Growth in the Amount of Literature Reviewed in a Meta-Analysis and Reviewer Resources

Michael Harwell
University of Minnesota

Accurate and consistent reviews of documents describing research studies are essential to valid and generalizable inferences in a meta-analysis. Traditionally a small number of reviewers screen studies using the title, abstract, and possibly the full text to determine a study's eligibility for a meta-analysis. This study explores whether reviewing loads and resources to support accurate and consistent reviewing have increased over time. A survey of N = 193 meta-analyses published between 1980 – 2019 showed that the average number of documents reviewed has increased, especially since 2010, but the typical number of reviewers has not changed over the past 40 years. The importance of meta-analysts providing information about reviewers and the review process to help readers evaluate the validity and generalizability of inferences is emphasized. This information would typically include the number of reviewers and their qualifications, number of titles, abstracts, and full-text documents reviewed, time spent reviewing documents, evidence of the accuracy and consistency of reviews, and the role of software in facilitating reviewing.

Introduction

Meta-analysis continues to be an important tool in educational research for quantitatively synthesizing study findings. In a typical meta-analysis a series of steps, each of which informs subsequent steps, are executed based on the general framework described in Cooper (1982):

1. The motivation for the meta-analysis and associated research questions are described.
2. The characteristics of a population of studies are specified and a sample of studies is generated by electronically searching databases, along with non-electronic strategies such as hand-searching relevant journals and references of previous meta-analyses, to collect documents that typically include published journal articles and book chapters, conference papers, technical reports, and dissertations.
3. The title and abstract of studies in the initial sample are screened to determine their eligibility for further review using general inclusion criteria.
4. The full texts of the remaining studies are reviewed using specific inclusion and exclusion criteria to identify the final sample of studies.
5. Relevant characteristics and the results of studies in the final sample are coded (e.g., nature of the treatment, type of research design, effect sizes).
6. Study data are analyzed and the results reported.

Meta-analysts sometimes add to or combine Steps 3 and 4 but the net effect is Step 3 categorizes each study as eligible for further review (yes, no), and Step 4 leads to a decision of whether to include a study in the final sample (yes, no).

Several guidelines for synthesizing literature via Steps 1 through 6 are available (APA Publications and Communications Board Working Group on Journal Article Reporting Standards [i.e., Meta-Analysis Reporting Standards or MARS], 2008; Appelbaum et al., 2018; Moher, Liberati, Tetzlaff, & Altman, 2009 [i.e., Preferred Reporting Items For Systematic Reviews and Meta-Analysis or PRISMA]; Rubio-Aparicio, Sánchez-Meca, Marín-Martínez, & López-López, 2018). A point of agreement among guidelines is the importance of transparency of the conduct of a meta-analysis to allow readers to assess the validity and generalizability of inferences.

Problem Formulation

Cooper's (1982) framework continues to provide guidance to meta-analysts but specific components of this framework have changed over the past 40 years, especially the scope of literature searches. For example, the meta-analysis of White (1982) investigated the relationship between socio-economic status (SES) and student achievement and reviewed 248 titles and abstracts in Step 3, Sirin (2005) reviewed 2,477 titles and abstracts on the same topic in Step 3. Whether increases in reviewing load (average number of document pages read by a reviewer) illustrated in these meta-analyses prompted the use of increased reviewing resources, which include additional reviewers and the use of computer software to facilitate reviewing, is unclear as is the impact on the accuracy and consistency of document reviews.

One important resource for supporting Steps 3 and 4 is computer software which can help to standardize the review process. These programs can support the downloading and screening of abstracts and full-text documents, produce aggregated results of the screening, and provide flexibility in formatting output. Covidence (Veritas Health Innovation, 2019), Distillers (Evidence Partners, 2016), EPPI - Reviewer 4 (Thomas, Brunton, & Graziosi, 2010), and Rayyan (Ouzzani, Hammady, Fedorowicz, & Elmagarmid, 2016) are examples of computer programs that can facilitate Steps 3 and 4 in a meta-analysis.

Still, the most important resource in a meta-analysis continues to be the reviewers who screen the titles and abstracts of an initial sample of studies (Step 3), review full-text documents (Step 4), and may also code study characteristics and effect sizes for the final sample of studies in Step 5. Despite the challenge of satisfactorily completing Steps 3 and 4, especially for large numbers of documents, the focus in meta-analysis guidelines (e.g., MARS, 2008; Orwin & Vevea, 2009; PRISMA, 2009) as well as published meta-analyses (e.g., D'Agostino & Powers, 2009; Ke, Whalon, & Yun, 2018; Kraft, Blazer, & Hogan, 2018), continues to be on Step 5. Empirical evaluations of published meta-analyses such as Ahn, Ames, and Myers (2012) and Valentine, Cooper, Patall, Tyson, and Robinson (2010) have similarly focused on Step 5. If information about reviewing documents in Steps 3 and 4 is provided (and it often is not) it usually takes the form of a statement that authors of a meta-analysis reviewed documents (e.g., Peng, Wang, & Namkur, 2018), graduate students reviewed documents after being trained (e.g., Graham et al., 2018), or unidentified individuals did the reviewing (e.g., Joksimovic et al., 2018).

It is important to emphasize that some meta-analyses provide information about reviewer load as well as evidence of the accuracy and consistency of categorizing studies in Steps 3 and 4. For example, Graham et al. (2018) used a meta-analysis to examine the impact of reading interventions on writing, and reported that when reviewing abstracts "Interrater agreement for

this initial screening was 98%, with the first and last authors differing on 433 entries. Disagreements were resolved by the first author, who had 40 years of experience conducting literacy research" (p. 252) and "Two trained graduate students read each document in full to determine if it met all inclusion and exclusion criteria (agreement was 92%). Disagreements were resolved by the first author." (p. 252) The percent agreement statistic provided evidence of consistency and using the first author's expertise to adjudicate disagreements provided evidence of accuracy. However, more details are needed for two reasons: Providing additional details of Steps 3 and 4 is consistent with recommendations for increased transparency in meta-analyses (MARS, 2008; PRISMA, 2009), and speaks to the concerns of readers who may be skeptical that large numbers of studies were accurately and consistently reviewed.

Purpose

The purpose of this study is to examine reviewing loads and resources to support accurate and consistent reviewing in published meta-analyses to learn whether they have increased over time and if so by how much. The premise is that increases in electronic search capabilities have dramatically increased the number of documents reviewed in Steps 3 and 4 in a meta-analysis, but it is unclear if the resources used to complete these steps (e.g., number of reviewers) have changed over time, which may have important effects on meta-analytic inferences. Increases in the number of documents reviewed may also be partly attributable to increases in the number of educational research studies.

To illustrate these issues consider two meta-analyses. Kulik and Kulik (1982) examined the impact of ability grouping among secondary students that began with four research questions:

1. What are the effects of ability grouping in the typical study?
2. Does grouping have different effects on different types of students?
3. Does grouping have different effects for different types of instructional outcomes?
4. Do the effects of grouping vary as a function of type of study, methodological features, types of settings, and type of grouping practice?

Kulik and Kulik (1982) performed an electronic search of three databases (ERIC, Comprehensive Dissertation Abstracts, Psychological Abstracts), followed by hand-searching the references of studies generated by the computer search. These search strategies produced an initial sample of 700 studies which were screened for their eligibility for the meta-analysis using inclusion criteria specified by the authors (Step 3). Step (3) reduced the number of studies eligible for further review to 180 which were subsequently retrieved and reviewed in their full-text form (Step 4). Ultimately 52 studies were included in the Kulik and Kulik (1982) synthesis. Who reviewed documents in Steps 3 and 4 was not stated.

Next consider Graham et al.'s (2018) meta-analysis of the impact of reading interventions on writing. These authors began by specifying the question(s) motivating the meta-analysis:

1. Does teaching reading enhance writing performance?
2. Does increasing students' interaction with words or text through reading or observing others read enhance writing performance?

Graham et al. (2018) then provided details of their literature search which included several databases (ERIC, PsychINFO, ProQuest Dissertations & Theses, Global, Linguistics and Language Behavior Abstracts, EBSCOhost) and hand-searching relevant journals, technical reports, and references in previous meta-analyses. These search strategies produced 17,301 titles and abstracts which were screened in Step 3 by two reviewers using several inclusion criteria (e.g., study used a reading intervention group that was part of a true- or quasi-experiment design, at least one writing assessment evaluated the impact of the reading intervention) to identify 677 studies for further review. Exclusion criteria (e.g., attrition exceeded 20% for the reading intervention condition) applied in Step (4) by two reviewers to the 677 full-text documents produced a final sample of 89 studies.

Perhaps the defining characteristic of a meta-analysis is its reading-intensive nature, which emerges in Steps (3) and (4). Kulik and Kulik (1982) stated that their initial sample of 700 studies was screened for eligibility for the meta-analysis but provided no details. Suppose the screening relied primarily on reading the abstract of each study to determine its eligibility for further consideration and that this was done by the authors. The number of words in an abstract varies somewhat across journals, for example, Educational Researcher (ER), Educational Evaluation and Policy Analysis (EEPA), the American Educational Research Journal (AERJ), and Review of Educational Research (RER) currently limit abstracts to 120, 120, 120, and 150 words, respectively. If the abstracts reviewed by Kulik and Kulik (1982) are assumed to consist of 120 words these authors read approximately 700 x 120 = 84,000 words, which is the equivalent of about 335 double-space pages (assuming an 8.5" x 11" page with 1" margins and about 250 words per page).

Kulik and Kulik (1982) then retrieved the full text of 180 of the 700 studies for further review. Manuscript length also varies across journals, for example, the current maximum manuscript lengths for ER, EEPA, AERJ, and RER are 20, 45, 50 (counting references), and 50 double-spaced pages, respectively (assuming an 8.5" x 11" page with 1" margins and about 250 words per page). If one-third of the manuscripts read by Kulik and Kulik were 20 pages, one-third were 45 pages, and one-third were 50 pages the total reviewing load consisted of approximately 335 + 6,900 = 7,235 pages in their meta-analysis. Naturally the exact reviewer load will depend on the length of full-text documents and their representation in the sample of studies. For example, the Kulik and Kulik (1982) value of 7,235 pages read is likely an overestimate if a significant percentage of full-text documents were shorter technical reports of 3,000 words and an underestimate if a significant percentage were dissertations.

Now consider the Graham et al. (2018) meta-analysis. If the 17,301 abstracts each consisted of 120 words the two authors who reviewed abstracts read approximately 8,300 pages. If the distribution of manuscript length was again trimodal (20, 45, 50 pages) in the sample of 677 retrieved studies the two graduate students reviewed approximately 26,000 pages, and for the meta-analysis as a whole the reviewing load was approximately 8,300 + 26,000 = 34,300 pages. The difference in electronic search capabilities almost certainly explains much of the difference in the reviewing load of the Kulik and Kulik (1982) and Graham et al. (2018) meta-analyses; what is less clear is how widespread this pattern is in educational meta-analyses and whether resources supporting reviewing in Steps 3 and 4 have increased with increased reviewer loads.

Increases in electronic search capabilities and reviewer loads illustrated by the Kulik and Kulik (1982) and Graham et al. (2018) meta-analyses suggest two important research questions for the present study:

1. Has the reviewing load in Steps 3 and 4 in educational meta-analyses increased over time and if so by how much?
2. Have resources that support accurate and consistent reviewing in Steps 3 and 4 increased over time and if so by how much?

Method

To explore these questions the author conducted an empirical review of published meta-analyses to examine reviewer load and resources for Steps 3 and 4 over time. Four journals sponsored by the American Educational Research Association (ER, EEPA, AERJ, RER) were searched for meta-analyses, systematic reviews, and research syntheses published between 1980 – 2019 (from here on the term meta-analysis is used for simplicity). The variables coded for each meta-analysis appearing in the four journals included the year a study appeared, number of study authors, journal (ER, EEPA, AERJ, RER), number of abstracts reviewed, number of reviewers of abstracts, number of full-text documents reviewed, number of reviewers of full-text documents, percent agreement or reliability statistics among reviewers provided for Steps 3 and/or 4, and whether software was used to manage documents and facilitate reviews. This information was used to estimate the mean reviewer load for abstracts (assuming 120 words per abstract and 250 words per page) and full-text documents (assuming 20 pages for meta-analyses appearing in ER, 45 pages for those appearing in EEPA, and 50 pages for meta-analyses appearing in AERJ and RER all of which assumed 250 words per page), which were added together to estimate mean reviewer load. Reviewer load was then computed for each journal across four time periods (1980 – 1989, 1990 – 1999, 2000 – 2009, 2010 - 2019) chosen to capture change.

Results

The findings for abstracts and full-text documents are summarized in Tables 1 and 2 by journal and time based on a sample of $N = 193$ meta-analyses. One distinct pattern in these tables is that ER has rarely published meta-analyses ($n = 3$), EEPA occasionally publishes meta-analyses ($n = 8$), AERJ used to publish meta-analyses ($n = 11$ between 1980 – 2009) but no longer does, and RER did not publish meta-analyses between 1980 – 1999 but since then typically does ($n = 171$). The findings in the tables should also be interpreted in light of the potential impact of missing data. Table 1 shows that for 1980 – 1989, 1990 - 1999, 2000 – 2010, and 2010 - 2019 approximately 50%, 100%, 67.3%, and 75.6%, respectively, of the meta-analyses reported the number of abstracts reviewed but across the four time periods 0%, 0%, 30.6%, and 44.9%, respectively, of these meta-analyses reported the number of abstract reviewers. Table 2 shows a similar pattern for full-text documents. Across the four time periods approximately 92.8%, 100%, 61.2%, and 74%, respectively, of the meta-analyses reported the number of full-texts documents reviewed but the percentage of these meta-analyses reporting the

number of full-text reviewers over the four time periods was 7.1%, 0%, 28.6%, and 45.3%, respectively. Hence reviewer load may be somewhat more or less than suggested by the findings in Tables 1 and 2.

The findings in Tables 1 and 2 suggest three patterns that speak to the research questions. First, there is a trend over time toward reviewing larger numbers of abstracts and full-text documents. For example, ER and EEPA saw the mean number of abstracts reviewed jump from 1,415 between 1980 – 2009 to more than 15,000 in 2010 - 2019, and RER saw the mean number of abstracts reviewed in 1990 – 1999, 2000 – 2009, and 2010 – 2019 increase from 899 to 1,808 to 3,883, respectively. The mean number of full-text documents reviewed in RER also grew from 37 in 1980 - 1989 to 111, 265, and 366 in the three subsequent time periods.

Second, the growth in numbers of abstracts and full-text documents reviewed increased the mean number of pages read in a meta-analysis, and, correspondingly, reviewer loads (second column from right in Table 2). For meta-analyses reporting the number of reviewers the results for RER show mean reviewer loads of 11,928 pages between 2000-2009 and 12,628 pages between 2010-2019. Assuming authors served as reviewers in meta-analyses not reporting this information (right-most column in Table 2) produces clear but less dramatic increases in reviewing load. For example, the mean reviewer load for an author in a meta-analysis appearing in RER between 1990 – 1999, 2000 - 2009, and 2010 - 2019 was 1,969, 4,209, and 6,446, respectively. Growth in the mean number of pages read in a meta-analysis in RER in 2000 - 2009 compared to 1990 - 1999 was about 53.2%, and for meta-analyses appearing in RER between 2010 - 2019 compared to 2000 – 2009 was 34.7%. It is important to note that assuming authors served as reviewers when the number of reviewers was not reported likely underestimates load because the median number of reviewers of abstracts and full-text documents in meta-analyses reporting this information was two, whereas the median number of authors of meta-analyses not reporting this information was three.

Table 1
Summary of abstracts reviewed in meta-analyses appearing in ER, EEPA, AERJ, and RER between 1980 - 2019

Year(s)	Journal	Total number of meta-analyses	Mean and median number of abstracts reviewed	Number of meta-analyses reporting the number of abstracts reviewed	Number of meta-analyses reporting the number of reviewers (*r*) of abstracts
1980-89	ER	1	150	1	0, ---
	EEPA	2	556	1	0, ---
	AERJ	9	287 (md =170)	5	0, ---
	RER	2	---	2	0, ---
1990-99	ER	0	NA	NA	NA
	EEPA	1	150	1	0, ---
	AERJ	0	NA	NA	NA
	RER	2	899 (899)	2	0, ---
2000-09	ER	1	450	1	0, ---
	EEPA	2	109	1	0, ---
	AERJ	2	410 (410)	2	0, ---
	RER	44	1,808 (757)	29	7, *r* = 1 7, = 2 1, = 3
2010-19	ER	1	9,530	1	0, ---
	EEPA	3	5,497 (7,926)	3	1, *r* = 2
	AERJ	0	NA	NA	NA
	RER	123	3,883 (1,704)	92	13, *r* = 1 33, *r* = 2 6, *r* = 3 3, *r* = 4 1, *r* = 12

Note: The table is based on *N* = 193 published meta-analyses appearing in Educational Researcher (ER), Educational Evaluation and Policy Analysis (EEPA), the American Educational Research Journal (AERJ), and Review of Educational Research (RER) between 1980 – 2019. The median (md) and the number of meta-analyses these statistics are based on are reported, e.g., 287 (md = 170) indicates the mean and median number of abstracts (287, 170) appearing in 5 meta-analyses in AERJ between 1980-1989; *r* indicates the number of abstract reviewers in a meta-analysis, e.g., 7 meta-analyses appeared in RER between 2000 – 2009 that used one reviewer for abstracts; --- no information provided; NA = not applicable.

Table 2

Summary of full-text documents reviewed in meta-analyses appearing in ER, EEPA, AERJ, and RER between 1980 - 2019 and mean and median of the total number of pages read

Year(s)	Journal	Number of meta-analyses	Mean and median number of full-text documents reviewed	Number of meta-analyses reporting the number of full-text documents	Number of meta-analyses reporting the number of reviewers (*r*) of full-text documents	Mean and median of total number of pages read (abstracts + full-text documents)	Mean and median of total number of pages read (abstracts + full-text documents) when authors are reviewers
1980-89	ER	1	20	1	0	---	353
	EEPA	2	112	1	1, *r* = 4	1,079	1,176
	AERJ	9	69 (md = 470)	8	0	---	1,817 (1,149)
	RER	2	37 (37)	2	0	---	---
1990-99	ER	0	NA	NA	NA	NA	NA
	EEPA	1	81	1	0	1,858	2,034
	AERJ	0	NA	NA	NA	NA	NA
	RER	2	111 (111)	2	0	---	1,969 (1,969)
2000-09	ER	1	30	1	0	---	848
	EEPA	2	199	1	0	---	---
	AERJ	2	119	1	0	---	1,004
	RER	44	265 (137)	27	5, *r* = 1 8, = 2 1, = 3	11,928 (7291)	4,209 (2,849)
2010-19	ER	1	621	1	0	---	7,041
	EEPA	3	440 (556)	3	1, *r* = 2	18,080	5,064 (4616)
	AERJ	0	NA	NA	NA	NA	NA
	RER	123	366 (183)	121	16, *r* = 1	12,628 (7,459)	6,446 (3,655)

31	= 2
6	= 3
4	= 4

Note: The table is based on *N* = 193 published meta-analyses appearing in Educational Researcher (ER), Educational Evaluation and Policy Analysis (EEPA), the American Educational Research Journal (AERJ), and Review of Educational Research (RER). The median (md) and the number of meta-analyses these statistics are based on are reported, e.g., 69 (md = 470) indicates the mean and median number of full-text documents appearing in 8 meta-analyses in AERJ between 1980-1989; *r* indicates the number of reviewers of full-text documents in a meta-analysis, e.g., one meta-analysis appeared in EEPA between 1980 – 1989 that used *r* = 4 reviewers for full-text documents; --- no information provided; NA = not applicable. The right-most column assumes authors served as reviewers in studies not reporting the number of reviewers.

A subset of the sampled meta-analyses provides especially compelling evidence of the growth in reviewer load. Figure 1 plots the year a meta-analysis was published against the mean (total) number of pages read in meta-analyses reporting the number of reviewers. This figure shows increases in reviewer load around 2010, with the mean number of pages read in 2010 - 2019 corresponding to the 70th, 80th, and 90th percentiles of 13,403, 21,802, and 30,445 pages, respectively. Figure 2 shows increases in reviewer load after 2000, although mean reading loads are compressed relative to Figure 1 because assuming authors served as reviewers in meta-analyses not reporting this information generally lowers reviewing load compared to basing these loads on the reported number of reviewers. This pattern, and the fact that medians in the two right-most columns in Table 2 are almost always less than the corresponding mean, indicate the distribution of mean (total) number of pages read was strongly positively-skewed (skewness statistics = 1.81 and 2.06 for the variables defined in the two right-most columns in Table 2). Hence for subsets of the sampled meta-analyses, almost all of which appeared in RER between 2010 – 2019, the reviewing load was much greater than suggested by the RER mean of 12,268 pages. For example, the mean (total) number of pages read in eight of these meta-analyses exceeded 26,000 and for six meta-analyses exceeded 30,000; assuming authors served as reviewers the mean reading load exceeded 15,000 pages in seven meta-analyses and 30,000 pages in three meta-analyses.

A third important pattern suggested in Tables 1 and 2 is that when the number of reviewers is reported it typically is one or two and is essentially unchanged over time. Among all meta-analyses published between 2000 – 2009 that reported the number of abstract reviewers 46.6% used one reviewer, 46.6% used two reviewers, and 6.6% used three reviewers; for full-text documents during this time 35.7% used one reviewer, 57.1% used two reviewers, and 7.1% used three reviewers. For meta-analyses published between 2010 - 2019 that reported the number of abstract reviewers 23.2% used a single reviewer, 60.7% used two reviewers, 10.7% used three reviewers, and 5.4% used four reviewers; for full-text documents these percentages were 27.6%, 55.2%, 10.3%, and 6.9%, respectively. If authors are treated as reviewers in meta-analyses not reporting this information the percentage of meta-analyses published between 1980 - 1999 with one, two, three, or four reviewers was 5.9%, 41.2%, 35.3%, and 17.6%, respectively; for meta-analyses published between 2000 - 2009 (assuming authors served as reviewers) the percentage reporting using one, two, three, or four reviewers was 12.2%, 38.8%, 16.3%, and 14.3%, respectively; for 2010 – 2019 these percentages were 7.9%, 24.4%, 20.5%, and 26%, respectively. Relatedly, the percentage of meta-analyses using one or two reviewers in 1980 – 1999 and 2000 – 2019 (assuming authors served as reviewers) was 47% and 51.9%, and the percentage using one to three reviewers in 1980 – 1999 and 2000 – 2019 was 82.3% and 78.7%.

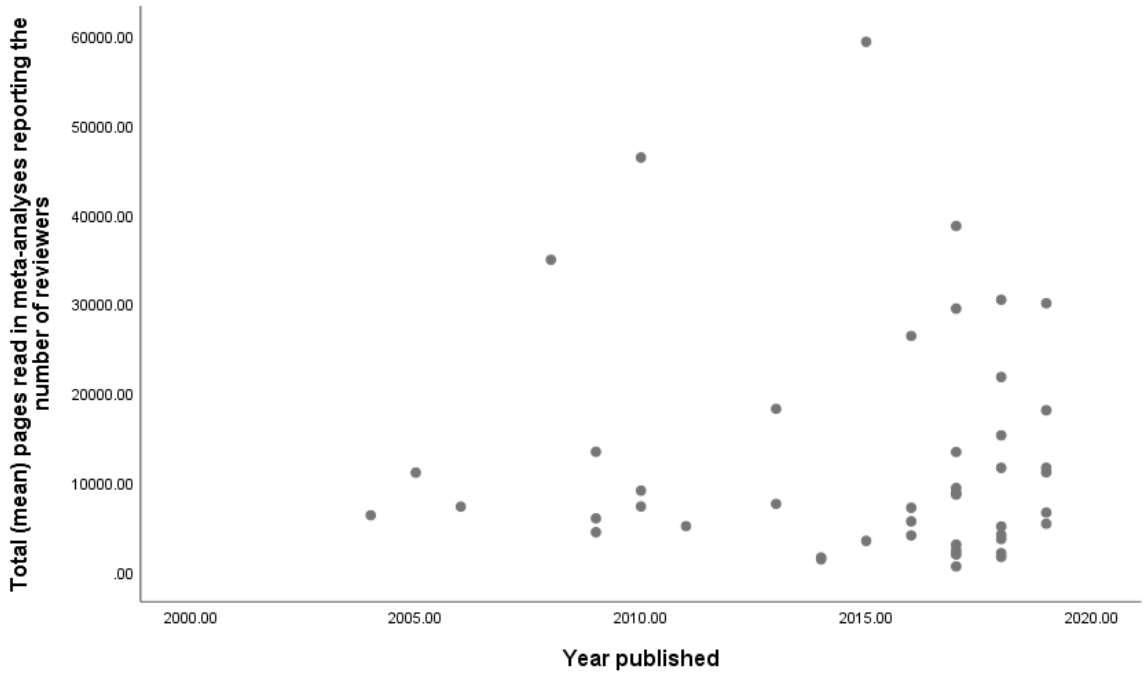

Figure 1. Plot of year published by mean (total) number of pages read in a meta-analysis for studies reporting the number of reviewers.

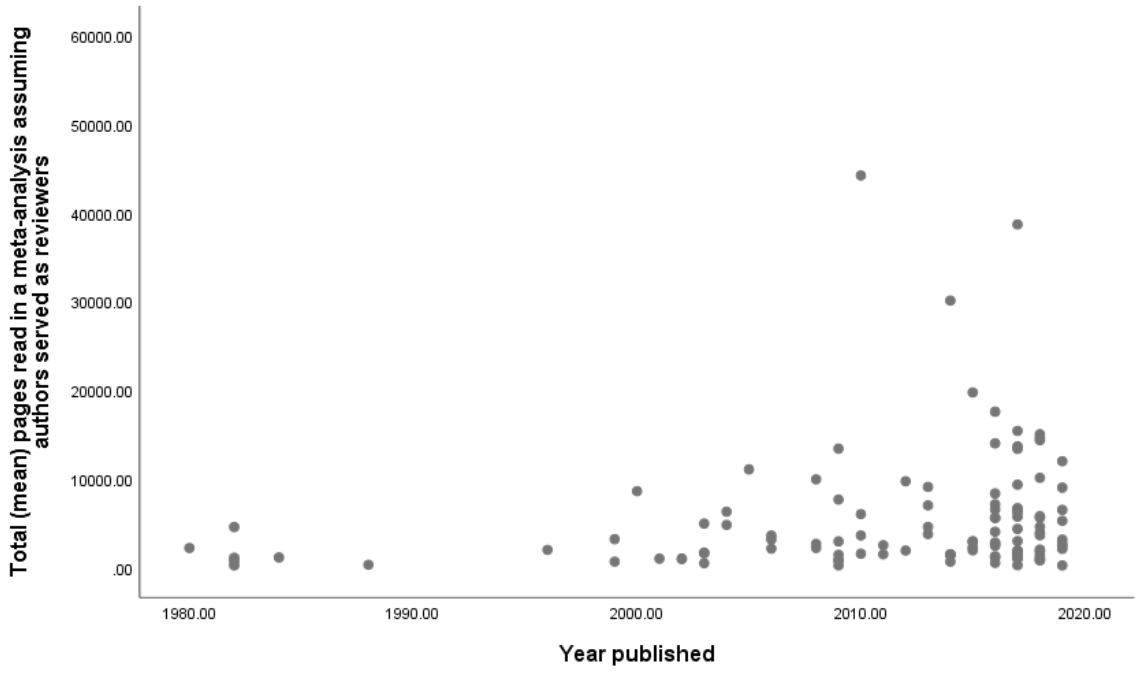

Figure 2. Plot of year published by mean (total) number of pages read in a meta-analysis assuming authors served as reviewers.

In sum, reviewing loads characterized by large numbers of pages read in meta-analyses published between 2010 - 2019 employed similar numbers of reviewers to those published between 1980 – 1999 (assuming authors served as reviewers in meta-analyses not reporting this information). The fact that recently published meta-analyses are more likely to report the number of abstract (44.7%) and full-text (47.1%) reviewers is a welcome trend but the impact on categorizing studies as eligible or not eligible for a meta-analysis (Steps 3 and 4) remains unclear. Moreover, only eight of the $N = 193$ meta-analyses (4.1%) reported using software to facilitate data management and reviewing and all eight used the Endnote software to manage references; none reported using more comprehensive software as Covidence, Distillers, EPPI - 4, or Rayyan. Finally, 36.6% of the meta-analyses in Tables 1 and 2 (all in RER with most appearing between 2010 - 2019) reported information on percent agreement or reliability among reviewers for Steps 3 or 4.

Discussion

An examination of educational meta-analyses provided evidence the reviewing load has increased over the past 40 years and especially since 2010. However resources that support increasing reviewer loads in the form of more reviewers and greater use of software to facilitate reviews do not appear to have increased. An examination of the number of reviewers, mean number of abstracts and full-text documents reviewed, and the mean (total) number of pages read by a reviewer in a sample of N = 193 published meta-analyses provides a basis for two complementary recommendations.

First, details of the initial screening of study abstracts and titles and reviews of full-text documents should be provided. These details would typically include the qualifications of reviewers, reviewer loads (number of titles and abstracts screened and the number, kind, and length of full-text documents reviewed), time spent reviewing documents, a description of any reviewer training that was provided, computer software used to help manage documents and facilitate reviews, evidence of the accuracy and consistency of reviewer categorizations guided by recommended practices (e.g., White, 2018), and how accuracy and consistency were maintained over time. Flow diagrams providing information about the number of studies reviewed in each step of a meta-analysis, such as those appearing in Adesope, Trevisan, and Sundararajan (2017) and Sheridan et al. (2019), could add many of these details. Collectively this information would provide evidence of the adequacy of the number of reviewers in a meta-analysis.

Second, software such as Covidence, Distillers, EPPI - Reviewer 4, or Rayyan should be routinely used to manage the process of screening titles and abstracts and the review of full text documents. This software should help to standardize the review process and give readers greater confidence in the consistency and accuracy with which studies were categorized as eligible or not eligible for a meta-analysis. Relatedly, it's also important for meta-analysts to report any software used to facilitate reviewing.

These recommendations should be implemented along with existing guidelines for conducting and reporting meta-analyses (MARS, 2008; PRISMA, 2009), and would almost certainly have

greater impact if added to the submission guidelines of journals such as Review of Educational Research that regularly publish meta-analyses.

Study Limitations

Two important limitations of the current study should be kept in mind. One is that study conclusions and recommendations are conditional on the sample of $N = 193$ meta-analyses reflecting current practices. Sampling all meta-analyses appearing in four AERA-sponsored journals over a 40-year period helps to ensure, but does not guarantee, an adequate sampling of current practices. A second limitation is the potential impact of missing data about the review process (e.g., number of reviewers not reported) which could distort the current study's conclusions.

Future Work

The growth in electronic search capabilities and the resulting increase in the number of titles, abstracts, and full-text documents reviewed in a typical meta-analysis is likely to continue and highlights the need for research in at least two areas. First is studying the ability of reviewers with different levels of expertise to accurately and consistently screen varying numbers of titles and abstracts and review varying numbers of full-text documents differing in substantive and methodological complexity. For example, in the Graham et al. (2018) meta-analysis the 17,301 titles and abstracts were reviewed by the first and last authors, and the 677 full-text documents were reviewed by two trained graduate students. The expertise of the first reviewer (40 years of experience in literacy research) speaks to this reviewer's qualifications, but the Graham et al. meta-analysis is silent on the qualifications of the other reviewer nor does it describe the nature or length of the training provided to the graduate students.

Second, the growth in electronic search capabilities highlights the importance of studying the relationship between increasing the number of documents screened or reviewed and generalizability. For example, are the results of an electronic search producing 10,000 abstracts to be screened and 400 full-text documents to be reviewed likely to be more generalizable compared to the same electronic search producing 5,000 abstracts and 200 full-text documents or 2,500 abstracts and 100 full-text documents? Evidence of the strength (or weakness) of this relationship would have an important impact on several components of a meta-analysis.

Author Notes

Michael Harwell is a professor at the University of Minnesota, Minneapolis, MN.

Correspondence concerning this article should be addressed to Michael Harwell at harwe001@umn.edu.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87 (3*), 659 –701. doi: 10.3102/0034654316689306

Ahn, S., Ames, A. J., & Myers, M. D. (2012). A review of meta-analyses in education: Methodological strengths and weaknesses. *Review of Educational Research, 82 (4*), 436-476. doi: 10.3102/0034654312458162

APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist, 63,* 839 – 851.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board Task Force Report. American Psychochologist, 73 (1), 3-25. doi: org/10.1037/amp0000191

Brunton, T. J., & Graziosi, S. (2010). *EPPI-Reviewer 4.0: Software for research synthesis*. EPPI-Centre Software. London: Social Science Research Unit, Institute of Education, University of London. Available at http://eppi.ioe.ac.uk/CMS/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4&.

Cooper, H. M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of Educational Research, 52,* 291-302.doi: 10.3102/00346543052002291

Covidence systematic review software. (2016). Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org.

D'Agostino, J. V., & Powers, S. J. (2009). Predicting teacher performance with test scores and grade point average: A meta-analysis. *American Educational Research Journal 46 (1*), 146-182. doi: 10.3102/0002831208323280

DistillerSR software (2016). Evidence Partners, Ottawa, Canada. Available at https://www.evidencepartners.com/

Graham, S., Liu, X., Bartlett, B., Ng, C., Harris, K. R., Aitken, A., Barkel, A., Kavanaugh, C., & Talukdar, J. (2018). Reading for writing: A meta-analysis of the impact of reading interventions on writing. *Review of Educational Research, 88 (2*) , 243-284. doi: 10.3102/0034654317746927

Joksimović, S., Poquet, O., Kovanović, V., Dowell, N., Mills, C. Gašević, D., Dawson, S., Graesser, A. C., & Brooks, C. (2018). How do we model learning at scale? A systematic review of research on MOOCs. *Review of Educational Research, 88 (1*), 43 –86. doi: 10.3102/0034654317740335

Ke, F., Whalon, K. & Yun, J. (2018). Social skill interventions for youth and adults with autism spectrum disorder: A systematic review. *Review of Educational Research, 88 (1*), 3-32. doi: 10.3102/0034654317740334

Kraft, M. A., Blazer, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88 (4*), 547-588. doi: 10.3102/0034654318759268

Kulik, C. L. C., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal, 19 (3)*, 415-428. doi: 10.3102/00028312019003415

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PloS Med, 6(7*). doi: org/10.1371/journal.pmed.1000097

Orwin, R. G., & Vevea, J. L. (2009). Evaluating coding decisions. In *The Handbook of research synthesis and meta-analysis*, Cooper H., Hedges L. V., & Valentine J. C., eds.), 177-203. New York, NY: Sage.

Peng, P., Wang, C., & Namkur, J. (2018). Understanding the cognition related to mathematics difficulties: A meta-analysis on the cognitive deficit profiles and the bottleneck theory. *Review of Educational Research, 88 (3)*, 434 - 476. doi: 10.3102/0034654317753350

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan — a web and mobile app for systematic reviews. *Systematic Reviews 5*, 2-10. doi: 10.1186/s13643-016-0384-4

Rubio-Aparicio, M., Sánchez-Meca, J., Marín-Martínez, F., & López-López, J. (2018). Guidelines for reporting systematic reviews and meta-analyses. *anales de psicología, 34 (2)*, 412-420. doi: 10.6018/analesps.34.2.320131

Sheridan, S. M., Smith, T. E., Kim, E. M., Beretvas, S. N., & Park, S. (2019). A meta-analysis of family-school interventions and children's social-emotional functioning: Moderators and components of efficacy. *Review of Educational Research, 89 (2*), 296 –332. doi: 10.3102/0034654318825437

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*, 417–453. doi:10.3102/00346543075003417

Thomas, J., Brunton, J, & Graziosi, S. (2010). *EPPI-Reviewer 4.0: Software for research synthesis*. EPPI-Centre Software. London: Social Science Research Unit, Institute of Education, University of London. Available at

http://eppi.ioe.ac.uk/CMS/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4&

Valentine, J. C., Cooper, H. M., Patall, E. A., Tyson, D., & Robinson, J. C. (2010). A method for evaluating research syntheses: The quality, conclusions and consensus of 12 syntheses of the effects of after-school programs. *Meta-analysis Methods, 1*, 20–38. doi:10.1002/jrsm.3

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*, 461–481. doi: org/10.1037/0033-2909.91.3.461

White, M.C. (2018). Rater performance standards for classroom observation instruments. *Educational Research, 47 (8)*, 492-501. doi: 10.3102/0013189X18785623