

## 교육·훈련프로그램 반응평가를 위한 평가범주와 질문지 개발준거

이성흠<sup>†</sup> (국민대학교)

### 〈요약〉

이 연구는 교육·훈련프로그램 반응평가를 위한 평가범주와 질문지 개발에 요구되는 필수요소를 탐색하고, 이들 반응평가 평가범주와 질문지 개발준거를 사용하여 우리나라 교육·훈련기관에서 사용하는 반응평가 질문지를 분석하고자 함이 그 목적이다. 이 연구는 반응평가를 위한 11영역의 평가범주와 질문지 개발에 필요한 5요소의 개발준거를 종합·제시하였으며, 이들 평가범주와 질문지 개발준거를 기준으로 우리나라 교육·훈련기관에서 사용하고 있는 반응평가 도구를 수집·분석하고, 반응평가 평가범주별 표본평가문항을 발췌·제시하였다. 반응평가에 사용되고 있는 질문지에 대한 분석결과는 대부분 교육·훈련기관은 프로그램 평가의 목적과 범위에 맞추기 위하여 3~5영역의 평가범주를 선택적으로 사용하고 있었으며, 소수의 반응평가 질문지가 질문지 개발준거에 따라서 개발·사용되고 있었다. 적절한 평가범주의 선정과 질문지 개발준거에 부합되는 반응평가 질문지 개발·활용은 반응평가가 교육·훈련프로그램 평가에서 '학습자 만족도' 확인을 위한 보다 유용하고 가치 있는 도구로 사용될 수 있을 것이다.

주요어 : 교육·훈련프로그램, 반응평가, 평가범주, 질문지 개발준거, 학습자 만족도

<sup>†</sup> 교신저자 : 이성흠, 국민대학교 교육연구소, 서울시 성북구 정릉동 861-1  
leesungheum@korea.com, 02-910-4725

## I. 문제 제기

21세기 지식기반사회는 여러 분야에서 가능성과 함께 이에 수반되는 해결하여야 할 많은 과제를 우리에게 던져주고 있다. 이들 가운데 인간의 수행향상(performance improvement)을 통한 성과 극대화를 위하여 설계된 교육·훈련프로그램은 수행과 관련되는 문제해결을 통하여 조직이 필요로 하는 요구를 충족하거나, 새로운 사업기회를 포착하여 수행향상에 공헌하여야 한다(Molenda, Pershing, & Reigeluth, 1996; Robinson & Robinson, 1989). 급속한 기술발달과 다방면에서 진행되고 있는 사회변화의 패러다임에서 “...지식은 하루가 다르게 변하고 있으며, 오래된 기능은 새로운 능력으로 대체되어야 한다”(Rosenberg, 1987, p. 57). 이들 사회변화와 관련되는 요인들은 조직이 교육·훈련 활동에 많은 자원을 소비할 것을 강요하고 있다. 교육·훈련 전문지 *Training*에 따르면, 100명 이상을 고용하고 있는 미국 기업들은 2004년도에 공식적인 교육·훈련경비로 514억 달러를 책정·사용하고 있음이 조사·보고 되고 있다(Dolezalek, 2004). 과거에는 대부분의 조직은 교육·훈련에 필요한 경비를 사업이 수행되는 과정에서 표출되는 문제 해결 또는 미래 신규사업을 위한 기회 포착에 영향을 미치는 것으로 생각하기 보다는 사업을 수행하는 데 필요한 높은 가격의 비용으로 간주하였다(Basarab & Root, 1992; Robinson & Robinson, 1989). 그러나 인간의 수행향상을 위하여 활동하고 있는 교육·훈련전문가들은 교육·훈련에 필요한 경비를 수행향상을 통한 성과 극대화를 위한 경영요구의 관점에서 투자로 생각하고 있으며, 기업을 포함한 공공조직의 관리자들도 교육·훈련 투자에 대한 수익회수 문제를 프로그램의 책무성과 정당성의 관점에서 심각하게 고민하기 시작했다. 그러나 핵심쟁점은 복잡하고 어려운 ‘프로그램 평가’ 문제이다.

교육·훈련프로그램 평가는 교육(공)학을 포함하여 인적자원개발 분야에서 직면하고 있는 가장 중요한 쟁점 분야 가운데 하나이다(Basarab, 1994; Brown & Seidner, 1998; Holton, 1996; Richey, 1995). 평가는 교육·훈련기관에서 끊임없는 수행향상을 통한 성과 극대화를 위한 필요·불가결한 요소로서 프로그램의 책무성과 정당성을 확인하는 주요 활동이다. 교육·훈련프로그램 효과를 평가하지 않는다면, 조직은 교육·훈련에 사용한 경비가 현명하게 사용되고 있는가, 성과향상을 위하여 설계·실행되고 있는 교육·훈련 프로그램을 계속하여 진행해야 하는가, 수정·보완하여야 하는가, 또는 중단하여야 하는가를 판단할 수 없다(Kaufman, Keller, & Watkins, 1995). 대부분의 프로그램 평가는 적절하게 설계·실행된 교육·훈련프로그램은 참가자들의 긍정적인 반응, 프로그램 내용의 학습, 프로그램 학습결과로 인한 직무현장에서의 행동변화, 그리고 조직에서 수행향상을 통한 결과(성과)산출에 공헌할 것이라는 가정에서 성립된다(Kirkpatrick, 1998;

Ostroff, 1991).

교육·훈련프로그램 참가자의 반응평가는 프로그램에 관한 ‘학습자 만족도(learner's satisfaction)’ 확인을 위한 준거로 가장 일반적으로 사용되고 있다. Kirkpatrick(1998)의 4수준 평가를 사용하는 프로그램 평가방법에 대한 미국훈련개발협회(American Society for Training & Development: ASTD)의 최근 조사연구(Sugrue & Kim, 2004)는 344개 조사기관의 8%가 제4수준 결과평가를, 14%가 제3수준인 교육·훈련프로그램 후 현장에서의 행동변화 평가를, 31%가 제2수준인 프로그램에서 습득한 지식, 기능, 태도와 관련된 학습평가를, 그리고 74%가 제1수준인 프로그램 종료 후 참가자의 프로그램 만족도에 관한 반응평가를 실시하고 있음을 보고하고 있다. 이들 결과는 조사연구에 참가한 많은 교육·훈련기관이 프로그램 평가에서 교육·훈련프로그램 종료와 함께 즉시 실시하는 프로그램 참가자의 반응과 의견을 수집하여 프로그램에 참가한 학습자 만족도를 평가하고 있음을 보여주고 있다.

교육·훈련프로그램 평가와 관련되는 많은 연구가 국내·외에서 수행되었지만(강이철, 2003; 정재삼, 2004; Brown & Seidner, 1998; Fisk, 1991; Keller, 1996; Kirkpatrick, 1998; McGehee & Thayer, 1961; Schouborg, 1993), 프로그램의 종료 후에 실시하는 참가자의 반응평가를 위한 체계적인 연구는 많이 이루어지지 않고 있는 실정이다(Faerman & Ban, 1993). 교육·훈련프로그램 종료 후 즉시 실시하는 평가로서 “반응평가는 교육·훈련프로그램에 대한 참가자의 태도와 감정을 포함하여 프로그램 내용의 유용성, 교(강)사의 효과성, 프로그램에서 사용한 자료의 질(質) 등에 대한 인식의 정도를 측정하기 위하여 설계되어야 한다”(Keller, 1996, p. 7). 반응평가로부터 의미 있는 정보를 얻기 위하여 인적자원개발 전문가는 프로그램 평가 설계에 있어서 엄격하여야 하며 과학적이어야 한다. 그러나 현재 사용되고 있는 반응평가 질문지는 교육·훈련프로그램에 관한 참가자의 반응을 측정하는 비공식적인 평가자료지(worksheet) 수준으로 구성되어 있으며(Carnevale, Gainer, & Villet, 1990), 프로그램 평가에서 사용되는 반응평가를 위한 질문지는 교육·훈련프로그램의 종료를 위한 하나의 의식적 기능(ritual function)으로 인식되고 있다(Easterby-Smith, 1985).

교육·훈련프로그램 평가연구에서 반응평가를 위한 유용한 범주 사용과 반응평가를 위한 질문지 설계·개발 방법에 대한 실제적인 지침 제시의 실패와 관련되는 연구결과는 프로그램 평가전문가들에게 반응평가를 위한 범주 선정과 질문에 사용할 수 있는 의미 있는 평가문항 개발에 많은 주의를 요청하고 있다. 프로그램 평가전문가는 반응평가로부터 유용한 정보를 획득하고 평가결과를 적절하게 사용하기 위하여 프로그램 반응평가와 관련되는 평가범주를 종합하고 반응평가 질문지 개발을 위한 준거와 지침을 확립할 필요성이 있다. 이와 같은 관점에서 수행향상을 통한 성과극대화에 공헌하는 교육·훈련프로그램의 평가에 대한 기본방향을 제시하기 위하여, 이 연구는 교육·훈련프

로그래밍 반응평가를 위한 평가범주와 질문지 개발의 필수요소에 대한 이론을 종합하고, 이들 반응평가 평가범주와 질문지 개발준거를 사용하여 우리나라 교육·훈련기관에서 사용하는 반응평가 질문지를 분석하고자 함이 그 목적이다.

이러한 연구목적 달성을 위한 구체적인 연구문제는 다음과 같다. 첫째, 효과적인 교육·훈련프로그램 반응평가를 위한 평가범주는 어떻게 분류되며 평가범주별 목적은 무엇인가? 둘째, 교육·훈련프로그램 반응평가를 위한 질문지를 설계·개발하는 과정에서 고려되어야 하는 개발준거는 무엇인가? 셋째, 우리나라 교육·훈련기관에서 사용하는 반응평가 질문지는 어떤 평가범주를 선택하여 사용하고 있으며, 반응평가 질문지 개발을 위한 개발준거에 따라서 개발·사용되고 있는가? 이상의 연구문제에 대한 결과는 교육·훈련프로그램 반응평가를 위한 적절한 평가범주의 선정과 효과적인 질문지 개발에 요구되는 기본 방향을 제시할 수 있을 것이며, 우리나라에서 사용되고 있는 교육·훈련프로그램 반응평가의 실태를 확인하는 데 도움이 될 것이다.

## II. 이론적 배경

### 1. 교육·훈련프로그램과 평가

수행향상을 위한 인적자원개발 활동으로 교육과 훈련은 미래와 현재의 담당업무를 잘 수행할 수 있도록 준비하는 학습활동이다. 프로그램(program)의 개념을 정의함에 있어서, 프로그램은 전통적인 교육과정(curriculum)의 개념과는 약간 다르다는 것을 언급할 필요가 있다(Grotelueschen, 1980). 성인을 위한 교육·훈련프로그램의 다양한 정의에 따르면(Boyle & Jahns, 1970), 프로그램은 근본적으로 개인 학습자의 요구에 반응하는 짧은 기간 동안의 학습경험을 함축하고 있으며, 전통적인 교육전달체제와는 다른 환경에서 실행되고 있다. 이 개념은 프로그램의 유연성(flexibility), 변화성(variability), 포괄성(all-inclusiveness)의 특성을 강조한다. 그러나 성인을 위한 교육·훈련프로그램은 비형식적인 프로그램에서부터 매우 엄격하게 구조화된 프로그램, 일반적인 프로그램에서 정교하고 전문화된 프로그램, 특정 분야의 단일 프로그램에서 여러 분야의 주제를 동시에 다루는 프로그램 등의 광범위한 연속체(continuum)에 따라 매우 다양하게 진행되고 있다(Grotelueschen, 1980).

교육·훈련프로그램은 기업과 공공기관 등의 조직 경영목표 달성과 사업기회 포착·실현에 가장 근접할 수 있도록 설계되어야 한다. 조직은 수행향상을 통한 조직목표 달성의 해결책으로 선정한 교육·훈련에 대한 투자비용이 시간, 금전, 노력, 자원 활용의

관점에서 가치 있는 활동인가를 알아야 할 필요가 있다. 조직의 경영과정에서 의사결정자도 교육·훈련프로그램의 가치에 관한 정확한 정보가 필요하다. 프로그램의 효과적인 평가체제는 교육·훈련의 효과성에 관한 적절한 정보를 제공할 수 있다. 과거에는 교육·훈련활동과 프로그램 평가는 서로 분리되는 과정으로 간주되었다. 교육·훈련이 책무성에 바탕을 두고 신뢰할 수 있는 인적자원개발 활동이 되기 위해서는 교육·훈련 과정과 프로그램 평가 과정은 수행항상체제에 통합되어 기능을 발휘하여야 한다(Gordon, 1994). 그러나 교육·훈련프로그램 평가가 학습자의 학습결과에 대하여 점수를 부여하여 판정하고 심판하여 학습자를 성공집단과 실패집단으로 유목화하는 기능으로 사용된다면, 교육·훈련 과정을 개선하고 학습자의 학습에 도움이 되는 방안제시에는 도움이 되지 못한다. 프로그램 평가의 기능은 교수·학습과정 개선에 최대한 도움이 되고, 학습자의 학습효과를 극대화시키는 역할을 하도록 하는 것으로 평가에 대한 관점의 변화가 필요한 시점이다(황정규, 1998).

교육·훈련프로그램의 평가는 교(강)사와 교육·훈련에 제공된 자원(resources) 등과 같은 프로그램의 특정 측면에 대한 의사결정을 하기 위하여 프로그램의 가치에 관한 유용한 정보를 찾아서 수집하고 제공하는 체계적인 과정이다. 평가는 교육·훈련프로그램의 역동적인 방향(flavor)을 포착하기 위하여 프로그램의 강점과 약점을 상세하게 기술할 수 있어야 한다. 프로그램 평가는 교육·훈련의 전체 과정의 일부분이며, 이 전체 과정의 모든 부분에서 정보를 수집하는 지속적인 과정이 되어야 한다(Phillips, 1993). 광의의 의미로는 교육·훈련프로그램 평가는 측정과 평가분야의 연구에 기초한 일반 원칙이 현실적으로 어떻게 프로그램 개선, 질(質)관리, 연구수행 등의 목적을 달성하기 위하여 통합되는가를 보여준다(Cangelosi, 1991). 이러한 교육·훈련프로그램의 목적과 관련하여, Kirkpatrick(1998)은 프로그램 평가를 하는 3가지 이유를 다음과 같이 설명하고 있다:

1. 조직의 목표 달성에 공헌도를 보여줌으로써 교육·훈련부서의 존재 이유를 정당화하기 위한 목적
2. 교육·훈련프로그램을 계속할 것인가 또는 중단할 것인가를 결정하기 위한 목적
3. 장래 교육·훈련프로그램 개선에 요구되는 필요한 정보를 얻기 위한 목적 (p. 16)

프로그램 평가에 관한 초기 연구에서 Cronbach(1963)은 교육적 상황에서 계몽적인(illuminating) 평가에 관한 설명을 하고 있지만, 그의 관점은 기업교육 환경에서도 동등하게 적용 가능하다. 그는 평가에서 사용하는 결정의 종류와 관련하여 평가 목적을 다음과 같이 3가지로 구분하고 있다. (1) 어떤 프로그램의 자료와 방법은 만족스럽고, 프로그램의 어떤 부분은 변화가 필요한가를 결정하기 위한 프로그램 개선의 목적; (2) 개별 학습자의 요구 확인, 선발과 반 편성을 위한 학습자의 장점 판단 및 학습자의 강점

과 약점의 획득을 통한 교육·훈련 참가자의 선발 목적; 그리고 (3) 교육·훈련의 여러 수준에서 다양한 지표로서 사용되는 교육·훈련프로그램의 내용, 훈련 교(강)사, 자원의 질(質) 등의 가치판단과 관련되는 행정규제(administrative regulation)의 목적이다.

## 2. 교육·훈련프로그램 평가의 관점

교육(공)학과 인적자원개발 분야에서 교육·훈련프로그램 평가를 위한 접근(approach) 또는 모형(model)은 여러 가지가 있다. 다양한 용어가 프로그램 평가의 여러 가지 양상(facets)을 설명하기 위하여 사용되고 있다. 그러나 학자에 따라 다르게 사용되는 용어가 동일한 현상을 기술하기도 하며, 상당히 다른 프로그램 평가활동이 학자에 따라 동일한 용어를 사용하여 논의되기도 한다. 프로그램 평가에 대한 여러 가지 관점 가운데 교육·훈련프로그램 평가를 선도하는 Kirkpatrick(1998)의 4수준 접근과 Phillips(1997)의 5수준 투자회수율 평가 체제에 대하여 간략하게 살펴본다.

### 1) Kirkpatrick의 4수준 접근

교육·훈련프로그램 평가 영역을 분류하는 데 가장 잘 알려진 것은 Kirkpatrick(1998)의 4수준-반응, 학습, 행동, 결과-평가 접근이다. '반응(reaction)'평가는 보통 반응질문지를 사용하여 프로그램 참가자가 프로그램에 대하여 어떻게 생각하는가를 평가하는 것이다. '학습(learning)'평가는 수행평가를 사용하여 프로그램의 구체적인 목적과 연결하여 학습자의 지식, 기능, 태도의 변화를 측정하는 것이다. '행동(behavior)'평가는 직무수행을 관찰하거나 생산량에 대한 자료를 통하여 프로그램의 결과로서 업무환경 직무에서 행동의 변화를 측정한다. '결과(results)' 평가는 조직의 목표와 절감된 비용, 향상된 제품의 품질, 투자수익률 등과 같은 효과성 기준과 비교하여 프로그램의 최종 순이익 공헌도를 측정한다. 미국에서는 대부분 교육·훈련기관에서는 프로그램 평가를 위하여 Kirkpatrick의 4수준 도식(scheme)을 채택하여 그들 자신의 상황에 적합하게 수정·보완하여 사용하고 있으며, (Dyer, 1994; Phillips, 1997), 우리나라에서도 대부분의 기업과 공공조직의 교육·훈련기관에서는 Kirkpatrick의 4수준 도식을 수정·보완하여 프로그램 평가에 활용하고 있다.

Kirkpatrick의 평가도식은 교육·훈련 실무자가 프로그램의 평가준거에 관한 생각을 정리하는 데 많은 도움을 주는 간단하고 실용적인 체계이기 때문에 설득력이 있다 (Alliger & Janak, 1989). 이와 동시에 Kirkpatrick의 평가도식 사용에 있어서 잘못된 오해와 과잉 일반화(overgeneralization)의 문제도 발생하고 있다. 이와 관련되는 하나의 오해는 평가의 수준(levels)이 제공되는 정보 가치의 중요도에 따라 올림차순으로 배열되어 있다는

것으로(Alliger & Janak, 1989), 평가의 수준이 높아짐에 따라 하위 수준의 평가 보다 평가절차 및 과정이 어렵고 많은 비용이 필요하다는 것이다. 프로그램의 평가 수준은 가치, 중요성, 바람직성 등의 위계를 구성하지 않는다. 각기 다른 평가수준은 교육·훈련 프로그램 평가의 성격, 범위, 목적 등에 따라서 선별적으로 사용될 수 있으며, 평가의 난이도와 평가에 소요되는 비용은 평가가 이루어지는 상황변인에 달려 있다(Molenda, Pershing, & Reigeluth, 1996). 또 다른 잘못된 오해는 평가 수준간에는 높은 일련의 상호 관련성이 있다는 것을 가정하는 것이다(Alliger & Janak, 1989). 일반적인 가정은 각각의 수준은 전 단계의 수준에 의하여 영향을 받는다는 것이다. Hamblin(1974)은 그의 5수준 평가모형의 결합방식은 교육·훈련 효과의 위계를 형성한다고 하였다. 그러나 이러한 생각은 잘못 이해된 것이다. Kirkpatrick 4수준 도식에 관한 Alliger와 Janak(1989)의 종합적인 연구에 의하면, 프로그램 평가를 위한 Kirkpatrick의 4수준 평가도식의 각 수준간의 인과관계는 연구결과에서 증명되지 않고 있음을 보여주고 있다.

## 2) Phillips의 5수준 투자회수율 접근

Phillips(1997)는 Kirkpatrick의 4수준 평가도식의 수정된 프로그램 평가체제를 개발하였다. 그의 평가체제에서 중요한 수정은 투자회수율에 초점을 맞추면서 프로그램 평가의 제5수준을 추가했다. 제5수준을 제외한 다른 4개 수준에서도 약간의 수정이 있지만, 대체로 Kirkpatrick의 4수준 평가도식과 유사하다. Phillips(1997)의 프로그램 평가체제는 반응과 계획된 행동, 학습, 직무적용, 경영성과, 투자회수율(Return on Investment: ROI)로 나누어진다.

수준 1, '반응과 계획된 행동'은 프로그램에 대한 참가자가 직무에서 프로그램에서 습득한 학습내용을 적용하기 위한 실천계획을 포함하여 프로그램에 대한 만족의 정도를 측정한다. 수준 2, '학습'평가는 다양한 평가 기법과 도구를 사용하여 프로그램 과정에서 학습한 내용의 평가에 초점을 두고 있다. 수준 3, '직무적용'은 다양한 추수평가방법(follow-up methods)을 활용하여 프로그램 참가자가 직무상황에서 프로그램에서 획득한 학습내용 적용의 정도를 판단하는 평가이다. 수준 4, '경영성과'는 프로그램 참가자가 프로그램에서 학습한 내용을 직무에서 적용한 결과 이루어지는 산출물, 질(質), 비용, 시간, 사용자 만족 등에 초점을 두는 평가와 관련이 있다. 수준 5, '투자회수율'은 프로그램에 투입된 비용과 프로그램 결과로부터 획득되는 금전적인 이익을 비교하는 평가이다.

Kirkpatrick의 4수준 평가 관점과 Phillips의 5수준 투자회수율 평가 관점에 추가하여 Warr, Bird와 Rackham(1970)의 맥락-투입-반응-결과(CIRO) 접근, Hamblin(1974)의 5수준 접근, Kaufman, Keller와 Watkins(1995)의 평가접근, Molenda, Pershing과 Reigeluth(1996)의 평가접근을 비교·분석하면 <표 1>과 같이 종합된다. <표 1>에서 제시하고 있는 것과

같이, 다양한 용어와 평가범주 또는 수준을 사용하고 있지만, 교육·훈련프로그램 평가의 체제는 제1수준에서 ‘반응, 만족 또는 계획된 실천계획’ 등에 대한 평가, 제2수준에서 ‘학습과 성취’의 평가, 제3수준에서 ‘행동과 학습결과의 직무에서 전이, 직무에서의 적용’ 등을 평가, 제4수준에서 ‘결과와 궁극적인 산출물’에 대한 평가, 제5수준에서 ‘사회적 영향’ 평가에 초점을 맞추고 있다.

<표 1>에 나타난 것과 같이 평가분류에 따른 평가 목적이 비슷하기 때문에 교육·훈련프로그램 평가 관련자는 특정 평가방법, 기법, 도구, 각각의 수준에서 자료 수집을 위한 설계방법 등을 효과적으로 선택하여 사용하여야 할 것이다. <표 1>에서 제시된 어떤 평가 체제도 모든 교육·훈련프로그램이 실시되는 상황에서 사용할 수 있는 것은 아니다. 그러나 평가를 위한 일반적인 지침은 대부분의 프로그램 평가에서 적용·가능하다(Robert, 1990). 교육·훈련프로그램 평가를 위한 관점과 수준의 선택은 프로그램의 성격, 교육·훈련프로그램 설계, 실행 및 평가 환경, 프로그램 평가목적과 결과의 활용 등 여러 가지 변인에 의존한다.

<표 1> 교육·훈련프로그램 평가의 관점 비교

평가분류	Kirkpatrick 4수준 접근	Warr, Bird와 Rackham의 CIRO 접근	Hamblin 5수준 접근	Kaufman, Keller와 Watkins의 접근	Molenda, Pershing과 Reigeluth의 접근	Phillips 5수준 접근
프로그램 관련 활동		• 맥락평가 • 투입평가			• 활동회계평가	
<b>프로그램 참가자 반응*</b>	• 반응평가	• 반응평가	• 반응평가	• 반응평가-투입 • 반응평가-과정	• 반응평가	• 반응 및 행동계획
프로그램 내용의 획득	• 학습평가	• 즉시결과 평가	• 학습평가	• 미시적 수준 평가	• 학습평가	• 학습평가
프로그램 내용의 적용	• 행동평가	• 중간결과 평가	• 직무행동 평가	• 미시적 수준 평가	• 전이평가	• 직무적용 평가
조직 성과	• 결과평가	• 최종결과 평가	• 조직결과	• 거시적 수준 평가	• 경영효과 평가	• 경영결과 평가 • 투자회수율 평가
사회적 가치			• 사회영향 평가	• 대규모 수준 평가	• 사회효과 평가	

\* 이 논문에서 다루는 내용은 표에서 진하게 표시된 부분의 ‘프로그램 참가자 반응’에 관한 것이다.

### 3. 교육·훈련프로그램 반응평가

#### 1) 반응평가의 평가범주

Kirkpatrick(1998)이 제안한 평가체제에서 제1수준 반응평가는 교육·훈련프로그램에 대한 참가자의 느낌, 태도, 의견 등과 관련된 수집 자료와 함께 이루어지는 평가이다. 반응평가는 프로그램이라는 일련의 사건(event) 그 자체와 프로그램에서 사용한 각종 자료의 효과성에 영향을 미치는 여러 가지 사항에 대한 송환(feedback)의 기초 정보를 제공하고, 참가자에게 프로그램의 학습내용에 대한 책무성을 공유하는 것을 지원하는 활동이다(Forsyth, Jolliffe, & Stevens, 1999). 반응평가는 프로그램에 관련된 교(강)사와 교육·훈련 환경시설을 포함하여 프로그램의 효과성, 사용매체의 가치, 교수·학습자료 전달의 속도, 교수·학습방법, 행정관련 사항 등에 관한 프로그램 참가자의 의견 또는 신념을 평가하기 위하여 사용되고 있다.

교육·훈련프로그램 반응평가는 프로그램 개선 목적의 형성평가로 사용할 수 있으며, 프로그램의 계속 또는 종료에 대한 결정 목적의 총합평가의 역할도 할 수 있다. 이와 같이 반응평가의 두 가지 뚜렷한 역할은 평가결과 사용의 목적이 다르지만, 성격상 명확한 차이점은 존재하지 않는다. 형성평가와 총합평가는 프로그램의 향상을 위하여 개발단계에서 필요할 뿐만 아니라 프로그램이 안정되었을 때는 최종 가치에 대한 판단 또는 프로그램의 미래를 결정하기 위한 의사결정이기 때문에 본질적으로 중요하다. 프로그램 평가의 목적 구분은 평가결과 사용에 따라서 판단되지만, 주요한 차이점은 평가 정보 자체가 아니고 평가정보를 어떤 목적으로, 언제, 그리고 누가 사용하는가에 있다(Beer & Bloomer, 1986; Keller, 1996; Worthen, Sanders, & Fitzpatrick, 1997).

인간의 수행향상을 위한 인적자원개발 전문가들은 프로그램 반응평가를 위한 여러 가지 평가범주를 제안·사용하고 있지만(Fisk, 1991; Forsyth, Jolliffe, & Stevens, 1999; Keller, 1996; Kirkpatrick, 1998; Mattoon, 1992; Phillips, 1997; Schouborg, 1993), 이들 대부분은 부분적이며 종합적인 평가범주가 아니다. 프로그램 반응평가를 위한 목적에 타당하고 적용 가능한 반응평가 범주의 확인과 선정은 기업교육에서 활동하는 프로그램의 평가실무자에게는 많이 부담되는 문제(formidable problem)이다(Mattoon, 1992). 그러나 반응평가를 위한 특정 목적과 관련되는 평가범주에 대한 종합은 평가실무자가 프로그램의 수정, 보완, 종료, 및 개선을 위한 유용한 정보를 획득하기 위한 반응평가 설계와 실행에 도움이 될 수 있을 것이다.

반응평가에 포함되는 평가범주는 교육·훈련기관에서 프로그램을 평가하는 목적에 따라 선정·사용될 수 있다. 반응평가의 목적과 관련하여 Sanderson(1995)은 프로그램에 참여하기 전의 예비강좌 안내에 대한 참가자의 견해, 목적, 내용, 방법, 자원, 교육·훈련환경·숙박시설, 직무와 의도된 변화의 기간과 관련성, 훈련 담당자, 일반 견해 등과

같은 반응평가를 위한 범주를 열거하고 있다. 전형적으로 반응평가는 프로그램 내용의 흥미와 유용성, 교(강)사의 효과성, 프로그램 자료의 질, 교육·훈련환경·시설, 프로그램 개선을 위한 건의사항 등에 관한 프로그램 참가자의 반응에 관한 문항으로 구성된다(Faerman & Ban, 1993; Keller, 1996; Schouborg, 1993). 이에 추가하여 반응평가를 위한 다른 영역으로는 교실, 온도, 그리고 시험이나 수행측정과 같은 여러 가지 방법론적 전략과 관심사항(logistical concerns)을 통합할 수도 있다(Wart, Cayer, & Cook, 1993). 그러나 반응평가에서 사용되는 송환의 자료가 필요한 평가범주 영역은 프로그램의 범위와 평가의 목적에 따라서 다르게 사용될 수 있다. 반응평가를 위한 공통 범주에 추가하여, Robinson과 Robinson(1989)은 반응평가는 프로그램의 내용과 과정에 관한 보다 정확한 정보를 확보하기 위하여 특정 프로그램 평가에 특화된 질문(specific questions)을 포함하여야 함을 지적하고 있다. 또한 그들은 프로그램 반응평가는 프로그램 자체에 초점을 맞추어야 할 뿐만 아니라 프로그램 참가자가 느끼는 교실에서 학습한 내용을 업무환경에서 전이 가능성에 대한 일반적인 질문과 학습내용의 전이를 위한 조직의 기능, 기법 또는 접근을 위한 준비도에 관한 질문도 가능함을 지적하고 있다.

## 2) 반응평가 질문지 개발준거

평가를 위한 설문조사 연구에서 질문지(questionnaire)는 "사람에게 질문을 하거나 각기 다른 견해와 관점을 표현하는 진술에 대한 '동의' 또는 '동의하지 않음'을 질문하는 자료수집의 한 가지 방법"이다(Babbie, 1995, p. 158). 질문지는 반응자와 면대면(face-to-face)으로 상호작용하는 것을 대체하는 간접적인 자료수집 방법이다. 질문지는 반응자가 이해할 수 있는 적절한 질문을 하고 질문에 대한 적절한 대답을 할 수 있는 경우에 가치가 있다. 질문지는 반응자에게 반응하기 위한 시간을 주고 반응자의 반응에 대한 익명성(anonymity)을 보장할 수 있다. 질문지는 자료수집 목적과 관련하여 풍부한 정보를 비교적 짧은 기간에 수집할 수 있으며, 저렴한 비용으로 쉽게 실행할 수 있는 자료수집 방법이다(Babbie, 1995). 반응평가 질문지는 교육·훈련프로그램에 관한 정보를 수집하기 위하여 작성된다. 대중성 및 융통성과 함께 반응평가 질문지는 의도된 목적을 충족하기 위하여 적절하게 설계하는 것이 중요하다. 부적절한 용어를 사용한 질문지는 조사연구에서 문제의 주요 원인이다(Neuman, 1997). 교육·훈련 참가자의 태도와 느낌을 측정하는 효과적인 평가도구로 사용하기 위하여, 반응평가 질문지 문항은 활용성, 책무성, 기술적 측면 등에서 높은 수준의 질(質)을 충족해야 한다(McBean & Al-Nassri, 1982). 이것은 질문을 위한 문항은 의도된 평가 목적에 적합해야 하며 통계적으로 높은 수준의 신뢰도와 타당도를 보여주어야 함을 의미한다.

Swisher(1980)는 양질의 질문지 작성에 사용되는 준거(criteria)를 질문지 형식 또는 배치(format or layout), 질문 작성, 질문 순서와 조직, 질문지에 대한 안내문(cover letter) 등으

로 구분하여 설명하고 있다. Covert(1984)도 제목, 안내문, 지시문, 인구통계 문항 부문(demographic section), 문항 작성, 구조와 형식 등의 질문지 개발을 위한 검목표(checklist)를 제안하고 있다. 이 검목표는 질문지의 기술적 속성에 대한 제안으로 질문지 작성의 가장 중요한 요소 가운데 하나이다. 보다 명료하고 이해하기 쉬운 질문은 보다 양질의 결과를 보장할 수 있다. 반응평가는 의미 있게 분석되고 해석되어야 하는 양적·질적 자료를 산출한다. 필요한 자료 분석의 형태는 평가계획 단계에서 결정된다(Phillips, 1997). 일반적으로 반응평가 질문지 설계자는 반응자료 분석을 위한 방법과 절차를 제공한다. 자료 분석의 설계는 (1) 대답되어야 하는 질문에 대한 상세한 설명, (2) 자료 분석을 위한 특정 개념의 조작적 정의, 그리고 (3) 자료 분석을 위한 적절한 방법의 선정 등 3단계로 구성된다. 이들 단계는 반응평가 질문지 개발의 초기 단계에서 신중하게 설계되어야 한다. 변인들은 질문지에서 사용되는 개념을 적절하게 조작적으로 정의하여야 하며, 통계적 기법은 변인과 측정수준에 적절해야 한다(Weisberg, Krosnick, & Bowen, 1996). 평가 설계를 위한 이들 3단계는 반응평가를 위한 자료 분석에서도 적절하게 사용되어야 한다.

### III. 연구 방법

이 연구의 내용 및 범위는 교육·훈련프로그램의 효과적인 반응평가를 위한 평가범주와 평가질문지 개발에 요구되는 필수요소를 종합·제시하고, 이들 이론적 기초에서 정립된 반응평가 평가범주와 질문지 개발준거를 사용하여 우리나라 기업과 공공기관의 교육·훈련센터에서 사용하고 있는 반응평가 질문지를 분석하는 것이다.

이 연구는 일련의 연속되는 단계를 거쳐 진행되었으며, 각 단계별 연구방법은 다음과 같다.

첫째, 교육·훈련프로그램 평가와 질문지 개발과 관련되는 심층적인 문헌연구를 통하여 반응평가를 위한 평가범주와 질문지 개발을 위한 필수요소를 종합·분류하였다.

둘째, 문헌분석을 통하여 종합·분류된 평가범주와 질문지 개발준거의 정확성(accuracy)과 완전성(completeness)을 확보하기 위하여 교육공학, 교육과정, 프로그램평가, 설문조사를 전공한 대학교수를 초빙하여 반응평가 범주분류 자문위원회(교육공학 1, 교육과정 1, 프로그램평가 1)와 질문지 개발준거 자문위원회(프로그램 평가 1, 설문조사 2)를 구성하였다. 이들 내용전문가(Subject Matter Expert: SME)에게 문헌연구를 통하여 일차적으로 연구자 단독으로 분류한 반응평가 범주와 질문지 개발 필수요소에 대한 정확성과 포괄성에 대한 확인과정(validation)을 거쳐 반응평가 범주와 반응평가 질문지 개발준거를 수정·보완하여 최종적으로 확정하였다. 이 확인과정에서 자문위원에게 개별적으

로 연구자가 분류한 평가범주 또는 질문지 개발준거에 대한 범위(coverage)와 중요성(importance)을 평가하게 하였으며, 추가의 평가범주 또는 개발준거를 제안하도록 요청하였다. 이 과정에서 전문가를 분야별로 3명으로 구성한 이유는 전문가 검토와 확인에 필요한 위원단(panel)의 수는 각 영역별로 2~3명의 전문가가 적절하다는 선행연구의 이론적 제안을 참고하였다(Denzin, 1988; Jick, 1979; Thiagarajan, 1978).

마지막으로 자문위원회의 확인과정을 거쳐 확정된 교육·훈련프로그램 반응평가 범주와 질문지 구성요소를 평가준거로 하여 기업과 공공기관 교육·훈련센터에서 사용하고 있는 반응평가 질문지를 수집·분석하였으며, 국내·외의 참고문헌과 교육·훈련기관에서 사용하고 있는 반응평가 질문지에서 표본평가문항을 발췌·예시하였다. 우리나라 교육·훈련기관에서 사용하고 있는 반응평가 질문지 분석에는 평가의 객관성을 확보하기 위하여 교육학을 전공하는 석사과정 학생 1명과 교육공학을 전공하는 박사과정 학생 1명이 연구보조원으로 참여하였다. 이 연구에서 최종적으로 평가된 반응평가 질문지는 22개 교육·훈련기관에서 현재 사용하고 있는 것을 수집·사용하였으며, 이들 반응평가 질문지는 2004년 11월에서 2005년 3월 사이에 수집되었다. 반응평가 질문지 분석에 사용된 평가도구는 12개 공공기관과 10개 일반기업 교육·훈련기관<sup>1)</sup>에서 사용하고 있는 것이며, 반응평가 질문지 분석결과는 빈도와 백분율을 계산하였다.

## IV. 연구결과 및 논의

### 1. 프로그램 반응평가 범주분류

#### 1) 평가범주 분류와 목적

교육·훈련프로그램 반응평가에 관한 연구 및 실제와 관련되는 문헌조사(Faerman & Ban, 1993; Fisk, 1991; Keller, 1996; Kirkpatrick, 1998; Mattoon, 1992; Phillips, 1997; Sanderson, 1995; Schouborg, 1993)의 결과에 따라 종합·분류한 반응평가를 위한 범주와 목적을 자문위원회의 확인과정과 제안사항을 추가하여 종합하면 <표 2>와 같다.

---

1) 반응평가 질문지 분석에 사용된 12개 공공기관은 광주광역시 지방공무원교육원, 국립환경연구원, 서울시공무원교육원, 자치인력개발원, 중앙공무원교육원, 철도인력개발원, 한국노동교육원, 한국생산성본부, 한국수자원공사, 한국전력중앙교육원, 한국표준협회, KT&G인력개발원이며, 10개 일반기업은 금호인력개발원, 대림교육연구원, 대우인력개발원, 롯데쇼핑인재개발원, 삼성인력개발원, SK아카데미, LG아카데미(인화원), 코오롱인재개발센터, 포스코인재개발원, 현대인재개발원이다.

〈표 2〉 반응평가의 범주분류와 목적

범 주	목 적
프로그램 목적·내용	• 참가자의 기대, 적절성, 구조, 수준, 내용 최신성 등의 측면에서 프로그램 내용이 프로그램 목적 달성에 미치는 영향 평가
프로그램 자료	• 교육·훈련프로그램에서 사용한 인쇄자료와 기타 학습보조자료의 효과성, 효율성 및 유용성 판단 평가
교육방법·활용매체	• 프로그램 내용 교수·학습에 사용된 방법 및 매체활용과 첨단기술에 기반을 둔 수업의 적절성과 효과성 판단 평가
교(강)사·학습촉진자	• 프로그램을 주도한 교(강)사·학습촉진자의 능력, 준비정도, 효과성 평가
교수·학습활동	• 교육·훈련프로그램에서 전개된 교실내·외 활동의 적절성과 도움의 정도를 평가
프로그램 시기·기간	• 프로그램 시기와 기간의 계획 변경과 고려사항을 위한 전체 교육·훈련 프로그램 기간 또는 단위 강좌의 시간과 기간 평가
교육·훈련 환경	• 교실, 식당, 숙박시설, 여가활동시설을 포함하는 물리적 교육·훈련환경의 적절성 평가
계획된 행동·전이기대	• 교육·훈련에서 학습한 내용을 직무에서 활용하기 위한 참가자의 계획·기대와 전이를 위한 장애 요소 파악에 관한 평가
방법론적 전략·행정사항	• 교육·훈련프로그램을 위한 시간계획, 등록, 기타 방법론적 전략과 행정사항의 유연성과 효과성 평가
종합평가	• 교육·훈련프로그램에 관한 참가자의 만족과 느낌에 대한 종합평가
프로그램 개선을 위한 건의사항	• 장래 교육·훈련프로그램 개선을 위한 건의 및 제안사항을 수렴하는 평가

<표 2>에서 제시되고 있는 것과 같이, 교육·훈련프로그램에 대한 프로그램 참가자의 반응, 태도 또는 느낌은 매우 다양한 영역에서 수집 가능하다. 교육·훈련프로그램에 대한 참가자의 반응은 교육·훈련프로그램 내용과 방법, 학습자(훈련생), 교육·훈련의 맥락, 교육·훈련 장소, 교육·훈련 목적을 성취하기 위한 훈련생의 인지된 성공요인 등과 같은 요인들 상호간의 역동적인 관계에서 파생되는 결과이다. 참가자의 반응은 성격상 일시적(transitory)이지만, 프로그램의 전체 과정을 통하여 다양하게 표출될 수 있다. 그러므로 프로그램 반응평가를 실시하기 위하여 평가자는 평가목적과 평가시기를 먼저 결정하여야 한다. 그리고 반응평가를 위한 평가도구가 개발될 것이다. 척도를 사용하는 반응평가 질문지는 의심의 여지없이 가장 일반적인 반응평가의 자료수집 방법이며, 이것은 관찰(observation), 면접(interviews), 비공식적인 토론(informal discussion) 등의

자료수집 방법에 의하여 수정·보완되어질 수 있다.

모든 종류의 교육·훈련프로그램에 사용하기 위하여 특정 평가범주로 구성된 하나의 반응평가 질문지를 만드는 것이 교육·훈련 부서가 자주 범하는 일반적인 실수 가운데 하나이다. 그러나 최소한 개별 교육·훈련프로그램의 반응평가를 위한 일정 부분의 문항은 반응평가를 위한 설계 단계에서부터 문항개발에 대한 계획과 관련 문항을 개발하여 최종 반응평가에서 사용하여야 한다(Robinson & Robinson, 1989). 가장 이상적인 프로그램의 반응평가는 개별 조직문화와 특정 프로그램 반응평가 요구를 충족시키기 위하여 특정 프로그램과 조직의 요구에 맞게 설계·개발되어야 한다(Dixon, 1996). ‘고객 만족도(customer satisfaction)’ 평가로 알려진 프로그램 반응평가는 학습자(훈련생)의 요구조건을 충족하는 보다 효과적인 교육·훈련프로그램을 설계·개발·실행하고 있는 거의 모든 최고의 교육·훈련프로그램 실행기관에서 사용되고 있다.

## 2. 프로그램 반응평가 질문지 개발준거

### 1) 반응평가 질문지 구성요소

교육·훈련프로그램 반응평가 질문지 개발에 관한 연구와 관련되는 문헌조사 결과(Babbie, 1995; Hayes, 1992; Maher & Kur, 1983; Neuman, 1997; Swisher, 1980)와 자문위원회의 확인과정과 제안사항을 종합하여 반응평가를 위한 질문지 개발준거는 (1) 소개문, 지시문(direction), 종료를 위한 문구, (2) 질문의 형태와 척도(rating scale), (3) 질문 작성, (4) 질문지 배치 및 형식(layout or format), (5) 자료 분석의 방법과 절차의 결정으로 요약된다. <표 3>은 반응평가 질문지 설계·개발에 요구되는 5가지 필수 요소 가운데 질문 작성을 제외한 4가지 기준에 대한 요약이다.

교육·훈련프로그램 반응평가를 위한 질문지 개발과정에서 개발담당자는 안내문, 지시문과 종료를 위한 안내문, 질문의 형태와 척도, 질문의 작성, 질문지의 물리적 배열 또는 형태, 자료 분석 등과 관련되는 주요한 질문지 개발준거를 고려하여야 한다. 비록 이들 각각의 요소에 가장 최선의 정답은 없지만, 반응평가 질문지 개발자는 반응평가 질문지를 타당하고 신뢰할 수 있는 측정도구로 만들기 위해서는 기본지침을 준수해야 한다. 이러한 관점을 견지한다면 프로그램 평가자는 의도된 평가목적에 적합한 반응평가 질문지를 개발할 수 있고, 프로그램 개선을 위한 의사결정의 도구로서 반응평가의 결과를 효과적으로 활용할 수 있을 것이다.

〈표 3〉 반응평가 질문지 설계 기준

준거	설명
<b>소개문, 지시문, 종료 안내문</b>	
소개문	<ul style="list-style-type: none"> <li>반응평가의 일반적인 목적, 협조관련 사항, 익명성(anonymity) 보장 절차, 평가담당자 소개, 평가결과 공유(debrief) 절차 등과 관련되는 사항의 진술; 질문지 완성에 소요되는 예상 시간과 우편으로 반응평가를 실시하는 경우에는 회송의 최종시간도 명시</li> </ul>
지시문	<ul style="list-style-type: none"> <li>질문을 완성하기 위하여 따라야 하는 일반적인 응답요령으로 선택형, 긍정/부정, 척도 사용인 경우에는 문항에 대한 응답요령을 설명하고, 개방형 질문에서는 반응자에게 얼마 정도의 길이로 답변하여야 하는가에 대한 일반지침에 대한 기술</li> </ul>
종료를 위한 안내문	<ul style="list-style-type: none"> <li>반응평가 질문지를 완성한 것에 대한 감사를 표시, 후속되는 자료수집의 관련사항 및 완성된 평가 질문지의 회수방법을 안내하는 설명</li> </ul>
<b>질문형태</b>	
개방형	<ul style="list-style-type: none"> <li>괄호 채우기, 짧게 답하기, 에세이 작성 등과 같이 반응자의 의견을 기술하도록 요청; 예상하지 않은 가치 있는 정보 수집이 가능</li> </ul>
폐쇄형	<ul style="list-style-type: none"> <li>선택형, 긍정/부정, 진/위, Thurstone, Likert, 의미변별법 등의 척도를 사용하여 반응의 통일성 유도</li> </ul>
<b>질문지 형식</b>	
페이지 양식	<ul style="list-style-type: none"> <li>질문지를 한쪽 면(front page)만 사용하거나 양쪽 면(front &amp; back)을 사용할 것인가를 결정</li> </ul>
질문배치	<ul style="list-style-type: none"> <li>반응평가 질문을 범주별 또는 전체 문항을 종합하여 배치할 것인가를 결정</li> </ul>
전체 배열	<ul style="list-style-type: none"> <li>질문지의 겉모양과 배열은 산뜻하고, 깨끗하며, 따라가기 쉽게 하여 대답의 정확성과 완성도를 향상시키고, 질문지의 자연스런 흐름에 도움을 주고, 높은 반응율을 확보할 수 있도록 국제적 보편 규약을 준수하여 작성</li> </ul>
자료분석	<ul style="list-style-type: none"> <li>수집된 반응평가 자료의 적절한 분석은 신뢰할 수 있고 신빙성 있는 결과를 보장하기 위하여 필수; 반응자료 분석을 위한 기술통계와 추리통계 방법을 문항별 또는 전체 질문에 대한 분석의 기본지침을 기술</li> </ul>

2) 반응평가 질문문항 개발의 유의점

질문지의 핵심인 반응평가를 위한 질문 작성은 중요하다. 사회과학에서 설문조사를

〈표 4〉 반응평가 질문문항 개발 유의점

문항개발지침	설 명
간결, 명확, 짧은 질문	<ul style="list-style-type: none"> <li>질문은 간단하고, 분명하며, 짧게 작성; 필요한 단어를 사용한 질문이 보다 효과적이고, 짧은 질문은 반응 비율이 높음</li> </ul>
구체적이고 정확한 질문	<ul style="list-style-type: none"> <li>구체적이고 정확한 어투(wording)의 사용은 모든 반응자에게 똑같은 의미로 의사소통이 가능; 자주(frequently), 가장 (most), 때때로 (sometimes), 규칙적으로(regularly) 등과 같은 부사 사용에 유의</li> </ul>
적절한 언어를 사용한 질문	<ul style="list-style-type: none"> <li>적절한 수준의 어법을 사용하여 표현; 전문용어(professional jargon), 속어 (slang), 전문기술용어(technical terms), 약어(abbreviations) 등의 사용에 유의</li> </ul>
대답 가능한 질문	<ul style="list-style-type: none"> <li>반응자가 질문에 대하여 대답할 수 있는 질문 작성; 소수의 응답자만이 질문에 답할 수 있는 경우는 대부분의 반응자를 실망시키고, 쓸모없는(poor-quality) 반응의 결과 초래</li> </ul>
단일 주제의 질문	<ul style="list-style-type: none"> <li>질문은 단일의 주제 또는 견해와 관련하여 작성; 한 문항에 두(2)가지의 질문을 조합하려는 문항은 '이중 목적(double-barreled)'의 문항으로 분리 작성의 필요성</li> </ul>
주요 단어에 적절한 강조	<ul style="list-style-type: none"> <li>강조를 위하여 주요 단어에 진하게, 이탤릭, 대문자, 밑줄 긋기 등을 사용; 단어사용의 적절한 강조는 질문의 명확성에 도움</li> </ul>
민감한 질문에 주의	<ul style="list-style-type: none"> <li>봉급, 종족, 인종 등과 관련된 정보제공을 요구하는 민감한 종류의 질문을 취급하는 경우에는 특별한 주의가 요청</li> </ul>
부정 질문과 이중부정 질문 사용의 금지	<ul style="list-style-type: none"> <li>부정 질문과 이중부정을 사용한 질문은 혼란스럽고 대답이 어려움; 이러한 상황은 "표현이 서투른 진술과 상당한 오류의 잠재적 원천 (potential source)'의 결과를 초래하므로 유의</li> </ul>
편견이 실린 또는 유도된 질문 사용의 금지	<ul style="list-style-type: none"> <li>'편견이 실린(biased) 또는 유도된 질문(loaded)'은 반응평가 문항개발에서 유의; 사회적 지위가 높은 사람과 관련이 있는 강한 감정적 의미 함축(emotional connotations)과 특별한 쟁점은 반응자에게 질문을 읽고 답하는 데 영향을 미침으로 문항개발에서 유의</li> </ul>
거짓 전제 또는 미래 의도와 관련된 문항 사용의 금지	<ul style="list-style-type: none"> <li>거짓 전제 또는 미래 목적(future intentions)과 관련된 문항은 반응평가 질문 작성에서 사용 금지; 반응평가를 위한 질문은 구체적이고, 실제적이며, 반응자의 경험과 관련이 있어야 함</li> </ul>
중첩되거나 불균형의 반응을 사용한 문항 방지	<ul style="list-style-type: none"> <li>반응분류 또는 선택은 상호간에 배타적(exclusive)이고, 망라적(exhaustive)이며, 균형을 유지해야 함; 중첩되는 반응을 피하기 위하여 반응분류(response categories)는 상호간에 배타적이어야 함</li> </ul>

위한 질문을 작성하는 데 쉽게 위반할 수 있는 공통 실수에 관한 연구의 종합적인 검토(Babbie, 1995; Labaw, 1980; Neuman, 1997; Newby, 1992; Rea & Parker, 1992; Sheatsley, 1983; Weisberg, Krosnik, & Bowen, 1996)와 자문위원회의 확인과정과 제안사항을 종합한 결과는 <표 4>와 같다.

반응평가 질문지를 위한 문항작성은 과학적인 성격보다는 예술적인 측면이 더 강하다(Neuman, 1997; Payne, 1951; Sheatsley, 1983). 반응평가를 위한 문항작성은 기술, 실제경험, 인내, 그리고 창의성이 요구되는 창조활동이다. Neuman(1997)은 “혼란을 예방하고 반응자의 관점을 유지하라”(p. 233)라는 문구를 사용하여 조사연구를 위한 질문의 두 가지 중요 원칙을 제안하고 있다. 타당한 평가도구가 되기 위하여 반응평가 질문지 문항은 반응평가자에게 신뢰할 수 있는 측정이 되어야 하며, 반응자가 질문을 이해하고 그들의 대답이 의미 있다는 것을 느낄 수 있도록 제작되어야 한다. 반응평가 질문지 문항작성에 많은 단어 사용과 문항을 추가하고 싶은 유혹이 있지만, 문항 개발자는 이러한 유혹을 억제할 것을 요구받는다. 교육·훈련프로그램 반응평가자는 모든 반응자가 제시된 질문에 똑같은 방법으로 이해하기를 바란다. 그러나 반응평가를 위한 문항개발자가 모든 응답자에게 똑같이 분명하고, 프로그램의 특정 측면과 관련이 있으며, 의미 있는 문항을 개발하는 것은 매우 어려운 작업이다(Neuman, 1997). <표 4>에서 요약된 질문 작성 유의사항의 준수는 반응평가를 위한 양질의 문항개발에 도움을 될 수 있을 것이다.

### 3. 반응평가 질문지 분석 결과

#### 1) 반응평가에 사용된 평가범주

교육·훈련프로그램 반응평가는 평가목적에 따라 평가범주를 선택하여 관련되는 적절한 문항을 설계·개발하여 사용할 수 있다. 이 연구에서 분석에 사용된 대부분의 반응평가 질문지들은 <표 2>에서 제안된 11영역의 반응평가를 위한 평가범주 가운데 3~5개의 평가범주를 사용하는 것으로 나타났으며, 반응평가 질문지에서 사용된 평가범주별 사용빈도를 분석한 결과는 <표 5>와 같다.

<표 5>의 분석결과에 기초하여 반응평가에서 사용되는 평가범주는 크게 3그룹으로 구분할 수 있다. 첫째, 대부분의 반응평가 질문지에서 가장 빈번하게(80% 이상) 사용되는 범주는 ‘프로그램 목적·내용’, ‘프로그램 개선을 위한 건의사항’, ‘교(강)사·학습촉진자’, 및 ‘총괄평가’의 4영역 평가범주이다. 이들 가운데 ‘프로그램 목적·내용’과 ‘프로그램 개선을 위한 건의사항’의 범주는 분석에 사용된 모든 반응평가 질문지에서 사용되고 있었다. 둘째, 반응평가에서 자주 사용(40~80%)되는 범주는 ‘계획된 행동·전이에

관한기대', '교육·훈련환경', '프로그램시간·기간', '교수·학습활동', '방법론적 전략·행정사항'의 5영역 범주이다. 마지막으로 가끔(40% 미만) 사용되는 평가범주는 '프로그램 자료', '교육방법·매체활용'의 2영역 범주이다.

<표 5>에 제시된 반응평가 질문지 분석결과는 대부분의 반응평가는 교육·훈련프로그램과 관련되는 다양한 평가범주를 포괄적으로 다루지 않으며 '프로그램 목적·내용', '프로그램 개선을 위한 건의사항', '교(강)사·학습촉진자' 등과 같은 특정 평가범주에 초점을 맞추고 있음을 보여 주고 있다. 이러한 제한적인 평가범주 사용은 반응평가를 프로그램 목적·내용, 교(강)사, 수업개선을 위한 제언 등과 관련되는 평가에 국한하는 것으로 잘못 인식할 가능성이 있다. 프로그램 반응평가는 교(강)사 개인에 대한 평가만이 아니고 전체 프로그램 과정의 일부로서 진행되기 때문에 프로그램의 특성을 여러 측면에서 분석하는 자료를 수집할 필요가 있다. 최근 기업과 공공기관에서 이루어지고 있는 교육혁신과 관련하여 교육현장에서 활용되고 있는 문제해결을 위한 실천학습(action learning)은 프로그램에 사용한 자료, 교육·훈련방법 및 활용매체, 교육·훈련환경 등과 관련되는 범주의 문항을 포함하는 포괄적인 반응평가가 요청되는 시점이다.

〈표 5〉 반응평가 질문지에서 사용된 평가범주별 빈도분포 (N=22)

평가에 사용된 범주	빈도수	백분율(%)
프로그램 목적·내용	22	100
프로그램 개선을 위한 건의사항	22	100
교(강)사·학습촉진자	21	95.5
총괄평가	18	81.8
계획된 행동·전이에 관한 기대	17	77.3
교육·훈련환경	15	68.2
프로그램 시간·기간	12	54.5
교수·학습활동	10	45.5
방법론적 전략·행정사항	10	45.5
프로그램 자료	7	31.8
교육방법·매체활용	6	27.3
학습평가관련 문항	5	22.7
기타		
인구통계관련 문항	3	13.6
선수학습정도	1	4.5

그리고 기타의 범주로는 평가관련 문항, 인구통계관련 문항, 선수학습정도 등과 관련되는 문항도 프로그램 반응평가에서 사용되고 있었다. 이들 범주의 문항은 프로그램 목적과 내용, 참가자의 선정, 학습내용의 역량평가 등과 연결하여 종합적으로 사용할 때, 유용하게 활용할 수 있는 반응평가 문항이다. 교육·훈련프로그램이 이루어지는 현장은 역동적인 과정으로 교육·훈련에 관련되는 많은 변인이 상호작용하여 프로그램의 효과성·효율성에 영향을 미친다. 프로그램 효과성을 제고하여 업무수행 능력향상을 통한 성과 극대화에 공헌하기 위한 반응평가도 프로그램의 일부로서 실행되어야 한다. 반응평가의 결과도 프로그램의 기능과 효과를 증진시키기 위하여 프로그램 설계·개발·활용에 환류(feedback)되어 프로그램 재설계(redesign)에 종합적으로 반영되어야 한다.

2) 반응평가 질문지 설계기준별 분석 결과

<표 3>과 <표 4>에서 설정한 반응평가 질문지 개발에 요구되는 5가지 필수 요소 가운데 자료분석에 관한 준거를 제외한 영역에서 프로그램 반응평가 질문지에 대한 평가결과를 요약하면 아래와 같다. 반응평가 자료분석에 대한 평가는 수집된 반응평가 질문지에는 나타나지 않으며, 반응평가를 설계하는 단계에서 작성된 지침서와 반응평가를 실시하고 난 다음 자료분석 결과를 통하여 분석·평가가 가능하다.

(1) 질문지 설계기준

첫째, 반응평가 질문지 작성에서 반응평가에 대한 소개, 질문지 작성에 대한 협조요청, 응답에 대한 익명성 보장절차, 반응평가 자료를 분석하고 난 다음 추가로 자료를 얻기 위한 협조요청, 질문에 대한 응답요령 등과 관련되는 질문지 구성요소에 대한 분석결과는 <표 6>과 같다.

<표 6>에 나타난 것과 같이 대부분의 반응평가 질문지는 반응평가의 목적(86.3%), 협조요청(72.7%), 문항응답요령(59.1%)과 관련되는 설명에 대한 진술은 하고 있으나, 응답에 대한 익명성 보장절차(9.1%)와 추가 협조요청(4.5%)에 대한 설명은 간과하고 있었다. 질문지 구성요소에 대한 진술 형태는 혼란스럽지 않고 비교적 명료하게 진술되어 있었다. <표 6>의 분석 결과는 반응평가 질문지 작성과정에서 신뢰할 수 있고 가치 있는 학습자의 반응을 수집하기 위한 장치로써 익명성 보장절차에 대한 약속과 추가자료

<표 6> 질문지의 소개문, 지시문 및 종료안내문의 평가결과 (N=22)

목적	협조요청	익명성절차	추가 자료 협조요청	문항응답 요령	서술형태	
					명료	혼란
19(86.3%)	16(72.7%)	2(9.1%)	1(4.5%)	13(59.1%)	22(100.0%)	0(0.0%)

〈표 7〉 반응평가 질문지에 사용된 문항 형태 (N=22)

폐쇄형 질문				개방형 질문	
긍정/부정	참/거짓	선다형	척도	괄호 채우기	단답형
1(4.5%)	0(0.0%)	14(63.6%)	22(100.0%)	1(4.5%)	22(100.0%)

제공자를 사전에 확보하는 절차에 대하여 언급할 필요성이 요청됨을 보여 주고 있다.

둘째, 반응평가 질문지에 사용된 질문 형태-폐쇄형 질문과 개방형 질문-에 대한 분석 결과는 <표 7>과 같이 요약된다.

분석에 사용된 반응평가 질문지는 폐쇄형 질문에서 Likert형 척도(100.0%)와 선다형 질문(63.6%)을 사용하고 있었으며, 극소수의 반응평가 질문지는 긍정/부정의 질문(4.5%) 형태를 사용하였으며, 참(true)/거짓(false) 형태의 폐쇄형 질문은 전혀 사용하지 않는 것으로 나타났다. 이와 반면에 개방형 질문은 모두 단답형 질문을 사용하고 있었다. 폐쇄형 질문에서 Likert형 척도는 척도의 연속체(continuum)에서 점수의 간격을 3~10점으로 다양하게 사용하고 있었다. 그러나 여러 척도 문항의 종합은 같은 종합 점수 또는 동일한 결과로 귀결되며 응답자 반응의 집합은 결과해석에서 잠재적 위험을 내포하고 있기 때문에(Neuman, 1997), 조잡한 배점 분류와 함께 수집된 자료는 나중에 자료분석에서 결과를 정확하게 해석할 수 없다. 이와 같은 이론적 기초를 참고하여 척도문항을 결합하는 경우에는 질문 문항은 논리적으로 각각의 다른 문항과 관계를 고려하면서 배점 분류를 결정하여야 할 것이다.

마지막으로 질문지 구성 형식과 관련하여 반응평가 질문지의 배치양식, 질문배치, 전체 배열 등과 관련되는 질문지 겉모양(appearance)을 평가한 결과는 <표 8>과 같다.

<표 8>에 나타난 것과 같이 다수의 반응평가 도구가 제작을 위한 표준선을 이용하여(grid-based)(59.1%), 범주별 표시와 함께 평가문항 집합(question clustered with label)을 이용하여(72.7%) 반응평가 질문지를 설계·개발하여 사용하고 있었다. 개방형 질문답지는 다소 협소한(36.4%) 것으로 평가되지만, 일반적으로 적절한 답지 공간을 확보(63.6%)하고 있는 것으로 나타났다. 일반적으로 반응평가 질문지는 복잡하지 않고 설문지 공간을 효과적으로 활용하여야 한다(Babbie, 1995). 이 연구의 분석에 사용된 반응평가 질문지는 일부의 컴퓨터용 질문지를 제외하고 일반 질문지와 같은 형태를 사용하고 있었다. 일반

〈표 8〉 반응평가 질문지의 겉모양 (N=22)

배열 혼잡	표준선 이용	설계	범주별 문항집합	질문사항에 대한 답지의 공간	
				적절	협소
4(18.2%)	13(59.1%)	16(72.7%)	14(63.6%)	8(36.4%)	

형태의 반응평가 질문지를 포함하여 효과적인 자료분석을 위한 컴퓨터용 반응평가 질문지 또는 인터넷을 활용하는 온라인용 반응평가 질문지 설계·활용을 위해서는 질문지 제작 형식에도 많은 관심을 가져야 할 것으로 판단된다.

(2) 질문지에 사용된 문항의 수

이 연구에서 분석된 대부분의 반응평가 질문지(80.0%)는 6~20문항을 사용하고 있었으며, 최저 4문항에서 최대 27문항으로 구성된 반응평가 질문지를 사용하고 있는 것으로 나타났으며, 구체적인 분석결과는 <표 9>와 같다.

<표 9>에 나타난 것과 같이 분석에 사용된 대부분의 반응평가 질문지는 1~2 문항의 개방형 질문과 함께 평가목적에 알맞은 평가범주의 질문을 Likert형 척도를 사용한 선다형 문항으로 구성되어 있었다. 이 연구에서 분석한 프로그램 반응평가 질문지는 대부분 일반적인 평가문항을 사용하는 총괄적인 평가도구를 사용하였다. 그러나 평가목적에 따라 선택된 평가범주에서 구체적인 평가문항을 사용하여 반응평가를 사용하는 경우에는 질문지에 사용하는 문항의 수는 평가목적과 사용되는 범주에 따라서 다양하게 나타날 수 있을 것이다.

<표 9> 반응평가 설문지에 사용된 문항의 수 (N=22)

문항수	문항의 형태별		전체문항(%)
	개방형(%)	선다형(%)	
1~5	20(90.1)	3(13.6)	1( 4.5)
6~10	2(9.1)	4(18.2)	3(13.6)
11~15	0	8(36.4)	9(40.9)
16~20	0	6(27.3)	6(27.3)
21~25	0	0	2( 9.1)
26~30	0	1( 4.5)	1( 4.5)

<표 10> 반응평가 질문지 문항의 부적절한 사용 요약 (N=22)

사용형태	수(number)	백분율(%)
이중목적 문항	15	68.2
편견이 실린/유도된 질문	7	31.8
미래의 의도관련 질문	5	22.7
부정과 이중부정	2	9.1
민감한 질문	1	4.5

## (3) 질문 문항의 평가

반응평가에서 사용하는 질문 작성에 사용되는 어법(wording of questions)은 질문에 대한 반응자의 해석에 결정적인 요소이다. 반응평가 문항 개발에 요구되는 유의점에 관한 준거를 사용하여 분석에 사용된 질문지 분석결과는 <표 10>과 같다.

<표 10>에서 나타난 것과 같이 반응평가 문항작성에서 가장 일반적으로 나타나는 오류는 '이중목적의 문항(double-barreled questions)'(68.2%)이었다. 이들 문항은 주로 '~와 (and) 혹은 ~, 또는(or)'를 질문 문항에 포함하는 것으로 분석에 사용된 질문지의 많은 문항에서 나타났다. 이와 같은 문항의 경우에 반응평가 자료분석에서 정확한 분석과 해석이 불가능하다. 문항작성에서 '~와(and) 혹은 ~ 또는(or) 같은 접속사 사용에 많은 주의가 요구된다. 다음으로 질문 작성에서 발견되는 문제점은 <표 4>에 제시되고 있는 편견이 실린 또는 유도된 질문(31.8%), 미래의 의도관련 질문(22.7%), 부정과 이중부정(9.1%), 민감한 질문(4.5%) 등의 문항지침 개발에서 유의하여야 할 영역과 관련되는 문항을 사용하고 있는 것으로 분석되었다. 문항작성 개발과정에서 요구되는 유의점은 양질의 문항개발을 위하여 반응평가 질문 작성에서도 엄밀하게 적용되어야 하는 준거 가운데 하나로서 이에 대한 실천이 필요하다.

이 연구의 분석에 사용된 반응평가 질문지에서 문항작성의 부적절한 기준을 수정·보완하고, 반응평가 관련 참고자료에서 발췌한 반응평가를 위한 평가범주별 표본평가문항은 <표 11>과 같이 종합·제시할 수 있다.

## V. 결론 및 제언

지식기반사회에서 인간의 수행향상을 통한 성과 극대화를 위하여 설계·실행되는 교육·훈련프로그램은 경쟁적인 국제환경에서 인적자원개발을 위한 중요한 요소의 하나로 부각되고 있다. 시간이 지나면 지날수록 교육·훈련해결책은 조직에서 인적자원개발을 위한 중요한 투자로 관점이 바뀌면서 수행향상을 담당하고 있는 관련자는 이익창출에 공헌하는 경영요구에 부응하기 위하여 효과적이고 효율적인 교육·훈련프로그램 설계·개발·운영에 최선의 노력을 경주하고 있다. 이들 노력과 관련하여 책무성(accountability)은 인적자원개발 분야에서 가장 중요하게 논의되고 있는 핵심쟁점 가운데 하나로서 교육투자에 대한 경영성과 향상을 확인하기 위하여 프로그램 평가와 연결하여 논의되고 있다. 교수·학습과정의 질(質)관리를 통한 교육·훈련기관의 책무성과 전문성 신장을 위한 일반 기업체 및 공공기관의 교육·훈련프로그램 반응평가는 프로그램 개선을 위한 의사결정에 필요한 타당하고 유용한 정보를 제공하기 위하여 사용되고 있다.

〈표 11〉 반응평가 질문지 개발을 위한 범주별 표본평가문항

평가범주	평가문항
프로그램 목적·내용	<ul style="list-style-type: none"> <li>☺ 교육·훈련프로그램 내용은 진술된 프로그램 목적과 합치했다.</li> <li>☺ 교육·훈련프로그램 내용은 효과적으로 배열되었다.</li> <li>☺ 프로그램 내용의 난이도에서 적절한 수준이었다.</li> </ul>
프로그램 자료	<ul style="list-style-type: none"> <li>☺ 사용된 교육·훈련자료는 프로그램 목적과 합치하였다.</li> <li>☺ 교수·학습을 위한 유인물의 내용은 이해하기 쉽게 구성되었다.</li> <li>☺ 프로그램 자료는 사용을 위한 명확한 안내문을 가지고 있었다.</li> </ul>
교육방법· 매체활용	<ul style="list-style-type: none"> <li>☺ 청각 교수·학습 보조물은 내용을 이해하는 데 도움이 되었다.</li> <li>☺ 시각 보조물은 프로그램 내용을 이해하는 데 도움이 되었다.</li> <li>☺ 교육·훈련방법은 프로그램 내용의 학습에 효과적이었다.</li> </ul>
교(강)사· 학습촉진자	<ul style="list-style-type: none"> <li>☺ 교(강)사는 프로그램 내용을 명료하게 제시하였다.</li> <li>☺ 교(강)사는 학습자의 질문에 적절한 반응을 하였다.</li> <li>☺ 교(강)사는 수업상황을 잘 통제하였다.</li> </ul>
교수· 학습활동	<ul style="list-style-type: none"> <li>☺ 교수·학습활동은 프로그램 목적과 관련성이 있었다.</li> <li>☺ 집단 토론은 참가자(훈련생)간의 의견교환에 도움이 되었다.</li> <li>☺ 학습과제는 프로그램 내용 이해에 도움이 되었다.</li> </ul>
프로그램 시기·기간	<ul style="list-style-type: none"> <li>☺ 교육·훈련프로그램의 기간은 프로그램 내용 학습에 충분하였다.</li> <li>☺ 교육·훈련프로그램의 시기는 프로그램 목표에 적절하였다.</li> <li>☺ 프로그램 내용의 연습을 위하여 배정된 시간은 적절하였다.</li> </ul>
교육·훈련 환경	<ul style="list-style-type: none"> <li>☺ 물리적 수업(교실)환경은 학습에 도움이 되었다.</li> <li>☺ 환경조건(예: 안락함, 온도, 소음 등)은 학습에 도움이 되었다.</li> <li>☺ 교실활동을 위한 공간은 적절하였다.</li> </ul>
계획된 행동· 전이에 대한 기대	<ul style="list-style-type: none"> <li>☺ 교육·훈련프로그램 내용은 직무와 관련이 있다.</li> <li>☺ 교육·훈련프로그램 내용을 직무에 전이하는 데 장애물이 있다.</li> <li>☺ 조직의 상사(또는 동료, 부하직원 등)에게 교육·훈련프로그램에서 학습한 지식(기술)사용을 위한 지원을 기대할 수 있는가?</li> </ul>
방법론적 전략· 행정지원 사항	<ul style="list-style-type: none"> <li>☺ 이 프로그램을 위한 등록과정은 쉽게 진행되었다.</li> <li>☺ 프로그램의 과외활동은 프로그램 내용의 이해에 도움이 되었다.</li> <li>☺ 프로그램에 관한 홍보용 설명서는 정보획득에 유익하였다.</li> </ul>
종합평가	<ul style="list-style-type: none"> <li>☺ 전반적으로 교육·훈련프로그램은 유익하였다.</li> <li>☺ 교육·훈련프로그램을 총괄적으로 평가한다면?</li> <li>☺ 교육·훈련프로그램은 의도된 요구를 충족하였는가?</li> </ul>
프로그램 개선을 위한 건의사항	<ul style="list-style-type: none"> <li>☺ 교육·훈련프로그램 개선을 위한 건의사항을 기술해 주십시오.</li> <li>☺ 교육·훈련프로그램 개선에 필요한 제안사항이 있습니까?</li> <li>☺ 교육·훈련프로그램 개선을 위한 건의사항을 기술해 주십시오.</li> </ul>

이 연구에서는 교육·훈련프로그램 반응평가와 관련되는 문헌연구를 통하여 반응평가에서 타당하고 신뢰할 수 있는 정보를 확보하기 위한 11영역에서 평가범주와 반응평가 질문지 개발에 필요한 5요소의 개발준거를 확인·종합하였다. 그리고 이들 평가범주와 질문지 개발준거를 기준으로 우리나라 기업과 공공조직 교육·훈련기관에서 사용하고 있는 반응평가 도구를 수집·분석하였으며, 반응평가 범주별 표본평가문항을 발췌·제시하였다. 이 연구에서 종합·제시하고 있는 11영역의 평가범주는 프로그램 운영의 역동성 때문에 다양한 변인을 망라한 완벽한 평가범주는 아니며, 기관의 특성과 프로그램 평가목적에 적절한 평가범주를 수정·보완하여 활용할 수 있을 것이다. 프로그램 평가를 포함하여 사회과학에서 연구(research)와 이론(theory)은 조직에서 성공적인 실제에 매우 중요하다(Stolovitch, 1998). 반응평가를 교육·훈련프로그램을 포함하여 수행향상을 위한 해결책에 대한 평가를 위한 유용하고 실제적인 도구로 만들기 위하여, 우리는 교육·훈련프로그램 평가 실제에서 연구와 이론에 많은 관심과 실제 노력을 기울여야 한다. 이러한 프로그램 평가실체는 이 분야의 연구와 이론의 발전에 공헌한다. 프로그램 평가에 대한 심화된 연구는 반응평가를 위한 적절한 범주사용과 반응평가의 목적을 정당화하고, 반응평가를 위한 대안적인 자료수집 방법의 설계에 도움을 주고, 그리고 평가범주간 상호관계의 의미를 발견하기 위하여 요청된다.

이 연구에서 교육·훈련프로그램 반응평가에서 사용되고 있는 질문지에 대한 평가결과는 대부분 조직에서는 교육·훈련프로그램 평가의 목적과 범위에 따라서(customize) 3~5영역의 평가범주를 선택적으로 사용하고 있는 것을 보여주고 있으며, 소수의 반응평가지가 질문지 개발을 위한 개발준거에 따라서 개발·사용되고 있었다. 교육·훈련프로그램 평가의 질(質)을 향상시키기 위하여 평가전문가는 교육·훈련프로그램의 직접적인 평가로부터 타당하고 신뢰할 수 있는 정보를 수집하기 위하여 반응평가를 이론에 기초한 설계에 초점을 맞추어 개발하고 실행할 필요가 있다. 실제적으로 이론에 기초하지 않은 반응평가 질문지는 조잡하게 설계·개발되어 사용되고 있는 것이 현실이다. 반응평가가 교육·훈련프로그램 평가에서 '학습자 만족도 확인'을 통한 프로그램 질(質)향상을 위한 보다 유용하고 가치 있는 도구로 사용되기 위해서는 적절한 평가범주의 선정과 질문지 개발준거에 부합되는 반응평가 질문지 개발·활용이 필요하다.

## 참 고 문 헌

- 강이철 (2003). 기업교육 프로그램 평가. 나일주, 임철일, 이인숙 (편저) (2003), 기업교육론 (pp.225-249). 서울: 학지사.
- 정재삼 (2004). 교육프로그램평가. 서울: 교육과학사.
- 황정규 (1998). 학교학습과 교육평가(개정판). 서울: 교육과학사.
- Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology*, 42(2), 331-342.
- Babbie, E. R. (1995). *The practice of social research* (7th ed.). Belmont, CA: Wadsworth.
- Basarab, D. J. (1994). Assessing the quality of training evaluation studies. *Performance & Instruction*, 33(3), 19-22.
- Basarab, D. J., & Root, D. K. (1992). *The training evaluation process: A practical approach to evaluating corporate training programs*. Boston, MA: Kluwer Academic.
- Beer, V., & Bloomer, A. C. (1986). Levels of evaluation. *Educational Evaluation and Policy Analysis*, 8(4), 335-345.
- Boyle, P. G., & Jahns, I. R. (1970). Program development and evaluation. In R. M. Smith, G. F. Aker, & J. R. Kidd (Eds.), *Handbook of adult education*. New York, NY: Macmillan.
- Brown, S. M., & Seidner, C. J. (Eds.) (1998). *Evaluating corporate training: Models and issues*. Norwell, MA: Kluwer Academic.
- Cangelosi, J. S. (1991). *Evaluating classroom instruction*. New York, NY: Longman.
- Carnevale, A. P., Gainer, L. J., & Villet, J. (1990). *Training in America: The organization and strategic role of training*. San Francisco, CA: Jossey-Bass.
- Covert, R. W. (1984). A checklist for developing questionnaires. *Evaluation News*, 5(3), 74-78.
- Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, 64(8), 673-683.
- Denzin, N. K. (1988). Triangulation. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: An international handbook* (pp. 511-513). Oxford: Pergamon.
- Dixon, N. M. (1996). New routes to evaluation. *Training & Development*, 50(5), 82-85.
- Dolezalek, H. (2004). Industry report 2004: Training magazine's 23rd annual comprehensive analysis of employer-sponsored training in the United States. *Training*, 41(10), 20-37.
- Dyer, S. (1994). Kirkpatrick's mirror. *Journal of European Industrial Training*, 18(5), 31-32.
- Easterby-Smith, M. (1985). Turning course evaluation from an end to a means. *Personnel Management*, 17(4), 25-27.
- Faerman, S. R., & Ban, C. (1993). Trainee satisfaction and training impact: Issues in training evaluation. *Public Productivity & Management Review*. 16(3), 299-314.
- Fisk, C. N. (1991). *ASTD trainer's toolkit: Evaluation instruments*. San Diego, CA: Pfeiffer.
- Forsyth, I., Jolliffe, A., & Stevens, D. (1999). *Evaluating a course: Practical strategies for teacher, lecturers and trainers* (2nd ed.). London, Kogan Page.
- Gordon, S. E. (1994). *Systematic training program design: Maximizing effectiveness and minimizing*

- liability. Englewood Cliffs, NJ: PTR Prentice Hall.
- Grotelueschen, A. D. (1980). Program evaluation. In A. B. Knox (Ed.), *Developing, administering, and evaluating adult education* (pp. 75-123). San Francisco, CA: Jossey-Bass.
- Hamblin, A. C. (1974). *Evaluation and control of training*. London: McGraw-Hill.
- Hayes, B. E. (1992). *Measuring customer satisfaction: Development and use of questionnaire*. Milwaukee, WI: ASQC Quality.
- Holton, E. F. III (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly*, 7(1), 5-21.
- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 24(4), 603-611.
- Kaufman, R., Keller, J., & Watkins, R. (1995). What works and what doesn't: Evaluation beyond Kirkpatrick. *Performance & Instruction*, 35(2), 8-12.
- Keller, J. M. (1996). *Evaluating diversity training: 17 ready-to-use tools*. San Diego, CA: Pfeiffer.
- Kirkpatrick, D. L. (1998). *Evaluating training programs: The four levels* (2nd ed.). San Francisco, CA: Berrett-Koehler.
- Labaw, P. J. (1980). *Advanced questionnaire design*. Cambridge, MA: Abt Books.
- Maher, J. H., & Kur, C. E. (1983). Constructing good questionnaires. *Training and Development Journal*, 37(6), 100, 102-108, 110.
- Mattoon, J. S. (1992). *Evaluating training and educational programs: A review of the literature*. Brooks, TX: Air Force Material Command.
- McBean, E. A., & Al-Nassri, S. (1982). Questionnaire design for student measurement of teaching effectiveness. *Higher Education*, 11(3), 273-288.
- McGehee, W., & Thayer, P. W. (1961). *Training in business and industry*. New York, NY: John Wiley & Sons.
- Molenda, M., Pershing, J. A., & Reigeluth, C. M. (1996). Designing instructional systems. In R. L. Craig (Ed.), *The ASTD training and development handbook: A guide to human resource development* (4th ed.) (pp. 266-293). New York, NY: McGraw-Hill.
- Neuman, W. L. (1997). *Social research methods: Qualitative and quantitative approaches* (3rd ed.). Boston, MA: Allyn and Bacon.
- Newby, A. C. (Tony) (1992). *Training evaluation handbook*. San Diego, CA: Pfeiffer.
- Ostroff, C. (1991). Training effectiveness measures and scoring schemes: A comparison. *Personnel Psychology*, 44(2), 353-374.
- Payne, S. L. (1951). *The art of asking questions*. Princeton, NJ: Princeton University Press.
- Phillips, J. J. (1997). *Handbook of training evaluation and measurement methods: Proven models and methods for evaluating any HRD program* (3rd ed.). Huston, TX: Gulf.
- Phillips, S. (1993). *Evaluation*. London: Library Association.
- Rea, L. M., & Parker, R. A. (1992). *Designing and conducting survey research: A comprehensive guide*. San Francisco, CA: Jossey-Bass.
- Richey, R. C. (1995). Trends in instructional design: Emerging theory-based models. *Performance*

- Improvement Quarterly, 8(3), 96-110.
- Roberts, A. (1990). Evaluating training programmes. *International Trade Forum*, 26(4), 18-23.
- Robinson, D. G., & Robinson, J. C. (1989). *Training for impact: How to link training to business needs and measure the results*. San Francisco, CA: Jossey-Bass.
- Rosenberg, M. J. (1987). Evaluating training programs for decision making. In L. S. May, C. A. Moore, & S. J. Zammit (Eds.), *Evaluating business and industry training* (pp. 57-74). Boston, MA: Clair Academic.
- Schouberg, G. (1993). *Flex: A flexible tool for continuously improving your evaluation of training effectiveness*. Amherst, MA: HRD.
- Sheatsley, P. B. (1983). Questionnaire construction and item writing. In P. H. Rossi, J. D. Wright, & . B. Anderson (Eds.), *Handbook of survey research* (pp. 195-230). San Diego, CA: Academic.
- Stolovitch, H. D. (1998). Human performance technology: Research & theory to practice. Paper presented at the International Society for Performance Improvement (ISPI) Conference and Expo, Chicago, IL, March 24-26.
- Sugrue, B., & Kim, K. (2004). American Society for Training & Development's 2004 state of the industry report. Alexandria, VA: American Society for Training & Development.
- Swisher, R. (1980). Criteria for the design of mail questionnaire. *Journal of Education for Librarianship*, 21(2), 159-165.
- Thiagarajan, S. (1978). Instructional product verification and revision: 20 questions and 200 speculation. *Educational Communication and Technology Journal*, 26(2), 133-141.
- Warr, P., Bird, M., & Rackham, N. (1970). *Evaluation of management training: A practical framework, with cases, for evaluating training needs and results*. London: Gower.
- Wart, M. V., Cayer, N. J., & Cook, S. (1993). *Handbook of training and development for the public sector: A comprehensive resource*. San Francisco, CA: Jossey-Bass.
- Weisberg, H. F., Krosnick, J. A., & Bowen, B. D. (1996). *An introduction to survey research, polling, and data analysis*. Thousand Oaks, CA: Sage.
- Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). *Program evaluation: Alternative approaches and practical guidelines* (2nd ed.), New York, NY: Longman.

<Abstract>

## Dimension Classification and Questionnaire Design Criteria for Reaction Evaluation of Educational/Training Programs

Sungheum Lee (Kookmin University)

The purposes of the present study were to classify the dimensions of information sought using reaction evaluations and to establish design criteria for developing reaction questionnaires of educational/training program. There are major phases of the study: (1) classification of the dimensions and questionnaire design criteria for using reaction evaluations; (2) validation of the classified dimensions and design criteria of questionnaire by subject matter experts; and (3) assessment of a sample of educational/training reaction questionnaires currently used in educational/training centers of Korea. The findings of the study were as follows: eleven dimensions for reaction evaluation of educational/training programs were identified and classified by purposes. Five overall design criteria, each consisting of several sub-criteria, were judged important factors in designing questionnaires for reaction evaluation. The criteria include: introduction, directions, and closing statements; question format; question construction; questionnaire layout; and data analysis. It was concluded that a well-designed questionnaire for reaction evaluation of educational/training programs integrates the proper application of design criteria with appropriate reaction dimensions. Most questionnaires for reaction evaluation of educational/training programs used in educational/training centers of Korea consisted of questions representing only 3-5 dimensions. The questionnaires varied in form and length. Few of them poorly utilized the established questionnaire design criteria. With appropriate dimensions of reaction evaluation and design criteria for constructing questionnaire of educational/training program evaluation, program evaluation professionals can make reaction evaluations of program more useful and valuable in educational/training programs.

*Key words : Educational/Training Program, Reaction Evaluation, Reaction Dimension, Questionnaire Design Criteria, Learner's Satisfaction*

---

† Correspondence : Sungheum Lee, Education Research Institute, Kookmin University, 861-1 Jeongneung-dong, Songbuk-gu, Seoul / E-Mail: leesungheum@korea.com