

교육학 연구방법으로서 메타분석(Meta-analysis)의 발전과정 고찰

장 덕 호(상명대학교 조교수)*

신 인 수(전주대학교 조교수)**

《 요 약 》

이 연구는 교육학 연구방법론으로서 메타분석(meta-analysis)의 발전과정을 고찰함으로써, 기존 문헌연구의 한계 및 메타분석의 의의를 이해하고, 메타분석을 통한 교육학연구의 학문적 발전을 위한 시사점을 찾고자 하였다. 이를 위해 우선 국내·외 메타분석연구의 증가 빈도를 시기별로 살펴보고, 통시적으로 메타분석이 발전된 과정을 연구결과의 종합화라는 메타분석의 본질적 특성 측면에서 고찰하였다. 우선 Glass가 메타분석이라는 용어를 사용한 1976년 이전의 개념적 발전기에서 출발하여 이후 연구결과를 종합하려는 학자들의 노력들을 조망해 보고, 또한 1976년 이후 메타분석에 대한 비판과 그러한 비판을 극복하는 과정을 살펴보았다. 아울러 1980년대의 다양한 교재들의 출현, 1990년대의 증거기반운동(evidence-based movement) 및 현재의 국제적인 네트워크 형성과 주요한 방법론적인 이슈들을 탐색하였다. 이상의 역사적 고찰을 바탕으로 앞으로 교육학 연구에서 메타분석의 발전을 위한 시사점을 몇 가지 제시해 보았다. 첫째, 메타분석과 유사 연구방법과의 개념적 차별화가 필요하고, 둘째, 기존 연구를 양적으로 통합하고 질적으로 종합하는(synthesizing qualitative research) 노력을 통하여 학술적으로도 검증된 지식을 축적하는 노력이 필요하며, 셋째, 경험적 교육학에서 많이 쓰이는 회귀분석 연구들을 종합하는 문제, 내재적 자료(nested data) 구조를 가진 연구물을 종합하는 문제, 메타분석에서의 독립성 가정 위반, 구조방정식 모형을 활용한 연구물들을 종합하여 이론 모형을 형성하려는 접근(model based meta-analysis, Becker 2009) 등 충분한 방법론적 발전이 부가되어야 함을 제시하였다.

주제어 : 메타분석, 역사적 발전과정, 문헌 연구, 효과 크기

* 제1저자

** 교신저자, s9065031@jj.ac.kr

I. 서론

본격적인 정보화 시대의 도래와 함께 학문 연구에서도 데이터베이스를 통한 각종 문헌과 연구 결과물들에 대한 검색의 용이함은 학문연구방법에 새로운 접근을 가져오고 있는데, 그 중 하나가 바로 메타분석(meta-analysis)이라고 할 수 있다(오성삼, 1995, 2009). Gene Glass(1976)가 메타분석을 “연구결과들을 통합하려는 목적으로 하는, 다수의 개별연구들의 분석결과물의 집합체에 대한 통계적 분석”이라고 명명한 이후 메타분석은 기본적으로 개별연구들의 연구성과를 종합하고 하는 통계적 방법론의 하나로 자리매김해 왔다. 학문은 기본적으로 협동적이고, 상호 의존적 특성을 가진다고 볼 때, 일 개인의 연구는 학문이 답하고자 하는 퍼즐판에 겨우 한 조각을 맞추는데 기여할 수 있을 뿐이다(Cooper, 2010). 메타분석은 세계 도처에서 쏟아지고 있는 이러한 단편적인 연구결과들을 놓고 상반된 주장을 펼치는 상황이 자주 목격되고, 방대한 연구 결과들의 양 또한 엄청난 증가 추세를 나타내고 있는 상황에서 보다 객관적이고, 신뢰로운 결론을 도출할 필요성에서 개발되었다고 할 수 있다. 특히, 경험과학적 연구자료를 다루는 교육학 및 사회과학 분야에서는 특정 연구주제, 연구질문을 위한 연구물들이 누적되어 가는데, 이들 연구들을 개별적으로 보면, 각각 다른 연구대상에 대해서 각각 다른 상황에서 연구가 이루어지고 있으며, 각각 서로 다른 연구결과를 제시하는 일이 자주 벌어지는 것이다. 이렇게 상이한 상황과 맥락 속에서 수행된 다양한 연구들을 하나의 일관되고 체계적인 틀 속에서 통합하여 분석해 봄으로써 연구결과의 누적을 단순화시킬 수 있는 경제적 연구방법이 바로 메타분석이라고 할 수 있다(황정규, 1988). 요컨대 기존 연구물들의 결과가 하나의 방향으로 수렴되지 않고, 서로 상반되어 하나로 통합되지 않을 때 기존 연구물을 연구대상으로 하여 통계적으로 종합하는 연구통합(research synthesis)의 기법이라고 할 수 있는 것이다(오성삼, 1995; 이희숙, 2011).

보다 구체적으로 메타분석은 다음의 상황에서 그 유용성을 확인할 수 있다. 첫째, 단일주제에 관한 두 개의 상반된 결론이나 논쟁이 야기되었을 때 이의 해결을 위한 보다 신뢰롭고 타당성 있는 대결론(big decision)을 내려야 할 필요성이 있을 때 효과적이다. 둘째, 현장으로부터 원자료(raw data)를 수집할만한 시간적 여유가 없거나, 경비와 노동력의 절약이 요구되는 상황에서의 2차 자료를 이용한 문제의 해결을 요할 때 메타분석방법이 유용하다. 셋째, 각종 학술정보의 범람 속에서 체계적으로 압축된 지식 또는 정보들을 필요로 할 때 효과적인 방법이라고 할 수 있다(오성삼, 1995).

메타분석 이전 단계의 전통적으로 연구통합의 방법으로는 화술적 종합법(narrative syntheses)과 투표식 방법(vote counting)이 있어 왔다. 화술적 종합법은 연구결과들을 연도순으로 이야기체로 기술해 놓은 방법으로 문헌 리뷰를 통하여 일련의 연구들 가운데 의미 있고, 연관성있는 정보들을 추출해나가는 기법이다. 이러한 문헌연구에서는 연구주제와 관련되면서도

쟁점으로 부각되는 연구결과들을 연구자가 주관적으로 나열하는 것이 일반적이다. 그러나 이는 대상 연구물 선택에서부터 분석과정에 이르기까지 그 과정이 대단히 주관적이며, 기존 연구결과를 과학적으로 이해하는데 많은 한계를 지닌다. 이러한 기법은 ‘관심이 되고 있는 변수들 상호간의 관계가 크기가 얼마인가?’ ‘처치(intervention)를 통해 대상이 얼마만큼 변화할 것인가?’ 또는 ‘이러한 관계 또는 처치가 다른 관계 또는 처치보다 얼마나 효과적인가?’ 등에 관해 답을 할 수 없다는 문제가 있다(Cooper, 2010). 즉, 문헌 연구 결과 변수들간의 관계에 대해 방향(direction) 및 크기(amount)에 대해 종합적인 결론을 내릴 수 없다는 데에 결정적인 문제가 있는 것이다. 투표식 방법은 유의수준을 이용하여 기존연구를 세 가지로 분류하여, 통계적으로 유의한 연구 중 연구결과 중 관심변수의 효과가 정적인 연구와 부적인 연구, 그리고 유의하지 않은 연구로 구분하여, 세 가지 유목 중 기존 연구결과들이 가장 많은 유목을 계산하여 그 결과를 종합하는 방법이다. 그러나 이 방식은 애써 얻은 유용한 기술적 정보를 상실하게 된다는 점, 연구결과의 종합의 과정에 표집의 크기를 도외시함으로써 사실과 다른 결론에 도달하거나 경우에 따라서는 적은 수의 표집에서 나온 결과와 많은 수의 표집의 사용한 연구결과를 동등하게 평가함으로써 실제와 정반대의 결론에 도달할 가능성이 있다는 점 등의 문제가 있다(이종승, 1983; 오성삼, 1995).

기존의 화술적 종합법 및 투표식 방법과 메타분석의 결정적인 차이점은 효과크기(effect size)를 이용해서 기존 개별연구결과들을 종합하여 관심 변수의 효과에 대해 그 방향(direction)과 크기(amount)를 알려준다는 데 있다. 예를 들면 학급당 인원수 감축이 학업성취증진에 효과가 있는지가 연구주제라면, 메타분석을 통해 개별 연구들을 종합해서 효과가 긍정적인지 부정적인지 방향을 알려주고, 효과가 있다면 얼마만큼 크기인지를 알 수 있다.

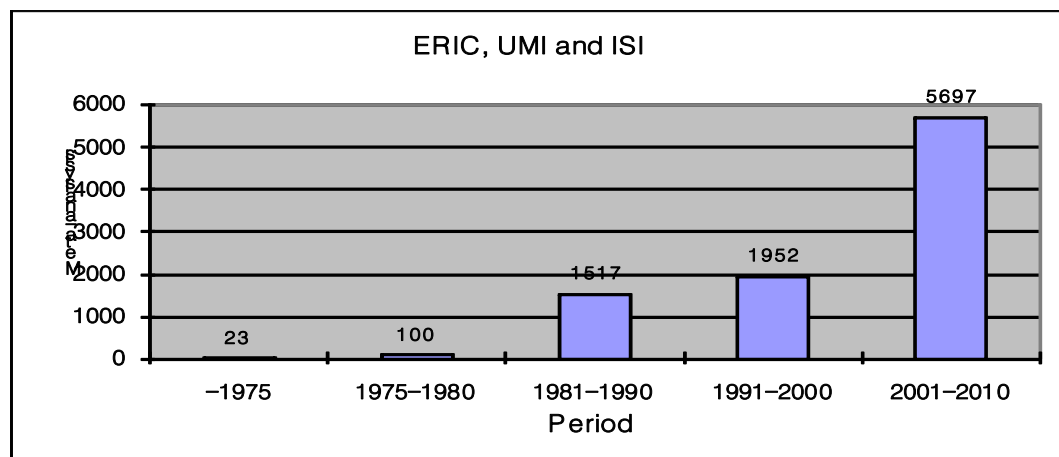
이러한 메타분석은 무수한 검증과 방법론적인 발전과정을 거쳐 현재는 교육학을 포함한 사회과학분야의 주요한 연구방법 중 하나로 널리 활용되고 있다. 그러나 북미권을 중심으로 메타분석이 교육학연구에 활발하게 적용되고 있음에도 불구하고, 이쉽게도 우리나라에서는 교육심리학 등 일부 전공분야를 제외하고는 교육학 분야에서의 메타분석에 대한 연구가 부족한 실정이다. 또한 효과크기에 대한 분석 없이 ‘메타분석적 연구’ ‘메타적 연구’ 등 각종의 이름으로 연구물들이 발표되고 있어서, 개념상의 혼란도 발생하고 있으며(안민호·고혜승, 2006; 황용석, 2006 등), 메타분석에 대한 역사적 고찰을 포함한 일반적 이론 연구는 교육학분야에서는 극히 제한적이다. 이러한 점에서 본 연구는 다음과 같이 두 가지 연구문제를 설정하였다. 첫째, 국내 및 국외 메타분석 연구들의 경향은 어떠하며, 메타연구의 역사적 발전과정에서 얻을 수 있는 새로운 해석은 무엇인가? 둘째, 메타분석의 발전을 위한 학술적 과제와 연구의 발전을 위한 시사점은 무엇인가? 이러한 질문에 대한 답을 위해 본 연구는 메타분석 발전 초기에 발생하였던 오해와 비판에서부터 계량적·통계적 분석 방법이 정착된 오늘날 메타분석의 주요한 쟁점들에 대한 연대기적 고찰을 시도하고자 한다. 이를 통해 교육학 연구의 한 방법론으로서 메타분석의 가치와

앞으로의 연구방향과 과제를 탐색하는데 도움을 줄 것이고, 특히, 메타분석의 주요한 이슈들을 종합적으로 개관해봄으로써 메타분석에 관한 학술적 이해의 지평을 넓히는데 일조하고자 한다.

Ⅱ. 메타분석연구의 국내외적 경향과 기존연구고찰

1. 메타분석연구의 양적 팽창

해외의 메타분석연구는 양적인 측면에서 비약적인 성장을 거듭하고 있다. 메타분석에 기반을 둔 연구물의 발전과정을 살펴보기 위해 교육학분야의 대표적인 데이터베이스인 ERIC, UMI, 그리고 사회과학분야 인용색인 ISI web of science의 데이터베이스에서 연구물 검색 기능을 이용하여 연구물 제목(title)에 ‘meta-analysis’ 혹은 ‘research synthesis’(1976년 이전시기에만 해당)를 입력하여 검색해본 결과, 1976년 이전에는 사회과학분야에서 23편의 연구물이 research synthesis와 관련해서 검색되었고, Glass가 1976년에 meta-analysis라고 명명한 이래, 1976년부터 1980년까지는 100편, 1980년대는 1,517편, 1990년대는 약 1,952여 편, 그리고 2001년 이후 2010년까지 사회과학분야에서 5,697편의 연구물이 검색되었다. [그림 1]은 이러한 메타분석연구의 비약적인 양적 팽창 과정을 보여주고 있다.

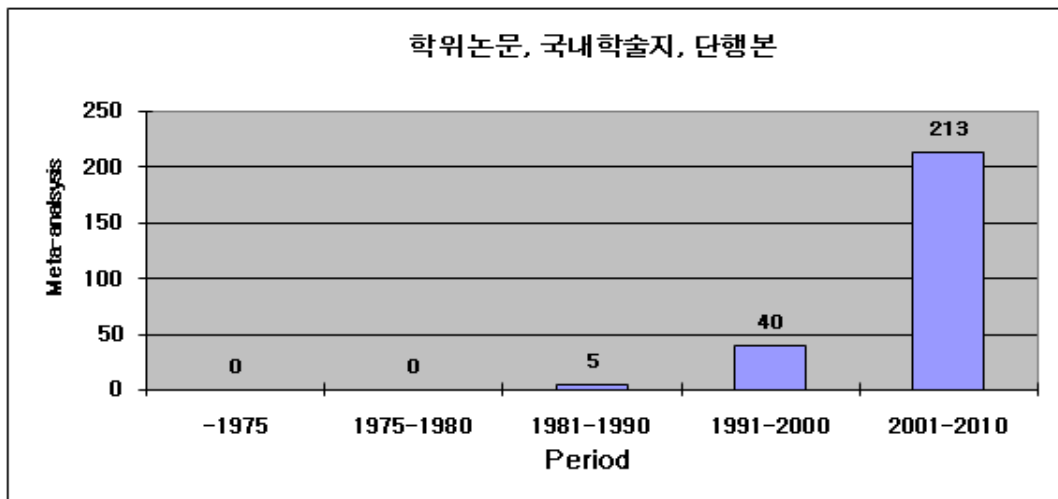


[그림 1] 메타분석연구의 양적 팽창

이상의 빈도수는 대표적인 교육학 및 사회과학분야의 데이터베이스 3개만을 대상으로 조사한 것이기 때문에 교육학 및 사회과학분야의 모든 메타분석연구의 빈도수라고 할 수 없고, 데이터

베이스 상호 간 연구물의 중복이 있을 수도 있지만, 시대 별로 메타분석연구의 양적 팽창에 대한 경향성을 파악하기는 충분하다고 판단된다. 연구자에 따라서는 이러한 메타분석의 양적 팽창에 따른 질적 저하를 염려하기도 하지만, 이러한 경향성은 메타분석이 교육학 및 사회과학 방법론으로서 정착되고, 비약적인 발전이 이루어지고 있음을 보여준다.

외국의 메타분석연구의 양적 팽창과는 달리, 국내 교육학 및 사회과학연구에서는 메타분석 연구가 그리 활발하지 못하다. 한국교육학술정보원 데이터베이스(RISS)에 검색어로 제목(title)에 '메타분석'을 입력하면 총 690건(학위논문 289, 학술지 372, 단행본 29)이 검색되는데, 여기서 메타분석연구물인지 아닌지, 그리고 교육학 및 사회과학 분야 연구인지를 제목과 초록을 분석·검토한 후 다시 정리해 보니, 단행본 6건, 학위논문 117건, 국내학술지 135건 등 총 258건이 교육학 및 사회과학 분야 연구물로 정리되었다. 물론, 연구물의 제목과 초록만을 검색하여 개별 메타분석 연구를 일일이 찾아내지 못할 개연성은 있으나, 개괄적으로 양적인 연구 성과의 경향성을 알아보는 데는 충분한 정보를 제공하고 있다.



[그림 2] 국내 메타분석연구의 양적 팽창: 교육학 및 사회과학

한국교육학술정보원에서 검색한 총 135편의 메타분석을 기반으로 한 국내학술지 연구물 중, 교육학 관련 메타분석 연구물은 총 101편으로 메타분석을 소개한 논문과 메타분석 방법을 적용한 논문으로 구분할 수 있다. 메타분석 연구방법을 적용한 논문이 92편, 메타분석을 소개한 논문 9편(이종승, 1982, 1983; 황정규, 1988; 송인섭, 1990; 오성삼, 1992, 1995, 2003; 박세훈, 오승희, 2003; 조지민, 2002)이 발표되었다. 이를 다시 교육학 세부 전공별로 92편을 구분해 보면, 교과교육학 22편, 교육과정 1편, 교육공학 및 방법 10편, 교육사회 및 평생교육 3편, 상담 13편, 교육심리 23편, 유아 및 아동 6편, 특수교육 10편, 교육평가 1편, 교육행정 3편 등이다.

이 중 메타분석을 적용한 논문 92편 중 2000년대 이후에 발표된 논문이 80편으로 학술지에 발
표된 논문의 대부분이 2000년대 이후에 발표된 것임을 알 수 있다.

국내 메타분석연구가 미진한 이유는 메타분석의 기초가 되는 양적(quantitative) 연구 방법
론을 이용해 데이터를 분석한 연구물이 부족한 것이 근본적 원인일 수 있고, 부가적으로는 메타
분석 방법론에 대한 이해의 부족 등도 한 원인일 수 있다.

2. 메타분석의 역사적 발전과정에 관한 연구동향 고찰

기존의 역사적 발전과정을 고찰한 논문들을 살펴보고, 이 연구의 차이점을 정리해 보면 다음
과 같다.

Becker(2007)는 다변량 메타분석의 이론적 발전과정을 설명하면서 Ingram Olkin과 그의
학문적 영향, 그리고 그의 학문적 가계도를 제시하였다. O'Rourke(2007)는 1,700년대 통계
학의 발달부터 교육학이나 사회과학분야뿐만 아니라 의학분야에서의 메타분석 발달과정을 통시
적으로 간략하게 설명하였지만 각 시기별로 특징적인 사항들에 대한 구체적인 설명은 미흡하였
다고 할 수 있다. Huberty(2002)는 메타분석에서의 핵심적인 개념인 효과크기(effect size) 지
수들(indices)들의 역사를 살펴보면서 관계지수(relationship indices), 집단차이지수(group
difference indices), 집단중첩지수(group overlap indices), 다변량지수(multivariable
indices)별로 역사적 발전과정을 소개하였다. Glass(1999)는 자신이 1976년 메타분석이라고
명명한 이후의 25년의 역사적 발전과정에 대해 자서전적인 설명과 함께 제시하였는데, 특히, 그
가 미국교육학회에서 메타분석에 대해 연설하게 된 계기를 설명하면서 그 시작이 통계적 유의성
의 한계에서 출발하게 되었다고 밝힌 점은 메타분석의 발전에서 매우 인상적인 장면이라고 평가
할 수 있다.

Olkin(1996)은 의료분야에서의 메타분석의 발전과정을 중심으로 다양한 분야에서 연구결과
를 종합해야 하는 필요성을 중심으로 서술하였다. 메타분석이 필요한 이유로서 모든 학문영역에
서 지식의 폭발적 증가임을 강조하였고, 개별 데이터(individual patient data)를 포함한 메
타분석을 어떻게 실행할 것인가의 문제와 그 과정에서 나타나는 다양한 편의문제 등을 논의하였
고, 특히 메타분석은 새로운 연구를 계획하는데 사용되어야 한다는 점을 강조하였다. Dalton
& Dalton(2008)은 메타분석을 적용한 연구물의 증가 추이를 심리학, 의료분야, 교육학, 조직
학 연구(Organizational studies) 별로 각각 1965-1974, 1975-1985, 1986-1995,
1996-2006 별로 제시하였으나, 각 시대별로 어떤 발전이 있었는지를 논의하기보다는 최근 조
직학연구방법에 투고된 메타분석 방법론적 연구물 4편에 대해 조망하는 데 중점을 두었다.
Rosenthal & DiMatteo(2001)은 메타분석의 장점과 역사적인 비평에 대해서 논의하면서,
메타분석의 절차와 이질성에 대한 설명과 효과크기에 대하여 논의하였다. 주로 상관계수 효과크

기의 장점과 서로 다른 종류의 상관계수, 그리고 상관계수 효과크기의 해석에 대해 설명하였다. Chalmers, Hedges, & Cooper(2002)는 메타분석의 역사를 관심 주제별로 구분하기도 하여 통계적인 불명확성 감소, 학계에서의 연구로서 인식, 정책과 실천을 위한 메타분석의 활용이라는 측면에서 정리하였다.

이상의 역사적 발전 과정을 종합하면 기존의 메타분석 발전과정에 대한 연구들은 의료분야 및 다른 학문 분야들을 중심으로 매우 간략하게, 혹은 부분적인 이슈들(효과크기, 절차, 장점 혹은 비판을 중심으로)로 논의되었다고 볼 수 있다. 따라서 메타분석연구의 전체적 발전 과정을 통시적으로 조망하여 교육학 연구방법론으로서 메타분석이 자리 잡는 과정을 살펴볼 필요가 있고, 이를 통해 메타분석에 대한 이해의 폭을 넓히고, 현재 논의되는 방법론적 이슈들에 대한 논의의 발전과정을 살펴봄으로써 향후 메타연구의 연구방향을 설정하는데 도움을 줄 수 있다.

이러한 점에서 본 연구에서는 우선 Glass가 메타분석을 명명한 1976년도를 기준으로, 그 이전인 1900년대 초부터 개별 연구들을 종합하려는 노력들이 어떤 방식으로 진행되어 왔는지 살펴보고자 한다. 이후 그러한 개별적 탐구 노력들이 Glass에 의해 메타분석이라고 명명되면서 1980년까지의 메타분석에 대한 인식의 확산 및 비판의 극복을 거쳐 1980년대 메타분석에 대한 교과서의 출현과 양적 발전, 1990년 증거기반운동(evidence-based movement)과 함께 발전되어 온 과정을 차례로 고찰해 보고자 한다. 이러한 접근은 메타분석이 시대별로 역사적 발전과정을 거치면서 기존에 논의된 이슈들이 현재의 메타분석 방법론의 발전에 어떻게 녹아들었는지, 그리고 현재 논의되고 있는 이슈들에 대한 과거의 논의들을 토대로 향후 발전방향을 제시하는 데 있어 인식의 지평을 확장하는 기능을 하게 될 것이다.

Ⅲ. 메타분석연구의 역사적 발전 과정

메타분석이 그 발전 초기에 많은 오해와 비판에 직면하였음을 상기해 볼 때, 메타분석연구가 아직 일천한 한국적 상황에서 메타분석의 역사적 발전과정을 고찰하는 것은 메타분석 방법론 적용 필요성에 대한 이해와 향후 연구방법론적 문제들에 대한 연구방향을 설정하는데 그 의미가 있을 것이다. 메타분석에 대한 역사적 발전과정을 이해하는 노력은 한용진(2007)이 교육학사의 필요성에서 밝힌 것처럼, 각 교육학 및 학문분야 전문가들이 그 교육학 및 전공 학문분야의 역사를 연구함으로써 지금까지의 연구의 발전과정 및 역사에 대한 인식의 공유와, 향후 학문적 연구 과제를 선정하는데 있어서 나름대로의 좌표 역할을 할 수 있을 것이라는 기대에서 비롯된다.

1. 연구 결과 통합에 대한 개념적 발전기(1976년 이전)

가. 1950년대 이전

1976년 이전에 메타분석연구와 관련하여 주목할 만한 학자는 바로 Karl Pearson(1904), Tippett(1931), Ronald Fisher(1932), Chochran(1937)이라고 할 수 있다. Karl Pearson(1904)은 장티푸스 예방접종의 효과에 대해 11개의 연구를 종합하기 위하여 평균 상관 계수(average correlation) 연구를 수행하였고, 이 연구가 통계적으로 개별연구를 종합한 최초의 시도로 간주 된다(O'Rourke, 2007). Tippett(1931)는 개별연구의 확률값(probability value)을 종합하려는 시도를 농업분야에서 하였다. Ronald Fisher(1932)는 처음으로 농업분야에서 유의수준(p-value)을 통계적으로 종합하는 기법을 제시했으며, 그의 방법론은 1960년대까지 10편 이상의 연구물에서 이용되었다(Olkin, 1990). Chochran(1937)은 처치와 효과의 상관관계의 종합보다 효과의 크기를 종합하는 방법을 제시했고(O'Rourke, 2007), 이 접근은 메타분석연구에서 매우 의미있는 접근 방법이었지만, 1930년대 후반 당시에는 주목을 받지 못했다.

나. 1950년대 이후

2차 세계 대전 이후, 1950년대와 60년대를 거치면서, 과학적 연구가 양적으로 팽창하게 되었고, 심리학, 사회학 및 의료분야에서 드물기는 하지만, 개별 연구들을 종합하려는 연구자들이 나타나기 시작하였다(Chalmers, Hedges, & Cooper, 2002). 우선 Hans Eysenck(1952)는 심리치료가 효과가 없다는 주장으로 임상 심리(clinical psychology) 분야에서 논쟁을 시작하는 계기를 만들었고, 이는 후에 교육행정 및 교육정책분야에서 학교에 대한 예산 투입이 학생들의 학업성취도를 향상시키는데 효과가 없다는 Hanushek(1981, 1986, 1989, 1991)의 주장과 함께 메타분석 발전과정에서 역사적인 논쟁을 유발하게 되고, 결국 Glass가 교육학분야에서 메타분석을 개발하는 계기가 되었다.

따라서 1960년대 이후에 본격적으로 연구의 통합을 위한 효과적인 방법에 관한 논쟁이 시작되었다. 우선 Kissinger(1960)는 새로운 연구의 추가는 단지 정책결정자들에게 새로운 정보를 주기보다는 단지 부담을 추가하는데 그치고, 미국 정책의 위기는 연구의 부족에서 비롯된 것은 아니며, 여러 가지 면에서 볼 때 정리되지 않은 정책연구가 너무 과다한데 직접적인 원인을 찾을 수 있다고 생각하였다. 이는 연구의 증가가 정책결정자들에게 종합된 정보를 제공해 주지 못하고, 또 하나의 불확실성을 추가하는 것에 그친다는 회의적인 시각에서 비롯된 것이다(Bergh, Button, Nijkamp, & Pepping, 1997, p. 36). 또한 Fieldman(1969)은 "Using the work of others: Some observations on reviewing and integration"라는 저술에서, 기존 연구를 중

합하는 것 그 자체를 연구의 일종으로 보고, 이를 위한 독자적인 통계적 기법을 가진 연구분야로서 인식하였다. 이는 단순한 문헌연구 차원이 아닌, 새로운 연구방법론으로 인식하는 노력이 벌써 이 시기에 개념적으로 주창되었음을 알 수 있다. 같은 해에, Light & Smith(1971)는 개별적 연구의 종합화를 위한 노력으로서 집락적 접근(cluster approach)을 제안하였고, 여러 연구들의 결과에서 변량이 존재하는 것은 연구자에게 혼란을 주기보다는 오히려 유용한 정보로 간주할 수 있다고 하였다. 여러 연구들의 변량을 정보로 인식하고자 한 노력은 이후 범주별 분석(categorical analysis)이나 임의효과 모형(random effect model) 및 혼합효과 모형(mixed effect model)의 발전을 가져오는 계기가 되었다. Taveggia(1974)는 개별연구의 결과를 확률로서 인식했고, 같은 주제에 대해 많은 연구가 존재한다면, 서로 상충되고 일관되지 않은 연구 결과들이 보고될 것이라고 했다. 상충되어 보이는 것이 결국은 각 연구결과들의 분포(distribution)의 일부라고 하였다. Taveggia의 이러한 입장은 이후 Olkin & Hedges(1985)에서 메타분석에 대한 통계적 추론을 가능하게 한 개념적 기초라고 생각된다. 이렇듯 메타분석에서 개별 연구물을 하나의 샘플로 이해하고 개별 연구들로부터 추출된 효과크기의 분포를 바탕으로 통계적 추론까지 이끌어내게 된 것도 바로 이러한 선구자들의 이론적인 논의와 개념화 시도로부터 발전되었음을 알 수 있다. Taveggia는 또한 문헌종합을 위한 여섯 가지 단계를 제시하였다. 즉, 연구찾기(selecting research), 자료추출(retrieving), 목록화와 코딩화(indexing and coding studies), 연구결과의 특성분석(analyzing the comparabilities of findings), 연구결과의 축적(accumulating comparable findings), 도출결과의 분포 분석(analyzing the resulting distributions), 결과의 보고(reporting the results) 등이 그것이다. 이는 후에 Cooper (1982, 1989, 1998)에 의해 정리되어 오늘날 메타분석을 위한 표준안으로 인정받고 있는 '메타분석의 5단계'와 개념적으로 거의 일치한다.

1976년 이전의 연구들을 검토해 보면 이미 개념적으로 1904년에 Pearson이 개별연구를 종합하려는 시도를 한 이래 Fisher의 유의수준(p-value) 종합 노력도 있어 왔고, 특히 Chochran(1937)은 개념적으로는 현재의 효과크기 개념과 일치하는 접근법을 이미 그 당시 제시했음을 알 수 있다. 종합적으로 볼 때, 1976년 이후에 본격적으로 인식되고 사용되기 시작한 메타분석은 그 이전의 연구들에 의해 개념적으로는 이미 주창되었지만, 학계나 실제 연구에서 활용되지 않았을 뿐이라는 것을 알 수 있다. 하지만 메타분석의 발전의 토대로서 개별연구의 결과를 확률로서 인식했다는 점, 상충되어 보이는 연구결과들이 결국은 그 연구결과의 분포의 일부라고 생각한 점은 이후 메타분석을 통한 연구결과들에 대한 통계적 추론이 가능하게 하는 이론적 토대가 되었음을 알 수 있다.

2. 메타분석 주창 및 비판 극복(1976-1980)

1976년을 기점으로 메타분석이라는 용어가 세상에 알려졌지만, 그즈음에 Glass 외에 다른 두 가지의 심리학적 접근방법 역시 유사한 방법으로 연구결과들을 종합하고자 하였다. 이러한 세 가지 접근법이 독립적으로 행해진 사실은, 1970년대 말에는 연구결과를 통계적·체계적으로 종합 하자는 시대적 요청이 성숙했음을 의미한다. 하지만 동시에 그러한 시도에 대한 비판적인 논쟁도 불가피하게 발생하였으며 이를 극복하기 위한 노력 역시 병행되었다(Cooper & Rosenthal, 1980).

가. 3개 연구팀

1970년대에는 메타분석 방법론에 대한 연구가 세 연구팀에 의해 독립적으로 진행되었지만, 각 연구팀들은 서로에 대해서 알고 있지는 않았다. 먼저 Glass와 Smith는 833개의 심리치료 (Psychotherapy) 효과를 종합하였고(Smith & Glass, 1977), 또한 725개에 달하는 학급규모(class size)와 학업성취도의 관계에 관한 연구들을 종합하였다(Glass & Smith, 1979). 특히, Glass가 메타분석을 발전시키게 된 계기는 Eysenck(1952)가 심리치료의 효과가 없다고 주장하는 것에 대한 반론을 준비하면서부터이다. 그는 1976년 전미교육학회(AERA, American Educational Research Association) 회장 취임 연설에서 “meta-analysis”라는 용어를 발표하게 되어 교육학 및 사회과학 분야에 미치는 파장이 컸고, 일반적으로 1976년의 연설문이 메타분석의 시초라고 간주된다. 하지만, 앞에서 살펴본 바와 같이 개별연구를 종합하려는 시도는 1900년대 초부터 있어왔고, 방법론적으로도 많은 부분이 이미 개념적으로 정리되어 왔기에, 1976년 Glass의 발표는 본격적으로 교육학 및 심리학 등 다른 사회과학분야에서 메타분석이 자리 잡는 역사적 계기가 되었다.

Glass와 비슷한 시기에 하버드 대학의 사회심리학자인 Rosenthal은 345개 연구의 개인 간 기대효과(the effects of interpersonal expectations)를 종합하였다(Rosenthal & Rubin, 1978). 그 이전에, Rosenthal은 실험연구자의 기대가 그들의 연구결과에 영향을 준다는 연구 결과에 대한 의문을 가지면서 메타분석에 관심을 갖기 시작하였다. 1950년대에 연구자의 기대 효과에 대한 연구들이 관심을 끌지 못했으나 많은 연구가 축적된 후 Rosenthal은 개별 연구결과들을 종합하고 비교해 보았고, 이미 1960년대부터 연구종합 절차를 가르치기 시작했다고 한다(Rosenthal, 1991). 조직 심리학자들(Organizational psychologists)인 Schmidt와 Hunter는 흑인과 백인 노동자의 채용 심사(employment test) 타당도에 관한 866개의 연구물들을 비슷한 시기에 종합하였다(Hunter, Schmidt, & Hunter, 1979).

세 연구팀에서는 각각 기존의 전통적인 문헌연구로는 심리학 및 교육학 분야의 폭증하는 개

별연구들을 종합하기에 미흡하다는 것을 인식하고, Pearson, Fisher, Chochran의 연구방법을 적용하여 독자적인 메타분석 방법론을 발전시켰다. Glass는 주로 심리치료 및 학급당 인원 수 감축 등에서 메타분석 방법론을 주장했고, Rosenthal은 주로 상관계수를 종합하는 문제를 중심으로 방법론을 발전시켰으며, Schmidt와 Hunter는 메타분석을 수행하는 데 있어서 장애물인 데이터의 변질 요소들(artifacts)을 제거하고 좀 더 정교한 효과크기를 측정하는데 주로 관심이 있었다. 대표적인 데이터 변질 요소에는 측정범위의 단축(range restriction)과 신뢰도가 낮은 측정치의 감소(attenuation) 등이라고 할 수 있다. 이는 주로 조직·산업심리학 분야에서 많이 쓰이고 있으며 측정학적 메타분석(psychometric meta-analysis)이라고도 한다. 이 세 방법론들 사이에 효과크기를 추정하는 데 이론적으로는 같은 접근법을 취하고 있고, 실제 추정치에 있어서도 큰 차이는 없고, 다만 현재 메타분석에서 가장 많이 쓰이고 있는 Hedges와 Olkin의 방법과 Hunter와 Schmidt의 방법을 이론(탐색연구 vs. 확인연구)과 연구 설계시 고려할 점(범위제한 및 측정 오차의 정도 등)에 따라 고려해야 한다(Aguinis, Sturman, & Pierce, 2008).

나. 메타분석에 대한 비판과 극복

Glass(1976)가全美교육학회(AERA) 회장 취임연설을 한 후, Eysenck(1978)는 이러한 그의 메타분석연구에 관해 엄청나게 어리석은 시도("an exercise in mega-silliness")라고 비판하였다. Glass, McGaw, & Smith(1981)에 따르면, 메타분석에 대한 비판은 크게 네 가지로 정리할 수 있다.

첫째, 비교 및 혼합이 불가능한 이른바 '사과와 오렌지의 비교 문제'(apples and oranges problem)로서 서로 다른 측정기법, 서로 다른 표본들을 측정한 것들을 비교하는 것은 논리적이지 않다는 것이다. 둘째, 질 낮은 데이터의 활용 문제(use of data from "poor" studies)로서 메타분석에서 연구의 질이 낮은 연구물들을 결합하는 것을 비판하고, 연구의 질이 높은 연구물만을 종합해야 한다는 주장이다. 셋째, 연구물 선택의 편차문제(selection bias in reported research)로서 메타분석은 연구물에 보고된 결과를 기초로 하여 통계적 분석을 통해 이루어지는데, 만약 학술지에 발표된 연구물과 교과서에 보고된 연구결과, 그 밖에 출판되지 않은 보고서의 결과물 사이에 체계적 편차가 존재한다면 메타분석에 의해 종합된 결과는 편차(bias)가 있을 것이라는 비판이다. 넷째, 독립성 문제(non-independence data)로서 메타분석에서는 일반적으로 한 연구에서 한 결과만 보고되지 않고 많은 결과들이 보고되고 그 결과들을 이용하여 메타분석이 이루어지는데 한 연구에서의 여러 가지 결과들은 독립성 가정 위반의 문제가 제기되고, 메타분석에 의해 종합된 결과물에 대한 최종적인 의문이 제기된다는 지적이다.

그러나 위의 네 가지 비판에 대한 반론도 타당성 있게 제기되고 있다. Glass, McGaw, &

Smith(1981)의 반론을 중심으로 차례대로 정리해 보면 다음과 같다. 첫째, 같은 주제에 대해 연구결과들이 같으면 종합할 필요가 없으며, 서로 다른 연구들이기 때문에 종합할 필요가 있다는 것이다. 서로 다른 측정치들에 대해서는 표준화를 통해서 단위를 통일하기에 종합이 가능하고, 서로 다른 결과들에 대해 종합함으로써, 기존 연구 성과들에 대한 연구결과를 종합적으로 이해하고 조망할 수 있다는데 메타분석의 장점이 있다는 주장이다. 서로 다른 샘플 사이즈에 대해서는 가중치를 적용함으로써 그 차이를 반영하고 서로 다른 형태 및 유목으로 보고된 결과물에 대해서는 연구주제에 맞는 유목변수(categorical variable)를 찾아 그에 맞는 유목별 분석(categorical analysis)을 하면, 특정 처치나 정책이 어떤 유형이나 대상에 효과가 있는지 종합적으로 고려할 수 있다. 둘째, 개별연구의 질(quality)을 어떻게 고려할까 하는 문제는 아직도 논란이 제기되고 있는 문제이다. Hedges(1981)는 이미 연구의 질이나 연구의 타당도를 고려한 통계기법을 제시해 왔다. 문제는 메타분석에서 연구의 질을 어떻게 포함해서 연구들을 종합할 수 있는지가 아니라 개별연구의 질을 어떻게 평가하며, 그 평가 결과를 어떻게 정당화(justify)할 수 있느냐가 보다 의미롭고 중요한 문제라는 것이다(연구의 질 이슈에 대해서는 Valentine(2009) 참조). 셋째, 출판된 연구물과 출판되지 않은 연구물의 차이에 의한 편의 문제(publication bias)에 관한 문제에 대해서 메타분석에서는 출판된 연구물뿐만 아니라 출판되지 않은 학회 발표물, 학위논문들을 모두 포함해서 분석하고 출판의 편의 존재 여부를 검증하고 있기 때문에 크게 문제 될 여지는 없다고 여겨진다. 넷째, 메타분석에서 한 연구물에서 여러 개의 효과크기를 보고 할 경우 한 표본을 여러 번 사용해서 효과크기를 측정하는 경우 독립성 가정이 문제 되는데 현재까지 Cooper(1998)의 분석단위의 이동(shifting unit of analysis) 등 독립성 가정 위반문제에 대한 많은 연구가 축적되어 연구자들이 분석 및 결과보고 단계에서 독립성 가정 위반 문제를 충분히 검토하고 있다.

요약하면 1976년 메타분석이라는 용어가 사용된 이래 1980년까지의 시기는 메타분석에 대한 비판과 그에 대한 극복의 노력으로 메타분석의 학문적 토대가 형성된 시기라고 할 수 있다. 서로 다른 연구그룹에서의 활발한 연구방법의 개발과 전개와 더불어 급격히 증가하는 학문적 성과물에 대한 통계적 종합의 필요성에 대한 사회적 요구 또한 메타분석이 자리 잡는 데 가장 큰 요인이 되었다고 할 수 있다.

3. 메타분석 방법론 연구 및 교과서 출현(1980년대)

1980년대 초기에 Jackson(1980)은 36개의 문헌연구 논문들(research review)을 검토한 후, 이러한 선행연구들이 적절한 방법론에 기반하지 않고 있음을 지적하고, 일반 개별 연구(primary research)와 유사한 6개의 문헌검토 작업을 제안한 바 있다. 2~3년 후에 Cooper(1982)는 메타

분석 5단계 모형을 제안하였고, 연구문제설정, 데이터 수집, 데이터 평가, 데이터 분석 및 해석, 발표 및 보고가 바로 그것이다. Light & Pillemer(1984)의 “Summing Up: The Science of Reviewing Research”는 사회과학뿐만 아니라 의료분야의 관심 있는 연구자들에게 중요한 메타분석적 연구자료로써 활용되었다. 특히, 기존의 문헌연구를 주관적이고 개인적인 판단에 의한 것으로 특징지을 수 있고, 어떤 연구들을 문헌연구에 포함시키며, 어떻게 전체적인 결론을 도출할지에 대해서도 객관적인 판단이 어려운 점을 지적하였다. 그래서 같은 페이지 셋을 분석하고도 서로 다른 결론을 도출해 내기도 하였다. 또한 연구 리뷰(research review)의 타당도 문제가 해결되어야 한다는 인식이 이 시기에 자리 잡기 시작하였다(Cooper, 1982; Jackson, 1980). 메타분석에서 출판의 편차(publication bias)에 대한 연구가 시작되어 편의(bias)없는 메타분석의 토대가 마련되었다(Rosenthal, 1979).

한편 1970년대 말 메타분석이 알려지면서 비판이 제기되고 그에 대한 극복과정과 함께 통계적으로 편의(bias)를 줄이려는 노력과 함께 통계적인 방법론을 체계화한 노력이 여러 교재 출판으로 이어졌다. 이 시기에 메타분석에 관한 사회과학분야에서 대표적인 다섯 가지의 교과서가 출간된 바 있다¹⁾. 그중에서 교육학을 중심으로 한 사회과학에서는 Hedges & Olkin이 ‘Statistical Methods for Meta-Analysis’를 1985년도에 출판하여 메타분석의 통계적 방법론의 기틀을 마련하였다. Becker(1998)는 1980년대 90년대의 메타분석 교과서 15개를 교육용 서적(instructional books)²⁾, 방법론용 서적(methodological books)³⁾, 그리고 메타분석적용 서적(Application-Oriented books)⁴⁾으로 구분하였다. 다양한 분야에 걸친 많은 교재의 출현은 메타분석 방법론의 통계적인 기초를 이루고, 연구자들 사이에 교육학 및 사회과학 연구방법론으로서 자리매김하는 기반이 되었다.

4. 증거기반운동(evidence-based movement) 및 양적 팽창(1990년대)

1970년대 말부터 1980년대를 지나면서 수천 개의 메타분석을 이용한 연구물이 출판되었고, 이는 1994년에는 그동안 이루어진 주요한 메타연구를 집약한 ‘Handbook of Research

1) Bickman & Rog(2008), Glass, McGaw, & Smith(1981), Hunter, Schmidt, & Jackson(1982), Rosenthal(1984), Light & Pillemer(1984), Hedges & Olkin(1985)이 그것이다.

2) Cooper(1989), Cooper & Hedges(1994), Glass, McGaw, & Smith(1981), Light and Pillemer(1984), Mullen(1989), Rosenthal(1991), Wolf(1986)를 예시하였다.

3) Cooper & Hedges(1994), Eddy, Hasselblad, & Schachter(1992), Hedge & Olkin(1985), Hunter & Schmidt(1990), Panel on Statistical Issues and Opportunities for Research in the Combination of Information(1992)을 예시하였다.

4) Cook, Cooper, Cordray, Hartmann, Hedges, Light, Louis, & Mosteller(1992), Hyde & Linn(1990), Wachter & Straf(1990)를 예시하였다.

Synthesis'가 출판되는 계기가 되었다(Cooper & Hedges, 1994). 1990년대에는 1970년대 말 Glass & Eysenck의 심리치료 효과 논쟁이 있었던 것처럼, 교육행정 및 교육정책 분야에서 교육재정 투입의 효과에 대한 논쟁이 있었다. 먼저 Hanushek(1981, 1986, 1989, 1991)은 교육생산함수(education production function)를 통한 연구들을 종합하여 볼 때 교육재정투입이 학업성취 향상에 기여한다는 증거가 없다고 결론지은 바 있는데, 이에 대해 Hedges와 그의 제자들은 Hanushek의 분석 기법은 단순한 투표식 방법(vote counting)에 의한 것이고, 60개의 연구들에 대해 메타분석을 이용해 통계적으로 종합해본 결과 교육재정 투입은 학업성취 향상에 긍정적이라고 반론을 제기하였다(Greenwald, Hedges, & Laine, 1996). 이러한 심리치료 효과 논쟁에 이은 교육재정 투입 효과 논쟁이 벌어지면서 메타분석은 정책효과 검증이나 교육실천(educational practice)과 관련한 의사결정에 자주 활용되기에 이르렀다. 이러한 정책적 활용 움직임과 함께 메타연구가 활발하게 적용된 학문분야 중의 하나인 의학 분야의 특기할 만한 사건이 생겼는데 바로 영국에 'Cochrane 센터'가 설립(1992년)되었다는 것이다. 'Cochrane 센터'는 여러 의료분야의 연구결과들을 체계적으로 종합(systematic synthesis)하는 국제수준의 네트워크라고 할 수 있고, 1993년 말에는 'the Cochrane Collaboration' 이라는 연구물들을 종합하기 위한 국제네트워크가 이어 설립되었다(www.cochrane.org).

메타분석 및 연구종합에 대한 갈증과 기대의 심화는 정책결정자들, 현장 전문가들(practitioners), 서비스를 이용하는 사람들 등이 현재까지 여러 연구결과들의 종합적인 정보가 지속적으로 갱신되기를 원하기 때문이다. 이러한 갈증과 기대는 연구자들이 새롭게 탐색해야 할 도전적 과제가 되었고, 이러한 도전적 과제에 대응하기 위해 많은 연구들이 출판되었고 국제적인 네트워크가 구성되어 체계적인 연구결과물을 집적하려는 움직임이 조직화되었다. Hunt(1997)는 연구결과(research evidence)들이 학교, 병원, 주 사회복지 프로그램, 법원, 감옥 그리고 다른 기관들에서 정책과 실천(practices)에 광범위하게 영향을 미치기 시작하였다고 평가하였다.

5. 현재의 이슈 및 향후 연구 방향(2000년 이후)

1990년대 말 메타분석에 대한 용어(terminology)가 혼란스러워져서 Chalmers와 Altman은 '메타분석(meta-analysis)은 통계적인 분석에 쓰이는 용어로 제한하고, 일반적으로 체계적인 절차를 통해 연구를 검증하고 종합하는 것은 체계적 문헌고찰(systematic review)'로 하자고 제안하였다(O'Rourke, 2007, p. 581). 이미 한국에서도 아직 메타분석이라는 용어가 통계적인 분석이 없는 연구물에도 쓰이는 등의 혼란이 발생하고 있다는 점에서(안민호·고혜승 2006; 황용석 2006 등), 체계적 문헌고찰(Systematic review)과 메타분석(meta-analysis),

연구종합(research synthesis) 등의 용어에 대한 학계의 개념정립이 필요한 상황이다. 2000년대 들어서 메타분석이라는 용어를 처음 사용한 Glass는 1976년 이래 25년간의 메타분석 발전 과정을 자전적으로 발표하기도 하였다(Glass, 2001). 의료분야의 Cochrane Collaboration의 빠른 발전에 힘입어, 2000년 초반에는 사회과학분야의 연구물들을 종합하자는 국제적인 네트워크인 Campbell Collaboration(<http://www.campbellcollaboration.org>)이 교육학, 사회복지, 형사정책분야에서 구축되었다. 중요한 정책결정 등의 의사결정을 위해, 연구결과들을 종합하고 축적하여 활용하려는 노력들이 사회과학분야에서도 국제적으로 조직화된 것이다.

향후 메타분석방법론의 발전을 위해 연구되어야 할 과제 목록을 정리해 보면, Cooper & Hedges(1994)는 메타분석에서의 출판의 편의(publication bias), 결측 데이터(missing data) 문제, 고정효과 모형과 랜덤효과모형 등의 분석모형 선택의 문제(Dunklin, 1996) 등을 들고 있다. 그 밖에 최근에 논의되고 있는 문제로는 독립성 가정 위반문제, 다변량 메타분석(multivariate meta-analysis), 이론모형에 기반한 메타분석(model driven meta-analysis), 회귀분석의 회귀계수를 종합하는 문제(synthesizing regression slope) 등이 현재 방법론적으로 논의되고 연구되어야 할 과제들이다. 이러한 방법론적 문제들에 대한 논의를 발전시키기 위해서는 메타분석이 활발하게 사용되고 있는 의료분야의 방법론적인 연구결과 및 논의들을 함께 고려해야 할 필요가 있다. 교육학 등 사회과학 분야의 메타분석 연구와 자연과학 분야인 의료분야의 방법론적인 노력들에 대한 학제적인 교류 협력이 활성화되어야 할 필요가 있다.

메타분석에 관심을 갖고 있는 많은 학자들이 어려움을 겪고 있는 것은 연구를 종합하는 데 활용하는 프로그램(software)에 관한 것이다. 왜냐하면 일반적인 통계 프로그램인 SPSS, SAS, R, Stata 등에서 기본적으로는 메타분석 수행을 지원하지 않기 때문이다(Borenstein et al, 2009). 메타분석을 위한 적절한 프로그램을 안내하는 것은 과학적인 메타분석 연구 뿐만 아니라 메타분석의 교육을 위해서도 매우 중요한 이슈이다. 서미옥(2011)은 국내학자들의 메타분석 프로그램 운영현황에 대해 정리하면서 국내연구들이 SPSS를 사용한 경우가 가장 많고, 그다음으로 SPSS와 Excel을 함께 사용한 연구가 많았고, 메타분석 전문 프로그램인 CMA(Comprehensive Meta-Analysis)를 사용한 연구들은 3편에 불과하다고 하였다. 해외에서는 Borenstein, et al (2009)이 메타분석시 활용가능한 프로그램을 소개하면서 Excel, CMA, Stata macro, RevMan에 대해 각각의 특징 및 장·단점을 소개하고 있다(pp. 391~403). Stata macro 프로그램의 경우 코드에 대한 이해가 필요하고, 데이터를 저장하는 포맷이 제한적이라는 단점이 있고, CMA는 서로 다른 종류의 데이터를 작업하는 데 융통성이 있고 사용이 쉬우며 forest plot을 쉽게 그려낼 수 있고 윈도우 버전이라는 장점이 있다. RevMan은 Cochrane Collaboration에서 Cochrane review를 위해서 만든 프로그램으로 메타분석만을 위한 프로그램이라기보다는 체계적 문헌 분석의 모든 단계들에 대한 메커니즘을 제공해 주는 프로그램이라고 할 수 있다.

IV. 결론 및 교육학 연구를 위한 시사점

이상에서 메타분석의 국내·외적 연구동향과 발전과정을 살펴보았다. 메타분석은 개별연구의 존재를 전제로 하는 점, 인과관계 분석에 대한 제약 등의 한계점에도 불구하고, 경험적 객관성이 요구되는 학문분야에 광범위하게 적용되어 왔음을 알 수 있었다. 다만 국내 교육학연구에도 2000년 이후 활발히 적용되고 있으나, 교과교육, 교육심리 및 상담 등 일부 학문분야에 그치고 있음도 파악할 수 있었다. 또한 메타분석은 1976년 이전 초기단계의 개념적 발전기를 거쳐 1976년 Glass가 메타분석을 ‘분석의 분석(analysis of analysis)’이라고 명명한 이후 많은 비판과 논쟁을 거듭하면서 발전해왔다. 흥미로운 사실은 Glass 뿐만 아니라 교육학에서 피그말리온 효과(pygmalion effect)로 유명한 Rosenthal, 조직 심리학분야의 Schmidt와 Hunter도 Glass와 거의 비슷한 시기에 각자 자기 분야에서 독립적으로 기존 연구를 종합하는 양적, 통계적 분석기법을 개발하였다는 것이다. 이는 1950년대 이후 연구의 양이 폭발적으로 증가하면서 더 이상 기존의 문헌연구(narrative review)로는 과학적인 분석과 연구결과를 체계적으로 종합하는 것이 불가능하고, 새로운 통계적 기법은 그 시대의 시대적 요청이었다고 생각된다. 그러한 비판을 극복하는 과정에서 독자적인 연구방법론으로서의 학문적 기반을 공고히 해왔으며, 특히, 정보화의 진전에 따른 데이터베이스의 확장 및 접근 가능성의 확대, 연구결과의 공유의 용이성 증가, 학술적 성과물에 대한 과학적 종합에 대한 요구가 크게 작용했다고 볼 수 있다. 1976년 이후의 비판 극복과정을 거쳐 1980년대 편의(bias)를 줄이기 위한 통계적 기법이 더 세련되어지기 시작했다. 또한 그러한 연구 결과들이 많은 교과서 형태로 출판되었다. 1980년대 부터는 의료분야에서 활발히 메타분석이 쓰이기 시작했고, 무엇보다 그러한 노력이 Cochrane collaboration이라는 국제적 네트워크를 완성시켰다는 점이고, 교육학 및 사회과학분야에서도 2000년대 이후 Campbell collaboration이라는 국제 네트워크조직이 만들어져서 교육학, 사회복지, 형사정책분야의 연구물들을 종합하여 각 분야의 실천(practice)과 학문적 발전에 기여하고 있다. 본 연구에서는 주로 양적 연구를 종합하는(synthesizing)하는 메타분석을 중심으로 발전과정을 고찰했지만, Campbell collaboration이나 Cochrane collaboration 등에서는 질적 연구를 종합하자는 움직임도 활발하다는 점도 참고할 필요가 있다. 이미 질적 연구를 종합하여 교육학 이론이나 의사결정, 실천에 기여하고자 하는 노력들이 그동안 계속 경주되어 왔고, 우리나라에도 소개되고 있다(Jensen & Allen, 1996; Mosteller, 1992; Noblit & Hare, 1988; 나장합, 2008).

이상의 결론을 바탕으로 앞으로 교육학 연구에서 메타분석의 시사점을 살펴보면 다음과 같다.

첫째, 메타분석과 유사 연구방법과의 개념적 차별화가 필요하다. 개념적으로 양적 연구 결과를 통계적으로 종합하는 메타분석(meta-analysis)과 기존의 화술적 종합법(narrative syntheses),

체계적인 문헌 고찰(systematic review) 및 내용분석(contents analysis), 질적 연구의 종합(Synthesizing qualitative research) 등 사이에는 엄밀한 구분이 필요함에도 불구하고, 기존의 연구들을 살펴보면 개념적 혼선과 방법론상의 혼란이 일어날 수 있는 여지가 많음을 발견할 수 있다. 이를 위해서는 메타분석방법 및 가정에 관한 충분한 교육과 함께 학계 차원의 다양한 형식의 의견 교환과 지식의 공유를 위한 노력이 이루어져야 할 것이다.

둘째, 기존 연구를 양적으로 통합하고, 질적으로 종합하는(synthesizing qualitative research) 노력을 통하여 학술적으로도 검증된 지식을 축적하는 노력이 필요할 것이며, 이러한 검증된 연구결과를 체계적으로 공유하는 노력도 필요하다. 또한 메타분석결과를 활용하는 교육 정책담당자들(policy makers)과 교육현장 실천가들(practitioners)이 메타분석의 중요성을 인식하고, 더욱 체계적인 과학적 메타분석을 요구하고 이를 지원하는 관행과 태도가 확산될 필요가 있다. 이러한 개념적 차별화 작업과 검증된 지식의 공유 노력, 그리고 메타분석 활용도의 증가를 통해 이른바 ‘증거에 기반한 실천 내지 정책(evidence-based practice)’을 구현하는 계기가 될 수 있을 것이며, 더 나아가 학교 교육의 실질적 변화를 이끌어낼 수 있는 기반도 형성될 수 있을 것이다.

셋째, 현재 메타분석과 관련하여 쟁점이 되고 있는 몇 가지 문제들에 대해서 국내 학계 나름대로의 해석과 정리가 필요하다. 특히, 교육학 및 사회과학 분야에서 많이 쓰이는 방법론인 회귀분석의 회귀계수를 종합하는 문제, 내재적 자료(nested data) 구조를 가진 연구물을 종합하는 문제, 메타분석에서의 독립성 가정 위반, 구조방정식 모형을 활용한 연구물들을 종합하여 이론 모형을 형성하려는 접근(model based meta-analysis, Becker 2009) 등에 대한 나름대로의 해석과 탐구를 통해 방법론적 발전을 이루어야 할 것이다.

넷째, 현재 메타분석을 적용하여 연구를 종합하려는 시도 중에서 방법론적으로 보완되어야 할 부분은 서미옥(2011) 등의 연구에서 지적한 바와 같이 동질성 검정, 분석모형의 선택, 메타분석에서의 독립성 가정문제, 연구의 질을 고려하는 문제 등이라고 할 수 있다. 교육학 분야의 개별 연구들은 다른 사회과학연구들처럼 다른 모집단으로부터 추출된 표본을 사용했을 가능성이 많아서 분석모형 선택 시 랜덤효과 모형을 이용해야 하는데, 현재 기존의 메타분석 연구물들을 조사한 서미옥(2011)의 연구에서와같이 고정효과 모형을 사용할 경우 효과크기를 과대추정할 수 있다. 또한 하나의 연구에서 여러 개의 효과크기를 보고했을 때 독립성 가정 위반이 문제 되나 이 문제를 적절한 방법으로 다룬 메타분석 연구물이 거의 없는데 이 또한 메타분석 연구결과의 타당성을 저해하는 큰 요인이 된다. 따라서 향후 교육학관련 메타분석 적용 연구에서는 랜덤효과 모형을 적용하는 것이 타당한 분석모형이 될 것이다(Borenstein et al., 2009). 또한 교육학과 같은 사회과학에서 연구의 질에 대한 판단은 정당성이 어려우므로 분석전에 판단하여 배제하기보다는 연구의 질과 관련된 특징(예를 들면, 무선헌당 여부, 출판 여부)들을 코딩하여 분석한 후, 의미를 고찰하는 게 바람직한 방법이라고 할 수 있다.

다섯째, Glass(1999)가 지적하듯이 향후 개별연구물을 보고할 때, 개별 연구 시 사용되었던 원본 데이터를 함께 보고해서 다른 연구자가 반복(replication)할 수 있다면, 개별연구의 타당성 확보와 함께 메타분석에서 요약 데이터만을 가지고 많은 정보를 포함하지 못하는 한계를 극복할 수 있다.

이상에서 살펴본 바와 같이 본 연구는 의료분야 등을 중심으로 한 기존의 메타연구의 대상을 탈피하고자 한 점, 그동안의 이슈별·주제별 접근에서 벗어나 메타분석의 전반적 발전과정의 조망을 시도한 점, 그동안 주목을 받지 못한 시대별 발전 과정에서의 맥락에 대한 고찰(장점의 계승과 단점의 극복 등)을 시도한 점 등은 기존의 메타연구와는 분명히 차별성을 갖는다고 하겠다. 그럼에도 불구하고 본 연구는 통시적인 조망을 통한 시각을 제시하는 데 주목적이 있으므로 메타분석 방법론 상의 개별문제에 대하여 심층적으로 고찰하지 못한 한계가 있다. 이러한 한계점들은 메타분석이 개별연구와 상보적인 관계를 가지고 지식을 형성해 가는 통계학적 절차라는 점을 고려한다면, 통계적 절차에 대한 이해와 메타분석을 수행하는 이론적 지식이 협력하여 기존 연구결과를 타당하게 종합하려는 연구자들의 통섭적 노력에 의해 보완되어야 할 것이다. 따라서 관련 학문분야간의 학제적 연구 및 독립성 가정위반 문제, 회귀계수의 종합 등 현재까지 논의되고 있는 방법론 문제 등에 대한 심도 있는 후속 연구 등을 통해 극복해야 할 것이다.

참 고 문 헌

- 나장함(2008). 질적 메타분석에 대한 고찰: 교육과정 연구에서의 적용가능성 탐색. **교육과정연구**, 26(4), 229-252.
- 박세훈, 오승희(2003). 교육연구에 있어서의 메타분석의 쟁점. **사회교육연구**(12), 1-16.
- 서미옥(2011). 메타분석 연구의 고찰과 연구방법론적인 제안. **교육학연구**, 49(2), 1-23.
- 송인섭(1990). 메타분석의 문제. **아동연구**, 5, 43-53.
- 안민호, 고혜승(2006). 커뮤니케이션 연구들의 일반적 특징과 언급된 주요이론들에 대한 메타분석. **사회교육과학연구**, 9(1), 175-198.
- 이종승(1982). 최근 연구방법의 동향. **교육연구방법론**. 한국교육개발원.
- 이종승(1983). 메타분석: 연구결과의 통계적 종합. **교육학 연구**, 21(1), 83-92.
- 오성삼(1992). 메타분석방법의 활용동향과 컴퓨터 프로그램을 이용한 메타분석방법. **교육평가연구** 5(2), 123-166.
- 오성삼(1995). 메타분석에 대한 고찰. **교육논총**, 25, 33-62.
- 오성삼(2003). 고전적 메타분석에 의한 보완적 접근방법의 탐색. **교육과정평가연구**, 6(2), 107-123.
- 오성삼(2009). 선행연구 결과의 통합과 재분석을 위한 메타분석의 이론과 실제. 건국대학교 출판부.
- 이희숙(2011). **교육행정연구에서 메타분석법 적용가능성 탐색**. 한국교육학회 2011 춘계학술대회 발표자료. 162-176.
- 조지민(2002). 유의수준 비교방법의 점근분포 유도과 정확성 검증. **교육방법연구**, 14(1), 19-34.
- 황용석(2006). 한국 온라인 저널리뷰 연구의 주제와 접근방법에 대한 메타분석. **한국기술교육학회지**, 9(1), 257-278.
- 황정규(1988). Meta-analysis의 이론과 방법론: 경험과학적 연구 결과의 종합을 위하여. **성공논총**, 19.
- 한용진(2007). 교육학사(教育學史) 서설, **교육문제연구 제29집**. 고려대학교 교육문제연구소.
- Aguinis, H., Sturman, M. C., & Pierce, C. A. (2008). Comparison of three meta-analytic procedures for estimating moderating effects of categorical variables. *Organizational research methods*, 11(1), 9-34
- Becker, B. J. (2007). Multivariate meta-analysis: Contributions of Ingram Olkin, *Statistical Science*, 22(3) 401-406
- Becker, B. J. (2009). Model based meta-analysis.. In H. M. Cooper, L. V. Hedges, & J. C. Valentile (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 377-395). New York: Russell Sage Foundation
- Bergh, J. C., Button, K. J., Nijkamp, P., & Pepping, G. C. (1997). *Meta-analysis in*

- environmental economics*. Kluwer Academic Publishers
- Bickman, L., & Rog, D. (Eds.). (2008). *The SAGE Handbook of applied social research methods, 2nd edition*. Sage Publications
- Borenstein, M., Hedges, L., Higgins., & Rothstein, H. R. (2009). *Introduction to meta-analysis*, West Sussex: John Wiley & Sons, Ltd.
- Chalmers, L., Hedges, L. V., & Cooper, H. (2002). A brief history of research synthesis. *Evaluation & The Health professions*, 25(1) 12-37
- Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society*, 4(Suppl.), 102-118.
- Cook, T. D., Cooper, H. M., Cordray, D. S., Hartmann, H., Hedges, L. V., Light, R. J., Louis, T. A., & Mosteller, F. (1992). *Meta-analysis for future explanation: A case book*. New York: Russell Sage Foundation.
- Cooper, H.M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin*, 87(3), 442-449.
- Cooper, H.M. (1982). Scientific principles for conducting integrative research reviews. *Review of Educational Research*, 52, 291-302.
- Cooper, H.M. (1989). *Integrating research: A guide for literature reviews (2nd ed.)*. Newbury Park, CA: Sage.
- Cooper, H.M. (1998). *Synthesizing research: A guide for literatitre synthesis (3rd ed.)*. Thousand Oaks, CA: Sage.
- Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York: Russell Sage.
- Dunkin, M.J. (1996). Types of errors in synthesizing research in education. *Review of Educational Research*, 66(2), 87-97.
- Eddy, D. M., Hasselblad, V., & Schachter, R. (1992). *Meta-analysis by the confidence profile method: The statistical synthesis of evidence*. San Diego, CA.
- Eysenck. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology*, 16, 319-324
- Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychology*, 33, 517.
- Feldman, K. A. (1969). Using the work of others: Some observations on reviewing and integrating. *Sociology of Education*, 44, 86-102.
- Fisher, R. A. (1932). *Statistical methods for research workers (4th ed.)*. London: Oliver and Boyd.
- Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational*

- Researcher*, 10, 3-8.
- Glass, G. V. (1999). *Meta-analysis at 25*. Paper presented to Office of Special Education Programs Research Project Directors' Conference, U.S. Department of Education. Washington D.C., July 15, 1999.
- Glass, G.V., McGaw, B. & Smith, M.L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Glass, G. V (2001). *Meta-analysis at 25*. Retrieved December 2001 from <http://glass.ed.asu.edu/gene/papers/meta25.html>
- Glass, G. V., & Smith, M. L. (1979). Meta-analysis of the relationship between class size and achievement. *Educational Evaluation and Policy Analysis*, 1, 2-16.
- Greenwald, R., Hedges, L. R., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66(3) 361-396.
- Jackson. (1980). Methods for integrative reviews. *Reviews of Educational Research*, 50, 438-460
- Jensen, L. A., & Allen, M. N. (1996). Meta-synthesis of qualitative findings. *Qualitative Health Research*, 6(4), 553-60.
- Hanushek, E. A. (1981). Throwing money at schools. *Journal of Policy Analysis and Management*, 1, 19-41.
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24, 1141-1177.
- Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher*, 18(4), 45-65.
- Hanushek, E. A. (1991). When school finance "reform" may not be good policy. *Harvard Journal on Legislation*, 28, 423-456.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107-128
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego: Academic Press.
- Huberty, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement*, 62, 227-240.
- Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin*, 86, 721-735.
- Hunter, J.E., Schmidt, F.L., & Jackson, G.B. (1982). *Meta-analysis: Cumulating research*

- findings across studies*. Beverly Hills, CA: Sage.
- Hunter, J.E., & Schmidt, F.L. (1990). *Meta-analysis: Correcting error and bias research findings*. Newbury Park, CA: Sage.
- Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. New York: Russell Sage.
- Hyde, J. S., & Linn, M. C. (1986). *The psychology of gender: Advances through meta-analysis*. Baltimore: Johns Hopkins Press.
- Light, R. J., & Pillemer, D. B. (1984). *Summing up : the science of reviewing research*. Cambridge. Harvard University Press.
- Light, R. J., & Smith, P. V. (1971). Accumulating evidence: Procedures for resolving contradictions among research studies. *Harvard Educational Review*, 41, 429-471.
- Mosteller, F. (1992). *Meta-analysis for explanation*. NY: Sage.
- Mulrow, C. D. (1987) The medical review article: State of the science. *Annals of Internal Medicine*, 106, 485-488
- Mullen, B. (1989). *Advanced BASIC meta-analysis*. Hillsdale, NJ: Erlbaum.
- Noblit, G.W., & Hare, D. (1988). *Meta-ethnography: Synthesizing qualitative studies*. Newbury Park: Sage.
- Olkin, I. (1990). History and goals. In K. W. Wachter & M. L. Straf (Eds.), *The future of meta analysis*, (pp. 3-10). New York: Russell Sage Foundation
- Olkin, I. (1996). Meta-analysis: Current issues in research synthesis. *Statistics in medicine*, 15, 1253-1257.
- O'Rourke, K. (2007). An historical perspective on meta-analysis: dealing quantitatively with varying study results. *Journal of the royal society of medicine*, 100, 579-582
- Oxman, A. D., & Guyatt, G. H. (1988). Guidelines for reading literature review. *Canadian Medical Association Journal*, 138, 697-703.
- Panel on Statistical Issues and Opportunities for Research in the Combination of Information (1992). *Combining information: Statistical issues and opportunity for research*. Washington, DC: National Academy Press.
- Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, 3, 1243-1246.
- Rosenthal, R., & Rubin, D. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, 3, 377-415.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological*

- Bulletin*, 86, 638-641.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annu. Rev. Psychol.* 52, 59-82
- Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.
- Valentine. (2009). Judging the quality of primary research. In H. M. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 129-146). New York: Russell Sage Foundation
- Taveggia, T. C. (1974). Resolving research controversy through empirical cumulation: Toward reliable sociological knowledge. *Sociological Methods & Research*, 2, 335-407
- Tippet, L.H.C. (1931). *The method of statistics*. London: Williams and Norgate.
- Winkelstein, W. (1998). The first use of meta-analysis? *American Journal of Epidemiology*, 147(8), 717.
- Wacher, K. W., & Straf, M. L. (1990). *The future of meta-analysis: A project of the Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council*. New York: Russell Sage Foundation
- Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Newbury Park, CA

· 논문접수 : 2011-09-01/ 수정본접수 : 2011-10-10/ 게재승인 : 2011-10-25

ABSTRACT

Historical development of meta-analysis as an educational research methodology

Deok-Ho Jang(Assistant professor, Sangmyung University)

In-Soo Shin(Assistant professor, Jeonju University)

This study attempted to understand the limits of existing research review and the importance of meta-analysis, and to suggest some implication for future development of the method by reviewing the historical development of meta-analysis as an educational research methodology. We compared meta-analysis with other methods of research review such as narrative review and vote counting. Glass coined the term of 'meta-analysis' in 1976 and we investigated the phenomena of synthesizing quantitative research from 1900s to present. Especially, we focused on methodological critiques and overcoming efforts made by several frontiers on the field. Lastly, we presented the current issues and future research implications in addition to the present problems of meta-analysis research in Korea. Conceptual differentiation between meta-analysis and other similar methods, diffusion of the importance and knowledge sharing on meta-analysis, and additional academic efforts to solve some disputes around the data analysis, assumption, and approaches should be made in further research and practices.

Key Words : meta-analysis, historical development, narrative review, effect size