

## Abstract Title Page

**Title: Investigating the File Drawer Problem in Causal Effects Studies in Science Education**

**Authors and Affiliations:**

**Joseph Taylor**

**Director of Research and Development**

**BSCS**

719.219.4104

[jtaylor@bscs.org](mailto:jtaylor@bscs.org) • [www.bscs.org](http://www.bscs.org)

Susan Kowalski, BSCS

Molly Stuhlsatz, BSCS

Christopher Wilson, BSCS

Jessaca Spybrook, Western Michigan University

Preferred Sections

**1. Mathematics and Science Education in Secondary Grades**

**2. Research Methods**

## Abstract Body

### **Background / Context:**

This paper seeks to investigate the existence of a specific type of publication bias. That is, the bias that can result when causal effect studies with small sample sizes (low statistical power) yield small effects that are not statistically significant and therefore not submitted or accepted for publication. In other words, we are investigating for science education studies what Robert Rosenthal originally coined as the “file drawer” problem (Rosenthal, 1979).

The importance of this work is highlighted by two specific problems that can result from this type of publication bias. First and foremost, the field loses the benefit of learning about interventions that may have a small or moderate effect, but an important effect nonetheless. Further, the field also benefits from learning about interventions had a near-zero or a negative effect in a particular study context.

Second, this type of publication bias poses threats to validity for meta-analyses. Specifically, when meta-analysts compute summary effect sizes for a particular group of studies, it is assumed that the weighted mean summary effect is based on a representative sample of all of the effects available for the intervention in question (Borenstein, 2009). If all or most of the smaller effects are excluded from the summary effect estimate because they were never published, the summary effect will be artificially inflated (Sutton, 2009).

### **Purpose / Objective / Research Question / Focus of Study:**

The purpose of this paper is to use both conceptual and statistical approaches to explore publication bias in recent causal effects studies in science education and to draw from this exploration implications for researchers, journal reviewers, and journal editors. This paper fills a void in the *science education* literature as no previous exploration of its kind can be located.

### **Setting:**

The studies in this publication bias analysis are taken from a larger meta-analysis that includes studies from the United States, Europe and East Asia.

### **Population / Participants / Subjects:**

In this paper the “subjects” are studies. We conducted a keyword search using the Social Science Index within the Web of Knowledge research database. Our string of Boolean operators was intended to detect all science education studies published in specific journals in 2010-11 that sought causal effects through experiments or quasi-experiments. This search yielded approximately 60 studies. From these studies, studies were eliminated because they did not report an effect size, did not report descriptive statistics on outcome measures that could be used to compute an effect size, or neither. From the original set of studies, 33 either had an effect size reported or provided outcome descriptive statistics that allowed us to compute an effect size. In summary, the 33 causal studies in this analysis have the following characteristics:

1. Studied a science education intervention,
2. Isolated causal effects using an experimental or quasi-experimental design,
3. Reported an effect size, a variance/standard deviation/standard error, and sample size
4. Were published in 2010 or 2011, AND
5. Were published in one of the following journals – *American Educational Research Journal, Educational Researcher, Educational Evaluation and Policy Analysis, Review of*

*Educational Research, Journal of Research on Educational Effectiveness, Journal of Research in Science Teaching, Science Education, International Journal of Science Education, Research in Science Education.*

**Intervention / Program / Practice:**

All of the interventions described in the included studies are *science education* interventions.

**Research Design:**

The general research design is random effects meta-analysis. Specific tests of publication bias were performed within the meta-analysis context. These include funnel plots, Galbraith plots and Egger's test of Asymmetry.

**Data Collection and Analysis:**

Publication bias was investigated using both intuitive and statistical techniques. The intuitive portion of the analysis involved inspection of both funnel plots and Galbraith plots. Each of these plots provide unique ways to visualize publication bias or lack thereof. In the funnel plot, standard error of a study is usually plotted on the vertical axis versus the standardized effect size on the horizontal (see Figure 1 for an example). Note that the standard error is inversely proportional to the sample size so the decrease in standard error as you go up the vertical axis is equivalent to increasing the sample size.

**Insert Figure 1 about here**

In inspecting the funnel plot, asymmetry about the vertical line (mean of effect sizes) would suggest publication bias. The second intuitive approach involves a second type of plot called the Galbraith plot (Galbraith, 1988). In the Galbraith plot, effect sizes are standardized and plotted on the vertical axis against *precision* where precision is defined as the reciprocal of the square root of the sample variance (see Figure 2 for an example).

**Insert Figure 2 about here**

Again, sample size is proportional to precision. As sample size increases, the sample variance tends to decrease thus increasing the precision (i.e., increasing the reciprocal of the square root of variance). In the Galbraith plot, one is looking at the symmetry of the distribution of standardized effects around the mean effect (the horizontal axis). This example shows some amount of asymmetry.

We also conducted statistical tests of publication bias. More specifically, we conducted statistical tests of the symmetry of standardized effects within the Galbraith plot. This statistical test, called *Egger's Test of Asymmetry* (Egger, et al., 1997) examines the intercept of the regression line on the Galbraith plot. This regression line has the equation:

$$z_i = B_0 + B_1(Precision)_i + e_i$$

Where:

$z_i$  = is the standardized effect of study  $i$ , and

$Precision_i$  is the reciprocal of the sample variance ( $1/\sqrt{v_i}$ ) for study  $i$

In this analysis, the magnitude of the intercept provides a measure of asymmetry. Similarly, the statistical significance test that tests the null hypothesis ( $H_0: B_0 = 0$ ) is the significance test for asymmetry. In the Galbraith plot example above, the positive intercept suggests that larger samples sizes (higher precision) are associated with smaller effects. This relationship could mean that small effect, small sample size studies are not fully represented in the meta-analysis.

### **Findings/Results:**

The shape of the funnel plot below (Figure 3) indicates a degree of symmetry in the studies about the mean effect size (vertical line).

**Insert Figure 3 about here**

An additional intuitive representation is the Galbraith plot. The Galbraith plot for these studies is in Figure 4 below.

**Insert Figure 4 about here**

This plot shows that most values of study precision (i.e., the reciprocal of the square root of sample variance) fall between 0 and 1.0 and that there is a degree of symmetry in effect sizes within this range of precision values (see rectangular box). Also, symmetric distributions are characterized by near-zero precision-ES regression slopes and near-zero y-intercepts. The null hypothesis of the Egger's test of Assymetry is that the Precision-ES slope = 0 and that the intercept of the regression line = 0. The results of our Egger's test are in Table 1 below.

**Insert Table 1 about here**

This output suggests that although the y-intercept is non-zero, as is the Precision-ES slope, these values are not statistically significant at the  $\alpha = .05$  level. We conclude that for the 33 studies that were included in this 2010-11 analysis, there is **not** strong evidence of systematic publication bias against studies with small effect sizes and small sample sizes. We acknowledge here that these 33 studies are a subset of the original 60 and there is a possibility that they are systematically different from the excluded studies. We will address this validity threat as described below by contacting authors of disqualified studies, requesting the necessary information.

### **Conclusions:**

#### *Implication for Researchers*

The implications of this study for researchers are encouraging. In meta-analyses where strong publication bias exists against small effect-small sample studies, many resources are put toward finding these studies in unpublished or "grey" literature. Grey literature often includes documents such as government technical reports, doctoral dissertations and conference proceedings. The results of this analysis suggest that meta-analyses may not require significant work in the grey literature to produce unbiased results. The caveat in this conclusion is that these conclusions are based in 2010 and 2010 publications. If these studies are merged with others conducted in the entire past decade, the inverse relationship between precision and effect size for

2010-11 may become stronger and statistically significant. If this paper is accepted, the research team will increase the size of the data set in two ways. First, we will contact authors of the approximately 30 studies from 2010-11 that could not be included in the study. The goal will be to obtain the needed effect size or variance information. Second, we will replicate our searching and coding process for additional prior years of science education publications, back to 2008 at a minimum.

### *Implications for Reviewers and Editors*

For some time now, the AERA has been encouraging researchers to submit small effect-small sample studies. This platform is part of a larger effort to avoid an overreliance on statistical significance testing as the driver for publication and to put emphasis on more substantive reporting that includes reporting exact p-values, confidence intervals, and practical significance expressed through effect size metrics (AERA, 2006).

The data from this 2010-11 study suggest that small effect-small sample studies are being *submitted* for publication and are being *accepted* for publication. This result is positive for the research community two important reasons. If this trend continues, we will learn about many more science education programs that produce small but meaningful effects. Further, future meta-analyses will yield summary effects that exhibit little or no upward bias.

## Appendices

*Not included in page count.*

### Appendix A. References

- American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, 35(6), 33–40.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to Meta-Analysis*. John Wiley: West Sussex.
- Egger, M., Smith, G., Schneider, M., Minder, C. (1997). Bias in meta-analysis detected by a simple graphical test. *British Medical Journal*, 315(7109): 629-634.
- Galbraith, R. (1988). Graphical display of estimates having differing standard errors. *Technometrics*, 30(3): 271–281.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin* 86 (3): 638–641.
- Sutton, A. (2009). Publication Bias. In Cooper, H., Hedges, L., & Valentine, J. (Eds.) *Handbook of Research Synthesis and Meta-Analysis*. Russell Sage: New York.

## Appendix B. Tables and Figures

### Funnel Plot of Standard Error by Std diff in means

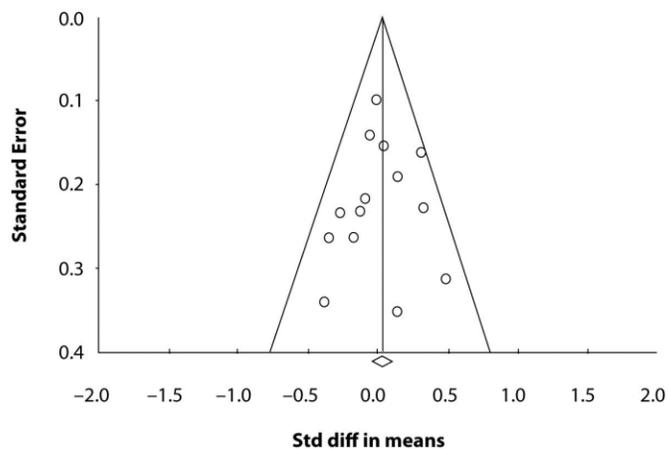


Figure 1. Example of a Funnel Plot

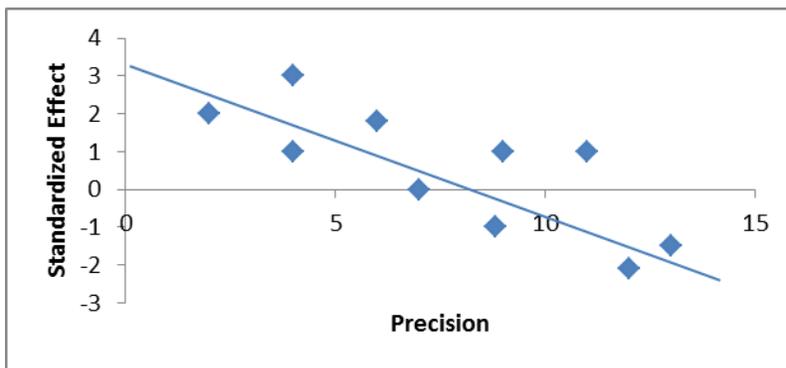


Figure 2. Example of a Galbraith Plot

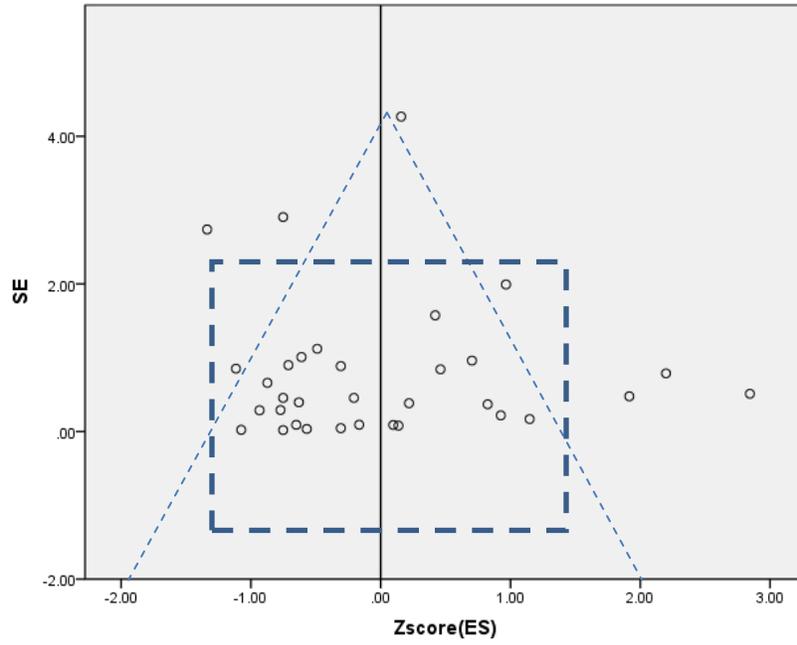


Figure 3. Funnel Plot of Standard Error vs. Standardized Effect Size

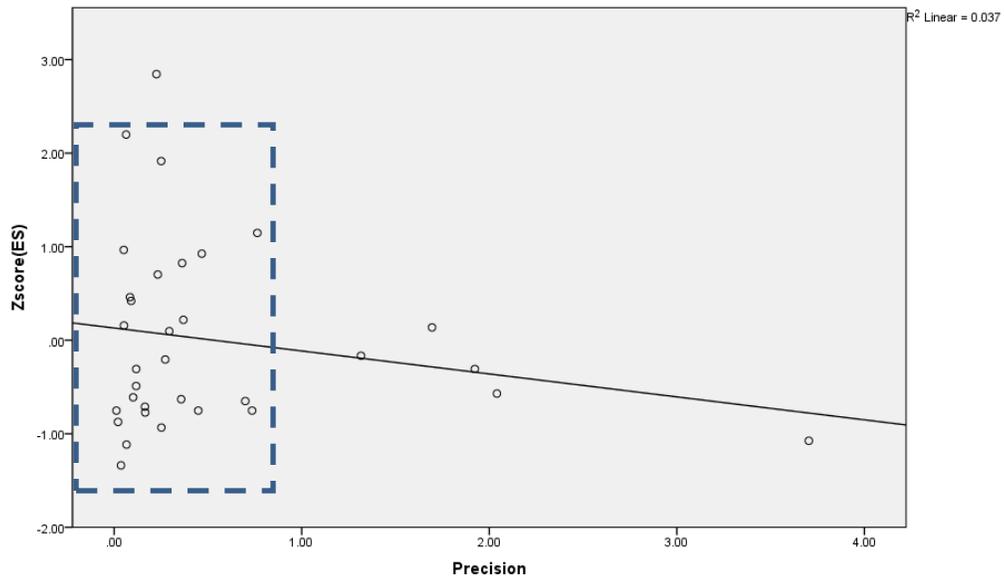


Figure 4. Galbraith Plot

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.131	.211		.620	.540
	Precision	-.245	.224	-.193	-1.095	.282

a. Dependent Variable: Zscore(ES)

*Table 1. Regression Output for Egger's Test of Asymmetry*