못하고 있다.

BARS 기법에서 척도점수설정을 위한 공통적인 기법은 한 행위 예가 특정의 차원을 얼마나 잘 나타내는가를 판단하는 것이다. 이것은 보통 형용사나 숫자로 척도기준이 설정되는 그림

고과척도의 형태와 유사하게 설정된다. 행위기준들의 최종 점수 선정은 항목들의 평균값과 표준편차에 기준하여 이루어진다. 의사결정 규칙은 대체로 임의적으로 이루어진다. 즉 연속선상에 가능한 많은 점이 표시된 항목이나 상대적으로 작은 표준편차를 가진 항목을 선정한다. Bernardin(1977)[13], Berardin, Alvares, 그리고 Cranny(1976)[14]는 전통적인 고과오류에 대한 고과척도의 효과성은 어느 정도는 척도개발과 척도기준설정의 엄격성에 기인한다는 것을 보여 주었다. 그는 이러한 엄격성이 각 항목에 대한 표준적인 분석과정을 통하여 이루어질 수 있다고 주장하였다.

따라서 BARS 양식이 다른 양식과 비교하여 상대적으로 안 좋은 결과를 보이는 것은 척도 기준의 선정과 척도화과정에서 엄격성이 결여됐기 때문이라고 할 수 있다.[15]

3. 고과양식의 유형에 따른 효과성

3.1 그림척도(Graphic scales)

그림척도는 1922년 Paterson에 의해 소개되었다. 그의 의견에 의하면, 이 방법은 다음의 두 가지 특징을 갖는다. 즉 고과를 할 때 고과자가 계량적 판단으로 부터 자유롭고, 바람직한 차별을 할 수 있다는 것이다. 이 척도는 특성명칭(trait labels)과 이 명칭에 대한 간단한 정의 그리고 다양한 유형 형용사들이 있는 척도로 되어 있다. 그림척도의 구체적인 예는 <그림 1>과 같다.

<그림 1> 그림척도의 예

성명 _____ 부서 _____ 일시 _____					
	매우 우수	우수	보통	부족	매우 부족
업무의 질	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
업무의 양	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
직무지식	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
협동성	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
적극성	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

그림척도를 사용할 경우 고과자의 주관적이고 임의적인 특성이 고과결과에 반영될 수 있기 때문에 이에 대한 비판이 집중하였다.[16] 즉 고과자들이 그림척도를 사용하여 고과할 경우 고과결과에 관대화, HALO 등과 같은 고과오류가 발생할 가능성이 많기 때문에 성과고과에서의 유용성이 적다는 것이다.

Barrett, Taylor, Parker, 그리고 Martens(1958)은 그림척도의 구조화 정도를 변화시킨 4가지의 다른 양식의 적절성을 검증하였다.[17] 양식(format) 1은 10인치(25.4 센치미터)의 선을 15구간으로 나누고 하나의 특성(trait) 이름이 있다; 그리고 특성에 대한 정의 혹은 이 척도에 대한 척도기준이 주어지지 않았다. 양식 2는 동일하게 분할된 선으로 구성되었으나 특성정의가 추가되었다. 양식 3에서는 분할된 선이 행위적 척도기준에 의해 정의되었다; 여기에서는 특성의 명칭이 주어졌으나 정의는 하지 않았다. 양식 4는 행위적 척도기준에 의해 정의된 선과 특성정의로 구성되며 특성의 명칭은 주어지지 않았다. 연구결과는 양식 3이 가장 높은 신뢰성, 낮은 관대화, 낮은 HALO를 보였다.

Madden과 Bourdon(1964)은 몇가지 다른 유형의 그림 고과척도 양식을 비교하였다.[18] 이

양식들은 척도 끝의 위치, 척도의 공간지향(수평 대 수직), 척도의 분할(분할 대 연속), 그리고 척도수준의 수치부여(1에서 9까지 대 -4에서 4까지) 등으로 변화시켜 고과결과에 미치는 영향을 연구하였다. 연구결과에 의하면 척도유형에 따른 효과는 작았다.

Blumberg, DeSoto, 그리고 Kuethe(1966)의 이후 연구에서 척도의 공간 지향이 다시 연구되었다.[19] 이들은 척도의 “좋은(Good)”쪽의 방향의 위치(위, 아래, 왼쪽, 혹은 오른쪽)가 유의한 차이를 보이지 않는다는 것을 발견했다. 이들은 고과자들이 다양한 양식을 선호할 수는 있지만, 이러한 선호는 실제 고과행위에 효과가 적거나 없었다고 주장했다.

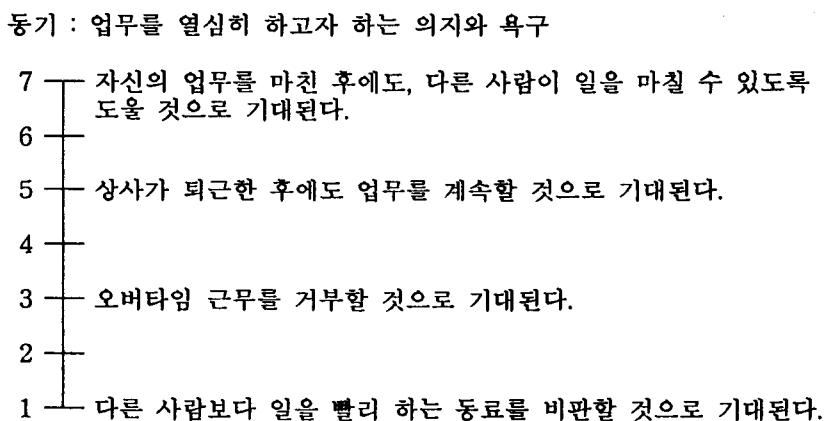
따라서 그림척도는 주관적이고 임의적인 특성으로 인하여 여러가지의 고과오류를 발생시키며, 그림척도의 개선에 의해 고과의 타당성을 증진하는 데는 한계가 있는 것으로 보인다.

3.2 행위기준인사고과(BARS: Behaviorally Anchored Rating Scales)

1963년에 Smith와 Kendall은 고과척도개발을 위한 새로운 기법을 소개하고 Behavioral Expectation Scaling(BES)라고 명명하였다.[20] 이 척도는 Behaviorally Anchored Rating Scales (BARS)이라고 불리워져 왔다. 실질적으로 BARS는 척도상의 간격에 따라 나타나는 척도기준(anchors)이 단순한 숫자나 특성명칭을 수정한 형용사가 아닌 실제 행위의 예라는 점에서 전통적인 그림척도와 다르다. BARS를 사용할 경우 고과자는 피고과자의 직무행위를 BARS의 해당차원의 행위적 척도기준과 비교하게 된다. 이것은 여러 수준의 척도점수에 대한 효과성이 어떠한 수준인가에 대한 설명이 없이 피고과자를 고과하는 것보다 우월한 것으로 보인다.

BARS의 아이디어는 양식 이상의 것이다. 어떤 학자들은 이것은 하나의 시스템 혹은 철학이라고 주장한다.[21] 예컨대, 이상적으로 볼 때 고과자들은 성과고과를 하기 위해서 고과기간을 통하여 종업원 작업행위의 예들을 기록해야 한다. BARS의 또 하나의 긍정적인 특징은 이 시스템의 사용자들은 척도의 개발에 참여하며, 따라서 양식에 대한 신뢰성이 강화된다. 더우기 BARS 개발단계는 중요 직무 성과차원을 식별하기 위한 훌륭한 수단을 제공한다.[22]

<그림 2> BARS의 예



자료원 : Gary P. Latham, Kenneth N. Wexley, "Increasing Productivity Through Performance Appraisal", Addison-Wesley Company, Inc, 1981.

성과를 직접적으로 고과하는 기법이라는 점에서 BARS시스템은 현재 가장 많은 관심을 받고 있다. 이 기법의 장점은 먼저 척도기준이 고과자들에게 친숙한 용어를 사용하여 행위적으로 설정되기 때문에 특성중심의 그림척도에서와는 달리 고과에 대한 모호성을 줄인다는 것이다. 둘째는 종업원들에게 강점과 약점에 대한 구체적인 피드백을 제공할 수 있기 때문에 상담과 동기부여가 용이하다는 것이다.[23] BARS의 구체적인 예는 <그림 2>와 같다.

Smith와 Kendall(1963)은 이 척도가 간호원의 성과기술에 유용하게 사용될 수 있다는 것을 보여주었다.[24] Maas(1965)는 BARS가 면접성과의 측정에서 효과적이라는 것을 보였다.[25] Landy와 Guion(1970)은 작업동기의 고과에서 이 기법이 유용하다는 것을 보였다.[26]

Borman과 Vallon(1974)는 BARS기법이 고과에 대한 고과자의 신뢰와 신뢰성이 우월한 고과를 가져왔으나 보다 단순한 수치적 양식이 관대화가 적고 피고과자들을 적게 차별했다는 것을 발견했다.[27]

Burnaska와 Hollmann(1974)은 세가지의 다른 양식을 비교하였다.[28] 첫번째 양식은 표준 행위기준 척도이다. 두번째 양식은 동일한 차원과 정의로 구성되었으나 행위적 척도기준 대신에 행동사적인 척도기준으로 대치되었다. 세번째 양식은 사전적인 차원을 가진 전통적인 그림 고과양식이었다. 비록 관대화, HALO가 모든 세 양식에서 나타났지만, BARS기법은 관대화를 줄였고, 피고과자의 차이를 나타내는 분산의 양을 증가시켰다. 따라서 각 양식은 각자 독특한 문제들을 가지고 있는 것으로 보인다.

Borman과 Dunnette(1975)는 표준 BARS양식과 동일한 차원명칭과 정의를 가졌지만 숫자 척도기준을 가진 고과척도, 그리고 특성 명칭과 숫자 척도기준(anchors)을 가진 전통적인 그림 척도와 비교하였다.[29] 이들은 표준 BARS 양식이 심리측정적으로 우수하다는 결론(HALO, 관대화, 신뢰성에서)대신에 양식의 차이는 고과분산의 적은 양(약 5%)만을 설명한다고 주장하였다.

Bernardin, Alvares, 그리고 Cranny(1976)는 엄격한 척도 개발이 양식의 척도에 관계없이 고과오류를 막는데 중요한 문제라고 주장하였다. Bernardin(1977)은 BARS기법에서 척도기준 선정에 사용되는 항목 분석절차가 사용되었을 경우, BARS 고과과 summated 고과사이에 차이가 없다는 것을 발견했다.

3.3 행위관찰인사고과(BOS:Behavioral Observation Scale)

BOS(Behavioral Observation Scale)는 기본적으로 BARS양식의 장점을 유지하면서 단점을 줄이고자 개발된 행위기준 유형의 고과수단이다.[36]

BARS와 BOS의 차이점은 태도측정치 개발에 있어서의 Trustone(1928)접근법과 Likert(1932)접근법의 차이와 동일하다. BARS는 고과자가 직무분석에 의해 식별된 중요사건(critical incidents)을 직무의 효과성과 관련하여 계량적으로 고과한다는 면에서 Trustone접근법과 유사하다.

BOS는 다음과 같은 면에서 Likert의 방법과 유사하다. 첫째, 직무성과와 관련된 행위서술문이 식별된다. 둘째, 종업원의 행위를 관찰하여 척도상의 행위를 어느 정도의 빈도로 행하였는가를 5점척도에 따라 고과한다. 셋째, 각 종업원의 총점은 각 행위항목에 대한 점수를 합하여 계산한다. 넷째, 종업원의 효과성을 식별할 수 있는 행위를 식별하기 위하여 통계적인 분석을 한다.

통계적인 분석이 용이하다는 것은 BARS와 크게 다른 점이라고 할 수 있다. BOS척도의 구체적인 예는 다음의 <그림 3>과 같다.

인사고과에 있어서의 BOS의 유용성은 다음과 같다. 먼저 BOS는 BARS와 마찬가지로 체계적인 직무분석을 기초로 하여 개발된다. 따라서 척도에 대한 고과자의 이해와 사용이 용이하

다. 둘째, BOS는 그 자체로 혹은 직무기술서의 보조자료로서 어떤 직무상에서 종업원에게 요구되는 행위가 무엇인가를 제공한다.

<그림 3> 관리자의 성과차원에 대한 BOS의 예

1. 변화관리

- 1) 변화의 내용을 종업원에게 상세히 설명한다.
거의 안한다 1 2 3 4 5 항상 그렇다
- 2) 변화가 왜 필요한가를 설명한다.
거의 안한다 1 2 3 4 5 항상 그렇다
- 3) 변화가 종업원에게 미치는 영향에 대하여 논의한다.
거의 안한다 1 2 3 4 5 항상 그렇다
- 4) 종업원의 의견을 경청한다.
거의 안한다 1 2 3 4 5 항상 그렇다
- 5) 변화가 잘 이루어질 수 있도록 종업원에게 도움을 청한다.
거의 안한다 1 2 3 4 5 항상 그렇다
- 6) 필요하다면, 종업원의 의견에 대한 실천을 확인하기 위해 사후토의 기회를 마련한다.
거의 안한다 1 2 3 4 5 항상 그렇다

합계 = _____

보완필요	보통	우수	월등	매우우수
6-10	11-15	16-20	21-25	26-30

자료원 : Latham, G.P. and Wexley, K.N., 1981. op. cit.

셋째, 종업원의 강점과 약점에 대하여 구체적인 피드백이 가능하다. 넷째, 척도상의 어떤 행위에 대하여 어느 종업원에 더 효과적인가에 대하여 고과할 수 있기 때문에 BARS에 비하여 보다 내용타당성이 있다. 다섯째, 타당성과 신뢰성에서 볼 때 평등고용법의 기준을 만족시킨다.

4. 고과양식의 비교

그림척도는 고과자가 계량적 판단으로 부터 자유로울 수 있고 고과자가 바람직한 차별을 할 수 있으면 개발이 용이하다는 면에서 장점을 갖는다. 그러나 고과자들이 그림척도를 사용하여 고과를 할 경우 고과자의 주관이 고과에 반영될 수 있기 때문에 관대화, HALO, 등과 같은 고과오류가 발생할 가능성이 많다. 이것은 그림척도가 주관적이고 임의적인 특성을 갖기 때문에 성과행위를 정확히 평가하거나 고과자의 인지적 과정에 도움을 주지 못하기 때문인 것으로 보인다. 따라서 그림척도의 개선에 의해 고과의 타당성을 증진하는 데는 한계가 있는 것으로 보인다.

그림양식과 BARS양식의 비교에 의하면 BARS는 그 장점이 개발절차에 있어서의 시간투자 증가를 정당화하기 어려운 것으로 보인다. 또한 Bernardin,[31] 그리고 Friedman과 Cornelius(1976)의 연구에 의하면, BARS가 우월한 것은 어떤 특이한 특성을 가지기 때문이 아

나라 개발과정에서의 업적성과 궁극적으로 고과를 하는 사람들이 어느 정도 참여한 결과라고 주장했다.[32]

Schwab, Heneman, 그리고 Decotiis(1975)가 지적했듯이 이 기법은 다음의 한계점을 가지고 있다.[33] 첫째는 직무분석에 의해서 식별된 많은 수의 중요사건(critical incidents) 중에서 일부만이 척도에 사용된다는 것이다. 예를 들면, 7개의 직무성과차원이 식별되었고, 차원별로 7개의 행위기준이 사용된다면, 직무분석에 의해 식별된 100여개 이상의 중요사건중에서 49개만이 사용된다는 것이다. 만일 직무분석에 의해 식별된 중요사건의 대부분이 직무의 성과를 잘 대표하고 성과를 고과하는 데 필요한 것들이라면, 이중에 일부만을 사용할 때 다음과 같은 문제가 발생할 수 있다. 첫째, 고과자가 관찰된 행위를 어떤 차원에 할당하는 데 있어서 어려움을 가져온다. 둘째, 고과자가 관찰된 행위를 척도의 예에 기준하여 고과하는 데 어려움을 가져온다.

두번째의 문제는 척도의 개발과정 중에서 중요사건의 분류에 주관성이 개입되기 때문에 범주들의 관계가 독립적이지 아닐 수 있다는 것이다.

세번째의 문제는 Borman(1979)이 지적한 것으로, 고과자들은 척도기준으로 사용된 구체적인 행위의 예와 종업원의 성과행위와의 유사성을 식별하는 데 어려움을 갖는다는 것이다.[34] 즉, 고과자들은 척도상의 기준을 자신들이 관찰한 종업원의 행위에 잘 적용시키지 못한다는 것이다.

네번째의 문제는 BARS의 결과를 종업원 상담과 개발에 사용하고자 할 경우, 고과자는 종업원의 행위를 고과기간 전체에 걸쳐서 체계적으로 기록해야 한다는 것이다. 그러나 대부분의 경우 고과자들은 이러한 작업을 할 시간을 갖지 못한다. 또한 특정 상황에서 개발된 척도가 다른 상황에 대해서도 타당성을 갖을 수 있는가에 관한 논쟁이 있다. Borman과 Vallon(1974)에 따르면 BARS는 척도가 개발된 상황내에서만 사용될 수 있는 한계가 있다고 주장했다.[35]

이 분야의 대부분의 연구자들은 BARS의 개발비용이 크다는 데 의견이 일치하고 있다. 일반적으로 행해지고 있는 개발절차는 독립된 고과집단들(궁극적으로 척도를 사용하게 될 고과자집단의 표본)이 필요하다. 이 집단들은 고과되는 차원과 이 차원들의 정의를 개발하고, 이 차원들의 다양한 수준에 대한 행위 예를 개발하고 소집단화한다. 그리고 마지막으로 척도 기준화(anchoring)과정의 부분으로서 이 예들에 척도의 값을 할당한다. 이 척도의 효과성은 적어도 부분적으로는 각 개발 단계에 참여하는 집단들의 독립성에 기초한다고 할 수 있다. 따라서 소요되는 시간이 많다고 보여진다. 결국 BARS에 대한 주요 난점은 이 척도가 산출하는 고과개발에 소요되는 비용을 정당화할 만큼 오류가 적은가하는 것이다.

Atkin과 Conlon(1978)은 논리적인 면에서 볼 때 BOS는 다음과 같은 면에서 BARS에 비하여 우월하다고 주장하였다.[37]

(1) BARS는 한 차원에 대하여 한 항목으로 고과하지만, BOS는 한 차원에 대하여 여러 항목으로 고과하기 때문에 차원에 대한 신뢰성이 높다.

(2) 행위들을 수집하고, 종업원이 각 행위를 보인 빈도를 고과한 후, 요인분석등을 실시하여 최종수단으로 사용하기 때문에 BARS에 비하여 행위항목의 설정에 있어서 주관성을 최소화할 수 있다.

(3) BARS를 사용하는 경우에 고과자가 평소에 종업원의 성과행위를 체계적으로 기록하지 않으면, 고과시점에서 종업원의 평균적인 성과행위를 기억하기가 어렵다. 그러나 BOS는 고과자와 피고과자에게 직무기능과 관련된 체크리스트의 역할을 하기 때문에 무엇을 고과하고 고과받는가를 명확히 한다. 따라서 실질적으로 고과자에 대한 인지적인 부담이 적으며, 평균적인 순위에 대하여 고과할 수 있다.

(4) BARS의 경우 고과자가 어떤 특정의 차원이 다른 차원보다 중요하다고 생각할 경우, 이 차원에 있어서 적절하다고 고과되는 행위범주를 보다 엄격하게 정의하게 된다. 따라서 부적

절한 수준의 행위범주는 증가하게 된다. 반면에 BOS에서는 고과해야 하는 행위들이 척도에서 주어지고 고과자는 단지 각 행위에 대한 발생빈도만을 고과하면 된다.

BOS와 BARS의 선택에 관한 문제는 Likert형의 척도와 Thurstone형의 척도 중에서 어느 것을 선택하는가와 동일한 문제이다.[38] 태도측정에 있어서 두 척도에 관한 실증적인 비교연구에 의하면, 신뢰성에 있어서 Likert형의 척도가 우수하다.[39] 이러한 것은 인사고과 분야에 있어서도 큰 차이가 없을 것으로 보인다.[40]

5. 결론

고과양식은 고과차원, 척도기준, 척도의 점수와 등의 특성에 따라 고과결과의 정확성에 영향을 미치는 것으로 보인다. 그러나 더욱 중요한 고려사항은 피고과자의 참여등을 통한 엄격한 척도 개발과정이 이루어져야 한다는 것이다. 이상적으로 볼 때 고과양식은 고과자가 성과정보를 처리하는 자연스런 인지적 과정을 효과적으로 도울 수 있어야 한다. 고과자들은 피고과자의 성과 행위에 대한 체계적인 조사, 관찰된 피고과자 행위의 해석, 각 차원에 대한 피고과자들의 효과성 수준에 대한 판단 등의 과정을 통하여 고과를 한다. 고과양식은 위의 각 단계에 대하여 고과자에게 도움을 줄 수 있어야 한다.

따라서 보다 효과적인 고과양식이 되려면 피고과자의 참여 등을 통한 엄격한 개발과정을 통하여 고과차원, 고과척도 등과 같은 고과양식의 특성이 고과자가 관찰된 작업행위를 고과하는 것을 도울 수 있는 명확한 표준을 제공할 수 있도록 해야 한다. 고과양식을 개발할 때에는 성과행위를 직무 특성 및 고과환경에 따라 효과적으로 관찰하고 고과할 수 있는 특성을 식별하여 설계하는 것이 필요하다. 이같은 관점에서 볼 때 그림척도는 효과성의 개선에 한계를 가지고 있는 것으로 보인다. BARS와 BOS는 엄격한 개발과정을 거칠 경우 보다 바람직한 대안이 될 수 있을 것이다.

참고문헌

- [1] Kavanagh, M. J., The Content Issue in Performance Appraisal: A Review. *Personnel Psychology*, 24, 653-669, 1971.
- [2] ① Campbell, J. P., Dunnette, M. D., Lawler, E. E. III and Weick, K. E. Jr. (1970). op. cit.
 ② James, C. R., Criterion Models and Construct Validity of Criteria. *Psychological Bulletin*, 80, 75-83, 1973.
 ③ Smith, D. E., Training Programs for Performance Appraisal: A Review. *Academy of Management Review*, 11, 22-40, 1986.
- [3] Kavanagh, M. J., 1971. op. cit.
- [4] Landy, F. J. and Farr, J. L., Performance Rating. *Psychological Bulletin*, 87, 27-107, 1980.
- [5] ① Bendig, A. W., A Statistical Report on a Revision of the Miami Instructor Rating Sheet. *Journal of Educational psychology*, 43, 423-429, 1952a.
 ② Bendig, A. W., The Use of Student Rating Scales in the Evaluation of Instructors in Introductory Psychology. *Journal of Educational Psychology*, 43, 167-175, 1952b.
 ③ Bendig, A. W., The Reliability of Self-Rating as a Function of the Amount of Verbal Anchoring and of the Number of Categories on the Scale. *Journal of Applied Psychology*, 37, 38-41, 1953.
- [6] Barrett, R. S., Taylor, E. K., Parker, J. W. and Martens, W. L., Rating Scale Content: I. Scale Information and Supervisory Ratings. *Personnel Psychology*, 11, 333-346, 1958.
- [7] ① Ibid.
 ② Bendig, A. W., 1952a. op. cit.
 ③ Bendig, A. W., 1952b. op. cit.
 ④ Maas, J. B., Patterned Scale Expectation Interview: Reliability Studies on a New Technique. *Journal of Applied Psychology*, 49, 431-433, 1965.
 ⑤ Smith, P. and Kendall, L. M., Retranslation of Expectations: An Approach To the Construction of Unambiguous Anchors for Rating Scales. *Journal of Applied Psychology*, 47, 149-155, 1963.
- [8] Finn, R. H., Effects of Some Variations in Rating Scale Characteristics on the Means and Reliabilities of Rating. *Educational and Psychological Measurement*, 32, 255-265, 1972.
- [9] Kay, B. R., The Use of Critical Incidents in Forced-Choice Scale. *Journal of Applied Psychology*, 43, 269-270, 1959.
- [10] Landy, F. J. and Farr, J. L., 1980. op. cit.
- [11] Barrett, R. S., Taylor, E. K., Parker, J. W. and Martens, W. L., 1958. op. cit.
- [12] ① Bendig, A. W., 1952a. op. cit.
 ② Bendig, A. W., 1952b. op. cit.
- [13] Bernardin, H. J., Behavioral Expectation Scales Versus Summated Scales: A Fairer Comparison. *Journal of Applied Psychology*, 62, 422-427, 1977.
- [14] Bernardin, H. J., Alvares, K. A. and Cranny, C. J., A Recomparison of Behavioral Expectation Scales To Summated Scales. *Journal of Applied Psychology*, 61, 564-570, 1976.
- [15] ① Schwab, D. P., Heneman, H. G. III and DeCotiis, T. A., Behaviorally Anchored Rating Scales: A Review of the Literature. *Personnel Psychology*, 24, 549-562, 1975.
 ② Bernardin, H. J., 1977. op. cit.
- [16] Ryan, F. J., Trait Ratings of High School Students by Teachers. *Journal of Educational Psychology*, 49, 124-128, 1958.
- [17] Barrett, R. S., Taylor, E. K., Parker, J. W. and Martens, W. L., 1958. op. cit.
- [18] Madden, J. M. and Bourdon, R. D., Effects of Variations in Rating Scale Format on Judgement. *Journal of Applied Psychology*, 48, 147-151, 1964.

- [19] Blumberg, H. H., DeSoto, C. B. and Kuethe, J. L., Evaluations of Rating Scale Formats. *Personnel Psychology*, 19, 243-259, 1966.
- [20] Smith, P. and Kendall, L. M., 1963. op. cit.
- [21] Bernardin, H. J. and Smith, P. C., A Clarification of Some Issues Regarding the Development and Use of Behaviorally Anchored Rating Scales. *Journal of Applied Psychology*, 66, 458-463, 1981.
- [22] Campbell, J. P., Dunnette, M. D. and Arvey, R. D., The Development and Evaluation of Behaviorally Based Rating Scales. *Journal of Applied Psychology*, 57, 15-22, 1973.
- [23] Latham, G. P. and Wexley, K. N., *Increasing Productivity Through Performance Appraisal*. Addison-Wesley Company. pp. 3-4, 1977.
- [24] Smith, P. and Kendall, L. M., 1963. op. cit.
- [25] Maas, J. B., , 1965. op. cit.
- [26] Landy, F. J. and Guion, R. M., Development of Scales for the Measurement of Work Motivation. *Organizational Behavior and Human Performance*, 5, 93-103, 1970.
- [27] Borman, W. C. and Vallon, W. R., A View of What Can Happen When Behavioral Expectation Scales Are Developed in One Setting and Used in Another. *Journal of Applied Psychology*, 59, 197-201, 1974.
- [28] Burnaska, R. F. and Hollmann, T. D., An Empirical Comparison of the Relative Effects of Rater Response Biases on Three Rating Scale Formats. *Journal of Applied Psychology*, 59, 307-312, 1974.
- [29] Borman, W. C. and Dunnette, M. D., 1975. op. cit.
- [30] Latham, G. P. and Wexley, K. N., 1981.
- [31] ① Bernardin, H. J., Behavior-based Versus Trait-oriented Performance Ratings: An Empirical Study. *Journal of Applied Psychology*, 60, 561-565, 1977.
② Bernardin, H. J., Alvares, K. A. and Cranny, C. J., 1976. op. cit.
- [32] Friedman, B. A. and Cornelius, E. T. III, Effect of rater Participation In Scale Construction On the Psychometric Characteristics of Two Rating Scale Format. *Journal of Applied Psychology*, 61, 210-216, 1976.
- [33] Schwab, D. P., Heneman, H. G. III and DeCotiis, T. A., 1975. op. cit.
- [34] Borman, W. C., Format and Training Effects on Rating Accuracy and Rater Errors. *Journal of Applied Psychology*, 64, 410-421, 1979.
- [35] Borman, W. C. and Vallon, W. R., 1974. op. cit.
- [36] Atkin, R. S. and Conlon, E. J., Behaviorally Anchored Rating Scales: Some Theoretical Issues. *Academy of Management Review*, 3, 119-128, 1978.
- [37] Latham, G. P. and Wexley, K. N., 1981. op. cit.
- [38] Seiler, L. H. and Hough, R. L., Empirical Comparisons of the Thurstone and Likert Techniques. In G. E. Summers(ed). *Attitude Measurement*. Chicago. Rand McNally, 1970.
- [39] Latham, G. P. and Wexley, K. N., 1981. op. cit.