

古典的 메타분석에 대한 補完的 接近方法의 探索¹⁾

오 성 삼
(건국대학교)

《 요 약 》

메타분석이 연구 방법으로 정착되기 시작한 1950년 초 이래로 50여 년의 세월을 거치면서 이 방법을 적용한 수많은 연구들이 쏟아져 나왔다. 사회과학뿐만 아니라 자연과학의 특정 주제들이 메타분석의 대상이 되었고, 미국을 중심으로 세계 여러 나라들에서 석·박사학위 논문들이 해마다 수적 증가를 보이고 있다. 근자에 이르러 국내에서도 메타분석 방법을 적용한 학위 논문들이 점진적 증가 추세를 보이고 있지만 메타분석 방법에 대한 일부 학자들의 비평 또한 만만치 않다. 초기 단계에서의 메타분석 방법이 통계적 절차와 방법상에 취약점을 지니고 있다는 지적이다. 본 연구는 메타분석 방법이 지니고 있는 일부 학자들의 문제점 제기에 대한 방어 논리와 고전적 메타분석 방법이 지니고 있는 취약점을 보완하기 위한 방법의 탐색을 목적으로 실시되었다. 보완적 접근방법으로는 역변량 가중치(inverse weight)와 통합표준편차(pooled standard deviation) 값을 사용하여 효과크기를 구하는 방식을 적용한 것이다. 뿐만 아니라 고전적 방법에서 지적되고 있는 효과크기의 동질성 검정, 표준오차와 신뢰구간 산출 등을 동시에 실시함으로써 메타분석 절차에 대한 신뢰성 제고와 결과에 관한 타당성 확보에 주력하였다.

주제어 : 메타분석, 연구방법

I . 서 론

오늘날 컴퓨터의 보급과 활용 그리고 정보통신의 발달은 국내외를 막론하고 세계 도처에 소장되어 있는 학술적인 자료들을 한곳에 모아 데이터 베이스의 구축을 가능케 하였다. 이처럼 선행 연구물들의 손쉬운 검색에 힘입어 메타분석으로 일컬어지는 선행 연구 결과물들의 통합분석 기법이 의학, 농학, 심리학 등 다양한 학문 분야에서 활성화되게 되었다. 오성

1) 본 논문은 2002년도 건국대학교 학술진흥연구비에 의해 연구되었음

삼(1992)은 유명한 과학사가 Derek J. de Solla Price의 말을 인용하여 1800년을 기준으로 세계 과학 잡지의 수가 매 50년마다 10배씩의 증가 현상을 보이고 있음을 소개한바 있다. 세계 도처에서 쏟아져 나오는 각종 출판문헌과 연구결과들은 새로운 지식 정보의 홍보 속에 살아가고 있음을 의미한다. 그러나 세계 도처의 연구소나 학술단체들로부터 폭증하는 연구 결과물들은 경우에 따라 저마다 상이한 발견점들을 내놓음으로써 어떠한 결론이 보다 타당하며 보다 신뢰로운 것인가에 대한 혼란을 초래하기도 한다. 이와 같은 상황 속에서 이제 지식을 접하는 사람들은 그 수많은 단편적 연구의 발견점들을 한데 통합하여 보다 객관적이며 신뢰롭고 강력한 결론을 도출해 내는 연구의 필요성을 요구하기에 이른 것이다. Hedges와 Olkin(1982)의 조사에 의하면 선행 연구의 결과들을 하나로 통합하는 방법은 이미 1930년대 농업분야에서 사용되기 시작했다고 한다. 그러나 정작 방법론으로 정착하게 된 것은 1952년도에 Eysenck(1952)의 임상심리학 분야에서였다. 그는 심리치료가 환자들에게 별다른 도움이 되지 못한다는 문제를 제기했다. 이를 계기로 1950년대 중반에서 1970년대에 이르기까지 심리치료 효과에 관한 수많은 논문들이 쏟아져 나오게 되었다. 그런데 그 결과는 긍정적인 것도 있었고 부정적인 결과들도 있었으며 또 어떤 연구 결과들은 심리치료 효과에 대해 판단을 내릴 수 없다는 결과도 나왔다. 이와 같은 상황에서 Eysenck가 제기한 논쟁에 대한 결론을 유도하기 위한 방안의 하나로 Glass는 375편의 심리치료 효과에 관한 선행연구 결과들을 수집하였고, 통계집단과 실험집단의 평균차에 대한 표준화 점수(standardized score)를 산출해냄으로써 이른바 ‘메타분석(meta-analysis)’란 새로운 연구의 통계적 방법을 개발하게 되었다. 이제는 고전적 방법이 되어버렸지만 Glass와 그의 동료들이었던 Marry, Lee, Smith 등에 의해 개발된 메타분석방법은 오늘날에 이르러 심리학, 사회학, 교육학, 그리고 의학, 경영학 분야 등에서 널리 쓰이게 되었다. 1976년에 제시된 Glass의 메타분석 방법의 가장 중심적인 기여는 선행연구결과들의 통합에 효과크기(effect size)란 개념의 도입이었는데, 이를 바탕으로 Hedges와 Olkin(1985)은 효과크기의 개념을 도입한 메타분석을 위한 통계적 방법(Statistical Methods for Meta-Analysis)을 저술하게 되었다. 이처럼 효과크기에 대한 새로운 개념의 통합 연구 방법의 출현으로 인해 기존의 유의도(p-value) 통합방법이나 투표식(vote count) 통합방법이 주춤해지는 경향을 보이게 되었다. 그러나 과거 50 여 년에 걸친 메타분석 방법에 대해 많은 사람들의 비판이 제기되었고 메타분석이 지닌 취약점을 극복하기 위한 노력들 또한 꾸준히 발전되어 왔다. 그럼에도 불구하고 아직도 메타분석을 시도하는 많은 연구자들이 기존의 메타분석이 지닌 취약점을 극복하기 위한 보완적 방법에 대한 이해 부족으로 고전적 방법 혹은 부분적인 취약점에 대한 보완 수준에 머물러 있음을 보게된다. 이와 관련하여 본 연구는 ①메타분석을 시도하려는 초기 이용자들에게 메타분석에 관한 올바른 이해를 제공하고, ②메타분석에 관한 비판적 지적 사항들에 관한 대응논리를 제시하고, ③통계적 적용과 관련하여 기존 방식에 대한 보완적 접근 방법을 소개하기 위한 취지와 목적을 가지고 시도되

있다. 따라서 본 연구에서 다루게 될 주요 내용은 ①메타분석에 관한 기존의 비판들을 종합적으로 정리하고 주요 현안들에 대한 각각의 대항논리를 제시하고자 하며, ②기존의 메타분석 통계와 관련하여 이 방법이 지닌 통계적 취약점을 보완하기 위한 방법의 제시, 그리고 ③아직 해결을 위한 방안이 정착되고 있지 않은 질적 연구결과들에 대한 자료처리 방안의 필요성과 대안제시에 본 연구의 주안점을 두고자 한다.

II. 메타분석에 관한 기존의 비판과 대항논리

일반적으로 메타분석에 관한 Eysenck(1978)와 Glass(1982)의 비판을 소개하면 다음과 같다. 메타분석 방법과 관련해서 가장 빈번하게 지적되어 온 문제는 동질성에 관련된 내용이다. 이와 관련한 학자들의 지적은 메타분석이 ‘사과와 오렌지들(apples and oranges)’을 합해 놓고 거기에서 통합된 결론을 추출하는 것은 비논리적이라는 비판이다. 이 같은 주장에 내재된 생각의 전제는 어떤 측면에서 “동일한” 연구들만을 통합할 수 있다는 것이다. 이러한 주장은 피상적으로 보기에선 일면 타당한 것 같아 보인다. 그러나 이 주장의 근거가 되는 동일하다는 것이 무엇을 의미하는지도 분명치 않을 뿐 아니라 만약 연구가 모든 측면에서 동일하다면 그것을 비교한다는 것 자체가 모순이다. 연구주제, 연구의 특징, 관련변인, 연구절차, 연구결과들이 “동일”하다면 비교할 필요가 없는 일이다. 역설적으로 표현하면 비교하고 통합해야 할 연구들은 “이질적”인 연구, 어떤 측면으로든지 “상이한(different)” 연구들이다. 메타분석의 강점과 특징은 상이한 연구들을 통합할 수 있는 이론과 공통된 수단을 제공함으로써 같은 기초 위에서 결론을 추출하고 해석할 수 있는 길을 모색해 주는 데 있다. 연구의 궁극적 목적은 일반화 가능성의 힘을 갖자는 데 있다. 만약 한 개의 연구결과에 의해 이 같은 일반화를 주장할 수 있는 개연성이 있다면, 여러 개의 많은 수의 연구결과를 통합한 메타분석의 결과를 기초로 주장하는 일반화는 더 강력한 힘을 가질 것은 자명하다. 만약 단일한 연구 내에서 이 같은 상이한 효과를 볼 수 있다면 연구들간(between studies)에서도 이러한 검증은 가능하다고 보아야 한다. 메타분석은 바로 연구의 특징에 따라 효과의 크기를 달리 추정하려는 것이 특색이다. 이것을 이질적인 것들의 잡탕이라고 비판하는 것은 잘못된 해석에 기인하거나 설득력 없는 주장일 뿐이다. 오히려 메타분석의 장점과 특징이 바로 여기에 있다고 해야 할 것이다(황정규, 1988).

둘째의 비판은 연구의 질에 관한 것이다. 메타분석은 질적으로 빈약한 연구를 질적으로 우수한 연구와 혼합함으로써 결론의 추출에 편파성을 초래할 가능성이 크다는 점을 지적하고 있다. 이 비판에는 두 가지 문제가 내포되어 있다. 첫째는 무엇이 “빈약한 연구”인가라는

문제이며, 둘째는 이 같은 연구의 질에 따라 효과의 크기에 차이가 실제 의미 있게 나타나느냐는 문제이다. “빈약한 연구”라는 개념은 대단히 주관적 판단일 가능성이 높다. 흔히 어떤 연구자가 빈약한 연구라고 판단하여 연구 리뷰에서 제외하는 것은 연구의 질의 본질적 특성에 관련되어 있는 경우보다 연구자 자신의 주관적 판단, 특히 연구자 자신이 상정하거나 의도하고 있는 연구방향이나 결론의 방향과 배치되는 경우에 “빈약한 연구”, “질이 낮은 연구”라는 판단을 내리는 경우가 많다. 또한, 연구의 질이 빈약하다, 우수하다는 판단은 대개 연구의 내적 타당도에 관련해서 평가하기 마련이다. 그러나 현실적으로 어떤 단일한 연구가 내적 타당도의 다양한 기준들을 완벽하게 만족시킬 수는 없는 일이다. 어떤 연구이건 몇 개의 기준에서는 빈약하기 마련이다. 메타분석의 또 하나의 장점은 이같이 부분적으로 빈약한 연구설계를 가지고 연구한 결과들을 통합함으로써 한쪽 연구가 가진 연구의 질의 빈약성을 다른 쪽 연구가 보상할 수 있도록 한다는 데 있다. Glass, McGaw & Smith(1981)는 20개의 메타분석 연구를 재분석하여, 연구의 질을 상, 중, 하로 나눈 뒤, 질의 수준에 따라 효과의 크기가 크게 차이가 나는지를 비교해 보았다. 그 결과 질의 수준에 따라 약간의 차이는 있으나 거의 일관되게 나타나고 있는 것을 제시한바있다. 이러한 결과는, 특히 연구의 질에 따라 실제 효과의 크기 추정치가 크게 차이가 날 것이라는 상식적 기대와 일치하지 않음을 알 수 있다. Rosenthal(1984)은 만약 연구의 질과 효과의 크기 사이에 어떤 관계가 존재할 것으로 상정되면, 통합과정에서 연구의 질에 따른 가중치 적용도 고려해 봄직하다는 대안을 제시하고 있다.

셋째, 연구의 독립성(independence)에 관한 비판이다. 메타분석에 이용되는 연구는 한 개의 연구에서 여러 개의 효과크기를 산출하고 이것이 메타분석을 하기 위한 각 유독에 통합될 때마다 마치 독립된 연구에서 얻은 효과의 크기처럼 다루고 있다. 실제 한 개의 연구에서 같은 피험자를 대상으로 한 효과의 크기는 서로 독립성이 보장되지 않으며, 엄밀한 의미에서 한 개의 자유도로 보아야 타당하다. 그러나 메타분석에서는 같은 연구에서 산출된 효과의 크기에 각각 1개의 자유도(df)를 부여함으로써 비독립적 자료를 독립적 자료인 것처럼 다루고 있다는데 문제가 있다는 지적이다. 그러나 이 문제는 각각의 효과의 크기가 같은 연구에서 나온 경우라도 효과의 크기 자체를 연구의 단위로 보는 것이 메타분석이기 때문에 크게 우려할 바는 못된다.

넷째, 유의도 수준에 따른 오류 가능성에 대한 비판이다. 대개의 연구에서 유의도 수준을 표시할 때 $p < .01$ ($z > 2.33$)인 경우 대개 “ $p = .01$ 에서 유의함”이라고 보고하는 관례를 따르고 있다. 앞에서 논의한 바처럼 연구보고 자체의 정보가 이같이 생략되어 있는 경우는 어쩔 수 없는 일이지만 $p < .01$ 로 나와있는 유의도를 $p = .01$ 로 결정하고 메타분석을 하는 경우도 있다. 그러나 이러한 연구들이 계속 누적되는 경우 추리에 심각한 오류를 발생시킬 수 있다. 또한, 메타분석에서 이용되는 “연구의 수가 증가하면 할수록” 영가설을 부정할 확률도 점점 커진

다. 물론 영가설이 정말 참인 경우에는 연구 수의 증가가 별다른 영향을 미치지 않지만, 그렇지 않을 경우에는 영가설을 부정할 확률을 증가시킬 개연성이 커지는 것은 사실이다. 메타분석에서 장차 해결해야 할 과제의 하나는 이 같은 문제를 통제할 수 있는 방법론을 개발하는 것이다. 그러나 연구의 수가 증가해도 효과의 크기 추정치를 증가시키지는 않는다는 점이며, 이 같은 점에서 확률의 가산법 보다는 표준점수 평균 차에 의한 추정법이 보다 신뢰할 수 있는 방법으로 평가할 수 있다.

다섯째, 정보의 상실에 관한 비판이다. 한 개의 연구를 평균효과의 크기와 같은 단일한 값으로 요약한다는 것은 그 연구가 지니고 있는 여러 가지 다양한 정보를 너무 단순화시킴으로서 많은 정보를 상실해 버린다는 비판이 제기될 수 있다. 이 비판은 실제 많은 점에서 타당하다. 그러나 동시에 메타분석은 많은 정보 중 가장 핵심적인 추정치를 토대로 성립한다는 방법론적 특징을 수락한다면, 우려하는 바처럼 그렇게 심각한 문제가 아닐 수도 있다. 마치 단일한 연구를 n , \bar{X} , σ 라는 대단히 단순한 추정치로 정리해 내듯 여러 개의 연구들을 평균 효과의 크기로 변환해 내었을 때 그 속에는 메타분석자가 원하는 정보의 정수가 농축되어 있다고 해석할 수도 있다.

이 외에도 Hedges & Stook(1983)은 메타분석에 대해 두 가지 문제점을 지적하고 있다. 하나는 통합된 연구들이 자료의 질과 자료의 이질성(heterogeneity) 문제를 들고 있으며, 또 하나는 메타분석에 대한 체계적인 통계이론과 형식적인 통계모형의 부재를 지적하고 있다. 이와 같은 지적에 대해 Hedges 와 그의 동료들은 각각의 연구를 통합하기 위한 통계모형을 제시하고 효과크기의 동질성 검증을 실시함으로써 이러한 비판을 일축하고 있다(Hedges, 1981, Hedges & Stock, 1983).

Ⅲ. 고전적 메타분석 방법의 보완

앞서 제기된 기존 메타분석 방법의 취약점 내지는 비판적 시각에 대한 보완적 방안으로 본 연구는 ①효과크기의 가중치 적용, ②효과크기들의 동질성 검정, ③이질분포의 효과크기, ④평균효과크기의 표준오차와 신뢰구간 산출, ⑤논문표집의 안전계수 확보, 그리고 ⑥메타분석에서 신뢰도와 타당도 확보 등에 관한 보완적 접근 방법을 제시하고자 한다.

1. 효과크기의 가중치 적용

메타분석에서 분석의 대상이 되는 각종 선행연구들은 각기 다른 수의 표집 대상자들을

지니고 있게 마련이다. 어떤 연구들은 대상자들의 표집수가 적어서 그 집단을 대표하기에 미흡한 경우도 있고 반면 다른 연구들에서는 많은 대상자를 갖고 있을 뿐만 아니라 너무나 훌륭하게 무작위 통제집단 설계를 사용한 경우를 보게 된다. 이처럼 다양한 형태의 연구들을 같은 비중으로 놓고 통합분석을 시도한다는 것은 메타분석을 통해 추출해 내려는 대표값을 추정하는데 문제가 따른다. 이를테면, 100명으로부터 얻어진 효과크기는 10명을 대상으로 얻어진 효과크기 보다는 보수적 값을 추정하는데 훨씬 정확한 값을 지니게 된다. 따라서 사례수가 큰 연구 결과는 사례수가 작은 연구에 비해 보다 큰 가중치를 부여해 주어야 한다는 논리가 적용된다. 이처럼 가중치를 적용하는 방법에는 사례 수를 가중치로 하여 효과크기를 산출하는 단순한 방법과 표준오차의 역수를 가중치로 사용하는 방법이 있는데 후자의 경우가 보다 안정된 가중치를 제공하는 방법으로 평가되고 있다. 표준오차는 효과크기의 정확성을 추정하는 하나의 지수임을 우리는 알고 있다. 그래서 표준오차의 값을 가지고 신뢰구간을 설정하게 된다. 결국 보다 적은 표준오차를 지닐수록 보다 정확한 효과크기를 얻게 되는 것이다. Hedges는 표준오차의 값을 사용하여 메타분석에서의 적절한 가중치를 산출해 내는 <공식 1>을 제안하고 있다.

$$w = \frac{1}{SE^2} \cdots \text{<공식 1>}$$

그러나 효과크기가 어떤 방법에 의해 산출된 것인가에 따라 역변량 가중치를 추정하는 공식이 달리 사용되게 된다.

1) 표준화 평균차에 따른 효과크기에서의 역변량 가중치 추정공식

$$se = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{ES_{sm}^2}{2(n_1 + n_2)}} \cdots \text{<공식 1-1>}$$

$$w = \frac{1}{se^2}$$

2) 표준화 상관계수(Zr)로 변환된 효과크기의 역변량 가중치 추정공식

$$se = \sqrt{\frac{1}{n-3}} \cdots \text{<공식 1-2>}$$

$$w = n-3$$

3) 로그화된 승산비 효과크기의 역변량가중치 추정공식

$$se = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \text{ <공식 1-3>}$$

$$w = \frac{1}{se^2}$$

여기에서 a, b, c, 그리고 d는 2x2 분할표에서의 각 셀에 대한 빈도수를 나타내는데, 단순한 효과크기의 값에 비해 가중치를 적용한 효과크기의 값을 통해 얻어진 평균효과 크기는 모수의 값을 추정하는데 훨씬 정확한 값을 지니게 된다.

2. 효과크기들의 동질성 검정

Glass(1976)에 의해 메타분석 방법이 소개된 이후 메타분석에 관한 여러 가지 문제점들이 지적되고 있는데 그 가운데 하나가 다른 추리통계 방법에서와 마찬가지로 파인 수집분석 대상이 되는 각각의 연구결과들의 효과크기들이 동일 모집단으로부터 추출된 나온 값인가 하는 의문점이다. 다시 말해서 각각의 효과크기들에 대한 모집단 동질성(동변량성)의 전제조건을 충족시키고 있는가 하는 점인데 분석 대상이 되는 효과크기들이 모집단의 효과크기를 잘 나타내 주고 있는가의 여부를 위해 효과크기의 동질성 검사(test for the homogeneity of effect size)의 실시를 권장하고 있다(Hedges & Olkin, 1983). 이때 효과크기의 동질성 검증을 위한 통계적 귀무가설을 세울 수 있다.

$$H_0: \delta_1 = \delta_2 = \dots \delta_m$$

이 경우, 대안가설은 최소한 δ_i 는 나머지 다른 효과크기의 변량과 차이가 있을 것이다. 따라서 만약에 이들 효과크기의 분산 추정치가 동일한 모집단으로부터 추출되었다는 가설(H_0)이 입증되고 대안가설이 부정된 경우)이 증명되면 메타분석에서 얻어진 연구의 결과는 추리통계에 선행되는 조건을 충족한 것으로 볼 수 있다는 것이다. 동질성 검정의 통계 Q의 값을 구하기 위해 다음 <공식 2>를 적용하면 된다.

$$Q = \sum(w \times ES^2) - \frac{[\sum(w \times ES)]^2}{\sum w} \text{ <공식 2>}$$

얻어진 동질성 검정 통계 Q의 값에 대한 해석은 Q값은 카이제곱 분포와 같다. 따라서

구해진 Q값이 주어진 자유도(df) 하에서 쿼무가설을 기각하지 못하면 효과크기들은 동질적인 것으로 간주되게 된다. 즉, 효과크기들의 분산이 표집오차를 초과하지 않고 있음을 의미한다.

3. 이질분포의 효과크기

앞서 제시한 동질성 분석이란 모든 효과크기들이 동일한 모집단 평균값을 추정하고 있는가의 여부를 검정하는 일이었다. 그러나 동질성 검정이 실패할 경우 효과크기들의 분포가 이질적이라는 추론이 가능하여 산출된 평균 효과크기의 값이 집단분포를 대표하는 값으로서 적절치 못하게 된다. 이와 같은 상황 하에서의 메타분석을 위해 몇 가지 대안적 방법을 생각해볼 수 있다.

첫째는 연구마다 모수가 임의로 변한다고 가정하는 랜덤효과 모형(Random Effects Model)을 생각할 수 있다. 이 모형이 전제하는 가정은 변량의 발생이 분석대상 개체들의 표집오차에서 비롯된 것이란 출발을 하고 있다. 메타분석의 대상이 되는 각각의 연구들이 하나의 모집단으로부터 랜덤 하게 추출된 표본연구라는 가정 하에 산출된 결과를 모집단으로 일반화시킬 수 있다는 점이다. 서로 다른 연구들의 이질성을 감안하고 있는 랜덤효과 모형에서는 표집오차 이외에 또 다른 랜덤오차 요인을 분석모형에 포함시켜야 한다는 점이다. 즉, 각각의 효과크기들이 표집오차에 의해 모집단 평균 효과크기로부터 차이를 지니게 되어 효과크기들에 존재하는 분산은 두 영역으로 구성되어 있다고 보는데, 그 하나가 효과크기들의 표본오차이고 다른 하나는 랜덤효과 변동인 것이다. 여기에서 이들 두 변동의 합을 v_i^* 로 표시하면 총 분산은 다음과 같이 표기 할 수 있다.

$$v_i^* = v_{\theta} + v_i$$

여기서 v_{θ} 는 랜덤 혹은 연구간(between-studies) 변동의 추정을 의미하며, v_i 는 측정대상의 표본오차 변동을 의미한다. 여기서 평균 효과크기, 신뢰구간, 유의도 검정, 그리고 동질성 검정 통계량 Q 등의 계산에서는 단지 표본오차만을 고려하였던 고정효과 모형에서의 계산된 값들로부터 새로 포함시킨 랜덤효과 변동의 값들을 빼주어야 한다. 따라서 기존의 역변량 가중치 $1/v_i$ 는 $1/v_i^*$ 의 값을 가지고 다시 계산되게 된다.

$$v_{\theta} = \frac{Q_T - k - 1}{\sum w - \left(\frac{\sum w^2}{\sum w} \right)} \dots \text{<공식 3>}$$

여기에서 Q 는 동질성 검정 통계량을, k 는 효과크기 관측치의 수를, w_i 는 각각의 효과크기에 대한 역변량 가중치를 의미한다. 만약 위의 공식에서 마이너스(-) 값이 나올 경우가 있는데 이는 곧 효과크기들의 분포가 동질적인 경우를 의미하게 되기 때문에 v_0 의 값을 0으로 놓고 계산하면 된다.

둘째 접근방식은 앞에서와 같이 효과크기의 이질적인 분포(연구논문들의 차)가 관찰이 불가능한 랜덤성으로부터 유래된 것이라는 가정보다는 코딩과 관련한 연구논문들의 특성과 중재변수 등 분산의 출처에 대한 규명이 가능하다는 가정 하에 접근하는 방식이다. 이처럼 연구논문의 독립변수로 인해 발생하는 분산에 초점을 맞추어 분석을 시도하는 것을 모수효과모형(fixed effects model)이라 한다. 효과크기 값들의 분포가 이질적인 분포를 이루고 있을 경우 이를 해결하기 위한 방법으로 표본대상의 표본오차분산이 체계적일 것이라는 가정을 추가하여 접근하는 방식은 분석대상이 되는 각각의 효과크기들의 이질성은 연구논문들이 지니고 있는 특성들로 인해 효과크기 값이 영향을 받은 것으로 추정하는 것이다. 따라서 이 모수효과 모형에 따른 분석을 시도할 때에는 효과크기 분산을 분할하는 방식, 즉 분산분석(ANOVA) 방식의 적용방법과 아울러 회귀분석(Regression Analysis) 방식을 적용하는 방법을 생각할 수 있다. 분산분석의 접근방식은 주로 Hedges(1982a)가 제안한 방법으로 독립변수가 범주형 변수(categorical variable)인 경우가 되겠고, 회귀분석 방법은 독립변수가 연속형(continuous) 혹은 이산형(dichotomous)인 경우에 적용되는데, Olkin's(1985)가 가중치를 적용한 중다회귀분석 방법을 변형하여 적용한 데에서 비롯된 방법으로 알려져 있다. 이 방법은 한 연구논문에 여러 개의 독립변수가 존재하는 논문들에 대한 메타분석에 적합하다.

세 번째의 접근방식은 초과 변량치의 근원이 일부는 체계적인 것이고 일부는 랜덤효과인 것이라는 추정 하에 혼합효과모형(mixed effects model)의 적용을 생각할 수 있다. 이 모형은 매우 복잡한 방법인 것 같으나 사실은 매우 단순한 논리를 지니고 있다. 즉, 분산분석(ANOVA)의 접근논리와 수정된 가중치를 적용한 회귀분석(Regression Analysis)의 논리를 혼합하여 접근하는 모형인 것이다. 분석방법은 중재변수를 고려한 후에 랜덤 변량 부분을 추정하게 되는데 SPSS 프로그램을 통해 메타분석에서 혼합효과 모형을 위한 ANOVA와 REGRESSION 프로그램에 별도의 "/model(mm)"선택사항을 두어 랜덤효과 변량부분을 포함시키면 된다. 그리고 최대우도추정(maximum likelihood estimate)을 위해서는 "/model(ml)"을 선택하면 된다. 이 프로그램은 랜덤효과 변동 부분의 추정을 위해서 랜덤효과 변동을 더한 역변량 가중치를 재계산하여 모형에 접합한 추정을 위해 설계된 것이다.

4. 평균효과크기의 표준오차와 신뢰구간의 산출

일반적으로 통계추정을 통해 우리가 얻게되는 값들에는 항상 표본오차(sampling error)가 존재하기 마련이다. 여기서 의미하는 표본오차란 모수값과 통계값과의 차이를 말하는데, 평균인 경우에는 $\mu - \bar{Y}$, 분산인 경우에는 $\delta^2 - S^2$, 비율인 경우에는 $\pi - P$ 등 모수값에 대응하는 통계값의 표집오차를 고려할 수 있다. 따라서 추정량의 전 표집분포를 고려하여 주어진 추정량에 포함될 가능한 오차를 추정하게 된다. 구간추정(interval estimation)은 주어진 추정량의 표준 오차를 고려하여 모수값이 포함될 추정값의 구간을 확률적으로 진술하게 된다. 평균효과크기의 신뢰구간을 설정하기 위해서는 <공식 4>, <공식 5>, 그리고 <공식 6> 등을 통해 필요한 값들을 산출하여야 한다.

$$\text{평균효과크기 } \overline{ES} = \frac{\sum(w \times ES)}{\sum w} \dots \dots \text{<공식 4>}$$

$$\text{평균효과크기의 표준오차 } se_{\overline{ES}} = \sqrt{\frac{1}{\sum w}} \dots \dots \text{<공식 5>}$$

$$\text{평균효과크기의 Z-검증 } Z = \frac{\overline{ES}}{se_{\overline{ES}}} \dots \dots \text{<공식 6>}$$

이상에서 구해진 값들을 가지고 모집단 평균의 신뢰구간은 <공식 7>을 통해 얻을 수 있다.

$$P(\bar{X} - Z \cdot \frac{\alpha}{2} \cdot \delta \bar{X} \leq \mu \leq \bar{X} + Z \cdot \frac{\alpha}{2} \cdot \delta \bar{X}) = 1 - \alpha \text{ <공식 7>}$$

위의 <공식 7>은 $1 - \alpha$ 의 신뢰도를 갖는 μ 에 대한 신뢰구간을 나타낸 것이다. 신뢰도는 이와 같이 구간으로 추정된 추정치가 실제 모집단의 모수를 포함하고 있을 가능성을 말하는 것이다. 신뢰구간을 구하는 공식과 관련하여 앞서 제시한 <공식 7>은 두 가지 가정을 전제로 성립되게 되는데 첫 번째 가정은 모집단의 표준편차를 알고 있다는 것이고, 또 다른 가정은 모집단이 정규분포를 형성하고 있다는 것이다. 따라서 이 두 가지 가정이 모두 충족될 때 표본의 크기에 관계없이 신뢰구간의 추정 공식은 성립되게 된다.

5. 논문표집의 안전계수 확보

연구자가 어느 한 분야의 선행연구결과들을 종합할 때, 그 분야에서 이루어진 모든 연구

를 망라하지 못하고 일부 연구 결과들만 가지고 종합한다면, 조사연구에서 표집의 대표성이 문제되듯, 종합하고자 표집한 연구들이 그 분야를 대표할 만한 연구들인가 하는 문제가 야기될 수 있다. 사실 어느 한 분야의 연구결과를 종합할 때 선행된 연구들을 모두 수집하기란 매우 어려운 일이다. 논문집이나 기타 적절한 매개체를 통하여 발표한 연구라 할지라도 연구결과를 종합하는 사람이 적기에 입수하지 못하여 빠뜨릴 수도 있다. 혹은 본래의 연구자가 연구를 끝내고서도 어떤 사정에 의하여 그 결과를 보고하지 않고 책상서랍 한구석에 버려 두었기 때문에 연구결과와 종합에서 제외되는 수도 있다. Rosenthal(1980)은 이것을 책상서랍의 문제(file drawer problem)라고 명명하고 연구결과와 종합과정에서 이러한 문제를 고려해야 된다고 주장한다. 연구를 마치고서도 그 결과를 보고하지 않은 것들 중에는 대개 영가설이 부정되지 않은 연구결과들일 가능성 높다. 그런데 어느 한 분야의 선행연구를 종합할 때 이처럼 영가설이 부정되지 않은 다수의 연구결과들을 제외시키고 종합한다면 그 종합결과는 해당 분야의 연구결과를 잘 대표한다고 보기 어려울 것이다. 만약 종합한 연구결과가 매우 높은 수준의 유의도를 나타내는 것이지만 여기에 소수의 유의하지 않은 연구들(null results)을 추가시키면 높은 유의수준을 보이던 종합결과가 유의하지 않은 것으로(예컨대, $p > .05$) 반복된다고 하자. 그렇다면 유의한 것으로 밝혀진 그 종합결과는 불안정한 상태에 놓여 있다고 할 것이며, 이럴 경우에 ‘책상서랍의 문제’는 큰 위협을 주는 요인이 될 수 있다. 안전계수(fail-safe number)란 바로 이와 같은 문제와 관련된 것으로서 유의하게 나타난 종합결과를 유의하지 않은 것으로 반복시키는데 요구되는 연구 총수(새로운 연구나 혹은 숨겨진 연구들으로써 전체적으로 합했을 때 영가설을 부정하지 못하는, 다시 말해서 Z점수의 합이 0이 되는 연구결과들)를 뜻한다. 안전계수는 앞에서 기술한 통합 유의도를 내기 위하여 표준점수 Z를 합산하는 공식을 이용하면 계산해 낼 수 있다.

$$N_{fs,.05} = \frac{\sum Z^2}{1.645} - N \cdots \text{<공식 8>}$$

위에 언급된 <공식 8>는 종합결과의 확률이 $p \geq .05$ 로 되려면 몇 편의 유의하지 않은 연구결과들이 추가되어야 하는가, 즉, 5%를 기준으로 했을 때 안전계수를 계산하는 일반공식이다. 물론 안전계수의 계산은 통합확률이 어느 일정한 유의수준, 예컨대, 5% 또는 1%와 같은 수준에 도달할 경우에 한하여 의미를 갖게 된다. 현재로서는 어느 정도의 계수가 ‘안전’하다고 여길 수 있는가에 대한 분명한 준거가 서있지 못하다. 각 분야에 따라 연구결과를 종합할 때 찾지 못하고 숨겨진 연구의 수가 제각기 다를 수 있기 때문이다. 다만 대략적인 기준으로서 종합한 연구의 수를 N이라고 할 때, 안전계수가 $5N+10$ 이상이면 대체로 안전하다고 본다. 이러한 생각의 배경에는 발견 못한 의의 없는 연구의 수가 종합결과에 반영된

연구수의 5배를 넘지는 않을 것이며, 단 1편의 연구결과인 경우에 최소 안전계수를 15정도로 하던 적당하리라고 가정한 것이다. 앞서도 기술한 바와 같이 메타분석을 위해 수집되어진 연구의 자료들이 대부분 기 발표된 것들에 의존하게 되는데 이처럼 기존에 발표되어진 연구의 결과들 이외에 메타분석을 시도하는 연구자가 미처 생각지 못했던 연구들이나 미 발표논문의 결과들을 수집하는데 실패함으로써 메타분석의 연구결과를 그릇되게 하는 경우를 책상서랍문제라고 한다. 따라서 메타분석에서 사용되어진 연구의 대상이 그 주제의 모집단을 대표할 만한 충분한 수집이 이루어지지 못한 경우 메타분석에서 얻어진 연구의 결과를 해석함에 매우 신중을 기해야 할 것임을 Strube & Hartmann(1983)은 강조하고 있다. Kraemer & Andrews(1982)에 따르면, 발표된 연구들은 긍정적인 결론 쪽으로 편포되는 경향이 있다고 한다. 이것은 실재보다 더욱 더 긍정적인 결론 속에 Type 1의 출판편견오류(publication bias error)를 띄게 된다.

6. 메타분석에서 타당도와 신뢰도 확보 문제

수많은 연구들이 메타분석과 관련한 신뢰도 문제에 관해 언급하고 있다(예를 들면 Glass 등, 1981 ; Green & Hall, 1984 ; Hunter 등, 1982 ; McGuire 등, 1985 ; Orwin & Cordray, 1985 ; Rosenthal, 1984 ; Stock 등, 1982). 다른 형태의 연구에서와 마찬가지로 메타분석을 할 때도 신뢰도를 고려할 필요가 있다. 여기서 의미하는 신뢰도란 주로 내적 합치도 혹은 일관성을 의미하는데 몇몇 사람들이 연구팀이 되어 분석대상이 되는 연구 결과물들을 각자 나름대로 코딩작업을 할 때 혹은 분석과정에서 포함되어 있는 연구들의 양상들을 부호화할 때 생기는 신뢰도가 주된 문제로 제기되곤 한다. Stock(1982)과 Rosenthal(1984)은 서로 다른 연구의 특성들을 부호화할 때 얻어지는 코딩작업자 상호간의 합치도 문제를 해결하기 위한 몇 가지 방안을 제시하고 있다. 즉, ①메타분석의 특성을 부호화하기 전에 예비검사(pilot-test)를 실시할 것 ②메타분석의 코딩작업 이전에 매우 구체적이고 명확한 개념을 지닌 코드책자(code book)를 개발할 것 ③코드책자와 코딩형태에 입각한 코딩작업자들에 대한 훈련의 실시 ④메타분석의 일부분으로서 코딩작업자 상호간의 신뢰도의 측정과 기록 ⑤필요한 경우 코드책자와 코딩형태를 수정하고 코딩작업자들을 재훈련 실시 ⑥새로운 코딩작업자들이 추가될 경우 그들에게 알려야 할 사항들을 숙지시킬 것 ⑦코딩규칙에 관한 사항을 결정할 때 코딩작업 종사자들의 적극적인 참여 권장을 제시하고 있다. 또 다른 신뢰도 문제는 계산과정에서 유의수준이나 연구의 효과크기를 잘못 계산하게 되거나 부호를 잘못 코딩하는 문제라 할 수 있는데 위에 제시된 내용은 계산과 코딩 상에서 일어나는 오류의 가능성을 최소화하기 위한 지침인 것이다. Hunter 등(1982)은 메타분석에서 측정상 발생하는 수많은 연구들의 표본오차를 교정하기 위한 방법들을 마련했고, Wortman(1983)은 메타분석의 타당도와 관련하여 외적

타당도, 구인 타당도, 내적 타당도 등 통계적 결론들에 관한 문제에 대해 언급하고 있다. 외적 타당도와 구인 타당도는 두 가지 모두가 통합된 연구들이 과연 동일한 속성을 지니고 있는가 하는 "사과와 오렌지"의 논리와 관련이 있다. 따라서 연구의 특성을 코딩하고 교정된 효과크기를 산출해 내고 동일성을 검정하는 것은 내적 타당도를 높이고 검토하는 데에 크나 큰 도움이 된다. Glass(1980)는 높은 타당도를 가진 실험연구와 낮은 타당도를 가진 실험연구의 평균 효과크기 사이에 표준편차 차이는 1/10을 넘지 않는다고 보고하고 있다. 그럼에도 불구하고 그는 각각의 메타분석이 내적 타당도를 조사하는 것이 바람직하다는 주장을 하고 있다. Green & Hall(1984)은 실험자의 무지의 정도, 무작위, 표본크기, 이기 잘못이나 속임수의 통제, 종속변수의 형태(예를 들면 자기 보고서 내 객관식)와 출판의 편파(bias) 등의 문제가 메타분석이 검토해야 할 과제들임을 지적하고 있다.

V. 향후 메타분석을 위한 제언

이제까지 본 연구는 기존의 고전적 메타분석 방법이 지닌 취약점들을 보완하기 위한 몇 가지 접근방법들을 제시해 보았다. 이 글을 맺으면서 향후 메타분석방법이 추구해야할 과제와 방향은 질적 연구결과(The Result of Qualitative Research)를 계량적 결과와 함께 메타분석에 통합화 할 수 있는 방법의 모색이란 생각이 든다. 통합연구로써 메타분석의 가치는 어느 선정된 주제에 관해 보다 광범위한 연구의 결과들을 한데 모음으로 해서 거기서 얻어진 결론을 통해 보다 신뢰롭고 타당한 연구의 결과를 도출해 내는데 근본 취지가 있음은 말할 나위가 없을 것이다. 그래서 메타분석의 성패 여부는 연구자가 그가 알아보고자 하는 주제에 관하여 얼마나 많은 연구결과를 수집했는가에 좌우된다고 할 수 있다. 이미 Rosenthal(1979)에 의해 지적된 'file drawer problem'이 아니더라도 출판 혹은 미 출판된 연구의 결과들이 수량적 접근을 시도한 것들뿐만 아니라 상당한 수의 논문들이 질적 방법에 의한 문제의 결론을 도출하고 있음을 간과해서는 안될 것이다. 정확히 꼬집어 사회과학 분야의 연구물들 가운데 얼마만큼이 수량적 방법을 사용한 것이고 어느 만큼이 질적 방법을 사용한 연구물들인가를 그 분량을 이야기할 수는 없겠으나 어느 한 쪽도 무시하고 넘어가도 좋을 만큼의 가벼운 비중은 아닐 것이다. 연구하고자 하는 주제의 선정이 끝나고 그 분야의 연구논문들을 수집하고 나서 각 편의 논문들을 리뷰하는 과정에서 연구대상에서 제외되는 대부분의 논문들이 메타분석에서 요구하고 있는 기본요건, 즉, 실험집단과 통제집단을 사용해야 하고 메타분석에서 요구하는 적절한 통계방법을 사용한 자료의 요건을 충족시키지 못하기 때문이다. 수많은 노력과 시간 그리고 경비를 들여 수집한 자료들이 수량적 방법에 의한 자료들이 아니라는 이유만으로 메타분석 대상에서 제외되어야 한다면 종합연구가 의도하는 본래의 효과를 절감시키는 결과를 초래하고 마는 것이다. 분야에 따라서 연구주제에 따라 애써 수집한

연구물들이 단지 수량화한 자료가 아니라는 이유로 인해 분석에서 제외되었다고 한다면 그 같은 종합연구는 제외된 분량만큼의 신뢰성과 타당성을 상실함은 물론 메타분석의 의미 또한 효과가 극소화시키는 결과를 초래하는 것이다. 따라서 앞으로의 효과적이고 의미 있는 종합화 연구방법을 위해서는 수량적 연구뿐만 아니라 질적 연구의 결과들도 함께 분석의 대상이 될 수 있도록 앞으로의 이 분야에 관한 연구기법의 개발이 요구된다. 이와 관련하여 오성삼(1991)은 언어적 표현의 표준편차 환산을 통한 질적연구 결과들에 대한 계량화를 제시한바 있다. 다시 말해서 종전에 수량화되지 않은 연구 결과라는 이유 때문에 메타분석 대상에서 제외되었던 연구 결과들을 언어적 표현의 표준편차 점수화를 통해 동일주제의 연구 결과들을 수량적 비수량적 연구를 함께 활용하는 접근 방법을 제시하고 있다. 그의 접근 방법은 공식 $X_{ij} = C_i S_j + K$ 를 응용하여 "Very"와 같은 부사형 어휘가 "Possible"과 같은 형용사 어휘와 함께 사용되므로써 곱셈의 효과(multiplicative effect)를 가져온다는 사실에 근거한 것인데 여기서 X_{ij} = i번째 부사형 어휘가 j번째 형용사와 결합하여 획득된 점수를 말하며 C_i 는 i번째의 부사형 어휘의 곱셈값, S_j = j번째 형용사의 심리적 척도상의 위치, K = 심리적 척도상의 0점과 획득된 척도 값의 인위적 영점값의 차이를 의미한다. 이를 통해 종전에 널리 사용되던 투표(voting)방법이나 부호(Sign)방법 등이 효과의 방향을 알아보는 방법에는 유용한 기법이었으나 효과의 크기를 표시하지 못한다는 제한점을 극복할 수 있는 장점을 지니게 된다는 점이다.

현재의 단계로는 질적연구의 결과들을 메타분석이 요구하는 요건을 구비하지 못하고 하여 기계적으로 분석대상에서 제외시키기보다는 수량적 연구물들의 통합을 통해 얻어진 현상들을 설명하는데 그리고 수량적 자료의 결론을 뒷받침하는데 설득력 있고 합리적인 이블테면 동일한 연구주제에 관해 집합된 자료들로서 충분히 활용되어질 때에 비로소 진정한 의미에서의 통합연구의 의도를 실현시킬 수 있을 것이다. 뿐만 아니라 Light와 Pillmer(1984)를 중심으로 한 일련의 학자들에 의해 질적 연구 결과들을 수량화하는 방법에 대한 연구가 시도되고 있고 Slavin(1986, 1995)의 소위 최적증거통합방법(best evidence synthesis²⁾)은 동일주제 하에서의 수량적 연구 결과와 질적인 연구 결과를 동시에 통합하는 방법을 제시하고 있어 향후 고전적 메타분석에 대한 보완적 접근 방법으로서의 사용을 제언하고자 한다.

2) Best-Evidence Synthesis란 메타분석과 전통적인 리뷰방법의 대안으로 제시된 통합연구방법으로 계량적 효과크기와 계량화가 불가능한 질적인 연구들로부터 연구를 통해 밝혀진 사실들을 함께 제시함으로써 내적타당도와 외적 타당도를 극대화 시킬수 있는 기존의 통합연구의 대안으로 Slavin에 의해 제시된 방법이다. 따라서 이 방법은 기존 선행 연구결과들의 통합과정에서 극단적으로 높거나 낮은 효과 크기들과 연구 결과의 특이성에 대한 논문들에 대해 단지 계량적 효과크기의 제시만으로 끝나는 것이 아니라 그와 같은 결과가 어떤 상황하에서 무슨 연유로 나왔는가 하는 점들을 서술형태로 밝혀주게 된다. 결과적으로 Best-Evidence Synthesis란 메타분석의 장점에다 전통적인 문헌 리뷰방법을 접목시킨 형태의 통합 연구방법이라 할 수 있다. 보다 자세한 내용을 원하면 Slavin, R. E. (1986). Best-Evidence Synthesis: An Alternative to Meta-Analytic and Traditional Reviews, Educational Researcher, 15, vol 9. Nov. 5-11. 참조바람.

참 고 문 헌

- 김억환, 오성삼, 구병두(1990). 학업성취와 관련된 가정환경변인, 학습풍토 변인, 교사변인에 대한 통합분석. 「건국대학교 교육연구소 논문집」 제14집.
- 송혜암(2001). 「의학, 간호학, 사회과학 연구의 메타분석법」. 청문각.
- 오성삼(2002). 「메타분석의 이론과 실제」. 건국대학교 출판부.
- 오성삼(1991). 비수량적 연구결과들의 통합분석 방법의 탐색. 「교육평가연구」 제 4권 제 1호. 한국교육평가학회.
- 오성삼, 구병두(1990). 학업성취관련변인에 관한 종합연구, 「건국대학교 대학원 논문집」 (제30집).
- 오성삼(1992). Meta 분석 기법의 활용동향과 컴퓨터 프로그램을 이용한 Meta 분석기법. 「교육평가연구」 제 5권 제 2호, 한국교육평가학회
- 오성삼(2002). 「선행연구결과들의 통합과 재분석을 위한 메타분석의 이론과 실제」. 건국대학교 출판부.
- 이지훈(1993). 「사회과학의 메타분석방법론」. 충북대학교 출판부.
- 황정규(1988). 「Meta-Analysis의 이론과 방법론(경험과학적 연구결과의 종합을 위하여)」. 성곡학술재단.
- Brewer, J. K. (1983). Meta-analysis and the planing of future studies. Florida Educational Research Journal. 5, 27-38.
- Class, G. McGAW, & Smith, M. L.(1981) Meta-Analysis in Social Research. Beverly Hills, CA: Sage.
- Cook, T. D., & Livition, L. C.(1980). Reviewing the literature: A comparison of tradition methods with meta-analysis. Journal of Persinality. 48. 449-472.
- Eysenck, H. J.(1978). An exercrise in mega-silliness. American Psychologist 33: 517.
- Fiske, D. W.(1983). The meta-analytic revolution in outcome research. Journal of Consulting and Clinical Psychology. 51. 65-70.
- Glass, G. (1976). Primary, secondary, and meta-analysis in social research.. Beterly Hills. CA: Sage.
- Glass, G. V. (1980). Summarizing effect sizes. In R. Rosenthal(Ed.), New Directions for Methodology of Social and Behavioral Science: Quantitative Assessment of Research Domains. San Francisco: Jossey-Bass.
- Glass, G. V. (1982). Meta-analysis: An approach to the synthesis of research results, Journal of Research in Science Teaching, 19. 93-112.
- Green, B., & Hall, J.(1984). Quantitative methods for literature review. Annual Review of

- Psychology, 35, 37-53.
- Hedges, L. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6: 107-128.
- Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, 92, 490-499.
- Hedges, L. & Olkin, I.(1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hedges, L. & Stock, W.(1983). The Effects of Class Size: An Examination of Rival Hypotheses. *American Educational Research Journal* 20, 63-85.
- Hunter, J. E., Schmidt, F. L. & Jackson, G. B.(1982). *Meta-analysis: Accumulating research findings across studies*. Beverly Hills, CA: Sage.
- Hunter, J. E., & Schmidt, F. L.(1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Kraemer, H. C., & Andrews, G.(1982). A non-parametric technique for meta-analysis effect size calculation. *Psychological Bulletin* 91, 404-412.
- Krauth, J. (1983). Non-parametric effect size estimation: A comment on Kraemer and Andrews. *Psychological Bulletin* 94, 190-192.
- Light, R. H., & D. B. Pillemer(1984). *Summing Up: The Science of Reviewing Research*. Cambridge, MA: Harvard University Press.
- Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, 3, 354-379.
- Pillemer, D. B., & Light, R. J.(1980). Synthesizing outcomes : How to use research evidence from many studies. *Harvard Educational Review*, 50, 176-195.
- Rosenthal, R.(1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Sharpe, D.(1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review*, 17, 881-901.
- Slavin, R. E.(1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, 15, 5-11.
- Slavin, R. E.(1995). Best evidence synthesis: An intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology*, 48, 9-18.
- Strube, M. J., & Hartmann, D. P.(1982). A critical appraisal of meta-analysis. *British Journal of Clinical Psychology*, 21, 129-139.
- Walberg, H. J., & Sung Sam Oh(1985). Effective schools: A quantitative synthesis of constructs. *The Journal of Classroom Interaction*, Vol.20, No. 2, Summer.
- Wolf, F. M.(1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage Publications.

• 논문접수 : 2003년 10월 4일 / 수정본 접수 : 2003년 11월 21일 / 게재 승인 : 2003년 12월 4일

ABSTRACT

A Repletion Approach for the Traditional Meta Analysis

Sung-Sam Oh

(Professor, Kon Kuk University)

Any statistical procedure or approach can be misused or abused. Most of the criticisms of quantitative approaches were objections to the misuse or abuse, real or potential, of meta-analysis. A number of problems in conducting a quantitative review have been pointed out. Meta-analysts have developed a number of ways to analytically approach many of these problems merit the thoughtful attention of the quantitative reviewer. But, some of these issues are not analytical but conceptual in nature.

Regarding this issue, this article focused on as a repletion study or make up study for the weak point of traditional meta analysis technique. As the study of repletion for the traditional meta-analysis, summarizing of the strengths and the weaknesses of meta -analysis were introduced to deal with analysis issues and strategies in this paper.

This study as a repletion approach for the traditional meta analysis focused how to deal with the effect size adjustments, analysis of heterogeneous distributions of effect size, analysis of statistically dependent effect sizes, confidence interval, and homogeneity test, analysis of heterogeneous distributions of effect size, weighted regression analysis and additional analysis issues such as validity and reliability on meta-analysis.

Finally, quantifying methods for the qualitative research results were introduced.

Key Words : meta-analysis