

ĐÁNH GIÁ HIỆU QUẢ TÁC ĐỘNG SỰ PHẠM
CỦA MỘT HƯỚNG NGHIÊN CỨU KHOA HỌC GIÁO DỤC:
TIẾP CẬN TỪ KỸ THUẬT META-ANALYSIS

GS.TS Trần Trung¹, TS. Phạm Đức Bình²

¹Học viện Dân tộc

²Trường Đại học Khoa học và Công nghệ Hà Nội

Tác giả liên hệ: trungt1978@gmail.com

Ngày nhận: 09/12/2024

Ngày nhận bản sửa: 14/12/2024

Ngày duyệt đăng: 24/12/2024

Tóm tắt

Đối với mỗi đề tài nghiên cứu khoa học giáo dục thường hướng đến mục tiêu đề xuất các tác động sự phạm nhằm đổi mới và nâng cao hiệu quả giáo dục, trong đó, kết quả thực nghiệm sự phạm sẽ là bằng chứng khoa học cho độ tin cậy và tính khả thi của kết quả nghiên cứu. Tuy nhiên, từ mẫu thực nghiệm nhỏ khi suy luận lên quần thể sẽ phải chấp nhận sai số nhất định, vì vậy, khi đánh giá hiệu quả tác động sự phạm của một hướng nghiên cứu cụ thể thì nhà nghiên cứu luôn mong muốn gộp số liệu thực nghiệm từ một số mẫu nghiên cứu khác nhau nhằm tăng hệ số tin cậy của các biện pháp sự phạm. Bài viết này trình bày quy trình bảy bước và các công cụ phân tích của kỹ thuật meta-analysis (phân tích gộp) để tổng hợp số liệu thống kê từ các kết quả nghiên cứu cùng một chủ đề nghiên cứu trong lĩnh vực khoa học giáo dục. Kỹ thuật meta-analysis áp dụng trọng số (weight) cho từng nghiên cứu độc lập được sử dụng dựa vào kích thước mẫu và độ ổn định của từng nghiên cứu, giúp giải quyết sự không đồng nhất trong kết quả từ các nghiên cứu độc lập nhằm tính toán chính xác hơn cỡ tác động (effect size) của một giả thuyết với cỡ mẫu lớn hơn so với cỡ mẫu được sử dụng trong các nghiên cứu thành phần, hỗ trợ loại bỏ sự sai lệch của các nghiên cứu độc lập, qua đó, tăng tính khách quan của kết quả nghiên cứu, nâng tiềm năng tái lập kết quả nghiên cứu và giúp xác định các giả thiết khoa học mới và tiềm năng cho các nghiên cứu trong tương lai.

Từ khóa: Khoa học giáo dục, meta-analysis, PRISMA, systematic review, tác động sự phạm.

The Evaluation of the Effectiveness of the Pedagogical Impact of a Research Direction in Educational Science: An Approach from the Meta-Analysis Technique

Prof. Dr. Tran Trung¹, Dr. Pham Duc Binh²

¹Vietnam Academy for Ethnic Minorities

²University of Science and Technology of Hanoi

Corresponding Author: trungt1978@gmail.com

Abstract

For each educational science research topic, the primary aim is often to propose pedagogical impacts aimed at innovating and improving the effectiveness of education, where the results of pedagogical experiments provide scientific evidence regarding the reliability and feasibility of the research outcomes. However, certain errors must be accepted when inferring from a small experimental sample to a larger population. Therefore, when evaluating the effectiveness of pedagogical impacts of a specific research direction, researchers often aim to combine experimental data from various research samples to increase the reliability of the pedagogical methods. This article outlines the seven-step process and the analytical tools of the meta-analysis technique to

synthesize statistical data from research results on the same topic in educational science. The meta-analysis technique applies weights to each independent study based on sample size and the stability of the findings, helping to address the heterogeneity in results from independent studies to more accurately calculate the effect size of a hypothesis with a larger sample size compared to the sample size used in the individual studies. This supports the elimination of bias in independent studies, thereby increasing the objectivity of the research results, enhancing the potential for replication of the research outcomes, and helping to identify new and promising scientific hypotheses for future studies.

Keywords: Educational Science, meta-analysis, PRISMA, systematic review, pedagogical impact.

1. Mở đầu

Đối với các nghiên cứu thực nghiệm sư phạm hoặc nghiên cứu can thiệp, một chủ đề nghiên cứu có thể được thực hiện bởi nhiều nhóm nghiên cứu (cùng thuộc một quốc gia, hoặc thuộc các quốc gia khác nhau). Những nhóm nghiên cứu này thực hiện nghiên cứu ở những thời điểm và những khu vực khác nhau, với những đối tượng nghiên cứu khác nhau, do đó, tuy chủ đề nghiên cứu là tương đồng nhưng kết quả từ những nghiên cứu này có thể không đồng nhất. Ví dụ, trong một nghiên cứu phân tích gộp về mức độ hiệu quả trong việc sử dụng các phần mềm hình học động (Dynamic Geometry Software) trong dạy học toán, Juandi và các cộng sự (2021) đã cho thấy sự không đồng nhất về kết luận trong các nghiên cứu độc lập đã được công bố. Trong khi nhiều nghiên cứu đã chỉ ra sự hiệu quả của việc sử dụng các phần mềm hình học động trong việc cải thiện khả năng học toán của học sinh so với các phương pháp học toán truyền thống (Baki và các cộng sự, 2011; Drigas & Pappas; Papadakis và các cộng sự, 2017, 2018; Sudirman và các cộng sự, 2020), một vài nghiên cứu khác lại có kết luận ngược lại (Hamidah và các cộng sự, 2020; Septian & Komala, 2019). Một ví dụ khác về sự không đồng nhất về kết quả nghiên cứu thực nghiệm sư phạm có thể được tìm thấy trong một nghiên cứu phân tích gộp tìm hiểu về mức độ hiệu quả của lớp học đảo ngược. Gillette và các cộng sự (2018) đã chỉ ra rằng một số nghiên cứu đã kết luận rằng sinh viên không thấy sự khác biệt giữa việc tham gia lớp học đảo ngược và lớp học truyền thống (Suda và các

cộng sự, 2014), ngược lại, một số nghiên cứu khác chỉ ra rằng lớp học đảo ngược cải thiện hiệu quả học tập của sinh viên và tối ưu hơn phương pháp học truyền thống (Cotta và các cộng sự, 2016; Hughes và các cộng sự, 2016). Từ hai ví dụ trên, chúng ta có thể thấy rằng sự bất đồng trong các kết quả nghiên cứu thuộc cùng một chủ đề sẽ gây khó khăn cho người đọc hoặc những nhà nghiên cứu khác khi họ không thể kết luận được kết quả từ nghiên cứu độc lập nào là chính xác và đáng tin cậy. Từ thực tế đó, phương pháp phân tích gộp được ra đời với mục đích tổng hợp và phân tích kết quả của tất cả các bài báo cùng nghiên cứu về một chủ đề, qua đó, rút ra được một kết luận duy nhất về tính hiệu quả hướng nghiên cứu chung của những nghiên cứu độc lập này. Phân tích gộp là một trong những phương pháp nghiên cứu chính để tổng hợp các kết quả của một câu hỏi nghiên cứu hoặc một lĩnh vực nghiên cứu, cung cấp những đánh giá định lượng về mức độ hiệu quả của các nghiên cứu thực nghiệm sư phạm hoặc nghiên cứu can thiệp thông qua việc tính toán cỡ tác động của từng nghiên cứu thành phần. Thông qua việc tính toán cỡ tác động, kết quả của mỗi nghiên cứu thành phần có thể được so sánh với nhau. Ngoài ra, phân tích gộp còn được sử dụng để kiểm tra các giả thuyết còn gây tranh cãi khi các kết quả nghiên cứu về giả thuyết này cho ra những kết quả không đồng nhất khi được thực hiện bởi những nhóm nghiên cứu khác nhau (Hansen và các cộng sự, 2022). Tuy có nhiều điểm tương đồng nhưng ý nghĩa và bằng chứng khoa học của các nghiên cứu phân tích gộp được xếp cao nhất, trên một bậc

so với các nghiên cứu phân tích tổng quan hệ thống trong mô hình tháp phân cấp mức độ quan trọng của các loại nghiên cứu khác nhau.

Trong lĩnh vực khoa học giáo dục, kết quả từ các nghiên cứu về phân tích gộp có thể hỗ trợ nhà nghiên cứu và giáo viên trong việc lựa chọn các phương pháp dạy và học phù hợp với học sinh. Người nghiên cứu có thể xem xét áp dụng các biện pháp can thiệp hoặc hỗ trợ học sinh đã được chứng minh có hiệu quả (nếu kết quả phân tích gộp chỉ ra giá trị cỡ tác động lớn). Ví dụ, một nghiên cứu phân tích gộp trên đối tượng là học sinh khuyết tật nhẹ đã chỉ ra rằng mô hình dạy học kết hợp giữa sự nhắc nhở của giáo viên và việc lặp lại thường xuyên các bài thực hành có thể giúp nâng cao kết quả học tập của học sinh (Swanson & Hoskyn, 1998). Do đó, giáo viên có thể áp dụng phương pháp học này để hỗ trợ học sinh khuyết tật nâng cao kết quả học tập. Tuy nhiên, kết quả của các nghiên cứu phân tích gộp không phải lúc nào cũng chính xác tuyệt đối, do đó, người nghiên cứu cần phải đánh giá được chất lượng của nghiên cứu phân tích gộp, và kết hợp với những kiến thức khác trước khi quyết định có áp dụng kết quả của phân tích gộp cho việc giảng dạy của giáo viên hay không.

Nhiều nghiên cứu đã thống nhất rằng phân tích gộp hiệu quả hơn phân tích tổng quan truyền thống, và phân tích gộp đã khắc phục được các nhược điểm của phân tích tổng quan truyền thống thông qua việc áp dụng một quy trình minh bạch nhằm giải quyết sự không đồng nhất trong kết quả của những nghiên cứu độc lập, qua đó, đưa ra được những kết luận quan trọng và đáng tin cậy (Lee, 2019; Stone & Rosopa, 2017). Những ưu điểm khác của phân tích gộp có thể được chỉ ra như sau: (i) Phân tích gộp áp dụng trọng số (weight) cho từng nghiên cứu độc lập được sử dụng, dựa vào kích thước mẫu sử dụng và độ ổn định của từng nghiên cứu, do đó, kết quả nghiên cứu từ phân tích gộp là kết quả tổng hợp của các nghiên cứu thành phần, và có độ chính xác, độ tin cậy cao hơn so với kết quả của từng nghiên cứu riêng lẻ; (ii) Phân tích gộp giúp

giải quyết sự không đồng nhất trong kết quả từ các nghiên cứu độc lập, và giúp tính toán chính xác hơn cỡ tác động tổng hợp (pooled effect) của một giả thuyết với cỡ mẫu lớn hơn nhiều so với cỡ mẫu được sử dụng trong các nghiên cứu thành phần. Lưu ý rằng độ chính xác của kết quả phân tích gộp sẽ tỉ lệ thuận với kích thước cỡ mẫu; (iii) Phân tích gộp loại bỏ sự thiên vị trong việc lựa chọn các nghiên cứu độc lập, qua đó, tăng tính khách quan của kết quả nghiên cứu; (iv) Tất cả các nghiên cứu độc lập được sử dụng cho phân tích gộp đều được trích dẫn rõ ràng, do đó, kết quả nghiên cứu từ phân tích gộp có khả năng được tái lập bởi các nhóm nghiên cứu khác; (v) Kết quả từ phân tích gộp giúp người nghiên cứu xác định các giả thiết mới và tiềm năng cho các nghiên cứu trong tương lai.

Mặc dù phân tích gộp là một phương pháp rất hiệu quả trong việc tổng hợp kết quả từ các nghiên cứu độc lập, tuy nhiên, khi thực hiện một nghiên cứu phân tích gộp, người nghiên cứu sẽ gặp phải một số thách thức như sau: (i) Thực hiện một nghiên cứu phân tích gộp thường tốn rất nhiều thời gian trong việc tìm kiếm và tổng hợp tài liệu. Ngoài ra, người nghiên cứu cần được trang bị tốt nhiều kỹ năng như: kỹ năng đọc hiểu, tổng hợp tài liệu, và kỹ năng lập trình. Do đó, không dễ dàng để hoàn thành một nghiên cứu phân tích gộp có chất lượng tốt; (ii) Phân tích gộp không thể tiến hành ở tất cả các ngành khoa học. Phân tích gộp chỉ phổ biến ở một số lĩnh vực khoa học như: y học, sinh học, khoa học sức khỏe, tâm lý học và giáo dục học; (iii) Nếu phân tích gộp tổng hợp các kết quả thuộc các chủ đề nghiên cứu quá khác nhau thì kết quả cỡ tác động tổng hợp sẽ không có nhiều ý nghĩa khoa học; (iv) Nếu phân tích gộp sử dụng đầu vào có một (hoặc một vài) nghiên cứu độc lập có chất lượng kém thì kết quả sẽ bị sai lệch và không chính xác (garbage in, garbage out). Do đó, trong bước thiết kế nghiên cứu, cần phải xác định rất rõ các tiêu chí lựa chọn nghiên cứu độc lập để làm đầu vào cho phân tích gộp để có thể tìm kiếm được những nghiên cứu có chất lượng tốt nhất. Nếu người nghiên

cứ không tìm được tất cả những nghiên cứu độc lập có cùng chủ đề nghiên cứu làm đầu vào, thì kết luận của phân tích gộp có thể bị sai lệch; (v) Một trong những yêu cầu khi thực hiện phân tích gộp là người nghiên cứu phải thu thập được tất cả những tài liệu hoặc kết quả nghiên cứu từ nhiều nguồn khác nhau để đánh giá một giả thuyết nghiên cứu. Mặc dù vậy, thực tế là một nghiên cứu phân tích gộp chỉ sử dụng một số ít các nghiên cứu độc lập (hầu hết là những bài báo khoa học được công bố trên các tạp chí chuyên ngành). Ngoài ra, còn có nhiều nghiên cứu cùng chủ đề nhưng chưa được công bố, hoặc không được công bố trên các tạp chí chuyên ngành. Do đó, khi thực hiện phân tích gộp, người nghiên cứu phải nỗ lực tối đa trong việc tìm kiếm những tài liệu (đã công bố và chưa công bố) có liên quan đến hướng nghiên cứu; (vi) Sai lệch xuất bản (publication bias) là hiện tượng các nhà xuất bản thường có xu hướng xuất bản những nghiên cứu chứng minh được tác động tích cực của một can thiệp. Những nghiên cứu chỉ ra tác động không rõ ràng của một can thiệp thường không được ưu tiên xuất bản, hoặc thậm chí không được tác giả gửi đến nhà xuất bản. Do đó, nếu chúng ta thực hiện phân tích gộp với dữ liệu đầu vào hoàn toàn là những bài báo khoa học đã được công bố thì kết quả của phân tích gộp có thể sẽ không phản ánh đúng và đánh giá cao hơn (overestimate) hiệu quả thật của can thiệp.

2. Quy trình thực hiện phân tích gộp trong đánh giá hiệu quả tác động sự phạm của một hướng nghiên cứu khoa học giáo dục

Nghiên cứu về quy trình các bước thực hiện một nghiên cứu phân tích gộp có thể được tìm thấy ở nhiều bài báo đã công bố (Basu, 2017; Tawfik và các cộng sự, 2019; Hansen và các cộng sự, 2022), tuy nhiên, chủ yếu trình bày về quy trình thực hiện phân tích gộp áp dụng cho lĩnh vực y tế và khoa học quản lý. Mặc dù không có quá nhiều khác biệt trong quy trình thực hiện một phân tích gộp ở các lĩnh vực khác nhau, trong bài viết này, chúng tôi đề xuất một quy trình 7 bước cơ bản khi thực hiện một nghiên cứu phân tích gộp, áp dụng cho ngành khoa học giáo dục như sau:

Bước 1. Xác định câu hỏi nghiên cứu cho phân tích gộp.

Tương tự như tất cả các nghiên cứu khoa học khác, việc đầu tiên khi tiến hành một nghiên cứu phân tích gộp là xác định câu hỏi nghiên cứu. Khi xác định câu hỏi nghiên cứu, có hai vấn đề mà người nghiên cứu cần lưu tâm. Thứ nhất, người nghiên cứu cần tìm hiểu xem hướng nghiên cứu đó có phát triển trong vài thập kỷ gần đây, có nhiều bài báo được công bố không? Một mặt, nhiều bài báo đã được công bố sẽ giúp tăng khả năng người nghiên cứu tìm được các công bố có liên quan, qua đó, tăng khả năng có thể thực hiện thành công nghiên cứu phân tích gộp. Mặt khác, việc tìm kiếm, phân loại và lựa chọn từ một số lượng lớn tài liệu để tìm ra các nghiên cứu có liên quan để sử dụng cho phân tích gộp là một công việc tốn rất nhiều thời gian của người nghiên cứu. Do đó, Steel và các cộng sự (2021) đã nhấn mạnh tầm quan trọng của việc phải cân bằng giữa số lượng tài liệu có liên quan và khả năng quản lý của người nghiên cứu trong việc định nghĩa câu hỏi nghiên cứu. Thứ hai, số lượng bài báo về phân tích gộp ở tất cả các lĩnh vực đều tăng mạnh trong những năm gần đây, do đó, có khả năng là một (hoặc một vài) nghiên cứu phân tích gộp có cùng chủ đề (hoặc chủ đề gần) với câu hỏi nghiên cứu mà tác giả định thực hiện đã được công bố. Do đó, người nghiên cứu phải xác định được hướng nghiên cứu về phân tích gộp mình đang quan tâm đã được những nhà khoa học khác triển khai hay chưa. Với những hướng nghiên cứu đã có các công bố về phân tích gộp, người nghiên cứu hoàn toàn có thể thực hiện lại để kiểm tra các kết luận từ các bài báo đã công bố (trong trường hợp nghi ngờ tính chính xác của các kết quả đã công bố), hoặc cập nhật các kết quả mới sử dụng những nghiên cứu độc lập mới được công bố. Việc cập nhật các kết quả phân tích gộp thường xuyên là rất quan trọng, vì nó cung cấp nhiều thông tin quan trọng cho cộng đồng khoa học.

Bước 2. Tìm kiếm các nghiên cứu có liên quan.

Quy trình tìm kiếm tài liệu khi thực hiện phân tích gộp cần phải có tính hệ thống và minh bạch để có thể tìm được tất cả các tài

liệu có liên quan, và đặc biệt là phải có khả năng tái lập kết quả nghiên cứu. Theo Harari và các cộng sự (2020), khi thực hiện việc tìm kiếm tài liệu cho phân tích gộp, người nghiên cứu có thể tham khảo một số chiến lược tìm kiếm sau. Thứ nhất, người nghiên cứu có thể sử dụng những kết quả phân tích gộp đã công bố có cùng chủ đề hoặc chủ đề có liên quan để lọc ra một danh sách những nghiên cứu độc lập có thể sử dụng, cũng như để hiểu kỹ hơn về chủ đề đang nghiên cứu. Tuy nhiên, chiến lược tìm kiếm này chỉ có thể áp dụng khi đã có những công bố phân tích gộp có cùng chủ đề nghiên cứu được xuất bản. Chiến lược tìm kiếm thứ hai, phổ biến hơn, đó là tìm kiếm tài liệu thông qua tìm kiếm từ khóa (keyword search) trên các cơ sở dữ liệu (CSDL) điện tử. Chiến lược tìm kiếm này sẽ giúp người nghiên cứu tìm được tối đa số lượng tài liệu có liên quan, bao gồm cả “tài liệu xám” (grey literature) chưa được xuất bản. Để loại bỏ sự sai lệch kết quả tìm kiếm gây ra do sự giới hạn phạm vi bao phủ của một CSDL, người nghiên cứu được khuyến nghị nên sử dụng ít nhất hai CSDL khác nhau khi tìm kiếm tài liệu (DeSimone và các cộng sự, 2020). Có 5 CSDL khoa học phổ biến nhất và thuận tiện nhất cho việc tìm kiếm những bài báo có liên quan ở hầu hết tất cả những lĩnh vực khoa học phổ biến trước khi thực hiện phân tích gộp bao gồm: Scopus, Web of Science (WoS), Google Scholar (GS), Microsoft Academic (MA) và Dimensions. Đồng thời, còn có các CSDL đặc thù cho các ngành khoa học khác như: y học, sinh học (PubMed, Medline), tâm lý học (PsycINFO), kinh tế học, và giáo dục học (ERIC). Để hạn chế việc có quá nhiều kết quả tìm kiếm, gây khó khăn cho công đoạn xử lý dữ liệu, người nghiên cứu phải xác định chính xác câu hỏi nghiên cứu và các từ khóa được sử dụng để tìm kiếm dữ liệu. Ngoài ra, việc thu thập những tài liệu có liên quan nhưng chưa được công bố, hoặc chỉ được công bố ở các hội thảo khoa học quốc gia, quốc tế, hoặc thu thập tài liệu thông qua mối quan hệ cá nhân của người nghiên cứu với những nhà khoa học khác cũng là một cách để có thể tìm kiếm

được tất cả những nghiên cứu có liên quan. Cần lưu ý rằng kết quả nghiên cứu của những luận án tiến sĩ có chất lượng tốt cũng có thể được sử dụng làm đầu vào cho phân tích gộp. Tuy nhiên, thông thường những kết quả của luận án sẽ được người nghiên cứu công bố trên các tạp chí khoa học chuyên ngành.

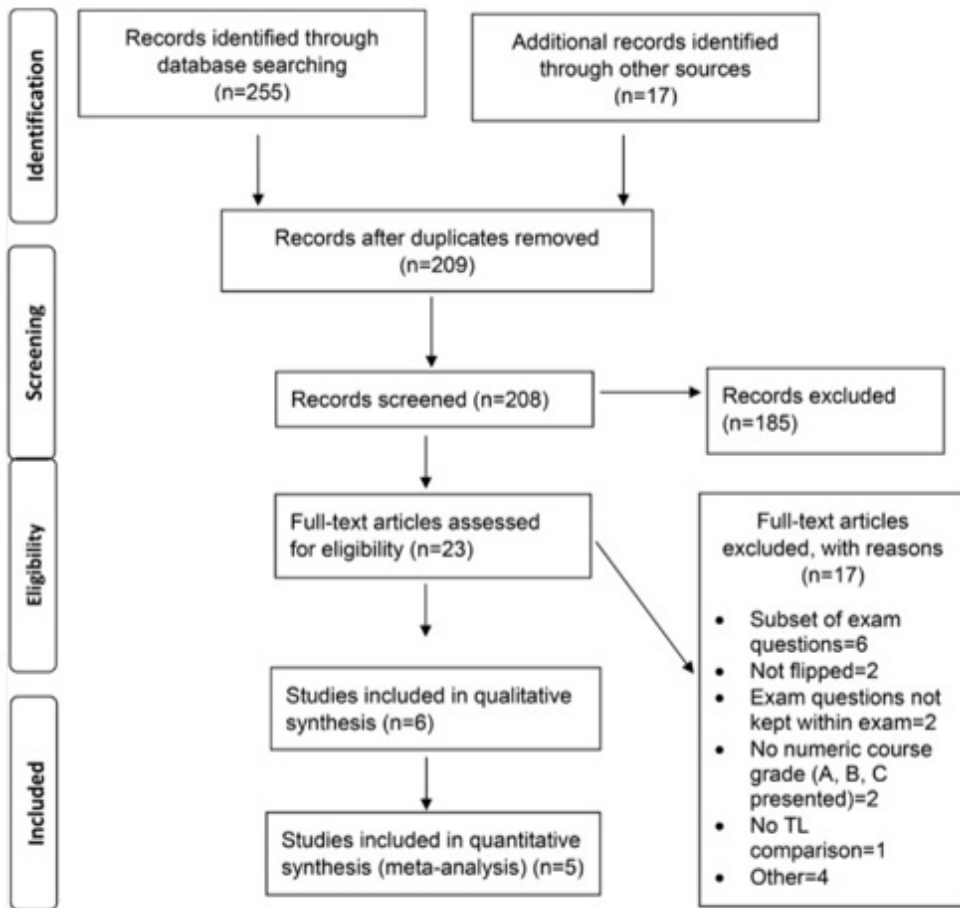
Bước 3. Xây dựng tiêu chí loại bỏ và lựa chọn nghiên cứu độc lập cho phân tích gộp.

Bước tiếp theo, người nghiên cứu phải xây dựng một bảng các tiêu chí để loại bỏ và lựa chọn các nghiên cứu độc lập để làm dữ liệu đầu vào cho phân tích gộp. Lưu ý rằng, bước này cần được thực hiện bởi tối thiểu hai nhà khoa học khác nhau một cách độc lập, sau đó, kết quả cần được so sánh và thảo luận để có thể đưa ra một danh sách cuối cùng những tài liệu sẽ được sử dụng. Các tiêu chí loại bỏ và lựa chọn bài báo đầu vào cho phân tích gộp không cố định, và phụ thuộc chủ yếu vào câu hỏi nghiên cứu và mục đích nghiên cứu. Tuy nhiên, một số tiêu chí sau đây sẽ giúp người nghiên cứu loại bỏ được đa phần những tài liệu không liên quan. Đầu tiên, phải loại bỏ những bài báo trùng lặp ở các CSDL được sử dụng để tìm kiếm. Tiếp theo, người nghiên cứu cần đọc tiêu đề, tóm tắt và từ khóa của những bài báo còn lại để loại bỏ những bài báo rõ ràng không có liên quan đến câu hỏi nghiên cứu (đa số tài liệu sẽ bị loại bỏ ở bước này). Ở bước này, nếu có sự mâu thuẫn trong quyết định giữa hai người thực hiện, cần phải đọc toàn văn tài liệu để quyết định giữ lại hoặc loại bỏ nghiên cứu. Với các bài báo còn lại, người nghiên cứu phải đọc toàn văn để nắm được nội dung của bài báo, sau đó, đưa ra quyết định giữ lại hoặc loại bỏ tài liệu. Cuối cùng, số lượng bài báo còn lại sẽ được sử dụng làm đầu vào cho phân tích gộp. Trong lĩnh vực khoa học giáo dục, một số tiêu chí sau thường được sử dụng để giữ lại hoặc loại bỏ bài báo: (1) *Thiết kế nghiên cứu*: bài báo phải là một nghiên cứu thực nghiệm sư phạm hoặc nghiên cứu can thiệp; (2) *Nguồn công bố*: bài báo phải được công bố trên các tạp chí khoa học uy tín, có phản biện; (3) *Đối tượng nghiên cứu của bài báo*: học sinh tiểu học,

trung học cơ sở, trung học phổ thông, hoặc sinh viên đại học; (4) *Cỡ mẫu*: cỡ mẫu của nghiên cứu phải đủ lớn để sai số lấy mẫu nằm trong ngưỡng cho phép, và kết quả của nghiên cứu có ý nghĩa thống kê. Theo nghiên cứu của Lin (2018), cỡ mẫu tối thiểu của một nghiên cứu sử dụng trong phân tích gộp cần nằm trong khoảng từ 5-10; Nghiên cứu áp dụng cho hai nhóm đối tượng: nhóm được tác động và nhóm không được tác động; Môn học được tiến hành thực nghiệm (Toán học, Vật lý, Hóa

học, Khoa học, Tiếng Anh, Ngoại ngữ, và các môn học không phải STEM); Thời gian tiến hành thực nghiệm sư phạm. Để tóm tắt các công đoạn thực hiện trong bước này, sơ đồ luồng PRISMA có thể được sử dụng. Hình 1 là một ví dụ về sơ đồ luồng PRISMA được sử dụng trong một phân tích gộp đánh giá về mức độ hiệu quả của lớp học đảo ngược (Gillette và các cộng sự, 2018). Từ sơ đồ này, chúng ta có thể hiểu được các bước nhóm tác giả đã thực hiện để tìm kiếm tài liệu:

Hình 1. Ví dụ về sơ đồ luồng PRISMA được sử dụng trong phân tích gộp về mức độ hiệu quả của lớp học đảo ngược (Gillette và các cộng sự, 2018)



Bước 4. Lựa chọn phần mềm thực hiện phân tích gộp.

Các phần mềm hỗ trợ thực hiện phân tích gộp rất đa dạng, từ những hàm tích hợp sẵn (built-in function) đến các gói hỗ trợ được cài đặt trong các phần mềm thống kê, và các phần mềm được thiết kế chuyên dụng

cho phân tích gộp (cả phần mềm thương mại và phần mềm mã nguồn mở). Việc lựa chọn phần mềm để thực hiện phân tích gộp phụ thuộc nhiều vào mức độ phức tạp của phương pháp phân tích và dữ liệu đầu vào sử dụng. Các chức năng và ứng dụng của các phần mềm là không giống nhau, do đó,

người nghiên cứu phải biết được phần mềm định sử dụng có khả năng thực hiện những phân tích mà họ muốn thực hiện hay không. Trong số những phần mềm thương mại, Stata (từ phiên bản v.16) cung cấp các chức năng thực hiện phân tích gộp và vẽ đồ thị. Phần mềm SPSS và SAS cũng cung cấp một số chức năng cho phép thực hiện phân tích gộp. Phần mềm mã nguồn mở như JASP và R cung cấp nhiều gói chức năng giúp thực hiện phân tích gộp. Ví dụ, gói công cụ meta và metaphor trong phần mềm R được trang bị hầu hết các chức năng cơ bản và cần thiết để thực hiện phân tích gộp, và thường được khuyến nghị cho những người mới tìm hiểu về phân tích gộp (Viechtbauer, 2010). Ngoài ra, Microsoft Excel cũng được trang bị một số chức năng giúp thực hiện một vài kỹ thuật phân tích gộp đơn giản, ví dụ như gói Meta-Essentials (Suurmond và các cộng sự, 2017). Cuối cùng, Comprehensive Meta-Analysis, Rayyan.AI và RevMan là những phần mềm thương mại, được phát triển riêng cho nghiên cứu phân tích gộp.

Bước 5. Mã hóa dữ liệu.

Sau khi lựa chọn được những nghiên cứu độc lập để sử dụng làm đầu vào cho phân tích gộp, người nghiên cứu cần mã hóa dữ liệu của từng nghiên cứu độc lập này. Mã hóa dữ liệu phục vụ hai mục đích chính. Thứ nhất, mã hóa dữ liệu giúp người nghiên cứu làm rõ được bối cảnh nghiên cứu, đối tượng được tiến hành nghiên cứu, và phương pháp nghiên cứu được sử dụng trong từng bài báo, qua đó, người nghiên cứu có thể hiểu được hạn chế của từng nghiên cứu. Ví dụ, nếu các nghiên cứu độc lập chỉ được tiến hành với học sinh tiểu học thì các kết luận có được khi thực hiện phân tích gộp chỉ đúng với học sinh tiểu học mà không thể áp dụng cho học sinh trung học cơ sở hoặc trung học phổ thông. Thứ hai, mã hóa dữ liệu giúp xác định được sự biến thiên của cỡ tác động, bối cảnh, số lượng mẫu, cũng như những đặc điểm khác của từng nghiên cứu. Người nghiên cứu cần sử dụng những số liệu này để tính toán cỡ tác

động tổng hợp. Nhiều nhà khoa học thường sử dụng phần mềm Microsoft Excel cho bước mã hóa dữ liệu.

Bước 6. Thực hiện các kỹ thuật phân tích gộp cơ bản.

Kỹ thuật phân tích đầu tiên cần thực hiện đó là phân tích các giá trị ngoại lai (nếu có) trong các nghiên cứu độc lập được sử dụng để đảm bảo chất lượng của kết quả phân tích gộp. Một vài mô hình thống kê có thể được sử dụng để phân tích các giá trị ngoại lai (Aguinis và các cộng sự, 2013; Viechtbauer & Cheung, 2010). Tuy nhiên, hiện nay, các nhà khoa học vẫn đang tranh luận về việc giữ lại hay loại bỏ những giá trị ngoại lai này khi thực hiện phân tích gộp. Trong một số trường hợp, các giá trị ngoại lai này là hợp lệ, tuy nhiên, nhiều trường hợp khác, các giá trị ngoại lai này lại không hợp lệ. Do đó, người nghiên cứu cần so sánh kết quả thực hiện phân tích gộp có sử dụng và không sử dụng các giá trị ngoại lai, sau đó, mới đưa ra kết luận giữ lại hoặc loại bỏ các nghiên cứu độc lập có các giá trị ngoại lai này.

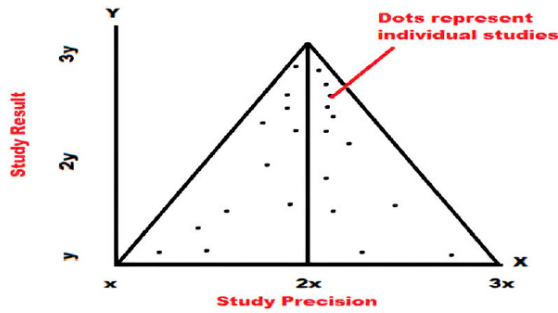
Tiếp theo, sai lệch xuất bản (publication bias) là hiện tượng khả năng được chấp nhận đăng của một nghiên cứu phụ thuộc vào kết quả đầu ra của nghiên cứu đó. Điều này xảy ra do các nhà xuất bản, các tạp chí có xu hướng ưu tiên chấp nhận đăng những nghiên cứu mà kết quả của nghiên cứu đó chứng minh được một giả thuyết ban đầu, hoặc cho thấy mức độ hiệu quả cao của một phương pháp can thiệp đến học sinh (Scherer RW & von Elm, 2007). Những nghiên cứu có kết luận mức độ hiệu quả thấp, hoặc phương pháp can thiệp không có tác động lên học sinh thường không được ưu tiên công bố, công bố chậm, hoặc thậm chí nhiều tác giả không gửi bản thảo của những nghiên cứu này đến tạp chí. Theo thống kê của Dickersin và các cộng sự (1987), trong lĩnh vực y học lâm sàng, những bài báo chỉ ra mức độ hiệu quả cao của một can thiệp sẽ có khả năng được xuất bản cao hơn ba lần so với những bài báo kết luận mức độ hiệu quả của một phương pháp can thiệp không có ý

ngiã thõng kê. Tuy nhiõn, sai lách xuấ bản chỉ là một trong số nhữg yếu tũ tác ðõng ðến kết quả và chất lượng của phân tích gõp. Các yếu tũ khác bao gồm: (i) Số lượng trích ðẫn (citation bias): Kẽ cả khi ðã ðược công bố, nhữg bài báo kết luận một phương pháp can thiệp không có tác ðõng, hoặc có tác ðõng tiêu cực lên học sinh sẽ ít ðược trích ðẫn trong các công bố sau này. ðiều này sẽ khiến việc tìm kiếm nhữg công bố này khó khăn hơn so với nhữg công bố có nhiều trích ðẫn. (ii) Thời gian xuấ bản (time-lag bias): Nhữg nghiên cứu chứng minh kết quả tích cực của một can thiệp thường ðược công bố sớm hơn so với nhữg nghiên cứu có kết quả không thuận lợi. (iii) Ngõn ngữ sử dụng (language bias): Phần lớn công bố khoa học ở tất cả các lĩnh vực nghiên cứu ðều ðược xuấ bản sử dụng ngôn ngữ tiếng Anh. Do ðó, người nghiên cứu thông thường sẽ không hoặc ít tiếp cận với nhữg công bố ðược viết bằng các ngôn ngữ khác tiếng Anh. (iv) Sai lách kết quả ðược ðưa vào công bố (outcome reporting bias): Khi tiến hành một nghiên cứu, người nghiên cứu có thể ðo lường và ðánh giá tác ðõng của một phương pháp can thiệp lên nhiều tiêu chí khác nhau của học sinh. Tuy nhiõn, trong một số trường hợp, người nghiên cứu chỉ báo cáo nhữg tác ðõng tích cực và không ðưa nhữg tác ðõng tiêu cực của can thiệp vào nội dung bài báo công bố.

Từ nhữg phân tích trên, chúng ta có thể thấy rằng sai lách xuấ bản và các yếu tũ khác có thể gây ảnh hưởng lớn ðến kết quả và sự chính xác của một nghiên cứu phân tích gõp. Do ðó, người nghiên cứu cần phải thực hiện ðánh giá sai lách xuấ bản. ðể ðánh giá có hay không sự sai lách xuấ bản trong tập hợp nhữg nghiên cứu ðộc lập ðược sử dụng, thông thường người nghiên cứu sẽ sử dụng biểu ðồ phễu (funnel plot). Biểu ðồ phễu là một biểu ðồ phân bố của các nghiên cứu ðộc lập, cỡ tác ðõng và ðộ chính xác của các nghiên cứu này. Hình 2 minh họa một biểu ðồ phễu, trong ðó, mỗi chấm tròn ðại ðiện cho một nghiên cứu ðộc lập ðược

sử dụng trong phân tích gõp. Trục hoành (trục x) thể hiện cỡ tác ðõng, trong khi trục tung (trục y) thể hiện sai số chuẩn của từng nghiên cứu. Nhữg nghiên cứu quan trọng, có tác ðõng lớn sẽ phân bố phía trên ðỉnh của hình tam giác, trong khi nhữg nghiên cứu ít quan trọng hơn sẽ phân bố ở phía ðưới ðáy của hình tam giác. Thông thường, trục tung của biểu ðồ phễu sẽ bị ðảo ngược, nghĩa là, nhữg nghiên cứu phân bố ở phía trên hình tam giác sẽ có giá trị sai số thấp. Trong trường hợp không có sai lách xuấ bản, các ðiểm dữ liệu trong biểu ðồ phễu sẽ tạo thành một hình phễu ngược và ðối xứng gần ðúng với nhau. Các nghiên cứu nằm ở phía trên (sai số thấp) sẽ nằm gần nhau, trong khi các nghiên cứu nằm ở phía ðưới (sai số tăng dần) ðược phân bố ðều và mở rộng ra cả hai phía của biểu ðồ phễu. Mặc ðù có tính trực quan cao, tuy nhiõn, việc phân tích sai lách xuấ bản chỉ ðựa vào ðánh giá ðịnh tính biểu ðồ phễu còn có nhiều hạn chế, ðặc biệt trong trường hợp số lượng nghiên cứu ðộc lập ít (nhỏ hơn 10 nghiên cứu) (Sedgwick, 2013). Ngoài ra, không có một quy tắc cụ thể nào ðể có thể ðưa ra một kết luận chính xác về sai lách xuấ bản trong trường hợp kết quả của biểu ðồ phễu quá bất ðối xứng. Sự bất ðối xứng này có thể bị gây ra bởi các yếu tũ khác như các yếu tũ liên quan ðến số lượng trích ðẫn, thời gian xuấ bản, ngôn ngữ sử dụng, hoặc sai lách kết quả ðược ðưa vào bài báo. Do ðó, bên cạnh việc sử dụng biểu ðồ phễu, người nghiên cứu cần ðánh giá sai lách xuấ bản sử dụng nhữg con số ðịnh lượng thông qua kiểm ðịnh hồi quy Egger (Egger's Regression Test), kiểm ðịnh Trim-và-Fill (Trim-and-Fill test), và kiểm ðịnh không an toàn N (fail-safe N). Người ðọc có thể tìm hiểu thêm về nhữg kiểm ðịnh này tại nghiên cứu của Nakagawa và các cộng sự (2022). Lưu ý rằng người nghiên cứu không nên chỉ sử dụng duy nhất một kiểm ðịnh, mà nên sử dụng cùng lúc nhiều kiểm ðịnh ðể có thể ðánh giá ðược chính xác nhất có hay không sai lách xuấ bản (Harrison và các cộng sự, 2014).

Hình 2. Ví dụ về biểu đồ phễu (funnel plot) giúp đánh giá sai lệch suất bản trong phân tích gộp



Nguồn: <https://www.statisticshowto.com/funnel-plot/>

Các nghiên cứu độc lập sử dụng cho phân tích gộp được thực hiện bởi những nhóm nghiên cứu khác nhau, ở những thời gian và địa điểm khác nhau, với những đối tượng nghiên cứu khác nhau, và thời gian chịu tác động khác nhau. Do đó, sự không đồng nhất về kết quả của những nghiên cứu này là điều được mong đợi. Tuy nhiên, người nghiên cứu cần xác định xem mức độ không đồng nhất giữa các nghiên cứu độc lập này có thể chấp nhận được hay không trước khi tiến hành các bước tiếp theo của phân tích gộp. Ngoài ra, việc đánh giá tính không đồng nhất của các nghiên cứu độc lập còn rất quan trọng vì điều này ảnh hưởng đến quyết định lựa chọn mô hình thống kê sẽ được sử dụng ở các bước tiếp theo (Huedo-Medina và các cộng sự, 2006). Một vài phương pháp thống kê có thể được sử dụng để đánh giá tính không đồng nhất, bao gồm: kiểm định Cochran's Q, kiểm định I², và phân tích hồi quy meta (meta-regression). Kiểm định Cochran's Q (kiểm định Q) là một kiểm định truyền thống, thường xuyên được sử dụng để xác định tính không đồng nhất trong phân tích gộp. Dựa trên phân bố chi bình phương (chi-square), kiểm định Q tính toán một giá trị xác suất. Giá trị xác suất này càng lớn cho thấy tính không đồng nhất càng cao giữa các nghiên cứu độc lập được sử dụng. Một hạn chế của kiểm định Q là kết quả sẽ không chính xác nếu số lượng nghiên cứu độc lập được sử dụng ít hơn 10. Do đó, người nghiên cứu được khuyến cáo nên sử dụng giá trị p lớn ($p > 0.05$) làm ngưỡng có ý nghĩa thống kê khi sử dụng kiểm định Q để xác định tính không đồng nhất. Kiểm định I² là một phương pháp mới được sử dụng

để ước lượng tính không đồng nhất trong phân tích gộp. Kiểm định I² ước tính tỷ lệ phần trăm thay đổi trong kết quả giữa các nghiên cứu độc lập dựa trên sự khác biệt thật sự chứ không phải do ngẫu nhiên. Kiểm định I² tính toán mức độ không đồng nhất bằng cách lấy kết quả của kiểm định Q chia cho số bậc tự do của giá trị Q. Nếu I² bằng 0%, sự không đồng nhất là do tình cờ. I² ít hơn 25% được coi là mức độ không đồng nhất thấp, I² từ 25-50% được coi là mức độ không đồng nhất trung bình, và I² lớn hơn 50% được coi là mức độ không đồng nhất cao. Điểm hạn chế của kiểm định I² là nó chỉ cung cấp một giá trị về tính không đồng nhất của toàn bộ các nghiên cứu độc lập chứ không cung cấp thông tin về yếu tố nào gây ra sự không đồng nhất này. Đây cũng là điểm hạn chế của kiểm định Q. Sau khi chạy kiểm định Q và kiểm định I², phân tích hồi quy meta (meta-regression) có thể được sử dụng để xác định các yếu tố gây ra sự không đồng nhất. Lưu ý rằng hầu hết các phần mềm hỗ trợ thực hiện phân tích gộp đều cung cấp các công cụ để tính toán giá trị của các kiểm định này.

Sau khi hoàn thành đánh giá tác động của giá trị ngoại lai, đánh giá sai lệch xuất bản, và đánh giá tính không đồng nhất, bước tiếp theo người nghiên cứu cần thực hiện là lựa chọn mô hình thống kê sẽ sử dụng cho nghiên cứu phân tích gộp, bao gồm mô hình thống kê cố định (fixed-effects model) hoặc mô hình thống kê ngẫu nhiên (random-effects model). Nếu các nghiên cứu độc lập chỉ khác nhau bởi sai số lấy mẫu (sampling error), nghĩa là, tính không đồng nhất thấp thì mô hình thống kê cố định sẽ được sử dụng. Ngược lại, nếu có các khác

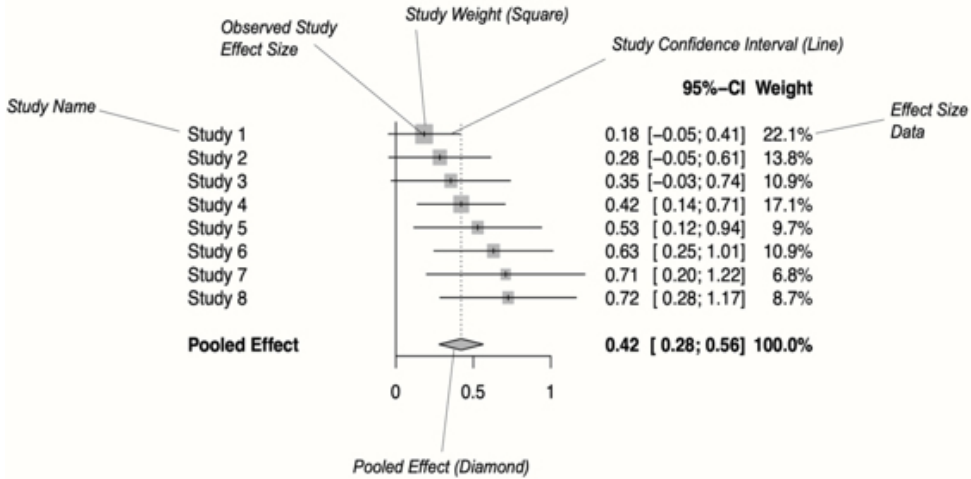
biệt khác ngoài sai số lấy mẫu, nghĩa là, tính không đồng nhất cao thì mô hình thống kê ngẫu nhiên sẽ được sử dụng. Như đã trình bày ở phần trên, người nghiên cứu thường kỳ vọng tính không đồng nhất cao giữa các nghiên cứu độc lập do những nghiên cứu này được thực hiện bởi những nhóm nghiên cứu khác nhau, ở những thời gian và địa điểm khác nhau, và đối tượng được nghiên cứu là khác nhau. Do đó, đa phần những phân tích gộp trong lĩnh vực khoa học giáo dục, và cả trong những lĩnh vực khoa học khác sẽ sử dụng mô hình thống kê ngẫu nhiên. Trong trường hợp tính không đồng nhất thấp, và mô hình thống kê cố định được lựa chọn, một số nhà khoa học vẫn khuyến khích nên sử dụng cả mô hình thống kê ngẫu nhiên (Borenstein và các cộng sự, 2007).

Bước tiếp theo người nghiên cứu cần thực hiện là tính toán cỡ tác động tổng hợp (pooled size) thông qua tổng hợp cỡ tác động (effect size) của từng nghiên cứu độc lập. Cỡ tác động là một giá trị định lượng thể hiện độ lớn về mức độ hiệu quả của một can thiệp. Cỡ tác động càng lớn thì mức độ hiệu quả càng cao. Cỡ tác động thường được biểu diễn bằng ký tự “p”. Theo Cohen (1962), $p \leq 0.1$ tương ứng với mức độ hiệu quả thấp, $p = 0.3$ tương ứng với mức độ hiệu quả trung bình, và $p \geq 0.5$ tương ứng với mức độ hiệu quả cao. Đôi khi, cỡ tác động của các nghiên cứu độc lập đã được các tác giả báo cáo trong bài báo, tuy nhiên, thông thường thì người nghiên cứu phải tự tính toán cỡ tác động của những nghiên cứu độc lập từ những dữ liệu được công bố trong những bài báo này (bao gồm: cỡ mẫu, giá trị trung bình, và độ lệch chuẩn của nhóm đối chứng và nhóm bị can thiệp trước và sau khi can thiệp). Người đọc có thể tìm hiểu thêm về công thức và cách tính cỡ tác động của các nghiên cứu độc lập trong nghiên cứu của Vo và các cộng sự (2017). Cỡ tác động tổng hợp là một trong những kết quả quan trọng nhất của phân tích gộp, do đó, phương pháp tính toán cỡ tác động cần phải thỏa mãn một số điều kiện sau (Harrer và các cộng sự, 2021): (i) Có thể so sánh được: Cỡ tác động của các nghiên cứu độc lập phải có khả năng so sánh được với nhau, và có ý nghĩa

khoa học; (ii) Có thể tính toán được: Cần phải có đủ các tham số cần thiết để có thể tính toán được cỡ tác động của tất cả các nghiên cứu độc lập sử dụng cho phân tích gộp; (iii) Đáng tin cậy: Ngay cả khi có thể tính toán được cỡ tác động của tất cả các nghiên cứu độc lập, sai số chuẩn của cỡ tác động thành phần phải được tính toán trước khi tổng hợp tất cả các tham số này. Một điểm quan trọng nữa cần lưu ý là định dạng của cỡ tác động cần phải phù hợp với kỹ thuật phân tích được sử dụng; (iv) Có thể giải thích được: Sau khi được tính toán từ các nghiên cứu độc lập, cỡ tác động tổng hợp phải phù hợp để trả lời những câu hỏi nghiên cứu đã được đặt ra.

Cỡ tác động của các nghiên cứu độc lập, cỡ tác động tổng hợp và một vài thông số quan trọng khác của phân tích gộp (khoảng tin cậy và trọng số của mỗi nghiên cứu) có thể được biểu diễn thông qua một biểu đồ rừng (forest plot). Hình 3 là một ví dụ về một biểu đồ rừng tiêu chuẩn và các thành phần chính của nó. Cột ngoài cùng bên trái thể hiện tên (và năm công bố) của các nghiên cứu độc lập được sử dụng cho phân tích gộp. Với mỗi nghiên cứu, cỡ tác động được biểu diễn bằng một hình vuông ở trung tâm của biểu đồ rừng. Đường thẳng song song với trục hoành, chạy qua tâm của mỗi hình vuông thể hiện khoảng tin cậy 95% (confidence interval). Kích thước của mỗi hình vuông tỉ lệ thuận với cỡ mẫu của nghiên cứu độc lập. Cỡ mẫu càng lớn thì hình vuông càng lớn, và nghiên cứu đó có trọng số càng cao. Ba cột còn lại trên biểu đồ rừng lần lượt thể hiện giá trị của cỡ tác động, khoảng tin cậy 95%, và trọng số của từng nghiên cứu. Ở phía dưới biểu đồ, hình thoi thể hiện giá trị của cỡ tác động tổng hợp. Hai đầu mút trái và phải của hình thoi thể hiện khoảng tin cậy 95%. Lưu ý rằng nếu hình thoi cắt đường thẳng đứng đi qua điểm 0 thì người nghiên cứu có thể kết luận rằng phương pháp can thiệp không có tác động. Tất cả các kết quả thể hiện trên biểu đồ rừng cần phải rõ ràng, chính xác, và phải có tính tái thiết lập (nghĩa là nếu người nghiên cứu khác sử dụng cùng một bộ dữ liệu đầu vào thì kết quả đầu ra phải giống nhau).

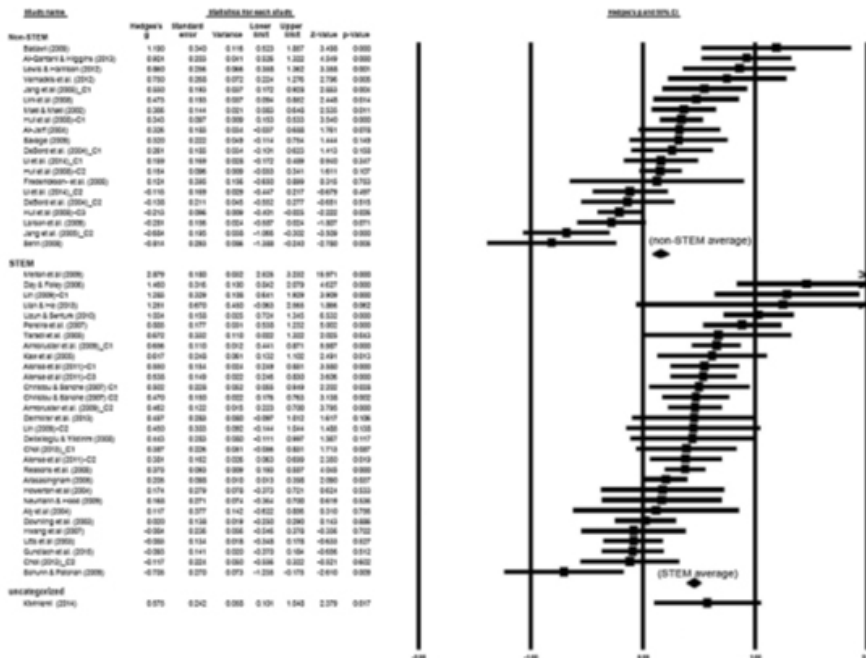
Hình 3. Ví dụ về một biểu đồ rừng tiêu chuẩn (Harrer và các cộng sự, 2021)



Ngoài ra, người nghiên cứu có thể tùy biến cách thể hiện một biểu đồ rừng cho phù hợp với nội dung nghiên cứu. Hình 4 là một ví dụ về một biểu đồ rừng thể hiện cỡ tác động của từng nghiên cứu độc lập, cỡ tác động tổng hợp được tính toán từ 40 nghiên cứu về mức độ hiệu quả của phương pháp dạy học kết hợp (blended learning) tới hiệu quả học tập của sinh viên đại học đang theo học

các ngành học có liên quan và không liên quan đến STEM (Vo và các cộng sự, 2017). Từ biểu đồ rừng ở Hình 4, chúng ta có thể kết luận rằng phương pháp dạy học kết hợp có tác động tích cực đến kết quả học tập của sinh viên đại học, và phương pháp này có tác động tích cực hơn tới sinh viên các ngành học có liên quan đến STEM so với sinh viên các ngành không liên quan đến STEM.

Hình 4. Ví dụ về một biểu đồ rừng trong một nghiên cứu phân tích gộp đánh giá mức độ hiệu quả của phương pháp dạy học kết hợp (blended learning) tới kết quả học tập của sinh viên đại học (Vo và các cộng sự, 2017)



Những kỹ thuật phân tích được trình bày trong bước 6 là những kỹ thuật cơ bản, cần phải có của một bài báo phân tích gộp. Sau khi hoàn thành những phân tích này, người nghiên cứu có

thể tiếp tục thực hiện những kỹ thuật phân tích nâng cao. Trong phạm vi của bài viết này, chúng tôi sẽ không đề cập đến những kỹ thuật nâng cao này. Người đọc có thể tìm đọc thêm tại sách

chuyên khảo về phân tích gộp của Harrer và các cộng sự (2021).

Bước 7. Tổng hợp kết quả và viết báo cáo.

Bước cuối cùng khi thực hiện phân tích gộp là tổng hợp các kết quả đã thực hiện ở các bước trên và viết báo cáo. Điều quan trọng nhất với báo cáo tổng hợp là phương pháp nghiên cứu và tất cả những bước đã được thực hiện phải được trình bày một cách tường minh và dễ hiểu. DeSimone và các cộng sự (2020) đã cung cấp một danh sách những tiêu chí cần phải có khi thực hiện phân tích gộp. Người nghiên cứu có thể dựa vào những tiêu chí này để đảm bảo rằng tất cả những tiêu chí quan trọng nhất khi thực hiện phân tích gộp đã được các tác giả thực hiện. Đối với các phân tích gộp định lượng, chúng ta có thể sử dụng bảng danh sách các tiêu chí được xây dựng bởi Levitt và các cộng sự (2018). Đối với phân tích gộp định lượng, bảng biểu cần có các thông tin quan trọng như cỡ tác động trung bình (mean effect size), sai số chuẩn (standard error) và khoảng tin cậy (confidence interval), cỡ mẫu của từng nghiên cứu độc lập (sample size), và mức độ đồng nhất của các nghiên cứu. Nếu số lượng nghiên cứu độc lập sử dụng cho phân tích gộp không quá nhiều, cần cung cấp một biểu đồ rừng nhằm tổng quan sự khác biệt và độ chính xác của từng nghiên cứu thành phần. Cần lưu ý rằng, việc sử dụng biểu đồ rừng là không khả thi khi có quá nhiều nghiên cứu độc lập trong bộ dữ liệu đầu vào. Tất cả những kết quả được thể hiện trên hình vẽ và bảng biểu cần phải được giải thích rõ ràng trong báo cáo. Quan trọng hơn, người nghiên cứu cần phải trả lời được câu hỏi nghiên cứu đã đưa ra. Ví dụ, phương pháp can thiệp được sử dụng có tác động tích cực, tiêu cực, hoặc không có tác động đến học sinh. Ngoài ra, từ kết quả của phân tích gộp, người nghiên cứu có thể trình bày thêm về

các ý tưởng và các hướng nghiên cứu tiềm năng trong tương lai.

3. Kết luận

Thông thường, các kết quả được trình bày trong một bài báo phân tích gộp có độ tin cậy cao, tuy nhiên, trong một vài trường hợp, kết quả nghiên cứu từ những bài báo phân tích gộp đã được xuất bản có thể không chính xác tuyệt đối, hoặc thậm chí cung cấp các kết luận sai lầm. *Thứ nhất*, kết luận từ một phân tích gộp không đáng tin cậy khi kết quả này được tổng hợp dựa trên nhiều dạng nghiên cứu khác nhau, hoặc phương pháp đo lường kết quả của những nghiên cứu độc lập này không giống nhau, hoặc chủ đề nghiên cứu của những nghiên cứu độc lập quá khác biệt. *Thứ hai*, nghiên cứu phân tích gộp sử dụng kết quả của cả những bài báo có chất lượng tốt và chất lượng không tốt. Điều này sẽ dẫn đến những kết luận sai lầm hoặc thiếu sót. *Thứ ba*, nếu một nghiên cứu phân tích gộp không báo cáo đầy đủ các bước tác giả tiến hành triển khai nghiên cứu, điều này sẽ gây khó khăn cho người đọc trong việc đánh giá chất lượng và độ tin cậy của kết quả nghiên cứu. *Thứ tư*, phân tích gộp sử dụng cùng một bộ dữ liệu các nghiên cứu độc lập làm đầu vào nhưng lại cho ra những kết quả khác nhau. Thứ năm, nếu một trong số những nghiên cứu độc lập được sử dụng làm đầu vào cho phân tích gộp có kích thước mẫu quá vượt trội so với những nghiên cứu khác thì kết quả phân tích gộp có thể bị sai lệch. *Cuối cùng*, nếu một phân tích gộp sử dụng những bài báo có mức độ sai lệch xuất bản lớn (high publication bias) làm dữ liệu đầu vào, thì sẽ dẫn đến những kết luận không chính xác. Từ những vấn đề trên, người nghiên cứu cần phải có đủ kiến thức chuyên môn để có thể đánh giá được chất lượng của một phân tích gộp trước khi quyết định có sử dụng các kết quả này.

Tài liệu tham khảo

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organizational Research Methods*, 16(2), 270–301. <https://doi.org/10.1177/1094428112470848>

Baki, A., Kosa, T., & Guven, B. (2011). A comparative study of the effects of using dynamic geometry software and physical manipulatives on the spatial visualisation skills of pre-service mathematics teachers. *British Journal of Educational Technology*, 42(2), 291–310. <https://doi.org/10.1111/j.1467-8535.2009.01012.x>

- Basu, A. (2017). How to conduct meta-analysis: a basic tutorial. *PeerJ Preprints*, 5:e2978v1. <https://doi.org/10.7287/peerj.preprints.2978v1>
- Borenstein, M., Hedges, L., & Rothstein, H. (2007). *"Meta-analysis: Fixed effect vs. random effects"*. Wiley. https://www.meta-analysis.com/downloads/Meta-analysis_fixed_effect_vs_random_effects_072607.pdf
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology*, 65(3), 145–153. <https://doi.org/doi.org/10.1037/h0045186>
- Cotta, K. I., Shah, S., Almgren, M. M., Macías-Moriarity, L. Z., & Mody, V. (2016). Effectiveness of flipped classroom instructional model in teaching pharmaceutical calculations. *Currents in Pharmacy Teaching and Learning*, 8(5), 646–653. <https://doi.org/10.1016/j.cptl.2016.06.011>
- DeSimone, J. A., Brannick, M. T., O'Boyle, E. H., & Ryu, J. W. (2020). "Recommendations for Reviewing Meta-Analyses in Organizational Research. *Organizational Research Methods*, 24(4), 694–717. <https://doi.org/10.1177/1094428120967089>
- Dickersin, K., Chan, S., Chalmers, T. C., Sacks, H. S., & Smith, H. (1987). Publication bias and clinical trials. *Controlled Clinical Trials*, 8(4), 343–353. [https://doi.org/https://doi.org/10.1016/0197-2456\(87\)90155-3](https://doi.org/https://doi.org/10.1016/0197-2456(87)90155-3)
- Drigas, A., & Pappas, M. (2015). A Review of Mobile Learning Applications for Mathematics. *International Journal of Interactive Mobile Technologies (IJIM)*, 9(3), 18–23. <https://doi.org/10.3991/ijim.v9i3.4420>
- Gillette, C., Rudolph, M., Kimble, C., Rockich-Winston, N., Smith, L., & Broedel-Zaugg, K. (2018). A Meta-Analysis of Outcomes Comparing Flipped Classroom and Lecture. *American Journal of Pharmaceutical Education*, 82(5). <https://doi.org/10.5688/ajpe6898>
- Hamidah, N., Afidah, I., Setyowati, L., Sutini, S., & Junaedi, J. (2020). Pengaruh Media Pembelajaran Geogebra Pada Materi Fungsi Kuadrat Terhadap Motivasi dan Hasil Belajar Peserta Didik. *Journal of Education and Learning Mathematics Research (JELMaR)*, 1(1), 15–24. <https://doi.org/10.37303/jelmar.v1i1.2>
- Hansen, C., Steinmetz, H., & Block, J. (2022). How to conduct a meta-analysis in eight steps: a practical guide. *Management Review Quarterly*, 72(1), 1–19. <https://doi.org/10.1007/s11301-021-00247-4>
- Harari, M. B., Parola, H. R., Hartwell, C. J., & Riegelman, A. (2020). Literature searches in systematic reviews and meta-analyses: A review, evaluation, and recommendations. *Journal of Vocational Behavior*, 118, 103377. <https://doi.org/10.1016/j.jvb.2020.103377>
- Harrer, M., Cuijpers, P., Furukawa, T. ., & Ebert, D. . (2021). *Doing Meta-Analysis with R: A Hands-On Guide (1st ed.)*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781003107347>
- Harrison, J. S., Banks, G. C., Pollack, J. M., O'Boyle, E. H., & Short, J. (2014). Publication Bias in Strategic Management Research. *Journal of Management*, 43(2), 400–425. <https://doi.org/10.1177/0149206314535438>
- Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index?" *Psychological Methods*, 11(2), 193–206. <https://doi.org/10.1037/1082-989X.11.2.193>
- Hughes, P. J., Waldrop, B., & Chang, J. (2016). Student perceptions of and performance in a blended foundational drug information course. *Currents in Pharmacy Teaching and Learning*, 8(3), 359–363. <https://doi.org/10.1016/j.cptl.2016.02.013>
- Juandi, D., Kusumah, Y. S., Tamur, M., Perbowo, K. S., Siagian, M. D., Sulastri, R., & Negara, H. R. P. (2021). The Effectiveness of Dynamic Geometry Software Applications in Learning Mathematics: A Meta-Analysis Study. *International Journal of Interactive Mobile Technologies (IJIM)*, 15(02), 18–37. <https://doi.org/10.3991/ijim.v15i02.18853>
- Lee, Y. H. (2019). Strengths and Limitations of Meta-Analysis. *Korean J Med*, 94(5), 391–395. <https://doi.org/10.3904/kjm.2019.94.5.391>
- Levitt, H. M., Bamberg, M. G., Creswell, J. W., Frost, D. M., Josselson, R., & Suárez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA Publications and Communications Board task force report. *The American Psychologist*, 73 1, 26–46.
- Lin, L. (2018). Bias caused by sampling error in meta-analysis with small sample sizes. *PLOS ONE*, 13(9), 1–19. <https://doi.org/10.1371/journal.pone.0204056>
- Nakagawa, S., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W. A., Parker, T. H., Sánchez-Tójar, A., Yang, Y., & O'Dea, R. E. (2022). Methods for testing publication bias in ecological and evolutionary meta-analyses. *Methods in Ecology and Evolution*, 13(1), 4–21. <https://doi.org/10.1111/2041-210X.13724>

- Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2017). Improving Mathematics Teaching in Kindergarten with Realistic Mathematical Education. *Early Childhood Education Journal*, 45(3), 369–378. <https://doi.org/10.1007/s10643-015-0768-4>
- Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2018). The effectiveness of computer and tablet assisted intervention in early childhood students' understanding of numbers. An empirical study conducted in Greece. *Education and Information Technologies*, 23(5), 1849–1871. <https://doi.org/10.1007/s10639-018-9693-7>
- Scherer RW, L. P., & von Elm, E. (2007). Full publication of results initially presented in abstracts. *Cochrane Database of Systematic Reviews*, 2. <https://doi.org/10.1002/14651858.MR000005.pub3>
- Sedgwick, P. (2013). Meta-analyses: how to read a funnel plot. *BMJ*, 346. <https://doi.org/10.1136/bmj.f1342>
- Septian, A., & Komala, E. (2019). Kemampuan Koneksi Matematik Dan Motivasi Belajar Siswa Dengan Menggunakan Model Problem-Based Learning (Pbl) Berbantuan Geogebra Di Smp. *Prisma*, 8(1), 1. <https://doi.org/10.35194/jp.v8i1.438>
- Steel, P., Beugelsdijk, S., & Aguinis, H. (2021). The anatomy of an award-winning meta-analysis: Recommendations for authors, reviewers, and readers of meta-analytic reviews. *Journal of International Business Studies*, 52(1), 23–44. <https://doi.org/10.1057/s41267-020-00385-z>
- Stone, D. L., & Rosopa, P. J. (2017). The Advantages and Limitations of Using Meta-analysis in Human Resource Management Research. *Human Resource Management Review*, 27(1), 1–7. <https://doi.org/10.1016/j.hrmr.2016.09.001>
- Suda, K. J., Sterling, J. M., Guirguis, A. B., & Mathur, S. K. (2014). Student perception and academic performance after implementation of a blended learning approach to a drug information and literature evaluation course. *Currents in Pharmacy Teaching and Learning*, 6(3), 367–372. <https://doi.org/10.1016/j.cptl.2014.02.017>
- Sudirman, S., Mellawaty, M., Yaniawati, P., & Indrawan, R. (2020). Integrating Local Wisdom Forms in Augmented Reality Application: Impact Attitudes, Motivations and Understanding of Geometry of Pre-service Mathematics Teachers'. *International Journal of Interactive Mobile Technologies (IJIM)*, 14(11 SE-Papers), 91–106. <https://doi.org/10.3991/ijim.v14i11.12183>
- Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison, and validation of Meta-Essentials: A free and simple tool for meta-analysis. *Research Synthesis Methods*, 8(4), 537–553. <https://doi.org/10.1002/jrsm.1260>
- Swanson, H. L., & Hoskyn, M. (1998). Experimental Intervention Research on Students with Learning Disabilities: A Meta-Analysis of Treatment Outcomes. *Review of Educational Research*, 68(3), 277–321. <https://doi.org/10.3102/00346543068003277>
- Tawfik, G. M., Dila, K. A. S., Mohamed, M. Y. F., Tam, D. N. H., Kien, N. D., Ahmed, A. M., & Huy, N. T. (2019). A step by step guide for conducting a systematic review and meta-analysis with simulation data. *Tropical Medicine and Health*, 47(1), 46. <https://doi.org/10.1186/s41182-019-0165-6>
- Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125. <https://doi.org/10.1002/jrsm.11>
- Vo, H. M., Zhu, C., & Diep, N. A. (2017). The effect of blended learning on student performance at course-level in higher education: A meta-analysis. *Studies in Educational Evaluation*, 53, 17–28. <https://doi.org/10.1016/j.stueduc.2017.01.002>