

# 실증적 연구에서 관련강도의 보고 필요성 \*

유 태 용

광운대학교 산업심리학과

Kathy A. Hanisch

Department of Psychology  
Iowa State University

실증적 연구논문에서 중요하고 의미있는 결과를 나타내는 지표로서 통계적 유의도(p값)를 오랫동안 사용해왔다. 본 논문은 연구결과의 의미를 이해하는데 있어서 통계적 유의도와 더불어 관련강도가 사용될 필요성을 논의하였다. 산업 및 조직심리학 분야의 세 학술지(Journal of Applied Psychology, Personnel Psychology, Organizational Behavior and Human Decision Processes)에 발표된 논문들 중, 변량분석을 사용한 21개 연구결과들을 관련강도의 보고여부와 저자들의 결과해석을 중심으로 살펴보았다. 21개 논문들 중에서 11개의 논문들은 여전히 통계적 유의도에 의해서만 결과를 해석하고 관련강도는 제시하지 않았다. 통계적 유의도와 관련강도를 함께 제시한 나머지 10개 논문들에서도 통계적 유의도와 관련강도에 관한 적절한 해석이 이루어지지 않은 경우도 있었다. 연구의 결과보고에서 통계적 유의도에 대한 지나친 의존과 통계적 유의도에 관한 잘못된 해석을 논의하고, 결과로부터 얻어진 관련강도가 연구문제와 관련하여 지니게 되는 실제적 중요성에 관하여 논하였다.

변량분석(ANOVA)과 중다회귀분석(Multiple Regression Analysis)은 동일한 계산근거를 가지고 있는 통계분석 기법이고 변량분석은 중다회귀분석의 특수한 경우라고 말할 수 있다. 하지만, 흥미로운 것은 변량분석을 사용한 연구와 중다회귀분석을 사용한 연구가 결과를 보고하는 방식에 있어서는 서로 다르다는 것이다. 변량분석을 사용하여 결과를 보고할 때는 독립변인이 종속측정치에 미치는 효과를 나타내기 위하여 흔히 F값과 p값에 초점을 두고,

중다회귀분석에서는 각 예언변인의 상대적 중요성을 나타내는 회귀계수와 중다상관계수의

---

자승(R<sup>2</sup>)에 초점을 두는 경향이 있다. 두가지 통계분

석 기법을 적용한 연구들의 중요한 차이는 결과에서 관심의 대상이 되는 연구변인들 간의 관련강도를 보고하느냐 안하느냐에 있다. 서서히 변화하고 있기는 하지만 변량분석을 사용한 연구에서 독립변인에 의해 설명되는 종속측정치의 변량인 관련강도를 보고하는 경우는 드물고, 중다회귀분석을 사용한 대부분의 연구에서는 관련강도를 나타내는 지수 중의 하나인 R<sup>2</sup>을 보고하고 있다. 즉, 사용된 분석방법에 따라 결과를 보고하는 방식이 서로 다르고 각각의 결과들이 상당히 다르게 논의된다.

변량분석을 사용한 연구들에서 전통적으로 보고되어 온 p값과 더불어 독립변인에 의해 설명되는 종속측정치의 변량을 나타내는 관련강도도 제시하고 그 의미를 해석하고 이해하는 것이 필요하다. 본 논문에서는 산업

\* 이 논문은 한국심리학회 1993년도 연차학술대회에서 발표되었음.

및 조직심리학 분야에서의 세가지 주요 학술지에 실린 변량분석을 사용한 논문들의 결과에서 p값과 더불어 관련강도를 함께 보고하였는지를 살펴보고 논문저자들이 통계적 유의도와 관련강도에 관하여 어떠한 해석을 내렸는지를 검토한 후, p값과 더불어 관련강도를 보고할 필요성과 그에 대한 적절한 해석의 필요성을 논의하였다.

## 통계적 유의도와 관련강도

영가설 검증에 기초를 두고 있는 통계적 유의도는 실증적 자료에서 계산된 검증통계치가 그 통계치의 표집분포에서 나타날 확률이 어느 정도나 되는지를 나타낸다. 검증통계치의 값이 표집분포 상에서 사전에 설정된 유의도 수준(보통 5%나 1%)에 상응하는 임계값을 넘으면 유의한 것으로 보고 얻어진 검증통계치가 그 표집분포로부터 나온 값이 아니라는 결정을 내린다. 영가설 검증은 처치효과가 우연한 표집오차로부터 나온 것인지 아니면 진정한 처치효과로부터 나온 것인지를 판단하는데 도움을 준다. 영가설 검증은 오랜 역사를 지니고 있고 추론 통계학의 발전에 큰 기여를 했다.

처치효과에 관한 측정치에는 크게 두부류가 있는데, 하나는 효과크기(effect size)이고 다른 하나는 관련강도(association strength)이다. 효과크기는 처치집단 평균들 간의 차이를 나타내고 관련강도는 독립변인에 의해 설명되는 종속측정치의 변량을 나타낸다. 집단 간의 차이인 효과크기를 나타낼 때 종속변인의 측정단위를 그대로 사용하기도 하고 종속측정치를 표준점수로 바꾸어 표준편차 단위로 나타내기도 하지만, 측정치를 표준화시키면 원래척도와 상관없이 처치효과 간의 비교가 가능하기 때문에 표준화된 차이가 효과크기의 지수로서 더 자주 사용된다.

관련강도는 0에서 1까지의 값을 갖는 비율의 개념이

다. 중다회귀분석에서 관련강도로 사용되는 R<sup>2</sup>과 변량분석에서 그에 대응되는 개념인 에타자승( $\eta^2$ )은 둘다 집단간변량(SSB)을 전체변량(SST)으로 나눔으로써 얻어진다. R<sup>2</sup>과 에타자승은 독립변인에 의해 설명되는 모집단의 변량을 체계적으로 과대추정하는 경향이 있지만 이러한 과대추정은 표본크기가 증가됨에 따라 감소된다(Maxwell & Delaney, 1990). 조정된 R<sup>2</sup>과 엡실론자승( $\epsilon^2$ )은 각각 R<sup>2</sup>과 에타자승의 과대추정을 교정하기 위한 대안적인 관련강도 지수로 사용된다. 관련강도의 또 다른 지수로서 오메가자승( $\omega^2$ )은 가장 보수적인 설명 변량 추정치이고 변량분석에서 관련강도를 나타내는데 가장 추천되는 지수이다(Hays, 1973; Maxwell, Camp, & Arvey, 1981). R<sup>2</sup>과 에타자승은 관련강도에 관한 기술적(descriptive) 목적으로 개발되었고 조정된 R<sup>2</sup>, 엡실론자승, 오메가자승은 추론적 목적으로 개발되었다(Maxwell, Camp, & Arvey, 1981).

동일한 자료로부터 실제 계산된 값들의 크기를 비교해 보면 다섯가지 지수 중에서 같은 값을 갖는 R<sup>2</sup>과 에타자승이 가장 크며 다음이 엡실론자승과 조정된 R<sup>2</sup>이고 가장 작은 값이 오메가자승이다. 관련강도로 사용되는 다섯가지 지수를 계산하는 공식(Keppel, 1982, pp. 91-93)과 다섯가지 관련강도 통계치들의 실제 계산을 위한 가상적인 변량분석표가 표 1에 제시되어 있다.

통계적 유의도에 대한 잘못된 해석

표 1. 관련강도 측정치들 간의 관계

(가상적인 변량분석 요약표)				
변량원	자승화 (SS)	자유도 (df)	평균자승화 (MS)	F
집단간 (Between)	90.00	2	45.00	18.00
집단내 (Within)	30.00	12	2.50	
전체 (Total)	120.00	14	8.57	

$2=SSB/SST=90/120=.75$   
 $R^2=SSB/SST=90/120=.75$   
 $2=SSB-(J-1)MSW/SST=90-2(2.50)/120=.71$   
 조정된  $R^2=MST-MSW/MST=8.57-2.50/8.57=.71$   
 $2=SSB-(J-1)MSW/SST+MSW=90-2(2.50)/120+2.5=.69$   
 그러므로,  $2=R^2 > 2=조정된 R^2 > 2$

많은 연구자들이 p값을 계속해서 "신성불가침한 (sacred)" 것으로 취급해 왔고 자료분석 후에 p값이 .05나 .01이하이면 무언인가 특별한 것이 일어났다고 믿어왔다. 연구에서 얻어진 .05보다 작은 p값은 바로 기쁨과 학위취득을 의미하고 .05보다 큰 p값은 실망, 좌절, 학위취득의 실패를 의미한다고 생각한다 (Rosnow & Rosenthal, 1989). 심지어 Bakan(1966)은 학술지의 편집인이 p값을 논문게재 여부를 결정하는 기준으로 사용하기도 했음을 지적했다. 관례적으로 .05나 .01수준을 경계로 하기는 하지만 통계적으로 유의한 결과와 유의하지 않은 결과 간에 명확한 구분이 있는 것도 아니고 p값은 제1종 오류로서 연속선 상에 존재하는 확률이다 (Cohen, 1990; Rosnow & Rosenthal, 1989).

p값은 표본의 수와 독립변인의 처치조건의 수에 따라 직접적인 영향을 받는다. Hays(1973)는 만일 연구자가 충분히 많은 피험자들을 사용한다면 실제 모집단에 존재하는 효과크기에 관계없이 (즉, 미소한 차이라도), 유의한 결과를 보일 수 있다고 언급하였다. 그런데도 연구에서 어느정도의 p값을 유의한 것으로 받아들일 것인가에 관하여 논란이 있었고 p값을 신비스러운 것으로 여기는 경향이 계속해서 있어왔다. 예를 들어, Cohen(1990)은

.05의 장벽을 넘은 결과는 통계적 유의미성이라는 명예로운 상태를 성취한 것으로 해석한다고 기술하였다. 그는 또한 많은 사람들이 통계적 유의미성을 연구에서 처치효과가 있는 것으로 이해한다고 언급하였다. Lykken (1968)은 통계적 유의성은 좋은 실험에 있어서 가장 덜 중요한 속성이라고 주장하였다. 많은 연구들은 통계적 유의성의 문제가 그다지 중요한 것이 아니고 심지어는 결과를 해석하는데 있어서 잘못을 저지룰 수도 있다는 것을 지적하였다 (예, Bakan, 1966; Meehl, 1978; Meehl, 1990). 하지만 많은 심리학자들이 여전히 영가설 검증과 통계적 유의성에만 지나친 강조를 두고 그들이 결론을 도출한 변인들 간에 존재하는 관계의 강도의 중요성은 소홀히 하는 경향이 있다 (Dunnette, 1966).

### 관련강도의 보고 필요성과 실제적 중요성

변량분석을 사용한 경우와 중다회귀분석을 사용한 경우의 예로 다시 돌아가서 변량분석을 사용한 연구자가 자신의 처치변인이 피험자들의 행동의 변량의 5%를 설명하고 있다고 이야기하면 우리는 어떤 생각을 갖는가? 심리학 논문들의 변량분석표로부터 설명변량의 크기를

계산해보면 (즉, 설명변량의 계산이 가능하도록 표가 충분한 정보를 제공하고 있다면) 5%라는 설명변량의 크기는 상당히 전형적인 값이다. 대조적으로 중다회귀분석을 사용한 연구자가 어떤 예언변인의 설명변량이 5%라는 것을 보고하면 우리는 그 결과를 무시하거나 적어도 그 변인이 준거를 설명하는 예언변인이 아니라는 것으로 해석할 가능성이 크다. 변량분석이나 중다회귀분석에서 설명변량의 크기가 5%일 경우에도 F값을 계산해보면 유사한 값을 가질 것이며 p값이 .01보다 작아서 통계적으로 유의한 결과를 얻을 수 있다. 그러나 각 통계기법으로부터의 결과들이 현재 연구자들 사이에서 상당히 다르게 판단, 평가되어지고 있다. 변량분석에서의 p값은 그 자체가 표본수와 같은 다른 요인과는 무관한 것처럼 독립적인 지표로서 여겨지고 마치 관련강도를 대신하여 취급되고 있다(Bakan, 1966; Cohen, 1990).

많은 학자들(Bakan, 1966; Cohen, 1990; Dunnette, 1966; Lykken, 1968; Meehl, 1978)에 의해 유의도를 나타내는 p값에 대한 대안적인 통계치가 필요하다는 것이 주장되어져 왔다. 연구에서 일차적이고 중요한 지표는 처치효과에 관한 측정치가 되어야지 p값이 되어서는 안된다(Cohen, 1965; Cook, Campbell, & Peracchio, 1990). 앞에서 언급한 것처럼 처치효과는 효과크기와 관련강도를 사용하여 나타낼 수 있다. 관련강도 통계치의 장점은 통계적으로 유의하거나 혹은 유의하지 않은 결과에서 독립변인이 종속측정치의 변량을 어느정도나 설명하고 있는지에 관한 정보를 얻는 데 있다.

Bakan(1966)은 통계적 유의도점증이 심리학적인 현상 그 자체에 관해서는 정보를 제공해 주지 않음에도 불구하고 연구자들이 p값에 관해 잘못된 이해를 하고 있다고 지적하였다. 나중에 논의부분에서 상세히 다루겠지만 바람직한 대안은 연구에 관하여 의미있는 결론을 내리는데 필요한 독립변인의 관련강도의 크기를 사전에 명시하

고 얻어진 자료로부터 계산된 관련강도의 크기를 p값과 함께 결과에 보고하는 것이다. 다시 말해서, 연구주제에 따라 실제적으로 의미를 지니는 적절한 크기의 관련강도를 연구를 수행하기 전에 명시할 필요가 있다. 심리학의 다양한 분야에서 다양한 관련강도값이 지니는 잠재적인 영향력이 다르기 때문에, 실질적으로 의미있는 관련강도의 크기는 연구주제에 따라 다를 수 있고 전적으로 연구자의 판단에 의존하는 주관적 문제이다. 예를 들어, 죽어가는 생명을 건질 수 있는 새로운 약의 효과를 검증하는데 어느정도의 관련강도를 적절한 것으로 간주할 것인가? 만일 그 약이 병의 호전의 1%라도 설명할 수 있다면 죽어가는 사람의 극히 일부라도 살릴 수 있기 때문에 1%의 관련강도라도 실제적 중요성을 지닌다. 즉, 실제적 중요성이란 통계적 중요성과는 달리, 경험자료가 연구주제와 관련하여 실제로 시사하는 중요성에 관한 연구자의 판단에 달려있다. 이러한 문제는 연구자들이 스스로 묻고 그들의 연구에서 정의내리고 방어해야 할 문제이다.

요약컨대, 연구자들이 p값에만 관심을 갖고 변인들 간의 관련강도나 연구주제에 따라 관련강도가 지니는 실제적 중요성은 소홀히 다룬 듯하다. 연구자들은 그들의 연구에서 기대하는 관련강도를 사전에 진술하지 않았고 어떤 변인에 의해 설명되는 변량이 실제적으로 어느정도의 중요성을 지니는지도 진술하지 않는 경향이 있다. 이러한 경우 연구자들은 오직 p값에만 근거하여 그들의 연구의 결론을 진술할 수 밖에 없고 결과의 영향이나 실제적 중요성은 말할 수 없게 된다. 따라서 p값뿐만 아니라 연구주제에 따라 관심의 대상이 되는 현상의 크기를 다른 연구자들에게 전달하기 위해서 관련강도 측정치들이 사용될 필요가 있다. 연구자는 t값과 F값을 계산하는데 사용되는 정보들을 이용하여 독립변인과 종속변인 간의 관련강도를 비교적 쉽게 계산해 낼 수 있다.

# 방 법

## 선택된 연구들

산업 및 조직심리학 분야에서 실증적 결과들을 주로 다루는 세가지 주요 학술지에 1988년과 1989년에 실린 논문들을 개관하였다. 세가지 학술지는 Journal of Applied Psychology(JAP), Personnel Psychology

(PP), Organizational Behavior and Human Decision Processes(OBHDP)였다. 세 학술지에 수록된 논문들 중에서 변량분석을 사용한 논문들이 본 연구를 위해 첫번째 고려대상이 되었다. 이 논문들 중, 변량분석을 사용한 논문의 결과부분에서 관련강도를 보고했거나 혹은 논문저자가 보고하지 않았더라도 변량분석표로부터 관련강도를 계산할 수 있었던 21개의 논문들을 최종적으로 선택하였다. 1988년의 논문들에서 12편,

표 2 본 연구에서 관련강도 검토를 위하여 포함된 기존연구들

저자와 연도	연구주제	학술지	검토된 p값들의 수 (p < .05, p > .05)	
Gist, Rosen 및 Schwoerer (1988)	교육훈련방법	PP	2	1
Binning, Goldstein, Garcia 및 Scattaregia (1988)*	면접	JAP	2	11
Schoorman (1988)*	수행평가	JAP	1	0
Hirst (1988)*	내적동기	JAP	4	2
Benedict와 Levine (1988)*	수행평가	JAP	14	3
Gordon, Rozelle 및 Baxter (1988)*	지원자 평가	OBHDP	4	10
Phillips와 Feedman (1988)*	목표설정	OBHDP	4	1
Paquette와 Kida (1988)	의사결정	OBHDP	4	2
Wright와 Aboul-Ezz (1988)	외적유인자	OBHDP	2	2
Weldon과 Mustan (1988)*	의사결정	OBHDP	4	9
Scott, Farh 및 Podsakoff (1988)	내적보강	OBHDP	8	27
Carnevale과 Conlon (1988)	의사결정	OBHDP	9	25
Bretz, Ash 및 Dreher (1989)	인사선발	PP	0	2
Gist (1989)	교육훈련방법	PP	2	0
Steiner와 Rain (1989)	수행평가	JAP	6	2
Gist, Schwoerer 및 Rosen (1989)*	교육훈련방법	JAP	5	1
Howell과 Frost (1989)*	리더쉽	OBHDP	14	19
Podsakoff와 Farh (1989)	목표설정	OBHDP	8	9
Northcraft와 Earley (1989)*	정보귀환(feedback)	OBHDP	5	1
Austin (1989)	목표설정	OBHDP	8	11
Hunt, Krzystofiak, Meindl 및 Yousry (1989)	의사결정	OBHDP	1	0
			전체수 107	138

주. \*는 관련강도 통계치를 보고한 연구를 나타냄.

1989년의 논문들에서 9편이 선택되었고, 학술지별로 분류해 보면 JAP로부터 6편, OBHDP로부터 12편, PP로부터 3편이 포함되었다. 각 논문에서 유의한 결과와 유의하지 않은 결과 모두가 검토되었다. 주효과와 상호작용효과에 관한 결과들만이 포함되었고 독립변인의 조작점검을 위한 결과들은 제외시켰다. 논문저자와 연도, 연구주제, 학술지, 각 논문으로부터의 유의하거나( $p < .05$ ) 유의하지 않은( $p > .05$ )  $p$ 값들의 수가 표 2에 열거되어 있다.

결과검토를 위하여 사용한 관련강도 지수

변량분석을 사용한 연구에서 관련강도의 지수로서 관행적으로 가장 많이 추천되는 것은 오메가자승과 엡실론 자승이고 회귀분석을 사용한 연구에서 추천되는 지수는  $R^2$ 이나 에타자승이다(Keppel, 1982; Maxwell, Camp, & Arvey; Maxwell & Delaney, 1990). 표본의 수가 클수록 에타자승과 오메가자승 간에 큰 차이를 보이지 않는다. 본 연구에서 검토한 21개의 논문들 중에서 10개의 논문들은 관련강도를 보고했는데 이들 중 다섯개는 에타자승을 이용하여 보고했고 나머지 다섯개는 오메가자승을 보고했다. 본 연구에서는 결과검토를 위하여 하나의 통일된 관련강도 지수로서 에타자승을 사용하였다. 오메가자승을 보고한 다섯개의 논문들로부터는 에타자승을 다시 계산하였고 관련강도를 보고하지 않은 11개 연구결과에 대해서는 변량분석 정보로부터 에타자승을 직접 계산하였다. 개별연구의 변량분석에서 모집단의 관련강도 추정을 주목적으로 할 때는 관련강도 지수로서 오메가자승을 사용하는 것이 바람직하지만 본 연구에서는 결과검토를 위한 다음의 편의적 이유때문에 에타자승을 관련강도 지수로 선택하여 사용하였다.

첫째, 본 연구의 목적은 연구에 포함된 각 주제에 관한 편기되지 않은 관련강도의 추정치를 구하고 각 연구

주제에서 얻어진 관련강도의 실제적 의미를 논의하는 것이 아니라 포함된 연구들에서 관련강도 지수 중 어떤 것이든 보고하지 않고 통계적 유의도에 의해서만 결과를 해석하는 경향이 있음을 보이려는 것이다. 전반적으로 오메가자승은 상당히 작은 값을 보이므로 포함된 여러 연구 결과로부터의 관련강도를 하나의 지수로 통일하여 비교검토해 보려는 본 연구의 목적 상, 동일한 자료로부터 가장 큰 값을 산출하고 가장 계산하기 편리한 에타자승을 편의적으로 선택하였다.

둘째, 변량분석도 중다회귀분석과 마찬가지로 일반적인 선형모형에 기초하고 있는데 중다회귀분석에서는 거의 항상  $R^2$ 을 이용하여 관련강도를 보고하고 있다. 두 가지 분석이 통계적으로 같은 계산근거를 가지고 있지만 결과보고 방식은 서로 다르다. 본 연구에 포함된 변량분석을 이용한 연구들의 관련강도의 크기를 중다회귀분석의 맥락으로 이해할 수 있는 것은 중다회귀분석의  $R^2$ 에 해당하는 에타자승이다.

## 결 과

세가지 학술지로부터의 21개 연구에 포함된 총 245개의  $p$ 값들 중에서 107개는 통계적으로 유의했고( $p < .05$ ) 나머지 138개는 통계적으로 유의하지 않았다( $p > .05$ ). 표본의 수는 32에서 354까지 있었으며 이들의 평균은 114, 표준편차는 대략적으로 38이었다. 141개의 주효과와 104개의 상호작용효과가 포함되었다. 집단 혹은 처치수준의 수는 2에서 12까지였는데 중앙치는 3이었다.  $F$ 값의 범위는 .01로부터 654.63까지였으며 대략적으로 80%가 10보다 작은 값이었다. 관련강도 통계치인 에타자승은 .00으로부터 .84까지의 값을 가졌으며 중앙치는 .02이었다.

통계적으로 유의한 107개의 결과들을 살펴보면,  $F$ 값

은 2.48에서 654.63까지의 값을 나타냈으며 대략적으로 F값들의 50%가 9보다 작은 값을 지녔다. 에타자승은 .00에서 .84에 걸쳐 있었으며 대략 75%의 값이 .20보다 작았다. 통계적으로 유의하지 않은 결과들에 있어서 F값은 .01에서 5.49의 값을 가졌으며 평균은 1.15이었다. 한편 에타자승은 .00에서 .09의 값을 지녔다.

통계적으로 유의한 결과들과 유의하지 않은 결과들에 있어서 F값과 에타자승 사이의 관계를 보다 자세히 검토해보기 위하여 각 결과에서의 F값과 에타자승으로 교차빈도표를 만들었다. 통계적으로 유의한 결과는 표 3에, 통계적으로 유의하지 않은 결과는 표 4에 각각 제시

되어 있다. 표 3과 표 4에서, F값과 에타자승값 사이의 관계를 검토해보기 위하여 F값들은 1부터 5까지의 각 범주에 대략적으로 동일한 값의 사례수가 들어가도록 편의 상 5등분하여 표를 구성하였다. 따라서 표 3과 표 4에서 1부터 5까지로 분류된 F값들의 다섯 범주에 사례수의 20%씩의 F값이 속해있다. 이러한 F값의 다섯 범주가 표 3에서는 .00에서부터 .84까지의 에타자승값과의 교차빈도표를 만들는데 사용되었고, 표 4에서는 .00에서 .09까지의 에타자승값과의 교차빈도표를 만들는데 사용되었다.

표 3. 통계적으로 유의한( $p < .05$ ) 결과들의 F값과 2값 간의 관계

		$\eta^2$																		
		.00	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10	.11	.12	.13	.14	.15	.16	.84	전체
F	1	1	1	1	5	1	3	1	4	1	0	0	0	1	0	0	0	2	21	
	2	0	1	1	2	2	4	3	3	0	1	0	1	0	0	1	1	1	21	
	3	0	1	0	0	3	4	0	5	1	0	0	1	0	1	0	2	4	22	
	4	0	0	2	0	0	1	0	1	2	1	2	2	1	2	0	1	6	21	
	5	0	0	0	0	1	0	0	0	0	0	1	0	1	0	1	0	18	22	
전체		1	3	4	7	7	12	4	13	4	2	3	4	3	3	2	4	31	107	

주. F값을 나타내는 범주로서 1은 2.48에서 4.97까지, 2는 5.02에서 6.82까지, 3은 6.84에서 13.20까지, 4는 13.35에서 30.48까지, 5는 31.43에서 654.63까지를 포함한다.

표 4. 통계적으로 유의하지 않은( $p > .05$ ) 결과들의 F값과 2값 간의 관계

		$\eta^2$										
		.00	.01	.02	.03	.04	.05	.06	.07	.08	.09	전체
F	1	27	0	0	0	0	0	0	0	0	0	27
	2	24	4	0	0	0	0	0	0	0	0	28
	3	12	14	0	1	0	0	0	0	0	0	27
	4	2	18	6	0	1	0	0	0	0	1	28
	5	1	7	7	7	4	1	1	0	0	0	28
전체		66	43	13	8	5	1	1	0	0	1	138

주. F값을 나타내는 범주로서 1은 .01에서 .13, 2는 .14에서 .52, 3은 .53에서 1.04, 4는 1.06에서 1.86, 5는 1.90에서 5.49까지를 포함한다.

표 3에서 볼 수 있듯이 에타자승값의 50% 이상이 .09와 같거나 .09보다 더 작았다. 극단적으로는 통계적으로 유의한 결과들 중에서 설명변량이 1%보다 작은 결과도 하나가 있었다. 표 4는 F값이 1보다 작은 경우는 일반적으로 .00에서 .01사이의 에타자승값들과 연관되어 있음을 보여준다. 전체적으로 이러한 에타자승값들과 교차빈도표로부터 학술지에 나타나 있는 통계적으로 유의한 결과들에 대하여 한가지 문제점을 지적할 수 있다. 통계적으로 유의한 결과들의 약 48%(51/107)가 관심의 대상이 된 종속측정치 변량의 7%이하를 설명하고 있다. 물론 작은 양의 설명변량도 실제적 중요성을 지닐 수 있지만, 문제점은 많은 연구들에서 저자들이 통계적 유의도와 설명변량 혹은 실제적 중요성과 연관된 논의를 하지 않았다는 것이다. 즉, 통계적으로는 유의하지만 관련강도는 매우 미미한 연구결과에 관하여 실제적 중요성에 바탕을 두고 논의하지는 않았다. 표 4에서 흥미로운 점은 관련강도가 상대적으로 큰 .09임에도 불구하고 통계적으로는 유의하지 않은 결과가 나온 것이다.

이러한 결과들을 보다 더 상세히 평가해보기 위하여 다음에는 개별연구들과 연구의 결과로부터 연구자들이 내린 결론들과 결과에 대한 해석을 살펴보았다. 이러한 목적으로 통계적 유의도와 더불어 관련강도를 보고한 연구들과 관련강도를 보고하지 않고 결론에서 전적으로 통계적 유의도에만 의존한 연구들로 나누어서 검토하였다.

## 관련강도를 보고한 연구들

관련강도를 보고한 논문들도 두 부류로 나누어 볼 수 있다. 한 부류는 관련강도를 논문 내에서 단순히 보고만 하고 결론을 내릴 때 그 의미에 관한 해석을 하지 않고 통계적 유의도에만 의존한 연구들이다. 다른 부류는 관련강도를 보고도 하고 결과해석에서 통계적 유의도와 함

께 논의한 연구들이다. 10개의 연구 중 관련강도를 보고한 4개(Gist et al., 1989; Northcraft & Earley, 1989; Phillips & Freedman, 1988; Weldon & Mustan, 1988)의 연구들은 관련강도에 관한 정보를 변량분석표에는 제시하였지만 본문에서는 전혀 언급하지 않았고, 단지 한 논문(Hirst, 1988)은 관련강도의 크기를 표와 본문 모두에 제시하였지만 본문에서 그 의미를 언급하지 않았다. 이러한 관련강도에 관한 정보는 분명히 독자들을 위해 제시되었지만 연구자들이 결과의 유의도에만 기초하여 결론을 내렸기 때문에 연구자들은 관련강도에 그다지 큰 의미를 부여하지 않았다.

관련강도를 보고한 나머지 5개 연구들은 관련강도의 제시와 더불어 그 크기에 기초하여 결과를 설명하는데 사용하였다. 이러한 면에서 가장 인상적인 논문은 집단간 설계를 이용한 일원변량분석에서  $F(2, 348)=12.36, p<.001, 2=.07$ (원래 논문에서는  $2=.06$ 을 보고하였다)을 제시한 Schoorman(1988)의 연구이다. 결과 부분에서 그는 “세가지 처치조건들 간에 통계적으로 유의한 차이가 있었다. 단계적 확대효과는 수행평가에 관한 평정의 변량 중 6%를 설명했다”라고 진술했다. 그는 또한 논문의 부분에서 “결과는 단계적 확대에 관한 가설을 강력히 지지한다. 단계적 확대가 수행평정 변량의 6%만을 설명하지 못하더라도 이는 통계적으로도 유의하고 실제적으로도 유의한 효과를 나타낸다. 그것의 유의성은 이러한 직무들에서 일할 종업원들을 선발하는데 사용되는 사무적 능력측정치보다 단계적 확대가 수행평가에서의 변량을 더 많이 설명하는 것으로써 예시될 수 있다”라고 결론내렸다. Schoorman의 그의 자료에 대한 설명은 그가 연구한 분야에서의 후속 연구들에 도움을 주고 그 분야가 발전해 가기 위해서 필요한 정보를 주는 것이다.

Gordon 등(1988)은 통계적으로 유의한( $p<.05$ ) 결과들이 .05에서부터 .13까지의 에타자승값을 지니고 있다



고 보고하고 논의 부분에서 “조작된 변인들에 의해 설명되어지는 변량의 양이 크지는 않지만 결과는 관련된 상황적 변인이 지원자의 중요한 속성들의 평가에 미치는 효과를 나타내고 있다”고 진술했다. 결과로부터 그들은 자신들의 연구의 실제적 중요성에 관한 평가는 추후 연구들에 의해 내려져야 한다고 결론내렸다.

관련강도를 제시하고 논의한 Binning 등(1988)은 통계적으로 유의한 상관( $r=.16, p<.03; r^2=.03$ )을 “무시될 수도 있는 공유변량의 양”이라고 인정하였다. 부가적으로 그들은 “여기에 보고된 주효과와 상호작용효과가 크지는 않더라도(그들의 논문에서 보고된 가장 큰 에타자승값은 .05였다.), 왜 우리가 이러한 효과들을 사소한 것으로 그냥 넘겨버리지 말아야 한다고 느끼는지에 관하여 몇가지 이유가 존재한다”라고 진술했다. 그들은 결과로부터 내린 결론에서 독자들의 이해를 돕기 위해 그 이유들을 기술하였다.

Howell과 Frost(1989)에 의한 연구는 14개의 유의한 F값을 보고하면서 오메가자승값도 함께 제시하였는데 단 하나의 오메가자승값에 관해서만 언급하였다. “2값은 리더쉽스타일의 주효과가 변량의 많은 부분을 설명했다는 것을 나타낸다”라고 하였다.

에타자승값을 보고한 마지막 논문은 Benedict와 Levine(1988)의 것인데 그들은 다양한 독립변인들이 종속측정치에 미치는 관련강도를 기술하면서 “상당히 큰(substantial)”, “중간정도의(moderate)”, “작은(slight)”이라는 용어를 사용하였다. 그들은 논문에서 .53과 .62의 에타자승값을 상당히 큰 것으로, .20과 .28을 중간정도로, .003을 작은 효과로 분류하였다. 이러한 값들과 연관된 모든 결과들은 통계적으로 유의하였다( $p<.05$ ). 처치변인의 효과에 관하여 이러한 용어들을 사용하기는 했지만 이러한 값들의 실제적 중요성에 관해서는 논의하지 않았다. 연구자들의 설명이 없이는 크거나 혹

은 작은 효과가 그들의 연구영역에서 어느정도나 중요하고 영향력이 있는 것인지 독자들이 이해하기란 어렵다. Benedict와 Levine(1988)은 그들의 가설검증에서 검증력 분석(power analysis)을 통하여 검증력을 계산함으로써 1종 오류와 2종 오류 모두에 관심을 기울였다. Bretz 등(1989)은 관련강도를 보고하지는 않았지만 본 연구에서 검토한 21개의 연구 중에서 논문에 검증력을 언급한 또 다른 연구이다.

## 관련강도를 보고하지 않은 연구들

앞에서 언급한 것처럼 11개 논문들은 연구에서 관련강도를 보고하지 않았고 결론을 도출하기 위하여 결과의 통계적 유의성에만 초점을 두었다. 하지만 이러한 연구들의 저자들이 결과를 보고하고 해석하는 방식들이 사회과학에서 오랜동안 내려오는 전통의 일부분이라는 것을 먼저 인식할 필요가 있다. 개별연구를 검토해보는 목적은 학술지의 결과보고와 해석에서 어떤 경향들이 있으며 심리학이 과학으로서 더욱 발전해 가기 위하여 무엇이 필요한지를 논의하기 위함이다.

Gist 등(1988)은 젊은 교육생들과 나이든 교육생들에게 모델링이 포함된 교육조건과 모델링이 포함되지 않은 교육조건을 비교하는 연구에서 두개의 통계적으로 유의한 결과들을 보고하였다. 연구의 표본수는 146이었고 모델링의 효과에 관한 결과는  $F(1, 143)=12.92, p<.001$  이었고, 젊은 사람들과 나이든 사람들 간의 수행에 관한 결과는  $F(1, 143)=20.09, p<.001$ 이었다. 그들이 보고하지는 않았지만 관련강도를 계산해본 결과, 두 결과들 각각의 에타자승값은 .05와 .08이었다. 결과에 관한 논의에서 그들은 “개인교사 방식의 교육에 비해 모델링의 전반적인 우월성은 인상적이었다”고 진술했으며 또한 컴퓨터 소프트웨어 교육기간 동안 나이든 교육생들이 유의하

게 낮은 수행을 보인 것은 주목할만하다고 했다. 나이든 교육생과 젊은 교육생 간의 평균적인 수행차는 대략적으로 7(34.59-27.67)이었다. 저자들은 교육생들의 수행점수의 범위나 수행에서 얻을 수 있는 최고점을 진술하지 않았을 뿐만 아니라 이러한 점수들이 어떻게 얻어졌는지에 관해서도 정보를 제공하지 않았다. 그들은 단지 수행이 객관적으로 채점된 25개의 과제들에서 얻어진 것이라고만 진술하였다. Gist 등(1988)은 논문에서 결과들이 잠정적이고 최종적인 일반화가 내려지기까지는 반복연구가 필요하다고 진술하였다.

Paquette와 Kida(1988)는 역시 관련강도를 보고하지 않았고 단지 통계적 유의성에만 의존했다. 통계적으로 유의한( $p < .05$ ) 그들의 결과로부터 에타자승을 계산해본 결과, .78, .57, .23, .05이었다. 그들의 결과들 중 두개는 통계적으로 유의하지 않았는데 그 중 하나는 에타자승값이 .09이었다. 통계적 유의도에 의해 저자들은 전반적으로 모든 효과들이 .01 수준에서 유의하다고 결론내렸다. 그들은 또한 "다양한 전략의 효율성에 관한 확고한 진술이 내려지기 전에 이러한 결과들의 일반화 가능성을 검토하기 위하여 부수적인 연구가 필요하다"고 진술하였다. 만일 그들이 통계적 유의도와 더불어 관련강도에 기초하여 그들의 결과들을 해석했다면 독자들은 실제적 중요성에 관한 정보를 얻을 수도 있었을 것이다.

결과의 중요성을 판단하는 기준으로 p값을 사용한 두개의 연구들은 Carnevale과 Conlon(1988)과 Austin(1989)이었다. Carnevale과 Conlon(1988)은 주요결과로서 34개의 F값과 p값을 보고했는데 9개는 통계적으로 유의했고( $p < .05$ ) 25개는 유의하지 않았다( $p > .05$ ). 부가적으로 그들은 24개의 부수적인 결과들을 보고했는데 그들의 논문에서 부가적인 변인들로 취급되었기 때문에 본 논문의 분석에서는 제외시켰다. 결과를 보

고하는데 있어서 p값을 나타내기 위하여 .07, .05, .01, .001, 그리고 .05( $p < .08$ )들이 제시되었다. 이것이 전통적인 유의도 수준의 보고와 다르기는 하지만 그들의 결과들에 대한 설명변량을 보고하고 이를 해석했다면 그들의 자료에 대한 독자들의 이해를 증진시켰을 것이다. 통계적으로 유의한 결과들에 관하여 계산된 에타자승값의 범위는 .02에서 .11이었다. 그들의 논문의 한 표에서는  $F(1, 130) = 14.24$  옆에  $p < .001$ 을 나타내기 위하여 4개의 별표(\*)를 달았다. 같은 표에서 상호작용 효과를 나타내면서  $F(1, 130) = 4.17$  옆에 두개의 별표( $p < .05$ )를 붙였다. 이러한 F값에 해당되는 에타자승을 계산해본 결과, 각각 .08과 .02였다. 그들 결과의 실제적 중요성에 관하여 그들이 적절한 설명을 했다면 단순한 4개의 별표나 통계적 유의도에 관한 정보 이상의 더 많은 의미를 독자들에게 제공했을 것이다.

통계적 유의도가 결과해석에 오해를 불러일으킬 수 있는 또 다른 예는 Austin(1989)의 연구다. 그는 19개의 결과들을 보고했는데 그중 8개는 통계적으로 유의했다. 마찬가지로 그들의 결과보고에서 F값과 p값과 더불어 관련강도를 포함했다라면 그들의 자료에 관하여 보다 나은 이해를 제공하고 보다 의미있는 설명이 가능했을 것이다. 통계적으로 유의한 결과와 연관된 에타자승값들은 .03에서 .10이었다. 그의 연구영역에서 10%라는 변량이 실제적으로 중요할지 모르지만 독자들은 통계적 유의도를 실제적 중요성과 동일하게 생각하지 않은 한 그것에 관해 이해하기 힘들다.

관련강도를 보고하지 않은 몇개의 논문들은 본 논문의 연구자들이 에타자승값을 주어진 정보로부터 계산해 보았을 때 상대적으로 큰 값을 가지고 있었다. 이는 그들의 연구의 실제적 중요성에 관하여 결론을 내리는데 독자들의 이해를 도울 수도 있었다는 것을 시사한다. 예를 들어, Wright와 Aboul-Ezz(1988)의 연구에는 .79

의 에타사승값을 갖는 결과가 포함되어 있었고 Scott 등 (1988)은 독립변인이 종속측정치 변량의 40%를 설명하는 결과를 지니고 있었다. Gist(1989)는 각각 22%와 37%의 변량을 설명하는 통계적으로 유의한 독립변인의 주효과를 발견했다. 또한 Steiner와 Rain(1989)은 통계적으로 유의한 두개의 결과와 관련하여 .20과 .35에 해당하는 에타사승값을 가지고 있었다. 다양한 연구주제에서 얻어진 이러한 관련강도가 지니는 실제적 중요성은 다를 수 있기 때문에, 이러한 결과들 각각으로부터 계산된 관련강도 정보에 근거하여 본 논문의 연구자들이나 독자들이 각 연구의 실제적 중요성을 판단하기는 힘들다. 논문의 저자들이 컴퓨터 결과물이나 통계책의 부록에 나와 있는 표로부터 얻었을  $p$ 값 이외에 관련강도에 의해 결과를 평가하고 해석했었다면 보다 유익한 정보를 독자들에게 제공할 수 있었을 것이다.

이러한 논문들을 검토하면서 얻은 부수적인 정보는 Harcum(1989)의 지적한 바대로 사회과학 분야의 경험적 논문들에서 많은 연구자들이 더 이상 사용되지 않는 것이 바람직한 "매우 유의한( $p < .001$ )", "거의 유의한( $p < .07$ )", "강력한 주효과( $p < .001$ )"와 같은 표현들을 여전히 사용하고 있다는 것이다. 이러한 표현들은 오해를 불러 일으킬 수 있는데 더욱 흥미로운 것은 관련강도치를 보고하지 않은 연구들에서 주로 사용되었다는 것이다.

## 논 의

이 연구의 목적은 변량분석을 이용한 비교적 최근의 산업 및 조직심리학 분야의 논문들에서 결과를 보고하는 양식과 결과에 대한 해석을 검토해 보는 것이었다. 그 결과, 논문들에서 통계적 유의도만을 보고하는 것은 결과가 의미하는 전반적인 내용을 독자들에게 전달하지 못

하고 실제로 오해를 불러일으킬 가능성조차 있다는 것을 강조하였다. 이러한 견해는 과거의 몇몇 논문들(예, Bakan, 1966; Cohen, 1990)에서도 진술되었지만, 비교적 최근에 발간된 세계의 주요 학술지에 포함된 변량분석을 이용한 연구들에서 반이 넘는 연구자들이 그들의 결론에 도달하기 위하여 통계적 유의성만을 사용하였다는 것은 흥미롭다. 일부 연구자들은 관련강도와 그것이 연구주제에서 지니는 실제적 중요성에 관한 언급없이 통계적 유의도를 마치 관련강도처럼 해석한 경우도 있었다.

본 논문에서 검토된 세계의 학술지로부터 얻어진 이러한 경향은 다른 심리학 학술지에 실리는 논문들에서도 흔히 볼 수 있는 현상들이다. 즉, 심리학에서 통계적 유의성에만 초점을 두는 경향이 있다. 가설검증에 있어서 통계적 유의도는 모집단에서 효과가 없음에도 불구하고 효과가 있다고 잘못된 결론을 내리는 1종오류( )에 관한 정보를 제공해주는 측면에서는 중요하지만, 이와 더불어 결과를 해석하고 설명하고 연구의 실제적 의의를 평가하기 위해서는 관련강도와 같은 효과에 관한 측정치들도 함께 고려되는 것이 필요하다. 관련강도가 큰 것이 항상 실제적으로도 중요하다고 말할 수는 없다. 앞에서 언급한 것처럼 종속측정치 변량의 6%를 설명하는 독립변인이나 60%를 설명하는 독립변인이나 연구영역에 따라 실제적 중요성 차원에서 평가해 볼 때는 똑같이 중요할 수 있다. 따라서 진정으로 요구되는 것은 연구자가 관심을 갖고 연구하는 영역에서 얻어진 결과를 실제적 중요성 측면으로도 해석하는 것이다.

과학으로서의 심리학을 보다 발전시키기 위해서 심리학의 모든 분야의 연구자들이 논문에서 통계적 유의도와 더불어 관련강도를 보고하고 연구주제에 따라 관련강도가 지니는 실제적 의미를 논의하는 것이 절대적으로 필요하다. 또한 연구에서 기대되는 결과의 실제적 중요성

이나 바라는 관련강도를 연구를 수행하기 전에 고려해 보는 것도 이상적일 것이다. 이러한 관습에 심리학자들이 일치된 견해를 갖기까지 오랜 시간이 걸릴지는 모르지만 심리학이 보다 실제적으로 우리생활에 기여하기 위해서는 거쳐야 할 길이라고 믿는다. Cohen(1990)은 변화에 시간이 걸리는 하나의 예로 1차 세계대전 10년 전에 발표된 Gosset에 의한 t검증이 2차 세계대전 후에야 심리통계 교재에 포함된 것을 들고 있다. 어느 누구도 심리학적으로 중요하다고 여겨지는 설명변량이 얼마인지에 관해서는 단언하기 힘들다. 그러나 우리는 p값에 부여하는 중요도를 상대적으로 줄이고 연구주체에 따른 효과크기와 관련강도에 주목할 필요가 있고 또 그래야만 한다. 연구의 결과 해석에서 효과크기에 관한 검토가 우선되어야 하고 p값은 그 다음에 고려될 문제다(Cohen, 1990). 왜냐하면 p값은 표본수에 의해 크게 좌우되기 때문에 똑 같은 효과크기에 관한 검증에서도 표본수에 따라 통계적으로 유의한 결과를 얻을 수도 있고 유의하지 않은 결과를 얻을 수도 있기 때문이다. 통계적으로 유의한 결과가 반드시 큰 효과를 의미하는 것도 아니고 통계적으로 유의하지 않은 결과가 반드시 작은 효과를 나타내는 것이 아니다.

큰 효과크기를 지니지만 통계적으로 유의하지 않은 결과는 표본수가 작아서 그 효과를 밝혀내기 위한 검증력(power)이 작은 경우이므로, 통계적 유의도에 기초하여 전혀 효과가 없다고 결론내리기 전에 표본의 수를 늘려서 다시 연구해볼 필요가 있다. Rosenthal과 Rosnow(1991)는 효과크기는 존재하지만 너무 작은 표본수를 사용하여 통계적으로 유의한 결과를 얻지 못하는 경우를 물체는 존재하지만 불빛이 너무 어두워서 찾아내지 못하는 것으로 비유하였다. 반면에 통계적으로는 유의한 결과이지만 효과가 아주 작은 경우에는 어떻게 할 것인가? 이때는 모집단에서 추정되는 작은 효과의 실제

적 중요성을 평가해보아야 할 것이다. 매우 작은 효과라도 그러한 발견의 파급효과가(예를 들어, 새로운 약의 개발로 인명을 건질 수 있다면) 실제적으로 중요하게 여겨질 수 있다. 그러나 효과크기가 작음에도 불구하고 많은 표본을 사용하여 통계적으로는 유의한 결과를 얻었지만 실제적 중요성은 매우 낮다면 그 결과는 과학의 발전에 그다지 기여하지 못할 것이다.

앞으로의 연구에서 통계적 유의도와 함께 관련강도를 보고할 필요가 있음을 논의하였지만 끝으로 관련강도가 지니는 제한점도 주목할 필요가 있다. 첫째, 원래 상관 연구의 맥락에서 개발된 관련강도의 개념을 변량분석을 사용하는 연구에 적용하는데서 발생할 수 있는 문제점이다. 변량분석을 이용하는 연구설계는 독립변인의 몇가지 처치수준의 효과를 알아보는데 변인의 처치수준의 선택이 관련강도 추정치에 큰 영향을 미칠 수 있다(Keppel, 1982). 예를 들어, 양적으로 큰 차이가 나도록 독립변인의 처치수준들을 선택하였다면 독립변인과 종속측정치간에 큰 관련강도를 얻게 될 가능성이 큰 반면, 차이가 적은 수준들을 선택하였다면 작은 관련강도를 얻게 될 가능성이 크다. 둘째, 큰 값의 관련강도가 과학으로서의 심리학의 발전에 반드시 크게 기여하는 것이라고 말할 수는 없다. 즉, 관련강도의 크기가 실제적 또는 이론적 중요성에 관한 정보를 직접적으로 전달해 주는 것은 아니다. 셋째, 관련강도 값은 모집단의 관련강도에 관한 추정치이기 때문에 표집오차에 의해 영향을 받을 수 있다(Maxwell, Camp, & Arvey, 1981). Carroll과 Nordholm(1975)은 몬테카를로 접근법을 이용하여 표본수가 상대적으로 작을 때 관련강도 추정의 표준오차가 크다는 것을 보였다. 마지막으로, 관련강도는 처치효과 크기를 종속측정치의 단위로써 나타내 주는 것이 아니다. 따라서 관련강도는 연구자가 진정으로 알고자 하는 정보를 제공할 수도 있고 그렇지 않을 수도 있다

(Maxwell, Camp, & Arvey, 1981). 어떤 연구문제에서는 연구자가 단순히 처치에 따른 집단들 간의 종속 측정치 평균차이에만 관심을 가질 수 있다.

요약컨대, 앞으로의 연구에서 통계적 유의도를 나타내는 p값에 의해서만 결과를 해석할 것이 아니라 관련강도도 함께 고려할 필요가 있다. 관련강도가 통계적 유의도 검증이 제공해주지 못하는 유용한 정보를 주지만, p값에만 의존하는 것이 바람직하지 않은 것처럼 관련강도에만 의존하여 결과를 해석하는 것도 바람직하지 않다. 앞으로 심리학 학술지에 논문을 제출할 경우에 p값과 더불어 연구목적에 적절한 관련강도 지수를 선택하여 보고하고 결과의 해석에서 그것의 실제적 중요성을 논의할 필요가 있다. 그 결과, 가까운 장래에 심리학 연구에서 또다른 발전이 기대될 수 있을 것이다.

## 참고 문헌

Austin, J. T. (1989). Effects of shifts in goal origin on goal acceptance and attainment. *Organizational Behavior and Human Decision Processes*, 44, 415-435.

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, 66, 423-437.

Benedict, M. E., & Levine, E. L. (1988). Delay and distortion: Tacit influences on performance appraisal effectiveness. *Journal of Applied Psychology*, 73, 507-514.

Binning, J. F., Goldstein, M. A., Garcia, M. F., & Scattaregia, J. H. (1988). Effects of preinterview impressions on questioning strategies in same- and

opposite-sex employment interviews. *Journal of Applied Psychology*, 73, 30-37.

Bretz, R. D., Ash, R. A., Dreher, G. F. (1989). Do people make the place? an examination of the attraction-selection-attrition hypotheses. *Personnel Psychology*, 42, 561-581.

Carnevale, P. J. D., & Conlon, D. E. (1988). Time pressure and strategic choice in mediation. *Organizational Behavior and Human Decision Processes*, 42, 111-133.

Carroll, R. M., & Nordholm, L. A. (1975). Sampling characteristics of Kelly's 2 and Hays's 2. *Educational and Psychological Measurement*, 35, 541-554.

Cohen, J. (1965). Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of clinical psychology* (pp.95-121). New York: McGraw-Hill.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45, 1304-1312.

Cook, T. D., Campbell, D. T., & Peracchio, L. (1990). Quasi experimentation. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (2nd ed.). Palo Alto: Consulting Psychologists Press.

Dunnette, M. D. (1966). Fads, fashions, and folderol in psychology. *American Psychologist*, 21, 343-352.

- Gist, M. E., Rosen, B., & Schwoerer, C. (1988). The influence of training method and trainee age on the acquisition of computer skills. *Personnel Psychology*, 41, 255-265.
- Gist, M. E. (1989). The influence of training method on self-efficacy and idea generation among managers. *Personnel Psychology*, 42, 787-805.
- Gist, M. E., Schwoerer, C., & Rosen, B. (1989). Effects of alternative training methods on self-efficacy and performance in computer software training. *Journal of Applied Psychology*, 74, 884-891.
- Gordon, R. A., Rozelle, R. M., & Baxter, J. C. (1988). The effect of applicant age, job level, and accountability on the evaluation of job applicants. *Organizational Behavior and Human Decision Processes*, 41, 20-33.
- Harcum, E. R. (1989). The highly inappropriate calibrations of statistical significance. *American Psychologist*, 44, 964.
- Hays, W. L. (1973). *Statistics for the social sciences* (2nd ed.). NY: Holt, Rinehart & Winston.
- Hirst, M. K. (1988). Intrinsic motivation as influenced by task interdependence and goal setting. *Journal of Applied Psychology*, 73, 96-101.
- Howell, J. M., & Frost, P. J. (1989). A laboratory study of charismatic leadership. *Organizational Behavior and Human Decision Processes*, 43, 243-269.
- Hunt, R. G., Krzystofiak, F. J., Meindl, J. R., & Yousry, A. M. (1989). Cognitive style and decision making. *Organizational Behavior and Human Decision Processes*, 44, 436-453.
- Keppel, G. (1982). *Design & Analysis: A researcher's Handbook* (2nd ed.). NJ: Prentice-Hall.
- Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, 70, 151-159.
- Maxwell, S. E., Camp, C. J., & Arvey, R. D. (1981). Measures of strength of association: A comparative examination. *Journal of Applied Psychology*, 66, 525-534.
- Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data*. CA, Belmont: Wadworth Publishing Co.
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Clinical and Counseling Psychology*, 46, 806-834.
- Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1, 108-141.
- Northcraft, G. B., & Earley, P. C. (1989). Technology, credibility, and feedback

- use. *Organizational Behavior and Human Decision Processes*, 44, 83-96.
- Paquette, L., & Kida, T. (1988). The effect of decision strategy and task complexity on decision performance. *Organizational Behavior and Human Decision Processes*, 41, 128-142.
- Phillips, J. S., & Freedman, S. M. (1988). The task-related competency and compliance aspects of goal setting: A clarification. *Organizational Behavior and Human Decision Processes*, 41, 34-49.
- Podsakoff, P. M., & Farh, J. (1989). Effects of feedback sign and credibility on goal setting and task performance. *Organizational Behavior and Human Decision Processes*, 44, 45-67.
- Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). McGraw-Hill Publishing Co.
- Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, 44, 1276-1284.
- Schoorman, F. D. (1988). Escalation bias in performance appraisals: An unintended consequence of supervisor participation in hiring decisions. *Journal of Applied Psychology*, 73, 58-62.
- Scott, W. E., Farh, J., & Podsakoff, P. M. (1988). The effects of "intrinsic" and "extrinsic" reinforcement contingencies on task behavior. *Organizational Behavior and Human Decision Processes*, 41, 405-425.
- Steiner, D. D., & Rain, J. S. (1989). Immediate and delayed primacy and recency effects in performance evaluation. *Journal of Applied Psychology*, 74, 136-142.
- Weldon, E., & Mustan, E. L. (1988). Felt dispensability in groups of coactors: The effects of shared responsibility and explicit anonymity on cognitive effort. *Organizational Behavior and Human Decision Processes*, 41, 330-351.
- Wright, W.F., & Aboul-Ezz, M.E. (1988). Effects of extrinsic incentives on the quality of frequency assessments. *Organizational Behavior and Human Decision Processes*, 41, 143-152.

# The Need of Reporting Measures of Association Strength in Empirical Studies: Is The P value Sacred?

Tae-Yong Yoo

Department of Industrial Psychology  
Kwangwoon University

Kathy A. Hanisch

Department of Psychology  
Iowa State University

The p value as an index of "important", "meaningful", and "scientific" results has long been used in empirical studies. This study suggested the need of reporting measures of association strength in addition to statistical significance(p value) in order to enhance understanding of study results. Different measures of association strength and their relations to each other were presented using a hypothetical example. The twenty-one articles using ANOVA in three refereed journals(Journal of Applied Psychology, Personnel Psychology, and Organizational Behavior and Human Decision Processes) published in 1988 and 1989 were reviewed to examine the number of articles reporting measures of association strength in addition to statistical significance. In these articles, the fashion of reporting and interpreting results was also examined. It was found that eleven articles among them did not report the measures of association strength and still interpreted the results by only the p value. Continued heavy reliance on the p value, the misinterpretation of the p value, the need of reporting the measures of association strength in the future studies, and the need of considering practical importance of study result were discussed.