DIGITALCOMMONS
@WAYNESTATE

Journal of Modern Applied Statistical Methods

Volume 15 | Issue 2 Article 7

11-1-2016

Longitudinal Stability of Effect Sizes in Education Research

Joshua Stephens
Cleveland State University, j.a.stephens@csuohio.edu

Recommended Citation
Stephens, Joshua (2016) "Longitudinal Stability of Effect Sizes in Education Research," *Journal of Modern Applied Statistical Methods*: Vol. 15 : Iss. 2 , Article 7.
DOI: 10.22237/jmasm/1478001900
Available at: http://digitalcommons.wayne.edu/jmasm/vol15/iss2/7

Journal of Modern Applied Statistical Methods
November 2016, Vol. 15, No. 2, 53-66.
doi: 10.22237/jmasm/1478001900

ISSN 1538 − 9472

Longitudinal Stability of Effect Sizes in Education Research

Joshua Stephens
Cleveland State University
Cleveland, OH

Educators use meta-analyses to decide best practices. It has been suggested that effect sizes have declined over time due to various biases. This study applies an established methodological framework to educational meta-analyses and finds that effect sizes have increased from 1970–present. Potential causes for this phenomenon are discussed.

Keywords: Effect sizes, meta-analysis, research methodology, publication bias.

Introduction

Effect sizes are commonly used in conducting meta-analyses, such as in educational research. Jennions and Moller (2002) suggested reliance on effect sizes has declined somewhat due to various sources of bias. The primary concern of this study is with the application of increased rigor to educational literature. It is important that educators and educational policy-makers use practices and policies based on the strongest empirical evidence. Because public school funding is a limited resource, it is important for that funding to be spent wisely and on effective innovations. This applies to other fields as well, such as social work (Shlonsky, Noonan, Littell, & Montgomery, 2011).

Meta-analysis

Effect sizes describe the magnitude difference between the null and alternative hypothesis. Effect sizes are calculated for each study, weighted by sample size and study quality, and then averaged to produce an overall effect size (Littell, Corcoran, & Pillai, 2008). Although typical data analysis uses multiple observations of a phenomenon as data points, meta-analysis uses multiple studies as data points (Wolf, 1986; Littell et al., 2008). The resulting literature synthesis

Dr. Stephens is a Faculty member in the Education and Psychology departments. Email him at j.a.stephens@csuohio.edu.

may become stronger than that provided in a qualitative or narrative fashion (Asher, 1990).

Unstable Effect Size

Ecologists discovered several examples of diminishing effect sizes (Alatalo, Mappes, & Elgar, 1997; Gontard-Danek & Moller, 1999; Poulin, 2000; Simmons, Tomkins, Kotiaho, 1999). An interpretation of why effect sizes apparently diminish over time has not emerged. The following are possible explanations. Alatalo et al. (1997) attributed diminishing effect sizes to changing belief systems. Palmer (2000) attributed the phenomenon to fads. Tregenza and Wedell (1997) attributed it to biased study design. Alatalo et al. (1997) suggested submitting findings for publication that support previously held ideas makes it easier to get published. Simmons et al. (1999) suggested that it is easier to publish confirmatory findings during early stages of research in a particular field, but it becomes more difficult as critique of that field narrows. This may be particularly emphasized in the social sciences, where it takes longer to publish non-significant results (Stern & Simes, 1997).

Social science researchers who study the phenomenon of diminishing effect sizes cite two primary potential causes: dissemination bias and citation bias.

Dissemination bias is a broad term encompassing many different sorts of biases related to the publication and dissemination process, including bias related to date of publication, language, multiple publication bias, selective reference citation, database index bias, media attributed bias, selective publication bias, familiarity of techniques, and the cost of research reports (Rothstein, Sutton & Bornstein, 2005; Song, Eastwood, Gilbody, Duley, & Sutton, 2000). "Dissemination bias occurs when the dissemination profile of a study's results depends on the direction or strength of its findings" (Song et al., 2000, p. 17). It refers to the notion that a given literature review does not represent a random sampling of all studies in a given field, and therefore is a type of non-random sampling error similar to that found when conducting primary research (Song et al., 2000).

Both indirect and direct evidence support the existence of dissemination bias (Sohn, 1996). Examples of indirect evidence include disproportionately high percentage of positive findings in journals, or larger effect sizes in small studies relative to large studies. Small studies are more vulnerable to dissemination biases, as the results of these studies will be more widely spread around the true results owing to greater random error (Begg & Berlin, 1988). Direct evidence includes such things as admissions by investigators and publishers and comparison of results from published and unpublished studies (Song et al., 2000). Rotton, Foos, VanMeek, and Levitt (1995) found that the most significant reason given by authors for not submitting their work for publication was the failure to find statistical significance.

The strongest evidence supporting the existence of dissemination bias comes from comparisons between published and unpublished studies (Song et al., 2000). Simes (1986) performed meta-analyses on both published and unpublished studies of a cancer treatment regimen and discovered that the published findings found that the treatment was effective, but when the published and unpublished studies were analyzed together, the treatment effect was not found.

There are specific types of dissemination bias. Biases in addition to those mentioned earlier include positive results bias, hot stuff bias, time-lag bias, grey literature bias, full publication bias, place of publication bias, outcome reporting bias, and retrieval bias (Song et al., 2000). These forms of bias may be prevalent in many disciplines and may account for observed decline in effect sizes in ecology and other fields.

Methodology

The purpose of this study is to analyze whether meta-analytically derived results are longitudinally stable in education research. To accomplish this task, a process similar to that used by Jennions and Moller (2002) will be invoked.

Study Selection

First, a set of meta-analyses, based on K–12 classroom interventions from the years 1970 to 2011, was selected from the EBSCOHost databases. Studies were included if they specifically provide effect size results based on meta-analytical techniques and provide a comprehensive list of studies used to generate effect sizes.

Table 1. Descriptive statistics of included studies

N	Year of Publication Range	Mean Year of Publication	Mean Number of Reported Effect Sizes Per Meta-Analysis
60	1984-2010	2002.3	42.7

The study involved a broad search for literature, which was then winnowed down through a rigorous paring process, resulting in a final set of 60 studies that were analyzed. Descriptive statistics of these studies are shown in Table 1. After final literature was selected, data analysis was initiated.

Statistical Procedure

The 60 selected studies were then analyzed using a process outlined by Jennions and Moller (2002), involving the use of four Spearman's ρ (rho) analyses on two levels. The first set of analyses dealt with the effect sizes reported in the selected studies. This will hereafter be known as the "study level" of analysis. The second set of analyses were conducted on the meta-analyses themselves. This is hereafter known as the "meta-analysis level."

On both the study level and the meta-analysis level, four relationships were analyzed: (i) the relationship between effect size and year of publication; (ii) the relationship between effect size and sample size; (iii) the relationship between standardized effect size and sample size; and (iv) the relationship between effect size and year of publication, after weighting for variation in sampling effort. The first three relationships were conducted using a Spearman's ρ (rho) test and were performed in SPSS.

The fourth relationship was conducted using MetaWin 2.0. This relationship was estimated by creating a random-effects continuous model meta-analysis with year of publication as the independent variable and the inverse of sampling variance as the weighting factor. Random-effects meta-analysis was selected over a fixed-effects model, as fixed-effects models become problematic when some studies have very large sample sizes. These studies then dominate the analysis, and the results from the studies with smaller sample sizes are largely ignored (Helfenstein, 2002).

MetaWin 2.0 was used to obtain a one-tailed ρ-value for year of publication generated by a randomization method with 999 replicates. A one-tailed ρ-value was chosen because the Jennions and Moller (2002) study used a one-tailed test, since they postulated that a declining effect size was more likely. The effect size generated by the meta-analysis was converted to a Spearman's ρ- (rho-) value so that all results were reported in a uniform manner. The formula to do this is as follows:

$$\rho = \sqrt{\frac{d^2}{d^2 + 4}}$$

All Spearman's ρ-values were then converted to standard normal deviates (Z-scores), using the formula:

$$\rho = \frac{\sqrt{Z^2}}{n}$$

This was done so that all results were normalized, thus diminishing the effects of outliers and providing a more robust answer to the research question.

Results

Results regarding the possibility of effect sizes diminishing over time are compiled in Table 2.

Table 2. Relationships (ρ) between effect size, standardized effect size, year of publication, and sample size.

	Method of Calculation			
Weighted meta-analysis of:	**Year v. Effect Size**	***n* v. Effect Size**	***n* v. Standard Effect**	**Year v. Effect Size** (after weighing for sampling variance)
Datasets	0.105*	-0.073**	-0.073**	0.440*
Original Meta-Analyses	0.317**	-0.148	-0.148	0.333*

Note: * Significant at the <0.001 level; **Significant at the <0.01 level

Beginning at the study level, these results indicate that there is a statistically significant positive relationship between year of publication and effect size ($\rho = 0.105$, $p < 0.001$, $n = 1167$). However, there was also a significant relationship between sample size and both effect size and standardized effect size, so the relationship was re-assessed after accounting for sampling variance. Still, however, a statistically significant positive relationship was observed ($\rho = 0.440$, $p < 0.001$, $n = 1167$). Figures 1 – 4 show scatterplots of these four relationships.

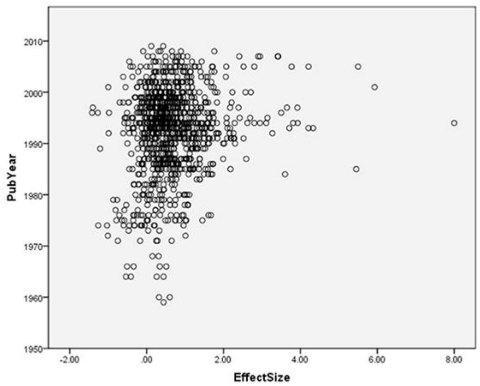

Figure 1. Publication year compared to effect size (g) at the study level

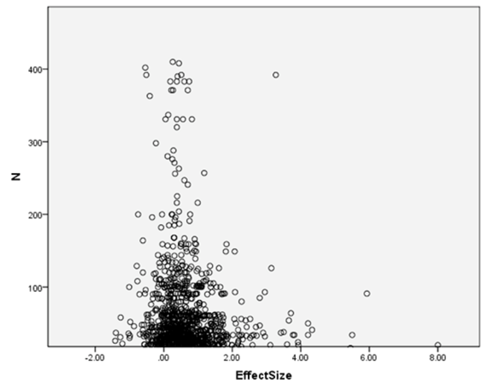

Figure 2. Sample size compared to effect size (g) at the study level

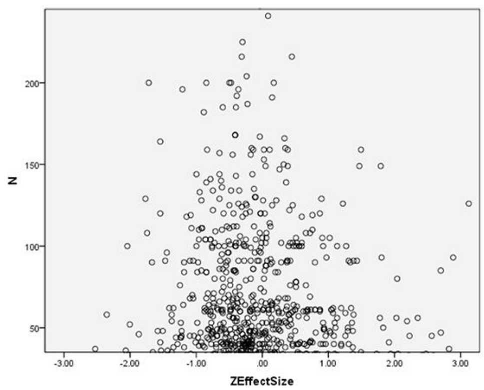

Figure 3. Sample size compared to standardized effect size (z-transformed) at the study level

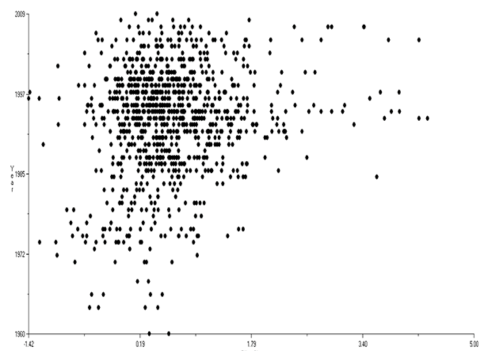

Figure 4. Year of publication compared to effect size (g) after weighting for sample size at the study level

A similar observation is found at the meta-analysis level. These results indicate that there is a statistically significant positive relationship between year of publication and effect size ($\rho = 0.317$, $p < 0.009$, $n = 60$). However, there was not a significant relationship between sample size and both effect size and standardized effect size. Still, however, a statistically significant, positive relationship was observed ($\rho = 0.333$, $p < 0.001$, $n = 60$) after accounting for sampling variance. Figures 5 – 8 below show scatterplots of the relationships from the meta-analysis level.

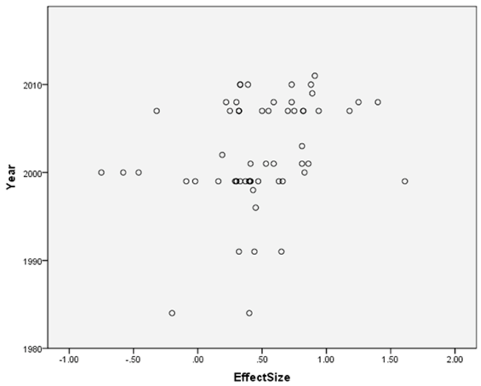

Figure 5. Publication year compared to effect size (*g*) at the meta-analysis level

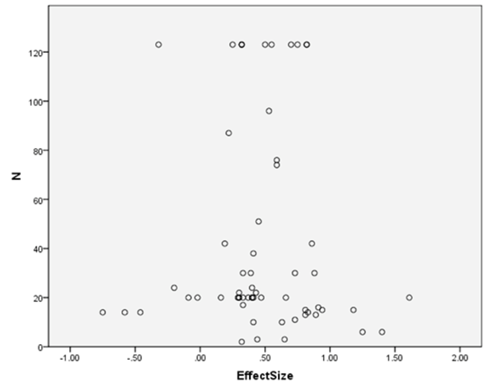

Figure 6. Sample size compared to effect size (*g*) at the meta-analysis level

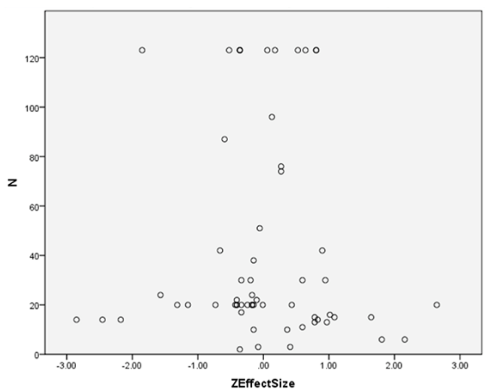

Figure 7. Sample size compared to standardized effect size (*z*-transformed) at the meta-analysis level

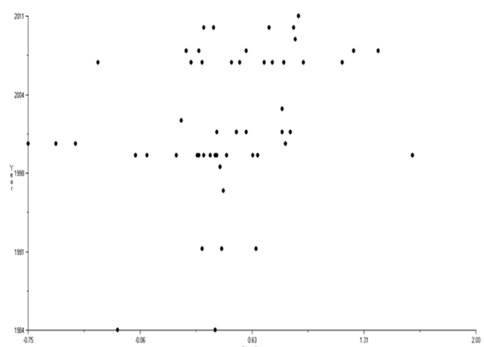

Figure 8. Year of publication compared to effect size (*g*) after weighting for sample size at the meta-analysis level

It is notable that effect sizes increase at both the study and meta-analysis levels. Data were parsed out to show mean effect sizes by decade to allow for simpler understanding of how effect sizes have increased over time. Table 3 shows this descriptive information.

Table 3. Mean effect sizes by decade.

	N	Mean effect size (*g*)	Range	Standard Deviation
1970s / 1980s	2	0.100	-0.20 – 0.40	0.424
1990s	21	0.424	-0.09 – 1.61	0.329
2000s	31	0.509	-0.75 – 1.40	0.506
2010s	6	0.595	0.33 – 0.91	0.276

Conclusion

It was found that education meta-analyses do not appear to follow the pattern seen in the natural sciences, because the effect sizes on which they are based did not decline. On the contrary, for the sample included in this study, they tended to increase over time.

This finding bears some consideration. If no statistically significant relationships had been observed between effect sizes and year of publication, then it could be assumed that meta-analysis provides a longitudinally stable measure, and a strong argument could have been made for wider use of this analytical technique. However, as measured effect sizes tend to increase over the time period 1970 – 2012, either there is some persistent set of biases that are impacting the conduct or publication of educational research, or effect sizes are indeed increasing over time as the field of education develops into a more complex and sophisticated science and leaves behind ineffective educational practices.

Persistent Bias in Educational Research

One explanation for the observed phenomenon of longitudinally increasing effect sizes is publication bias. Given the findings of this study, it seems reasonable to conclude that it is possible that some of these forms of bias may be more active than others. In particular, the following forms of publication bias are possible explanations for the findings of this study: positive results bias; hot stuff bias; grey literature bias; and confirmation bias.

Positive results bias Positive results bias refers to the tendency of authors to submit—and for editors to publish—positive or significant research results while ignoring non-significant results (Song et al., 2000). This seems to be a likely cause of increasing effect sizes. Since researchers generally will find statistically significant results when they are searching for literature to use to conduct meta-analyses, they will find ever-increasing effect sizes across time. Then this effect becomes multiplied, as other researchers use published meta-analyses to generate effect sizes for new research and duplicate biases from past research.

Hot stuff bias Another form of bias that could account for the phenomenon of increasing effect sizes is hot stuff bias. This refers to the phenomenon of journal publishers tending to publish topics that are timely or popular but which may only have relatively weak results (Sackett, 1979). This seems to be a likely form of publication bias in education where fads and trends dominate pedagogical practice. These trends may be pushed by textbook publishers looking to profit from a product, or politicians who make educational policy with little understanding of educational systems and processes.

Hot stuff bias may account for increasing effect sizes through publishers choosing articles to publish based on what they believe will promote their journal's readership. Publishers choose articles that may be methodologically unsound; these articles are then indexed in electronic indexes and used to conduct meta-analyses, thereby creating the appearance of increasing effect sizes over time. When the particular timely trend ends, no researcher bothers to fully repudiate it or no journal chooses to publish these repudiations, so it appears that these effect sizes are significant and increasing over time.

Grey literature bias Grey literature refers to things such as conference presentations, dissertations, working papers, and other pieces of literature that are difficult to obtain as they are not electronically indexed in any systematic manner (Auger, 1998). Grey literature bias refers to the notion that these pieces of literature tend to show non-significant or statistically weaker results and that excluding these from meta-analyses produces an artificially high effect size (Song et al., 2000). McAuley et al. (1999) sampled 135 meta-analyses, 38 of which included grey literature, found that those meta-analyses that included grey literature showed a diminished effect size of approximately 12%.

Grey literature bias would appear to be a significant problem in the field of educational research where many universities have large numbers of master's and doctoral students who are producing volumes of research that is never published. While it is difficult to quantify specifically how much research is conducted and never included in any sort of meta-analysis, it is safe to assume it must be a large amount every year. When one includes classroom research done by practicing teachers, the amount of grey literature skyrockets. While not all of this research would meet methodological criteria for publication or for inclusion in properly conducted meta-analyses, some certainly would. The exclusion of this grey literature could be a significant factor in the observed phenomenon of increasing effect sizes. If established researchers get their statistically significant findings published while student researchers or others who find non-significance do not, then effect sizes would tend to increase over time as no one individual or organization reputes earlier findings.

Confirmation bias Confirmation bias refers to the psychological phenomenon whereby humans tend to subconsciously look for ideas and information that confirms their earlier beliefs. This information tends to be more readily assimilated and utilized than does information that contradicts what an individual believes (Bushman & Wells, 2001).

Confirmation bias seems like a likely cause of increasing effect sizes. As researchers look for studies to help them build the case for their study, they will naturally begin by searching for studies that confirm what they already believe. As they find increasing numbers of these studies, it seems that the results of the study are a foregone conclusion. This may lead researchers to discount or ignore studies that may disagree with what they believe is true about a research question. In a meta-analysis, this may take the form of a researcher applying more stringent selection criteria to studies that don't confirm his or her hypothesis, leading to effect sizes that increase across time.

Increasing effect sizes represent educational reality

There is another explanation for the phenomenon of longitudinally increasing effect sizes in educational research: it is possible that effect sizes seem to be increasing because they actually are. This is a hopeful notion that as educational researchers have begun to more rigorously conduct research and educational practitioners have received better training in the utilization of research-based educational techniques, that educational practices have become more effective. This would be supported by the fact that, over the past 40 years in the sample considered in this study, many states have implemented tougher teacher training and licensure laws, and departments of education at universities have taken a more rigorously quantitative approach. However, when the outcomes of large-scale assessments of student learning are observed across this time period, no similarly significant gains are apparent. It is beyond the scope of this study to adequately assess the growth of students in comparison to the perceived growth of teacher effectiveness. However, it does seem less likely that this is the case and more likely that the correct explanation for the phenomenon of longitudinally increasing effect sizes is publication bias.

Potential solutions for addressing increasing effect sizes

If, as the results of this study suggest, effect sizes are in fact increasing over time, then this potentially indicates that there is a problem in the publication process that should be corrected by researchers and publishers. Failure to do so may cause misperceptions regarding the efficacy of a host of educational interventions that may diminish the impact of schooling for students which is a patently undesirable outcome.

Educational researchers should strive to conduct meta-analyses and other research in the most methodologically sound manner possible. Narrative literature reviews should be only used when a research question is either very limited in scope or is so new that very little literature is available such that it would be possible for a researcher to adequately summarize findings from the literature base without quantitative methods. It may also be useful to provide narrative literature reviews as an element of a meta-analysis. Meta-analytic techniques should be included in most literature reviews and these techniques should follow the guidelines set forth by the Cochrane and Campbell Collaborations (Pfeffer & Sutton, 2006). These organizations have initiated programming to assist researchers with developing the most accurate summarizations of literature possible. Following their recommendations globally would create a less biased body of educational literature that would be more useful to practitioners and researchers alike.

To further ameliorate this phenomenon, there would need to be a change in the way education research is published. First and foremost, there must be a journal dedicated to publishing only null or statistically insignificant findings. This journal must be indexed properly in major educational research databases and should draw from as many countries and languages as possible. By doing so, researchers who wish to properly conduct meta-analyses will be able to more readily access these results and then conduct a more methodologically sound and less biased meta-analysis.

Additionally, a comprehensive effort should be made to index the wide body of grey literature that is generated globally each year. Conference presentations, dissertations, theses, working papers, action research and other forms of grey literature may provide important insight into research questions and should not be ignored. Moreover, publishers should be conservative when announcing special issues or accepting papers on topics that are very new. Although this is difficult to do and may not always be advisable, this would help alleviate the problems associated with hot stuff bias, as described above.

Limitations of the present study

There are two limitations in this study that require comment. First, component studies came from a limited subset of education studies. Hence, a more inclusive literature search may invalidate or temper the results found here. Second, it has been opined that meta-analysis be conducted using a team of reviewers who make decisions regarding which studies to include. Presumably, that process creates a less biased set of inclusion criteria. It is possible that, had this research been conducted utilizing a team of researchers or assistants to help determine which studies should be included, the results of this project may have been different.

The larger question remains as to the cause of the observed phenomenon. Is it caused by pervasive publication biases that should be immediately addressed and remedied, or have effect sizes increased because educators have become better at their jobs over the past 40 years? This causal question is truly vexing and should be a primary focus of future research. In general, publication biases are not widely studied in education, and should be a source of concern for the community of educational researchers and for those who utilize that research.

Acknowledgments

This paper was developed in part from the author's dissertation (Stephens, 2013).

References

Alatalo, R. V., Mappes, J. & Elgar, M. (1997). Heritabilities and paradigm shifts. *Nature, 385*, 402 – 403. doi: 10.1038/385402a0

Asher, W. (1990). Education psychology, research methodology, and meta-analysis. *Educational Psychologist, 25*(2), 143. doi: 10.1207/s15326985ep2502_3

Auger, C. P. (1998). *Information sources in grey literature*, 4th ed. London: Bowker Saur.

Begg, C. B. & Berlin, J. A. (1988). Publication bias: a problem in interpreting medical data. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 151*(3), 419 – 463. doi: 10.2307/2982993

Bushman, B. J. & Wells, G. L. (2001). Narrative impressions of literature: The availability of bias and corrective properties of eta-analytic approaches. *Personality and Social Psychology Bulletin, 27*(9), 1123 – 1130. doi: 10.1177/0146167201279005

Gontard-Danek, M. C. & Moller, A. P. (1999). The strength of sexual selection: a meta-analysis of bird studies. *Behavioral Ecology, 10*(5), 476 – 486. doi: 10.1093/beheco/10.5.476

Helfenstein, U. (2002). Data and models determine treatment proposals – an illustration from meta-analysis. *Postgraduate Medical Journal, 78*(917), 131 – 134. doi: 10.1136/pmj.78.917.131

Jennions, M. & Moller, A. (2002). Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution. *Proceedings of the Royal Society of London B: Biological Sciences*, *269*(1486), 43 – 48. doi: 10.1098/rspb.2001.1832

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. New York, NY: Oxford University Press.

McAuley, L., Moher, D., Pham, B., Tugwell, P. (1999). Evaluation of the impact of grey literature in meta-analysis. In: *Proceedings of the 7th Cochrane Colloquium; Oct 5 – 9; Rome*. Rome: Universitas Tommaso D' Aquino: 17.

Palmer, A. R. (2000). Quasireplication and the contract of error: Lessons from sex ratios, heritabilities and fluctuating asymmetry. *Annual Review of Ecology and Systematics, 31*, 441 – 480. doi: 10.1146/annurev.ecolsys.31.1.441

Pfeffer, J. & Sutton, R. I. (2006). Demanding proof: engineers are poised to lead an evidence-based management movement. *Industrial Engineer, 38*(6), 43-47.

Poulin, R. (2000). Manipulation of host behaviour by parasites: a weakening paradigm? *Proceedings of the Royal Society of London B: Biological Sciences*, *267*(1445), 787 – 792. doi: 10.1098/rspb.2000.1072

Rosenberg, M., Adams, D., & Gurevitch, J. (2000). *Meta-win: statistical software for meta-analysis, v. 2.0.* Sunderland, MA: Sinauer.

Rothstein, H., Sutton, A. J., & Bornstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. Chichester, UK: Wiley.

Rotton, J., Foos, P. W., Vanmeek, L., & Levitt, M. (1995). Publication practices and the file drawer problem – a survey of published authors. *Journal of Social Behavior and Personality*, *10*(1), 1 – 13.

Sackett, D. L. (1979). Bias in analytic research. *Journal of Chronic Disease, 32*(1-2), 51 – 53. doi: 10.1016/0021-9681(79)90012-2

Shlonsky, A., Noonan, E., Littell, J., & Montgomery, P. (2011). The role of systematic reviews and the Campbell Collaboration in the realization of evidence-informed practice. *Clinical Social Work Journal*, *39*(4), 362-368. doi: 10.1007/s10615-010-0307-0

Simes, R. J. (1986). Publication bias: The case for an international registry of clinical trials. *Journal of Clinical Oncology*, *4*(10), 1529 – 1541.

Simmons, L. W., Tomkins, J. L., Kotiaho, J. S. & Hunt, J. (1999). Fluctuating paradigm. *Proceedings of the Royal Society of London B: Biological Sciences, 266*(1419), 593 – 595. doi: 10.1098/rspb.1999.0677

Sohn, D. (1996). Publications bias and the evaluation of psychotherapy efficacy in reviews of the research literature. *Clinical Psychology Review, 16*(2), 147 – 156. doi: 10.1016/0272-7358(96)00005-0

Song, F., Eastwood, A. J., Gilbody, S., Duley, L. & Sutton, A. J. (2000). Publication and related biases. *Health Technology Assessment, 4*(10), 1 – 125. doi: 10.3310/hta4100

Stephens, J. (2013). *Longitudinal stability of effect sizes in educational research* (Doctoral dissertation). Retrieved from EngagedScholarship@CSU (http://engagedscholarship.csuohio.edu/etdarchive/281/)

Stern, J. M. & Simes, R. J. (1997). Publication bias: Evidence of delayed publication in a cohort study of clinical research projects. *British Medical Journal, 315*(7109), 640 – 645.

Tregenza, T. & Wedell, N. (1997). Natural selection bias. *Nature, 386*, 234. doi: 10.1038/386234b0

Wolf, F. (1986). “Meta-analysis.” *Sage University Paper series on Quantitative Applications in the Social Sciences*. Beverly Hills: Sage Publications.