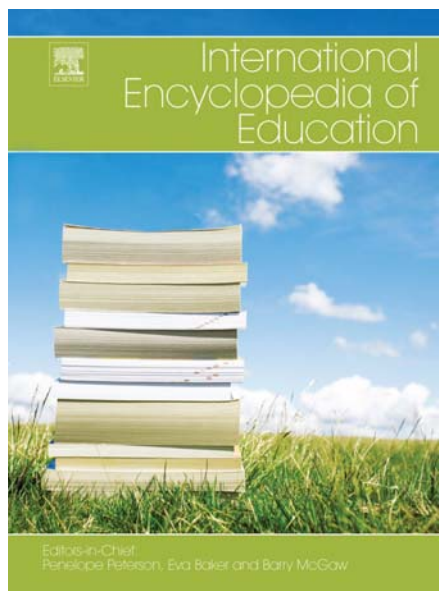

Sánchez-Meca J and Marín-Martínez F (2010), Meta Analysis. In: Penelope Peterson, Eva Baker, Barry McGaw, (Editors), International Encyclopedia of Education. volume 7, pp. 274-282. Oxford: Elsevier.

Meta Analysis

J Sánchez-Meca and F Marín-Martínez, University of Murcia, Murcia, Spain

Glossary

***d* family** – This is a set of effect-size indices which calculates the difference between the means of two groups (e.g., experimental vs. control, pre-test vs. post-test) standardized or not by an estimate of the within-study standard deviation. These effect indices are specially indicated for primary studies where subjects are assigned to different groups or conditions.
Effect size – This quantifies the extent to which the phenomenon investigated is present in the study results, regardless of the sample size and the result of the statistical tests.
Fixed-effects model – In meta-analysis, a fixed-effects model is a statistical model in which it is assumed that the effect sizes calculated from a set of primary studies estimate the same population effect size. The only error source is that produced by sampling error.
Meta-analysis – This is a research methodology that aims to quantitatively integrate the results of a set of primary studies about a given topic in order to determine the state of the art on that topic.
Mixed-effects model – When a meta-analysis aims to search for moderator variables that can explain the variability in the effect-size estimates, a mixed-effects model is applied where the studies compose a random variable and the moderator variables are fixed variables.
Moderator variable – This is a differential characteristic of the primary studies that is coded in a meta-analysis with the purpose of examining its possible relationship with the study results. The moderator variables can be substantive, methodological, or extrinsic.
Publication bias – This occurs when, in a given research field, the publication of research results depends on its direction. Meta-analyses that have included only published studies may suffer a bias in their effect estimates.
***r* family** – This is a set of effect-size indices that groups all of the correlation coefficients between two variables (e.g., Pearson correlation coefficient, biserial correlation, and point-biserial correlation). These effect indices are specially indicated for correlational studies.
Random-effects model – In meta-analysis, a random-effects model is a statistical model in which the effect size obtained from each primary study estimates a different population effect size and, as a consequence, the effect estimates have two error sources: sampling error and variability among the population effect sizes.
Standardized mean difference – This is an effect-size index from the *d* family that reflects the magnitude of the differences between the means of two groups (e.g., experimental vs. control) divided by an estimate of the within-study standard deviation.

Introduction

From the second-half of the last century, research in education has grown exponentially. From this explosion of evidence has arisen a need to develop systematic and objective methods for accumulating the scientific knowledge obtained from the primary studies. With this aim, meta-analysis has emerged as a methodology that is able to quantitatively combine the results of a set of primary investigations on a given topic in order to figure out the state of the art on that topic. Unlike traditional reviews, which are qualitative or narrative and characterized by subjectivity, meta-analysis considers the review of the past research to be a scientific enterprise that, like primary research, should be guided by the principles of objectivity, systematization, and replicability. The main characteristic of meta-analysis is the use of statistical methods to quantitatively integrate the results of the studies.

Although its origins go back to the decade of the 1930s, it was not until the end of the 1970s that G. V. Glass (1976) coined the term meta-analysis to refer to this research methodology. Meta-analysis began to gain popularity in psychology with the pioneer meta-analyses of M. L. Smith and G. V. Glass on the effectiveness of psychotherapy, R. Rosenthal and D. B. Rubin on the effects of interpersonal expectancy on the research results, and with the meta-analytic validity generalization approach of F. L. Schmidt and J. E. Hunter. In education, the pioneer meta-analysis was that carried out by M L. Smith and G. V. Glass on the effects of class size on the attitudes and academic performance of students. Since then, meta-analysis has been

International Encyclopedia of Education (2010), vol. 7, pp. 274-282

applied in education in many different research areas such as the effects of instructional programs on student performance, gender differences in academic performance, predictors of academic performance, or the validity of screening and diagnostic instruments in detecting psychoeducational problems.

The phases in developing a meta-analysis, an overview of the effect-size indices, the statistical models most usually applied in meta-analysis, an example to illustrate the meta-analytic calculations, and some final remarks are presented in this article. Throughout the article, the meta-analysis of Ginns (2006) on the instructional effects of spatial and temporal contiguity of learning materials is used as an example.

Phases in a Meta-Analysis

To carry out a meta-analysis, the researcher must follow several steps: (1) formulating the problem, (2) searching for the literature, (3) coding the studies, (4) statistical analysis and interpretation, and (5) publication of the meta-analysis.

Given below are the details of the steps to be followed:

1. *Formulating the problem.* As in any primary research, the first step in a meta-analysis consists of defining its purpose, which generally will be to examine the relationship between two or more (psychological and/or educational) constructs. In this phase, the researcher must define the constructs implied, both theoretically and empirically, as well as review the existing theoretical models and formulate the concrete objectives of the review.

   In Ginns' (2006) meta-analysis, the purpose was to review the experimental studies about the effects of instructional learning by manipulating the spatial or temporal contiguity of disparate but related elements of information. In this meta-analysis, Ginns (2006) tried to determine the extent to which the effects of instructional learning could be generalized across different students, learning materials, and testing contexts.
2. *Searching for the literature.* Once the objectives are formulated, the next step involves defining the selection criteria that the empirical studies must fulfill and carrying out as complete a literature search as possible. The selection criteria will depend on the purpose of the meta-analysis, but there are several criteria that should be present in any meta-analysis, such as specifying: (1) the time period of the studies, (2) the design type in the empirical studies, and (3) the language in which the study was written. In order to search for the studies that fulfill the selection criteria, several searching strategies should be used, combining both formal and informal search procedures. Formal procedures consist of consulting electronic bibliographic databases (e.g., ERIC, PsycINFO), relevant journals, bibliographic reference volumes, and references in studies. Informal sources enable us to find fugitive literature, that is, papers that have not been published (e.g., dissertations, technical reports, papers presented at congresses) or papers published in journals or books that cannot be found through formal sources.

   In Ginns' (2006) example, in order to be included in the meta-analysis the studies had to apply a between-group design with random assignment of the subjects to the groups and to report statistical data about the effects of manipulating spatial or temporal contiguity of instructional materials. The search strategy included consulting the ERIC and PsycINFO databases, Science Citation Index, and the references of the papers located.
3. *Coding of studies.* The main purpose of a meta-analysis is to explain the variability found in the study results on a given topic by examining the influence of differential characteristics among them. To accomplish this objective, the studies are subjected to a coding process in which relevant moderator variables of the results are identified. Moderator variables are study characteristics that can influence the outcome. Although the moderator variables to be coded depend on the purpose of the meta-analysis, it is usual to classify them into three main clusters: methodological, substantive, and extrinsic characteristics.

   Methodological characteristics refer to those related with the methodology of the study, such as the design type or the amount of attrition in the groups. Substantive characteristics are those directly related with the research topic investigated in the meta-analysis, as well as with the sociodemographic characteristics of the sample. In Ginns' (2006) example, some of the substantive characteristics coded were the type of split-attention effect (spatial or temporal contiguity), the field of study (science, or engineering/technical), and the educational level of the sample. Finally, extrinsic variables are those characteristics that should not be related with the study results because they have nothing to do with the research enterprise. Examples of this type of variables are the publication year of the study, the publication source (published or unpublished), and characteristics of the researchers (e.g., gender, affiliation). In addition to the moderator variables, the main result of each study is summarized by calculating an effect-size estimate.
4. *Statistical analysis and interpretation.* In a meta-analysis a mean effect is calculated from the set of effect sizes, as well as a confidence interval and an assessment of the extent of heterogeneity exhibited by them around the mean effect. If there is more heterogeneity than sampling error can explain, then additional analyses

are done to search for moderator variables of the effect estimates.

5. *Publication.* The final step in a meta-analysis is its publication. The sections that a meta-analytic report should include are basically the same as those in a primary study (introduction, method, results, and discussion), but in the method section, the epigraphs are not the same as those in primary research: literature search, coding of studies, effect-size index, and statistical analysis. A meta-analysis that is to be published must include all the information needed for its potential replication by another researcher.

Effect-Size Indices

An essential requisite to carry out a meta-analysis is to obtain a quantitative index that summarizes the results of each study. Moreover, the quantitative indices obtained from the studies must be in the same metric for their quantitative integration to be possible. The best strategy for accomplishing this requisite is to calculate an effect-size index, which represents the extent to which the phenomenon of interest is manifested in the study results. There are many different effect-sizes indices, the choice depending on the design type of the studies and on the nature of the variables involved in the studies (continuous, dichotomous, etc.). In education, the most frequently applied effect sizes in meta-analysis are those grouped into two families: the d family, for experimental and quasi-experimental designs, and the r family, for correlational studies (Hedges and Olkin, 1985).

The *d* Family

When the study design involves assigning subjects to different groups (e.g., experimental vs. control) and the dependent variable is continuous, the most appropriate family of effect sizes is the d family, which includes a set of effect sizes defined as the difference between two means divided by a within-group standard deviation. When the study comprises a two-group design, the effect size most usually applied from the d family is the standardized mean difference, which is calculated by

$$d = c(m)\frac{\bar{y}_E - \bar{y}_C}{S} \qquad [1]$$

$\bar{y}_E$ and $\bar{y}_C$ being the estimated means of the experimental and control groups, respectively, S being the estimated pooled within-study standard deviation obtained by

$$S = \sqrt{\frac{(n_E - 1)S_E^2 + (n_C - 1)S_C^2}{n_E + n_C - 2}} \qquad [2]$$

with n_E and n_C being the sample sizes of both groups, and S_E^2 and S_C^2 the respective estimated variances. When the study design includes a control group, Glass *et al.* (1981) proposed dividing the mean difference by the standard deviation of the control group, S_c, instead of S, because sometimes applying a treatment can alter the variability in the scores of the dependent variable. But provided homoscedasticity is met, S is a more efficient estimate of population standard deviation. The $c(m)$ factor is needed to correct a positive bias of the d index for small sample sizes and is obtained by

$$c(m) = 1 - \frac{3}{4(n_E + n_C) - 9} \qquad [3]$$

The d index is an estimate of the population standardized mean difference, $\delta = (\mu_E - \mu_C)/\sigma$, with μ_E and μ_C being the population means of the experimental and control groups, and σ being the common population standard deviation. As an estimate of δ, the d index is approximately normally distributed with mean δ and sampling variance σ_d^2, which is estimated by

$$\hat{\sigma}_d^2 = \frac{n_E + n_C}{n_E n_C} + \frac{d^2}{2(n_E + n_C)} \qquad [4]$$

Dividing the difference in mean by the standard deviation of the groups in [1] makes it possible to homogenize the metric of studies that have used different dependent variables to measure the outcome.

Several adaptations of the d index have been devised for studies that include pre-test and post-test measures for only one group or for a two-group design (Morris, 2008). Other effect-size indices that can be applied when the design is composed of two groups and the dependent variable is dichotomous, or has been dichotomized, are the difference between the success (or failure) proportions for the experimental and the control groups (or risk difference), the rate between both proportions (or risk ratio), and the odds ratio (Deeks and Altman, 2001).

The *r* Family

When the study results are obtained by applying a correlational design, that is, studies where the researcher does not manipulate independent variables, the most appropriate effect size consists of calculating a correlation coefficient. Depending on the type of variables involved in the relationship (continuous, dichotomous, dichotomized, ordinal variables, etc.), different correlation coefficients are used as the effect size: Pearson correlation coefficient, point-biserial correlation, Spearman's rank order correlation, etc. All of these correlation coefficients make up the r family of effect-size indices.

As the correlation coefficients do not always follow a normal distribution, their transformation into the Fisher's Z has been proposed to normalize the sampling distribution and to stabilize the variance. Thus, it is very usual in meta-analysis to transform the correlation coefficients, r, into Fisher's Z by means of

$$Z_r = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \quad [5]$$

where ln is the natural logarithm. However, there is no clear consensus on this topic, and some authors advise against transforming correlation coefficients into Fisher's Z (e.g., Hunter and Schmidt, 2004). The sampling variance of Fisher's Z is a function of the sample size, N:

$$\sigma^2_{Z_r} = \frac{1}{N-3} \quad [6]$$

Sometimes, a meta-analysis integrates experimental and correlational studies, so that the meta-analyst has been able to obtain a d index from the experimental studies and an r index from the correlational ones. In order to put all of the effect sizes into the same metric, formulas have been devised to transform them. Thus, to transform a d index into an r index, or vice versa, we can apply the equation

$$r = \frac{d}{\sqrt{d^2+4}} \quad [7]$$

There are also formulas for transforming an odds ratio into a d index (Sánchez-Meca *et al.*, 2003). Thus, most of the effect-size indices usually applied in meta-analysis can be transformed into one another.

Statistical Procedures in Meta-Analysis

Once we have coded a set of moderator variables in the studies and we have calculated an effect-size estimate from each of them, the statistical analyses carried out in a meta-analysis aim to answer several questions: (1) What is the average effect magnitude throughout the studies? (2) Is the average effect size statistically significant? (3) Is there heterogeneity among the effect-size estimates? and (4) If the effect estimates are not homogeneous, which of the moderator variables can explain the variability?

There is currently a consensus on the convenience of applying weighting procedures in meta-analysis in order to give more weight to the effect sizes obtained from the studies with a lesser sampling variance, that is, with a larger sample size. Depending on the assumptions made by the meta-analyst, the weighting procedures can be approached from different statistical models, basically, fixed-, random-, and mixed-effects models.

Calculating an Average Effect Size

The statistical analyses in a meta-analysis are guided by a statistical model that must be previously assumed. The main task of the statistical model is to establish the properties of the effect-size population from which the individual effect-size estimates have been selected. To accomplish the first purpose in a meta-analysis, that is, to calculate an average effect size, two statistical models can be assumed: the fixed- and the random-effects models.

Suppose there are k independent empirical studies about a given topic and T_i is the effect-size estimate obtained in the ith study (here T_i refers to any of the different effect-size indices presented above, both from the d and the r families.) In a fixed-effects model, it is assumed that all of the effect-size estimates come from a population with a common parametric effect size, θ, and as a consequence the only error source is that produced by sampling error, e_i. Thus, the model can be formulated as $T_i = \theta + e_i$, the sampling errors, e_i, being normally distributed with mean 0 and sampling variance σ_i^2, $e_i \sim N(0, \sigma_i^2)$. Therefore, the effect-size estimates, T_i, are also normally distributed with mean θ and sampling variance σ_i^2, $T_i \sim N(\theta, \sigma_i^2)$.

In a random-effects model, it is assumed that the effect-size estimates, T_i, estimate different population effect sizes, θ_i, that is, $T_i = \theta_i + e_i$, and θ_i pertains to a distribution of parametric effect sizes with mean μ and variance τ^2, usually called between-studies variance. The parametric effect sizes can be modeled as $\theta_i = \mu + \varepsilon_i$, ε_i being the errors of the parameters around its mean, μ. Therefore, the random-effects model is formulated as $T_i = \mu + \varepsilon_i + e_i$. Assuming normality, T_i has as mean μ and variance $\tau^2 + \sigma_i^2$, $T_i \sim N(\mu, \tau^2 + \sigma_i^2)$. Thus, the fixed-effects model can be considered a particular case of the random-effects model when the between-studies variance is zero ($\tau^2 = 0$) and, as a consequence, all the parametric effect sizes are equal ($\theta_1 = \theta_2 = \ldots = \theta_{i\ldots} = \theta = \mu$).

To calculate an average effect size from a set of studies, each effect-size estimate must be weighted by its precision. Both in a fixed- and a random-effects model, the uniformly minimum variance unbiased estimator (UMVUE) of the average effect size, μ, is that obtained by weighting each effect-size estimate by its inverse variance:

$$T_{\text{UMVUE}} = \frac{\sum_{i=1}^{k} w_i T_i}{\sum_{i=1}^{k} w_i} \quad [8]$$

where w_i is the optimal weight for the ith study and, depending on the statistical model assumed, it is defined as $w_i^{\text{FE}} = 1/\sigma_i^2$ or as $w_i^{\text{RE}} = 1/(\tau^2 + \sigma_i^2)$, for the fixed- and the random-effects models, respectively.

In practice, the optimal weights cannot be used, because the within-study sampling variances, σ_i^2, and the between-studies variance, τ^2, are unknown. For each effect-size index, formulas have been devised to estimate σ_i^2 and τ^2. Thus, the estimated weights are defined as $\hat{w}_i^{\text{FE}} = 1/\hat{\sigma}_i^2$ and $\hat{w}_i^{\text{RE}} = 1/(\hat{\tau}^2 + \hat{\sigma}_i^2)$ for fixed- and random-effects models, respectively. (Another option consists of simply weighting each effect-size estimate by its sample size (Hunter and Schmidt, 2004).) Here $\hat{\sigma}_i^2$ is the estimated within-study sampling variance for the ith study (e.g., eqn [4] for the d index), and $\hat{\tau}^2$ can be calculated by applying one of the different estimators of the between-studies variance proposed in the meta-analytic literature. The τ^2 estimator most usually applied in meta-analysis is that based on the moments method:

$$\hat{\tau}^2 = \frac{Q - (k-1)}{c} \qquad [9]$$

where k is the number of studies, Q is a heterogeneity statistic defined in eqn [17], and c is obtained by:

$$c = \sum_{i=1}^{k} \hat{w}_i^{FE} - \frac{\sum_{i=1}^{k} (\hat{w}_i^{FE})^2}{\sum_{i=1}^{k} \hat{w}_i^{FE}} \qquad [10]$$

When $Q < (k-1)$, then $\hat{\tau}^2$ is negative and must be truncated to zero. Other τ^2 estimators can be consulted in Viechtbauer (2005).

With the respective estimated variances, the population effect size, μ, is then estimated by:

$$T_{FE} = \frac{\sum_{i=1}^{k} \hat{w}_i^{FE} T_i}{\sum_{i=1}^{k} \hat{w}_i^{FE}} \qquad [11]$$

$$T_{RE} = \frac{\sum_{i=1}^{k} \hat{w}_i^{RE} T_i}{\sum_{i=1}^{k} \hat{w}_i^{RE}} \qquad [12]$$

for fixed- and random-effects models, respectively. When a fixed-effects model is assumed, T_{FE} is approximately normally distributed and its sampling variance defined as:

$$V(T_{FE}) = 1 / \sum_{i=l}^{k} \hat{w}_i^{FE} \qquad [13]$$

Thus, a confidence interval for the average effect size can be obtained by (cf. e.g., Cooper *et al.* 2009):

$$T_{FE} \pm z_{\alpha/2} \sqrt{V(T_{FE})} \qquad [14]$$

where $z_{\alpha/2}$ is the $100(\alpha/2)$ percentile of the standard normal distribution and α is a significance level.

Under a random-effects model, a better approach for obtaining a confidence interval for the overall effect size consists of assuming a Student t reference distribution with $k-1$ degrees of freedom, instead of the standard normal distribution:

$$T_{RE} \pm t_{k-1,\alpha/2} \sqrt{V(T_{RE})} \qquad [15]$$

where $t_{k\text{-}1,\alpha/2}$ is the $100(\alpha/2)$ percentile of the Student t distribution with $k-1$ degrees of freedom, and $V(T_{RE})$ is an estimate of the sampling variance for T_{RE}, which is obtained by (cf. Sánchez-Meca and Marín-Martínez, 2008)

$$V(T_{RE}) = \frac{\sum_{i=1}^{k} \hat{w}_i^{RE} (T_i - T_{RE})^2}{(k-1) \sum_{i=1}^{k} \hat{w}_i^{RE}} \qquad [16]$$

Alternative approaches to those presented in eqns [14] and [15] have been proposed (cf. e.g., Sánchez-Meca and Marín-Martínez, 2008).

Assessing Heterogeneity in the Effect Sizes

One of the main purposes of meta-analysis is to examine the variability among the effect-size estimates. If the heterogeneity exhibited by the effect estimates can be explained by random sampling alone, then the average effect size represents a good summary of the general trend in the study results. If, on the contrary, the effect estimates reflect true heterogeneity, that is, variability among the effect sizes due to differential characteristics of the studies, then the average effect size does not represent the effect estimates, and the meta-analyst should search for moderator variables that can explain (at least part of) the heterogeneity.

To assess whether a set k of independent effect estimates show true heterogeneity, the most usual strategy in meta-analysis is to apply the heterogeneity Q statistic which, under the null hypothesis of homogeneity (H_0: $\theta_1 = \theta_2 = \ldots = \theta_i \ldots = \theta = \mu$), follows a chi-square distribution with $k-1$ degrees of freedom, and is obtained through

$$Q = \sum_{i=1}^{k} \hat{w}_i^{FE} (T_i - T_{FE})^2 \qquad [17]$$

By assuming a given significance level, α, it is possible to make a decision about whether the effect-size estimates are homogeneous or not. However, due to the low statistical power of the Q statistic when the number of studies is small (about $k < 30$), it is advisable to complement this statistical test with the I^2 index, which quantifies the extent of true heterogeneity present in the studies (Higgins and Thompson, 2002):

$$I^2 = \frac{Q - (k-1)}{Q} \times 100 \qquad [18]$$

When $Q < (k-1)$, then I^2 is truncated to zero. Values of $I^2 = 25\%$, 50%, and 75% can be interpreted as reflecting small, medium, and high heterogeneity among the effect sizes, respectively.

Analyzing the Influence of Moderator Variables

If the effect-size estimates exhibit more heterogeneity than random sampling can explain, then the meta-analyst has to search for moderator variables that can account for that heterogeneity. In order to accomplish this objective, linear models are assumed, where the effect-size estimates are taken as the dependent variable, and moderator variables coded in the studies act as predictor variables. The most appropriate statistical model for dealing with this task is the mixed-effects model, in which the studies (or the effect-size estimates) are taken as random-effects variables, and study-level moderator variables are taken as fixed-effects variables. When the moderator variable is categorical (e.g., type of population), an analysis of variance by weighted least squares is applied, whereas regression models are used to assess the influence of continuous moderator variables (e.g., duration of the program, mean age in the sample).

In matrix notation the regression coefficient vector of the model is estimated by $\hat{\boldsymbol{\beta}} = (\mathbf{X'WX})^{-1}\mathbf{X'WT}$, where $\mathbf{X}$ is a kxp design matrix that includes p predictors, $\mathbf{T}$ is the effect-sizes vector, with range k, and $\mathbf{W}$ is a kxk diagonal weighting matrix, with the element in the diagonal defined as the estimated weights: $\hat{w}_i^{\mathrm{ME}} = 1/(\hat{\tau}^2_{\mathrm{ME}} + \hat{\sigma}_i^2)$. The between-studies variance, $\hat{\tau}^2_{\mathrm{ME}}$, is estimated by:

$$\hat{\tau}^2_{\mathrm{ME}} = \frac{Q_E - (k-p)}{tr\mathbf{W} - \; tr[\mathbf{WX}(\mathbf{X'WX})^{-1}\mathbf{X'W}]} \qquad [19]$$

where tr represents the trace of a matrix and Q_E is the weighted residual sum of squares of the regression model, that is defined as:

$$Q_{\mathrm{E}} = \mathbf{T'WT} \qquad [20]$$

Model misspecification can be tested by using the Q_E statistic. Under the null hypothesis that the regression model is not misspecified (H_0: $\mathbf{T} = \mathbf{X}\boldsymbol{\beta}$), the Q_E statistic follows a chi-square distribution with $k - p - 1$ degrees of freedom.

Once the model misspecification has been examined, the significance of the full regression model is tested with the weighted regression sum of squares, Q_R, which is obtained by

$$Q_R = \hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}} \qquad [21]$$

where $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}$ is the estimated variance–covariance matrix of the coefficient vector. Under the null hypothesis of no relationship between the predictors and the effect estimates (H_0: $\boldsymbol{\beta} = \mathbf{0}$), the Q_R statistic follows a chi-square distribution with p degrees of freedom. Finally, the partialized effect of each predictor can be assessed by applying a Z test for its regression coefficient, $\hat{\beta}_j$:

$$Z = \hat{\beta}_j/\hat{\sigma}_{\hat{\beta}_j} \qquad [22]$$

where $\hat{\sigma}_{\hat{\beta}_j}$ is the estimated standard error of the regression coefficient, taken from the jth diagonal element of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}$. Under the null hypothesis of no effect for the predictor, Z follows a standard normal distribution.

An Example

To illustrate the meta-analytic calculations presented in the previous epigraph, we have selected part of the data reported in Ginns' (2006) meta-analysis about the instructional effects of manipulating the temporal contiguity of learning materials. The effect-size index was the standardized mean difference, d_i, defined as the difference between the means of the experimental (temporally integrated materials) and control (non-integrated condition) groups divided by the pooled standard deviation (see eqn [1]). **Table 1** shows the meta-analytic database for this example, which includes information relative to 13 studies. (As Ginns' (2006) meta-analysis did not report the sample size for each study, we have invented them in order to carry out the statistical analyses. Therefore, our results have only illustrative purposes and do not coincide with those presented in Ginns (2006).) The moderator variables reported in **Table 1** were the publication year, the sample size, n_i (here we assumed that the sample sizes of the experimental and control groups were equal in each study, i.e., $n_{\mathrm{E}i} = n_{\mathrm{C}i} = n_i$), the type of testing (group vs. individual), and the field of the study (science vs. engineering/technical). **Table 1** also shows the estimated within-study variance, $\hat{\sigma}_i^2$, for each study, obtained by eqn [4], and the estimated fixed- and random-effects weights: $\hat{w}_i^{\mathrm{FE}} = 1/\hat{\sigma}_i^2$ and $\hat{w}_i^{\mathrm{RE}} = 1/(\hat{\tau}^2 + \hat{\sigma}_i^2)$, respectively (**Figure 1**).

A graphical representation typical in meta-analysis is the forest plot, which consists of showing each individual estimated effect size and its 95% confidence interval and, at the bottom of the graph, the average effect size for all of the studies. This graph helps to capture the general trend and the variability of the effect sizes.

Table 2 presents the average effect size obtained withthe 13 effect estimates and the 95% confidence interval for the fixed- and the random-effects models. Thus, assuming a fixed-effects model, the average effect size was $d_+ = 0.768$ (by applying eqn [11]), with confidence limits, 0.637 and 0.899 (by applying eqn [14]). Assuming a random-effects model, the average effect size was $d_+ = 0.888$ (with eqn [12]), with confidence limits, 0.574 and 1.202 (with eqn [15]). In both cases, the average effect size was statistically significant, as the null effect was not contained in the confidence intervals. The confidence interval for the random-effects model was larger (interval amplitude: 0.628) than that of the fixed-effects model (interval amplitude: 0.262), due to the fact that the random-effects model assumes that the effect sizes of the studies estimate different population effect sizes.

Table 1 Meta-analytic dataset of the effects of the temporal contiguity of instructional materials on student performance

Study	*Year*	*Testing*	*Broad field of the study*	n_i	d_i	$\hat{\sigma}_i^2$	$\hat{w}_i^{FE}$	$\hat{w}_i^{RE}$
1	1999	Group	Science	12	2.06	0.2551	3.9204	2.4352
2	1983	Group	Science	22	1.40	0.1132	8.8353	3.7209
3	1983	Group	Science	45	0.27	0.0448	22.2968	4.9896
4	1991	Individual	Science	40	1.10	0.0576	17.3724	4.6920
5	1994	Individual	Science	65	0.58	0.0321	31.1885	5.3296
6	1999	Group	Engineering/Technical	22	1.31	0.1104	9.0571	3.7597
7	1984	Group	Engineering/Technical	52	0.65	0.0405	24.6958	5.1005
8	1984	Group	Engineering/Technical	33	0.50	0.0625	16.0000	4.5857
9	1992	Individual	Engineering/Technical	30	1.64	0.0891	11.2259	4.0875
10	1992	Individual	Engineering/Technical	35	1.40	0.0711	14.0562	4.4109
11	1991	Individual	Engineering/Technical	35	0.87	0.0625	15.9874	4.5847
12	1994	Individual	Engineering/Technical	52	0.49	0.0396	25.2424	5.1234
13	2000	Individual	Engineering/Technical	48	0.33	0.0422	23.6777	5.0556

n_i – sample size for each group of the ith study, the total sample size of the study being $N_i = 2 \mathrm{x} n_i$; d_i – standardized mean difference for the ith study; $\hat{\sigma}_i^2$ – estimated within-study variance for the ith study (calculated by eqn [4]); $\hat{w}_i^{\mathrm{FE}} = 1/\hat{\sigma}_i^2$ is the estimated fixed-effects weight for the ith study; and $\hat{w}_i^{\mathrm{RE}} = 1/(\hat{\tau}^2 + \hat{\sigma}_i^2)$ is the estimated random-effects weight for the ith study.
From Ginns P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction* **16**, 511–525.

Study name	Statistics for each study			
	d	Variance	Lower limit	Upper limit
Study 1	2,060	0,255	1,070	3,050
Study 2	1,400	0,113	0,741	2,059
Study 3	0,270	0,045	-0,145	0,685
Study 4	1,100	0,058	0,630	1,570
Study 5	0,580	0,032	0,229	0,931
Study 6	1,310	0,110	0,659	1,961
Study 7	0,650	0,041	0,256	1,044
Study 8	0,500	0,063	0,010	0,990
Study 9	1,640	0,089	1,055	2,225
Study 10	1,400	0,071	0,877	1,923
Study 11	0,870	0,063	0,380	1,360
Study 12	0,490	0,040	0,100	0,880
Study 13	0,330	0,042	-0,073	0,733
Fixed-effects mean	0,768	0,004	0,637	0,899

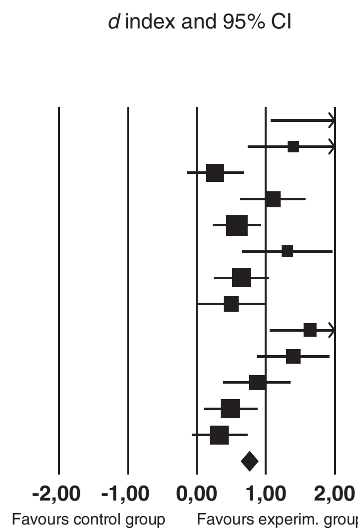

Figure 1 Forest plot of the *d* indices for 13 studies on the temporal contiguity effects. The variances were calculated by eqn [4], the 95% confidence intervals by eqn [11], and the mean effect size from a fixed-effects model by eqn [13].

Table 2 Summary results of calculating an average effect size and a 95% confidence interval from the fixed- and random-effects models

Statistical model	*k*	d_+	$V(d_+)$	*95% C. I.* d_l	d_u
Fixed-effects	13	0.768	0.0045	0.637	0.899
Random-effects	13	0.888	0.0208	0.574	1.202

K – number of studies; d_+ – average effect size calculated by eqns [11] and [12] for the fixed- and random-effects models, respectively; $V(d_+)$ – sampling variance of d_+, calculated by eqns [13] and [16], respectively; d_l and d_u – lower and upper confidence limits of the 95% confidence interval around d_+, calculated by eqns [14] and [15], respectively.

The heterogeneity Q statistic, calculated by eqn [17], was statistically significant [$Q(12) = 43.584$, $p < .001$]; the I^2 index, by eqn [18], was of high magnitude: $I^2 = 72.5\%$; and the between-studies variance was $\hat{\tau}^2 = 0.155$ (by eqn [9]). Therefore, the set of effect sizes exhibited more variability than random sampling can explain and, as a consequence, the search for moderator variables is justified. In order to do this, analyses of variance and regression models by weighted least squares and assuming a mixed-effects model can be applied.

Table 3 presents an example of how to analyze the influence of a qualitative moderator variable on the effect size by applying an analysis of variance. The moderator

Table 3 Analysis of variance by weighted least squares and assuming a mixed-effects model of the *d* indices as a function of the field of the study

			95% C. I.				
Field of study	*k*	d_+	d_l	d_u	Q_{wj}	*DF*	*p*
Science	5	0.947	0.499	1.395	6.336	4	.175
Engineering/ technology	8	0.523	0.523	1.205	6.762	7	.454
			$Q_B(1) = 0.084$, $p = .772$				
			$Q_W(11) = 13.098$, $p = .287$				

K – number of studies for each category; d_+ – average effect size for each category; d_l – and d_u – lower and upper confidence limits of the 95% confidence interval around d_+ for each category; Q_{wj} – within-category homogeneity *Q* statistic which, under the null hypothesis of within-category homogeneity, follows a chi-square distribution with $k - 1$ degrees of freedom (*DF*); Q_W – global within-category homogeneity *Q* statistic which, under the null hypothesis of global within-category homogeneity, follows a chi-square distribution with $k - c$ degrees of freedom, *c* being the number of categories; Q_B – between-categories *Q* statistic which, under the null hypothesis of equal average effect sizes, follows a chi-square distribution with $c - 1$ degrees of freedom; and *p* – probability level of the corresponding *Q* statistic.

Table 4 Simple regression models by weighted least squares and assuming a mixed-effects model of the *d* indices as a function of the publication year and sample size

Predictor variable	*k*	$\hat{\beta}_j$	$SE(\hat{\beta}_j)$	$Q_R(DF)$	*p*	$Q_E(DF)$	*p*	R^2
Year	13	0.018	0.024	0.604(1)	.437	12.771(11)	.308	.045
Sample size	13	–0.027	0.008	11.886(1)	< .001	10.630(11)	.475	.528

K – number of studies for each category; $\hat{\beta}_j$ – unstandardized regression coefficient for each predictor variable; $SE(\hat{\beta}_j)$ – standard error of $\hat{\beta}_j$; Q_R – weighted regression sum of squares (by eqn [21]); *DF* – degrees of freedom; Q_E – weighted residual sum of squares (by eqn [20]) *p* – probability level of the corresponding *Q* statistic; and R^2 – proportion of variance accounted for.

variable is the field of study (science vs. engineering/technical). The average effect sizes for both categories were very similar: $d_+ = 0.947$ for science and $d_+ = 0.864$ for engineering/technical, leading to a nonstatistically significant result for the Q_B statistic, which assesses whether the mean effect sizes are equal [$Q_B(1) = 0.084$, $p = .772$]. The Q_W statistic reveals that the model is not misspecified [$Q_W(11) = 13.098$, $p = .287$]. Therefore, the field of study does not seem to be a moderator variable of the effect sizes.

Finally, **Table 4** presents the results of two simple regression analyses, one for the publication year and the other for the sample size. The sample size exhibited a statistically significant relationship with the effect sizes, as the Q_R statistic reveals [$Q_R(1) = 11.886$, $p < 0.001$], whereas publication year did not reach statistical significance [$Q_R(1) = 0.604$, $p = 0.437$]. Moreover, the regression model for sample size was not misspecified [$Q_E(11) = 10.630$, $p = 0.475$]. Hence, it seems that the effect sizes were associated with sample sizes.

Concluding Remarks

In this article, we have focused on the meta-analytic approach most usually applied in education, which consists of calculating an effect-size index from each study and putting it into relation with study-level moderator variables that can explain the variability usually exhibited by the effect sizes. In addition to the statistical methods presented here, other analytic approaches have been devised in the meta-analytic literature. Thus, Hunter and Schmidt (2004) proposed the validity generalization approach with the purpose of integrating a set of validity coefficients of a given measurement instrument obtained from applying the instrument to different samples and contexts. In this meta-analytic approach, also named psychometric meta-analysis, different statistical and measurement artifacts that affect the empirical validity coefficients can be corrected, such as the measurement error, the range restriction, or the effects of dichotomizing variables. In the same vein, Vacha-Haase (1998) has proposed the meta-analytic reliability generalization approach, with the aim of integrating a set of reliability coefficients obtained in successive applications of the same measurement instrument, and assessing whether the reliability of the test scores can be generalized across different samples, contexts, and applications. Other specific approaches devised in meta-analysis enable us to apply multivariate models to integrate a set of dependent effect-size estimates obtained from each study (cf. e.g., Becker, 2000), or to integrate correlation matrices obtained from the studies in order to meta-analyze factor analyses (e.g., Becker, 1996).

Without a doubt, meta-analysis is contributing in education to better accumulate the scientific evidence disseminated across studies by applying the same scientific rigor demanded of primary research. Meta-analysis has become a necessary link between the past research and the future research efforts in any field. The main characteristic of this research methodology is the use of statistical methods to integrate study results but, as it is a relatively new field, more methodological work is needed in order to adapt the current meta-analytic techniques to address more complex problems.

See also: Measure of Association; Statistical Significance Versus Effect Size.

Bibliography

Becker, B. J. (2000). Multivariate meta-analysis. In Tinsley, H. E. A. and Brown, S. D. (eds.) *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, pp 499–525. New York: Academic Press.

Becker, G. (1996). The meta-analysis of factor analyses: An illustration based on the cumulation of correlation matrices. *Psychological Methods* **1**, 341–353.

Cooper, H., Hedges, L. V., and Valentine, J. C. (eds.) (2009). *The Handbook of Research Synthesis and Meta-analysis*, 2nd edn. New York: Russell Sage Foundation.

Deeks, J. J. and Altman, D. G. (2001). Effect measures for meta-analysis of trials with binary outcomes. In Egger, M., Smith, G. D., and Altman, D. (eds.) *Systematic Reviews in Health Care,* 2nd edn., pp 313–335. London: BMJ Publication Group.

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction* **16**, 511–525.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher* **5**, 3–8.

Glass, G. V., McGaw, B., and Smith, M. L. (1981). *Meta-analysis in Social Research.* Beverly Hills, CA: Sage.

Hedges, L. V. and Olkin, I. (1985). *Statistical Methods for Meta-Analysis.* Orlando, FL: Academic Press.

Higgins, T. P. T. and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**, 1539–1558.

Hunter, J. E. and Schmidt, G. L. (2004). *Methods of Meta-Analysis: Correcting Error and Bias in Research Synthesis,* 2nd edn. Newbury Park, CA: Sage.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-group designs. *Organizational Research Methods* **11**, 364–386.

Sánchez-Meca, J. and Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods* **13**, 31–48.

Sánchez-Meca, J., Marín-Martínez, F., and Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods* **8**, 448–467.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement* **58**, 6–20.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics* **30**, 261–293.

Further Reading

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2009). *Introduction to Meta-analysis*. Chichester, UK: Wiley.

Cooper, H. (2009). *Research Synthesis and Meta-analysis: A Step-by-Step Approach,* 3rd edn. Thousand Oaks, CA: Sage.

Hedges, L. V. and Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods* **3**, 486–504.

Kalaian, H. A. and Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods* **1**, 227–235.

Lipsey, M. W. and Wilson, D. B. (2001). *Practical Meta-Analysis.* Thousand Oaks, CA: Sage.

Petticrew, M. and Roberts, H. (2006). *Systematic Reviews in the Social Sciences: A Practical Guide.* Malden, MA: Blackwell.

Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research,* revised edition. Newbury Park, CA: Sage.

Rothstein, H. R., Sutton, A. J., and Borenstein, M. (eds.) (2005). *Publication Bias in Meta-Analysis: Prevention, Assessment, and Adjustments.* Chichester, UK: Wiley.

Sánchez-Meca, J. and Marín-Martínez, F. (1998). Testing continuous moderators in meta-analysis: A comparison of procedures. *British Journal of Mathematical and Statistical Psychology* **51**, 311–326.

Schmidt, F. L. (2008). Meta-analysis: A constantly evolving research integration tool. *Organizational Research Methods* **11**, 96–113.

Schmidt, F. L., Oh, I.-S., and Hayes, T. L. (2009). Fixed versus random effects models in meta-analysis: Model properties and an empirical comparison of difference in results. *British Journal of Mathematical and Statistical Psychology* **62**, 97–128.

Schulze, R. (2004). *Meta-Analysis: A Comparison of Approaches.* Cambridge, MA: Hogrefe and Huber.

Sutton, A. J. and Higgins, J. P. T. (2008). Recent developments in meta-analysis. *Statistics in Medicine* **27**, 625–650.

Viechtbauer, W. (2007). Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis. *Zeitschrift für Psychologie/Journal of Psychology* **215**, 104–121.

Whitehead, A. (2002). *Meta-Analysis of Controlled Clinical Trials.* Chichester, UK: Wiley.

Relevant Websites

http://www.vanderbilt.edu – Center for Evaluation Research and Methodology, Vanderbilt University (Nashville, TN).

www.campbellcollaboration.org – The Campbell Collaboration.

www.duke.edu – The Campbell Methods Group.

www.cochrane.org – The Cochrane Collaboration.

http://www.um.es – The Meta-Analysis Unit, University of Murcia, Murcia, Spain.

http://mason.gmu.edu – Website of Prof. David B. Wilson, George Mason University.

http://faculty.ucmerced.edu – Website of Prof. William R. Shadish, University of California, Merced.