anales de psicología / annals of psychology
2026, vol. 42, nº 1 (january-april), e11, 1-11
https://doi.org/10.6018/analesps.654891

Published by Editum, Servicio de Publicaciones, Universidad de Murcia (Spain), in https://revistas.um.es/analesps
ISSN online: 1695-2294. License Creative Commos 4.0 BY

The meta-analytical random effects model with *g* tends to underestimate the parameters: an alternative model

Mario Calabria-Sen*, Juan Botella, and Juan I. Durán

Universidad Autónoma de Madrid, Spain

Título: El modelo meta-analítico de efectos aleatorios con *g* tiende a infra-estimar los parámetros: un modelo alternativo.
Resumen: El modelo clásico de efectos aleatorios (CREM) presenta limitaciones al utilizar como índice del tamaño del efecto la diferencia de medias tipificada. Suero et al. (2025) han reformulado el CREM como un Modelo de Mezclas (MM) y han desarrollado estimadores insesgados de los parámetros fundamentales μ_δ y τ^2. Compararon sus estimadores con dos procedimientos clásicos ampliamente utilizados, Máxima Verosimilitud Restringida (REML; Viechtbauer, 2005) y el estimador de DerSimonian y Laird (DL; 1986), encontrando pequeñas, pero sistemáticas, infraestimaciones por parte de los procedimientos clásicos. El objetivo de este trabajo es comprobar si los resultados encontrados por Suero et al. (2025) son extrapolables fuera del ámbito de simulación. Para ello se crearon tres bases de datos con meta-análisis (MA) reales pertenecientes al ámbito clínico, experimental y educativo de la psicología. Los resultados son concordantes con los encontrados por Suero et al. (2025) siendo las estimaciones medias del MM superiores a las de REML y DL. Se discuten también los casos atípicos encontrados en MA reales como tamaños del efecto bizarros, tamaños muestrales desmesurados o MA con pocos estudios primarios.
Palabras clave: Modelo de mezclas. Modelo de efectos aleatorios. Meta-análisis. Diferencia de medias tipificada.

Abstract: The Classic Random Effects Model (CREM) has some limitations when using the standardized mean difference (SMD) as effect size (ES) index. Suero et al. (2025) have reformulated CREM as a Mixture Model (MM) and have developed unbiased estimators of the main parameters μ_δ and τ^2. They compared their performance with two classic methods widely used: Restricted Maximum Likelihood (REML; Viechtbauer, 2005) and DerSimonian & Laird estimator (DL; 1986), finding small but systematic underestimations yielded by the classic procedures. The aim of this study is to check if Suero et al.'s (2025) results can be found beyond simulation contexts. For this purpose, we created three databases with real meta-analyses (MA) from clinical, experimental and educations fields of psychology. The results found are consistent with those found by Suero et al. (2025) being the mean estimates of MM higher than those of REML and DL. We also discuss about outliers found in real MAs such as bizarre ES, disproportionated sample sizes or MAs with small numbers of primary studies.
Keywords: Mixture Model. Random Effects Model. Meta-analysis. Standardized Mean Difference.

Introduction

Due to the nature of the constructs involved in psychology, the random effects model is typically considered the most realistic meta-analytical model. The aim with this model is to estimate the hyperparameters that define the distribution of parametric effects: the mean and variance, μ_δ and τ^2. To estimate μ_δ a weighted mean is calculated using the inverse of the variance of the effect sizes (ES) observed in the primary studies (Hedges, 1981). On the other hand, there are many procedures for estimating τ^2 (Blázquez-Rincon et al., 2023; Langan et al., 2019; Veroniki et al., 2016). Most of them are based on the proposals of Hedges (1983; Hedges & Olkin, 1985) and his linear decomposition of the variance of *g* that makes a distinction between a within-study and a between-study variances. However, this linear decomposition of variance in the classic random effects model (CREM) can lead to dependency problems for certain ES indices. When calculating the ES for a primary study, the CREM considers the parametric ES value for the primary study (θ) and the estimation error concerning that parametric value (e_i):

$$ES_i = \theta + e_i \tag{1}$$

*** Correspondence address [Dirección para correspondencia]:**
Mario Calabria-Sen. Universidad Autónoma de Madrid (Spain).
E-mail: mario.calabria@uam.es
(Article received: 17-03-2025; revised: 15-06-2025; accepted: 23-06-2025)

In a random effects model, the value of θ itself is a random variable with an expected value μ_θ and variance τ^2 that varies for each primary study. Therefore, the true value of θ_i, associated with the *i-th* primary study in the MA, can be decomposed as:

$$\theta_i = \mu_\theta + u_i \tag{2}$$

where μ_θ is the expected value of θ_i alongside the estimation error of the true value associated with that primary study. Combining and substituting (2) into (1):

$$ES_i = \mu_\theta + u_i + e_i \tag{3}$$

This determines that the observed ES values are the result of two sources of variation: one associated with the parametric ES values (u_i) and the other associated with the random sampling of the estimator (e_i). Thus, the variance of the ES is decomposed as:

$$\sigma^2_{ES} = \sigma^2_{sampling} + \sigma^2_{parameter} \tag{4}$$

where $\sigma^2_{sampling}$ refers to the variance associated with sampling error, and $\sigma^2_{parameter}$ refers to the specific variance (τ^2) associated with the parameter θ_i. It is in this equation (4) where dependency issues arise when those two variances are added as if they were independent, when they are actually dependent on some ES indices, such as Pearson's correlation or standardized mean difference (SMD). This happens because there is a dependency between the variance of the estimator and the parameter itself. In Pearson's correlation, this problem is fixed by using Fisher's Z transformation, as the formula utilized to estimate the variance is an approximation based only on the total sample size [1 / (N-3)]. However, for SMD no transformation solves these dependency problems. Following the proposal made by Hedges (1983), the variance of the ES estimate is defined as:

$$\sigma_i^2(\delta_i) = \frac{a_i}{\tilde{n}_i}\left[1 + \tilde{n}_i \delta_i^{\,2}\right] - \delta_i^2 \qquad (5)$$

where $\tilde{n}_i = \frac{n_i^E n_i^C}{n_i^E + n_i^C}$, $a_i = \frac{m_i[c(m_i)]^2}{(m_i - 2)}$ and $m_i = n_i^E + n_i^C - 2$. As can be noted in the expression, the estimator's variance depends directly on the value of the parameter, δ_i.

Many authors have pointed out these limitations, but their proposals incur into the same problems as mentioned above (Bakbergenuly et al., 2020; Böhning et al., 2002; Friedman, 2000; Hamman et al, 2018; Lin & Aloe, 2021). This problem has recently been discussed by Suero et al. (2025) who propose to use a Mixture Model (MM) as an alternative model to the classic meta-analytical random effects model for *g*. MMs are statistical models where the distribution of a continuous random variable is defined as a mixture of a finite or infinite number of subpopulations. Focusing on the SMD, each observed g_i value found in an MA would come from a subpopulation with a specific pair of values for the sample sizes ($N_{1,i}$ and $N_{2,i}$) and a parametric ES, δ_i. In turn, δ is a random variable with its own distribution, common across all studies, with mean μ_δ and variance τ^2. As can be seen, one of the parameters of the distribution of *g*, δ, has its own distribution with its own parameters (μ_δ, τ^2).

Under this approach, Suero et al. (2025) have developed estimators of μ_δ and τ^2 and have demonstrated they are unbiased. Additionally, they conducted a simulation study to compare the bias and efficiency of their estimators with two commonly used in CREM: DerSimonian & Laird estimator (DL; 1986) and Restricted Maximum Likelihood (REML; Viechtbauer, 2005). The results found in their simulation study show that the REML and DL main parameters estimations have a small, but systematical, bias. Indirect comparisons between the estimates of the three procedures show a particular ordering where the MM estimations are on average larger than those of REML and DL. Estimates of the μ_δ parameter were unbiased with the MM while there was a mean underestimation of 0.02 with REML and DL. For the τ^2 parameter, the estimates were again unbiased with the MM and the differences with the classical procedures increased as the parametric value of τ^2 increased, finding a range of underestimation from 0.02 to 0.07 with REML and from 0.01 to 0.19 with DL.

That ordering found in simulation studies could be a subproduct of the special characteristics of simulation studies where the conditions are much more controlled than in real-world MAs. Thus, the main objective of this study is to check whether the results found by Suero et al. (2025) in simulation studies, that is higher average estimations with the MM than with the CREM procedures, also appear in real MA databases. For this purpose, the estimates obtained with the three procedures (MM, DL and REML) will be compared using data from real MA published in three different fields of psychology: clinical, experimental, and educational. In addition, Suero et al. (2025) also suggested that the key to the underestimate the parameters relies specifically on the negative correlations found between weights and *g* values in the CREM procedures. We expect to find a pattern of coefficients congruent with this explanation in real MAs.

Method

Search strategy

The search strategy aimed to obtain multiple primary studies from different fields of psychology to systematically code three variables: ES, experimental group sample size, and control group sample size. We focused our search on clinical, experimental, and educational fields, as these are three representative areas of this discipline and are good potential fields to find MAs that use the SMD as ES index. The procedure to create each database was different according to the search field but all of them aimed for a common objective: to have a sample of real MAs with prototypical data from that field. For this purpose, we chose journals with high impact and relevance in their field and searched for MAs published in the last 10 years.

The *clinical field* database is composed of MAs studying the effectiveness of psychological interventions. This database comes from the second study by López-Nicolás (2023). The *experimental field* database focuses on MAs included in experimental psychology published in *Psychonomic Bulletin & Review*. Lastly, the *educational field* focuses on the educational psychology area. We searched for MAs published in *Review of Educational Research,* a journal specialized in review studies in the educational field.[1]

Inclusion and exclusion criteria

Common inclusion criteria across the three fields were: a) the ES index used is the SMD (Cohen's *d* or Hedges' corrected *g*) and is reported for every primary study. That is, we did not access to the primary studies; instead, we copied the values reported in the MA, typically in the forest plot. If Cohen's *d* is used, we corrected it into Hedge's *g* (1981); b) the MA specifies that the primary studies have two experimental conditions, an experimental or treated group and a control group, or specify a table with the same sizes of two different groups; c) the MAs specify the sample size of each group. If they do not distinguish the sample sizes of the experimental conditions the total sample was divided by two to impute

1 The databases used for this study are available upon request. Please contact the authors for access.

half of the sample to each condition; d) if the ES is not specified, the MAs provide enough data to calculate the SMD (sample sizes, means and standard deviations for each experimental condition of every primary study) e) the MAs have to be written in English or Spanish.

Common exclusion criteria across the three fields were: a) the ES index was other than the SMD; b) the MAs analyze repeated measures (pre-post) on a single sample; c) the MAs do not specify the estimated ES for each primary study; d) the MAs do not specify the sample sizes of the primary studies; e) if one primary study reports more than one ES one of them is chosen randomly to respect the assumption of independence.

It is essential to acknowledge that the absolute accuracy of the data collected from the studies is not critical for this research. The aim is to compare the estimates obtained by the three procedures (MM, REML, DL) with a range of variety in *g* values and sample sizes that roughly represent the values that typically appear in three different fields of psychology. In the simulation study of Suero et al. (2025) the generating parameters were set with rational criteria but may not be representative. In this study, data from real MAs are collected.

Data Analysis

Once the databases were created, estimates were obtained for each MA using three procedures: MM, DL and REML. We used R Statistic software with MixtureREM package (Durán et al., 2024, available for download at osf.io/rwt2q) to estimate the parameters with the MM and the Metafor package (Viechtbauer, 2010) for DL and REML. The combined ES of each MA was calculated before making any comparison. If it was negative, we inverted the sign of all *g* observed in the primary studies of that MA. This facilitates the interpretation of differences between estimates which is our focus of interest. The differences between estimates were formulated such that positives values indicate a higher value of the estimate with MM than with the other procedure (REML or DL). Because of the ordering found in the simulation studies of Suero et al. (2025) we expected average positive differences. We also compared the correlations between the weights assigned by each procedure and the *g* observed in each primary study. Suero et al. (2025) suggested that the use of weights negatively correlated with *g* could be the reason for the underestimation of the parameters with REML and DL. We therefore expect a difference between the correlations calculated with the three methods. In the simulations of Suero et al. (2025) the correlations with REML and DL showed mean values around -.15, while they were practically zero with the MM.

The simulation study by Suero et al. (2025) was conducted under a range of simulation conditions that covered different typical MA scenarios in psychology. However, in real MAs, atypical situations can be found that are not usually considered in a simulation study. One of these situations is the publication of MAs with a very limited number of primary studies. Jackson & Turner (2017) recommend having at least 5-10 primary studies to improve statistical power when using a random effects model, and Guolo & Varin (2017) caution against using traditional methods like DerSimonian & Laird when the number of studies is small. Taking this into consideration, and as a test of sensitivity, we did a reanalysis after removing from analysis all the MAs with fewer than 10 primary studies to check for differences.

There is enough evidence to suggest that there might be an underlying factor affecting the correlations between *g* and the weights. Marks-Anglin & Chen (2020) discuss the effect of small studies where studies with smaller sample sizes report different ES from the other studies leading to biases in the estimation of the main parameters. In their REML simulation studies Hamman et al. (2018) concluded that using the inverse of the variance as weight with ES larger than 1 combined with small sample sizes increase estimation bias and reduce the coverage to 85%. Also, correlations between *g* and weights, or in other words, between *g* and sample sizes, could be interpreted as an indicator of publication bias (Kühberger et al., 2014, Sterne et al., 2000). In the simulation studies conducted by Suero et al. (2025) they did not study the influence of publication bias. However, in the present databases, composed of real MAs, publication bias might be involved. To support this interpretation, we have applied the Begg & Mazumdar (1994) asymmetry test to all the MAs included in this study.

Results

Clinical field

The clinical field database was composed of 92 MA and 1372 primary studies. The number of primary studies ranged from 3 to 88 ($\bar{k} = 14.91;\ sd_k = 11.3$). The results (*Table 1*) showed higher average estimations in the parameters with MM than REML and DL. All average correlations between weights and *g* were negative but higher with the classic procedures than with MM.

Table 1
Results from clinical field: mean estimates μ_δ and τ^2 values, mean differences between the Mixture Model (MM) estimates and the other procedures and mean correlations between weights and g.

Procedure	$\bar{\hat{\mu}}_\delta$	$\bar{\hat{\mu}}_{\delta_{MM}} - \bar{\hat{\mu}}_{\delta_i}$	$\overline{\hat{\tau}^2}$	$\overline{\hat{\tau}^2}_{MM} - \overline{\hat{\tau}^2}_i$	$\bar{r}_{w;g}$
MM	0.490	--	0.2749	--	-.223
REML	0.468	0.022	0.2099	0.0650	-.316
DL	0.464	0.026	0.1467	0.1282	-.321

Note. MM = Mixture Model; DL = DerSimonian & Laird estimator; REML = Restricted Maximum Likelihood

The estimates of both parameters for each method and MA in the clinical field are included in *Table S1*. In 62 out of the 92 MAs analyzed (67.39%) the estimation of both parameters with the MM were higher than with the classic procedures. Out of those 92 MAs there were 13 MAs where the

μ_δ estimates for both REML and DL were higher than MM estimates. This occurred in 25 MA when estimating τ^2 with REML and 26 with DL. The largest differences in μ_δ estimations were found, both positive and negative, between the MM and REML with a difference of 0.288 (MA10) and a difference of -0.028 (MA58). For τ^2 estimates by MA the largest difference both positive and negative was found between the MM and DL with a difference of 2.177 (MA15) and -0.16 (MA11)

Experimental field

The experimental field database was composed of 32 MA and 1099 primary studies. The number of primary studies ranged from 2 to 138 ($\bar{k} = 34.34;\ sd_k = 38{,}48$). The results (*Table 2*) were similar to those found in the clinical field, finding slightly higher mean estimates in both parameters with the MM compared to REML and DL. Average correlations between weights and *g* were close to zero with the MM while they are negative with the classic procedures REML and DL.

Table 2
Results from experimental field: mean estimates μ_δ and τ^2 values, mean differences between the MM estimates and the other procedures and mean correlations between weights and g.

Procedure	$\bar{\hat{\mu}}_\delta$	$\bar{\hat{\mu}}_{\delta_{MM}} - \bar{\hat{\mu}}_{\delta_i}$	$\overline{\hat{\tau}^2}$	$\overline{\hat{\tau}^2}_{MM} - \overline{\hat{\tau}^2}_i$	$\bar{r}_{w;g}$
MM	0.693	--	0.4517	--	.002
REML	0.657	0.035	0.3432	0.1085	-.210
DL	0.653	0.040	0.2559	0.1958	-.210

Note. MM = Mixture Model; DL = DerSimonian & Laird estimator; REML = Restricted Maximum Likelihood

The estimates of both parameters for each method and MA in the experimental field are included in *Table S2*. In 28 out of the 32 MAs analyzed (87.50%) the estimation of both parameters with the MM were higher than with the classic procedures. There was only one MA (MA22) where we found μ_δ estimates of REML and DL higher than MM estimates. This happened in 3 and 4 MAs when estimating τ^2 with REML and DL, respectively. The higher differences in μ_δ estimation between the MM and classic procedures were found, positively, in MA20 between MM and DL, showing a difference of 0.254. The largest negative difference was found in MA22 where the difference between MM and REML is -0.026. For τ^2 estimates the largest positive difference was found between MM and DL with a difference of 1.837 (MA20) and the largest negative difference was found between MM and REML with a difference of -0.0369 (MA22).

Educational field

The educational field database was composed of 54 MA and 2418 primary studies. The number of primary studies ranged from 2 to 165 ($\bar{k} = 44.78;\ sd_k = 36.23$). The results (*Table 3*) were similar to those found in the other two fields, finding higher average estimates with the MM for the main parameters μ_δ and τ^2. Average correlations between *g* and the weights were lower with the MM than those obtained with the classic REML and DL procedures.

Table 3
Results from educational field: mean estimates μ_δ and τ^2 values, mean differences between the MM estimates and the other procedures and mean correlations between weights and g.

Procedure	$\bar{\hat{\mu}}_\delta$	$\bar{\hat{\mu}}_{\delta_{MM}} - \bar{\hat{\mu}}_{\delta_i}$	$\overline{\hat{\tau}^2}$	$\overline{\hat{\tau}^2}_{MM} - \overline{\hat{\tau}^2}_i$	$\bar{r}_{w;g}$
MM	0.563	--	0.4553	--	-.188
REML	0.537	0.025	0.3692	0.0860	-.302
DL	0.523	0.039	0.2315	0.2238	-.300

Note. MM = Mixture Model; DL = DerSimonian & Laird estimator; REML = Restricted Maximum Likelihood

The estimates of both parameters for each method and MA in the educational field are included in *Table S3*. In 39 out of the 54 MAs analyzed (72.22%) the estimation of both parameters with the MM were higher than with the classic procedures. Out of those 54 MAs analyzed there were six of them were the μ_δ estimates of REML and DL are higher than the estimates of MM. This happens in 12 and 10 cases when estimating τ^2 with REML and DL, respectively. In this context, in one case (MA22) the estimation of μ_δ using the classic procedures was considerably higher than the one estimated with the MM, finding a difference of -0.189 in favor of the REML procedure. The largest positive difference was found in MA43 between MM and DL, with a difference of 0.344. For τ^2 estimates the largest positive and negative difference was found between the MM and DL with a difference of 1.972 (MA20) and -0.16 (MA47), respectively.

Sensitivity analysis

The sensitivity analyses after removing the MAs with less than 10 primary studies are included in *Table 4*. The MAs excluded from the analyses were 25 in the clinical field, 10 in the experimental field and 10 in the educational field, resulting in 67, 21 and 43 MAs analyzed, respectively. The average estimates of the three procedures for μ_δ are slightly lower in the clinical and experimental fields but slightly higher in the educational field. In τ^2 parameter estimations the results differ depending on the procedure. The average estimates for MM decreased almost by half in the clinical field ranging from 0.2749 to 0.1695, remained stable in the experimental field and doubled in the educational field. As for the REML procedure, the average estimates in the experimental and educational fields remain stable while they decrease in the clinical field from 0.2099 to 0.1306. Under the DL procedure the average estimations decreased in the clinical field but increased in the experimental and educational fields. The largest average estimation difference found under this procedure was in the experimental field, increasing from 0.2599 to 0.3989. Finally, the mean correlations found between *g* and the weights assigned by each procedure did not result in many changes (the largest difference found was .05).

Table 4
Overall results aggregated by procedure including only meta-analysis (MAs) with 10 or more primary studies.

Procedure	Field	$\bar{\hat{\mu}}_{\delta}$	$\bar{\hat{\mu}}_{\delta_{MM}} - \bar{\hat{\mu}}_{\delta_i}$	$\overline{\hat{\tau}^2}$	$\overline{\hat{\tau}^2}_{MM} - \overline{\hat{\tau}^2}_i$	$\bar{r}_{w_i g}$
MM	Clinical	0.439 (0.490)	--	0.1695 (0.2749)	--	-.212 (-.223)
	Experimental	0.678 (0.693)	--	0.4441 (0.4517)	--	-.065 (.002)
	Educational	0.581 (0.563)	--	0.4709 (0.2315)	--	-.172 (-.188)
REML	Clinical	0.421 (0.468)	0.018	0.1306 (0.2099)	0.0389	-.287 (-.316)
	Experimental	0.649 (0.657)	0.029	0.3532 (0.3432)	0.0909	-.263 (-.210)
	Educational	0.558 (0.537)	0.023	0.3989 (0.3692)	0.0719	-.303 (-.302)
DL	Clinical	0.420 (0.464)	0.019	0.1095 (0.1467)	0.0601	-.291 (-.321)
	Experimental	0.643 (0.653)	0.034	0.3989 (0.2599)	0.0719	-.260 (-.210)
	Educational	0.543 (0.523)	0.038	0.2407 (0.2315)	0.2301	-.300 (-.300)

Note. MM = Mixture Model; DL = DerSimonian & Laird estimator; REML = Restricted Maximum Likelihood. In parentheses the estimations using all the MAs codified.

The results of the asymmetry test to determine the presence of cues of publication bias are shown in *Table 5*. The percentage of significant tests obtained using the MM method were 11.96%, 9.38%, and 22.22% in the clinical, experimental, and educational fields, respectively. In contrast, the results from the traditional REML and DL procedures resulted in 21.74% in the clinical field, 21.88% in the experimental field, and 37.03% in the educational field.

Table 5
Frequency and percentage of studies that had significant tests in Begg & Mazumdar (1994) asymmetry test.

	Clinical	Experimental	Educational
MM	11/92 (11.96%)	3/32 (9.38%)	12/54 (22.22%)
REML	20/92 (21.74%)	7/32 (21.88%)	20/54 (37.03%)
DL	20/92 (21.74%)	7/32 (21.88%)	20/54 (37.03%)

Note. MM = Mixture Model; DL = DerSimonian & Laird estimator; REML = Restricted Maximum Likelihood

Discussion

The aim of this paper is to check if the results obtained with real MA databases are congruent with those obtained by Suero et al. (2025) in their simulation studies. These authors compared their model estimation procedure (MM) with two classic procedures widely used in MA (REML and DL), finding a clear tendency of the classic procedures to obtain slight underestimates of the two main parameters μ_δ and τ^2. To check whether their results can be extended beyond the simulation context we looked for real MAs in three different areas of psychology and compared the three procedure estimations. The results of this study are consistent with Suero et al. (2025) results, showing similar procedure rankings: slightly higher average MM estimations compared to REML and DL in both parameters. In short, the bulk of published MAs in psychology using SMD may be underestimating both the mean and the variance of true effects.

In relation to the parameter μ_δ these average differences range between 0.02 and 0.04 when comparing the MM with the other two procedures. In their simulation study the authors found similar average differences (ranging from 0.01 to 0.04). These results are therefore consistent with those found in a real MA context. We have further explored the explanation given by Suero et al. (2025) for this discrepancy between the procedures. These authors pointed out that with the two classic procedures there is a trend of small but persistent negative correlations between *g* and the weights. In contrast, the weights within the MM are independent of *g*. We have provided evidence that this pattern changes significantly for the three databases of real MAs.

Correlations found in real MAs are also congruent with the findings of Suero et al. (2025). The classic REML and DL procedures show negative correlations between *g* and the weights. However, these authors did not find correlations with the MM, and we found them in clinical and educational contexts. Following the ideas of previous studies we explored the possible influence of the presence of publication bias, and we applied the Begg & Mazumdar (1994) asymmetry test. The percentage of significant tests with the variances obtained with the MM is 12%, 9.4% and 22% in the clinical, experimental, and educational fields, respectively. As in all cases the expected 5% is exceeded in the absence of publication bias, we can conclude that in general this phenomenon affects all three fields. However, when evaluating the results yielded by classic procedures REML and DL the percentages of significance rise to 21.7%, 21.9% and 37.0% (the percentages are identical with the two procedures, REML and DL, because it is an ordinal coefficient and with these two procedures the weights are necessarily ordered in the same way). As expected, the percentages of significance with REML and DL are higher than those with MM. These correlations found in the classic REML and DL procedures would be the combined effect of two factors: the usage of formulas that involve dependance between *g* and the variance, and publication bias. The negative correlations found in the MM procedure could be exclusively due to publication bias. This is why the combination of these two sources produces a stronger negative correlation in classic procedures than with the MM.

Differences found in τ^2 parameter are also congruent with the results found by Suero et al. (2025) where slightly higher estimates were found with the MM. However, these differences are higher than those found by the authors in the simulation studies. In this study, the differences found range between 0.0650 and 0.2238, whereas in the simulation studies they ranged between 0.01 and 0.19. This may be due to the major differences between the MM and the classic procedures in specific MAs.

The fact that in all three fields there are cases where the MM estimations of the main parameters are lower than REML and DL estimations indicates that there are no structural aspects in the formulae whereby the MM estimates necessarily lead to larger magnitudes. This is simply a tendency that occurs most of the time, but not always. In their simulation studies the authors also found a small percentage of MAs where the MM estimates were lower than those obtained by the classical procedures, but on average the MM estimates were higher. This may also be due to the variability of the estimates. Any estimation method has a variability associated with the error in the parameter estimation, which can be randomly higher or lower. The problem with classical procedures is that REML and DL estimators are biased, resulting in systematic underestimations.

Practical implications

We must acknowledge that the practical implications of our work are small. Together with the work of Suero et al. (2025) in this article we show that over the long run, the CREM tends to show underestimations of the two model parameters. However, the magnitude of these underestimations is rather small. The mean difference is between 0.02 and 0.04 in the mean effect, which is a relatively small error for typical values of this parameter (according to Cohen's criteria 0.20 is considered a small effect and 0.80 a large effect). Few main conclusions and few important decisions would be affected by a correction of this magnitude in μ_δ. In the case of the variance of the effects, the mean difference in τ^2 is around 0.03, which is somewhat larger in relative terms, given that typical values of this parameter do not usually exceed 0.20. The error in estimating this parameter could have somewhat greater consequences when obtaining the prediction interval and making inferences based on it (Botella & Sánchez-Meca, 2024) but depends on several factors.

Clearly, the magnitude of the mean errors does not warrant a review of the estimates generally obtained in MAs published in psychology. However, looking to the future, we suggest paying attention to it. The problem addressed is well known and has been recognized for decades by multiple authors. When the first estimators of specific variance were developed in the 1980s, the limitation of analytical and computational resources made it more natural to accept approximations that allowed us to overcome obstacles that seemed insurmountable in exchange for small errors. But today, we believe we should introduce more precise models and resources if we have them at our disposal. A gain in precision should always be welcome, no matter how small. Therefore, we recommend using the MM to make more accurate and unbiased estimates of the fundamental parameters, μ_δ and τ^2, when using SMD as ES index. This is one of the limitations of the MM, as it has so far only been proposed and developed for the SMD. It should also be noted that the distribution of the τ^2 estimator and its variance are lacking, making it impossible by now to obtain confidence intervals based on the actual shape of its distribution.

Trend reversals

Although the general trend of the estimates is congruent to the simulation results (slightly higher estimates of both parameters with the MM than with REML and DL), in real MAs we found results that could be considered atypical or discordant with this trend. While the simulation study covered the 'typical' conditions of real MAs in a wide range of circumstances, unexpected combinations sometimes occur that may lead to an inversion of the trend resulting in lower estimates with the MM. We will now mention some theoretical concepts that can provide a brief theoretical explanation of those trend reversals. We want to comment a set of scenarios or circumstances that occur in the MAs included in this study with 'atypical' results in the above-mentioned sense. However, we must point out that also in the simulations the authors obtained a few cases with inverted estimates to the general trends. There is nothing in the structure of the estimation formulae of the three procedures that makes them necessarily order in a particular way. It is therefore not surprising that there are MAs where the order of the estimates is reversed.

In the simulation studies of Suero et al. (2025) the ES values of *g* were considered as small, medium, or large following Cohen values 0.20, 0.50 and 0.80 as reference (Cohen, 1988). Although Cohen himself advised not to use these references blindly, in no case seems reasonable to take as valid *g* values as those reported in some MAs. For example, there are some *g* values larger than 3 in MA number 20 in educational field, reaching 6.96 the highest of them. It does not seem reasonable to use the meta-analytical statistical model in these circumstances as the assumptions of Cohen's *d* definition (two homoscedastic normal distributions) are probably not met or there are transcription or calculation errors. An anomalous distribution of *g* is not a good starting point to carry on an MA.

On the other hand, an MA combines a sample of independent studies that are representative of a field of research. But if a study has a disproportionately large sample size the weighting scheme will make the results of that study almost definitely determine the main result of the MA. For example, in the MA21 in the educational field there is one study with almost 200,000 participants, being the total of the MA about 258,000 participants; i.e., almost 77% of the participants in that MA are in that study (two other studies integrate another 48,000 of the remaining studies). The contribution of the other studies to the average estimation is minimal, preventing the sampling of conditions and participants from a variety of research context from having its effect.

These two characteristics (extremely large ES values and primary studies with very high and disproportionate sample sizes, plus MAs conducted with few primary studies) lead to higher differences to those found in simulation studies. Simulation studies are conducted in a controlled artificial context that also try to include atypical situations but do not capture situations as extreme as those that can be found in real MAs (such as ES of 6, studies with 200,000 participants or MAs conducted with 3 primary studies). In any case, our main conclusion remains: the results found in these three fields are in the same direction as those found by Suero et al. (2025), that is, the average estimates of the MM are slightly higher than the estimates of the classic REML and DL procedures.

Complementary information

Conflict of interest.- The authors declare no conflict of interest.

Financial support.- This work was supported by the Ministerio de Ciencia e Innovación of Spain [project reference: PID2021-122404NB-I00]

References

Bakbergenuly, I., Hoaglin, D. C., & Kulinskaya, E. (2020). Estimation in meta-analyses of mean difference and standardized mean difference. *Statistics in Medicine*, *39*, 171–191. https://doi.org/10.1002/sim.8422

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101.

Blázquez-Rincón, D., Sánchez-Meca, J., Botella, J., & Suero, M. (2023). Heterogeneity estimation in meta-analysis of standardized mean differences when the distribution of random effects departs from normal: A Monte Carlo simulation study. *BMC medical research methodology*, *23*(1), 19. https://doi.org/10.1186/s12874-022-01809-0

Böhning, D., Malzahn, U., Dietz, E., Schlattmann, P., Viwatwongkasem, C., & Biggeri, A. (2002). Some general points in estimating heterogeneity variance with the DerSimonian–Laird estimator. *Biostatistics*, *3*(4), 445-457.

Botella, J., & Sánchez-Meca, J. (2024). Meta-analysis: Confidence intervals and Prediction intervals. *Anales de Psicología*, *40*(2), 344-354.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

DerSimonian, R., &Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*, 177-188. https://doi.org/10.1016/0197-2456(86)90046-2

Durán, J. I., Suero, M., & Botella, J. (2024, November 4). *MixtureREM*. Retrieved from osf.io/rwt2q

Friedman, L. (2000). Estimators of random effects variance components in meta-analysis. *Journal of Educational and Behavioral Statistics, 25*(1), 1-12.

Guolo, A., & Varin, C. (2017). Random-effects meta-analysis: the number of studies matters. *Statistical methods in medical research*, *26*(3), 1500–1518. https://doi.org/10.1177/0962280215583568

Hamman, E. A., Pappalardo, P., Bence, J. R., Peacor, S. D., & Osenberg, C. W. (2018). Bias in meta-analyses using Hedges'd. *Ecosphere*, *9*(9), e02419.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128. https://doi.org/10.2307/1164588

Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin, 93*, 388-395

Hedges, L. V. & Olkin, I. (1985). *Statistical Methods for Metaanalysis*. San Diego, CA: Academic Press.

Jackson, D., & Turner, R. (2017). Power analysis for random-effects meta-analysis. *Research synthesis methods*, *8*(3), 290-302. https://doi.org/10.1002/jrsm.1240

Kühberger, A., Fritz, A. & Scherndl, T. (2014). Publication Bias in Psychology: A Diagnosis Based on the Correlation between Effect Size and Sample Size. *PLoS ONE, 9*(9): e105825. https://doi.org/10.1371/journal.pone.0105825

Langan, D., Higgins, J. P. T., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., Viechtbauer, W., & Simmonds, M. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods*, *10*(1), 83–98. https://doi.org/10.1002/jrsm.1316

Lin, L., & Aloe, A. M. (2021). Evaluation of various estimators for standardized mean difference in meta-analysis. *Statistics in Medicine*, *40*, 403–426. https://doi.org/10.1002/sim.8781

López-Nicolás, R. (2023). *Research Synthesis: Challenges, Reproducibility and Future Directions*. [Unpublished doctoral dissertation] Universidad de Murcia.

Marks-Anglin, A., & Chen, Y. (2020). Small-study effects: current practice and challenges for future research. *Statistics and Its Interface, 13*(4), 457-484.

Sterne, J. A., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *Journal of clinical epidemiology, 53*(11), 1119-1129.

Suero, M., Botella, J., Durán, J. I., & Blázquez-Rincón, D. (2025). Reformulating the Meta-analytical Random Effects Model as a Mixture Model. *Behavior Research Methods, 57*, 7. https://doi.org/10.3758/s13428-024-02554-6.

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P., Langan, D., & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, 7(1), 55-79. https://doi.org/10.1002/jrsm.1164

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261-293. https://doi.org/10.3102/10769986030003261.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1-48. doi:10.18637/jss.v036.i03.

Supplementary materials

Table S1 includes the 92 MAs analyzed in the *clinical field* ordered by the number of primary studies (k). The average number of primary studies in this field is 14.91, with an average of 92.37 participants per study. The first 25 MAs are composed of less than 10 primary studies. Excluding these MAs the average number of primary studies per MA increases to 17.88 and the average number of participants to 96.38.

Table S1
Estimations of μ_δ and τ^2 and correlations (Cor_) g – weights by procedure for each meta-analysis (MA) in the clinical field.

MA	k	Total Sample	Average Sample	μ REML	τ^2 REML	μ DL	τ^2 DL	μ MM	τ^2 MM	Cor_MM	Cor_REML	Cor_DL
1	3	239	79.67	0.153	0.0000	0.153	0.0000	0.153	0.0000	.942	.942	.942
2	4	320	80.00	0.901	0.0711	0.901	0.0716	0.912	0.0376	.028	-.144	-.144
3	5	206	41.20	0.162	0.1444	0.163	0.1499	0.154	0.1111	-.825	-.836	-.835
4	5	222	44.40	1.542	1.7981	1.480	0.8494	1.618	2.1838	-.114	-.871	-.854
5	6	484	80.67	0.424	0.0000	0.424	0.0000	0.425	0.0000	-.540	-.564	-.564
6	6	327	54.50	0.995	1.9310	0.973	1.1918	1.022	2.0216	-.866	-.980	-.980
7	6	779	129.83	0.186	0.0517	0.184	0.0576	0.184	0.0574	-.043	-.032	-.041
8	6	655	109.17	0.421	0.0000	0.421	0.0000	0.421	0.0035	.455	.433	.433
9	7	1037	148.14	0.871	1.2797	0.858	0.7530	0.884	1.3413	-.592	-.881	-.880
10	7	433	61.86	0.503	0.0000	0.596	0.1095	0.791	0.7535	-.699	-.458	-.733
11	7	519	74.14	0.479	0.2777	0.477	0.3880	0.489	0.2283	.045	-.060	-.099
12	7	289	41.29	0.730	0.5984	0.685	0.3084	0.810	1.0330	-.614	-.924	-.898
13	7	289	41.29	0.747	0.3741	0.729	0.2626	0.804	0.6047	-.622	-.857	-.839
14	7	289	41.29	1.038	0.6006	0.984	0.3066	1.137	1.1129	-.687	-.958	-.937
15	7	289	41.29	0.688	2.1331	0.563	0.6725	0.774	2.8495	-.658	-.961	-.951
16	8	683	85.38	0.113	0.0000	0.113	0.0000	0.113	0.0000	.205	.203	.203
17	8	259	32.38	0.341	0.0205	0.347	0.0416	0.376	0.0933	-.228	-.321	-.330
18	8	1639	204.88	0.528	0.0179	0.527	0.0163	0.503	0.0014	-.188	-.130	-.133
19	8	1804	225.50	0.220	0.0000	0.220	0.0000	0.221	0.0000	-.080	-.086	-.086
20	8	709	88.63	0.850	0.0220	0.851	0.0256	0.863	0.0350	-.115	-.231	-.235
21	8	551	68.88	0.607	0.2310	0.604	0.2043	0.625	0.2754	-0.239	-.457	-.451
22	9	893	99.22	0.664	0.3386	0.648	0.2037	0.684	0.4017	-0.334	-.675	-.648
23	9	498	55.33	1.338	0.4729	1.328	0.3833	1.378	0.5502	-0.318	-.702	-.700
24	9	602	66.89	0.036	0.1963	0.039	0.1655	0.032	0.2145	.166	.206	.218
25	9	405	45.00	0.298	0.0000	0.298	0.0000	0.319	0.0218	-.464	-.473	-.473
26	10	1306	130.60	0.251	0.0178	0.252	0.0191	0.255	0.0238	-.406	-.438	-.440
27	10	1741	174.10	0.371	0.0131	0.371	0.0113	0.373	0.0033	-.108	-.157	-.153
28	10	844	84.40	0.295	0.1017	0.293	0.1178	0.285	0.1890	.504	.442	.460
29	10	1177	117.70	0.272	0.0093	0.276	0.0175	0.298	0.0522	-.514	-.384	-.438
30	10	643	64.30	0.280	0.0455	0.280	0.0467	0.290	0.0507	-.321	-.374	-.373
31	10	613	61.30	0.233	0.0004	0.239	0.0360	0.274	0.1412	-.357	-.255	-.362
32	10	1079	107.90	0.122	0.0993	0.121	0.0718	0.123	0.1406	-.109	-.089	-.080
33	10	1019	101.90	0.164	0.0354	0.168	0.0420	0.180	0.0707	-.297	-.300	-.302
34	10	1706	170.60	0.203	0.0185	0.225	0.0506	0.363	0.5293	-.895	-.675	-.794
35	10	1375	137.50	0.173	0.0469	0.173	0.0454	0.174	0.0631	-.048	-.060	-.061
36	10	1192	119.20	0.576	0.2424	0.559	0.1593	0.593	0.2643	-.409	-.565	-.563
37	10	1192	119.20	0.154	0.2275	0.157	0.1527	0.151	0.2518	.084	.106	.110
38	11	1459	132.64	0.308	0.2159	0.308	0.2358	0.312	0.2068	.177	.084	.085
39	11	661	60.09	0.256	0.1122	0.256	0.1150	0.251	0.0707	-.129	-.122	-.122
40	11	784	71.27	0.514	0.1510	0.510	0.1405	0.548	0.2139	-.837	-.891	-.890
41	11	697	63.36	0.384	0.1074	0.384	0.1034	0.420	0.2526	-.235	-.365	-.362
42	11	751	68.27	0.635	0.0276	0.644	0.0463	0.682	0.1227	-.766	-.785	-.802
43	12	1634	136.17	0.419	1.5597	0.416	0.8351	0.430	1.6039	.384	-.232	-.234
44	12	770	64.17	0.198	0.0705	0.199	0.0785	0.205	0.0655	-.064	-.128	-.124
45	12	599	49.92	0.509	0.0817	0.509	0.0842	0.521	0.0789	.059	-.060	-.061
46	12	1499	124.92	0.391	0.0700	0.391	0.0660	0.397	0.0821	-.002	-.065	-.061
47	12	1138	94.83	0.425	0.1360	0.424	0.1217	0.427	0.1169	-.055	-.095	-.094
48	12	564	47.00	0.707	0.1749	0.708	0.1803	0.770	0.3676	-.290	-.551	-.552
49	12	2254	187.83	0.345	0.0689	0.345	0.0463	0.349	0.0780	.129	.069	.058
50	12	887	73.92	0.044	0.0453	0.044	0.0447	0.047	0.0484	-.660	-.669	-.669
51	12	1785	148.75	0.253	0.1701	0.221	0.0819	0.317	0.4521	-.709	-.744	-.684

MA	k	Total Sample	Average Sample	μ REML	τ^2 REML	μ DL	τ^2 DL	μ MM	τ^2 MM	Cor_MM	Cor_REML	Cor_DL
52	13	2581	198.54	0.252	0.0800	0.250	0.0633	0.256	0.0871	-.074	-.135	-.153
53	13	1006	77.38	0.643	0.0884	0.643	0.0872	0.650	0.0699	-.132	-.231	-.231
54	13	4445	341.92	-0.002	0.0048	0.002	0.0039	0.016	0.0016	.348	.372	.367
55	13	636	48.92	0.252	0.0542	0.252	0.0600	0.258	0.0842	.001	-.053	-.057
56	13	1941	149.31	0.301	0.2570	0.257	0.1136	0.361	0.5300	-.678	-.791	-.752
57	14	950	67.86	0.126	0.0479	0.126	0.0446	0.099	0.0000	-.148	-.027	-.031
58	14	827	59.07	0.445	0.2477	0.436	0.1959	0.481	0.3665	-.579	-.674	-.665
59	15	1945	129.67	0.481	0.1470	0.480	0.1097	0.491	0.1682	.156	-.014	-.026
60	15	1329	88.60	0.358	0.0156	0.358	0.0147	0.361	0.0111	-.490	-.532	-.530
61	15	1329	88.60	0.326	0.0468	0.326	0.0440	0.329	0.0353	-.376	-.422	-.421
62	15	1515	101.00	0.482	0.1482	0.478	0.1231	0.495	0.1751	-.380	-.551	-.549
63	16	1046	65.38	0.517	0.0600	0.517	0.0601	0.524	0.0807	.352	.268	.268
64	16	1046	65.38	0.762	0.0982	0.763	0.1119	0.771	0.0572	.001	-.086	-.079
65	16	943	58.94	0.061	0.0387	0.061	0.0315	0.059	0.0073	.044	.089	.078
66	16	1486	92.88	0.632	0.1763	0.629	0.2297	0.643	0.1318	.334	.239	.239
67	16	1685	105.31	0.904	0.2520	0.905	0.2612	0.922	0.2709	-.498	-.619	-.620
68	17	1093	64.29	0.795	0.1269	0.796	0.1348	0.811	0.0968	.106	-.029	-.027
69	17	1223	71.94	0.268	0.0000	0.267	0.0140	0.286	0.0934	-.189	-.140	-.172
70	18	714	39.67	0.695	0.0775	0.702	0.0948	0.758	0.1911	-.512	-.631	-.643
71	18	706	39.22	0.691	0.1512	0.690	0.1461	0.729	0.2078	-.419	-.590	-.589
72	19	1600	84.21	0.292	0.0472	0.291	0.0446	0.288	0.0334	-.409	-.425	-.427
73	20	2188	109.40	0.335	0.0410	0.335	0.0395	0.340	0.0436	.024	-.026	-.026
74	20	825	41.25	0.399	0.0000	0.399	0.0000	0.406	0.0000	-.201	-.248	-.248
75	20	825	41.25	0.259	0.0000	0.259	0.0000	0.264	0.0000	-.369	-.396	-.396
76	20	918	45.90	0.764	0.1042	0.767	0.1115	0.806	0.1398	-.406	-.545	-.548
77	21	4798	228.48	0.368	0.0862	0.367	0.0765	0.371	0.1045	-.147	-.176	-.182
78	21	2679	127.57	0.631	0.4232	0.630	0.3297	0.640	0.4281	.033	-.144	-.144
79	21	1329	63.29	0.509	0.0821	0.509	0.0824	0.524	0.1135	-.218	-.312	-.312
80	21	1307	62.24	0.389	0.0909	0.389	0.0889	0.400	0.1059	.067	-.034	-.035
81	21	821	39.10	0.275	0.0425	0.274	0.0401	0.278	0.0272	-.290	-.340	-.340
82	23	1156	50.26	0.575	0.0712	0.580	0.0792	0.587	0.0626	-.631	-.703	-.705
83	24	1300	54.17	0.702	0.2494	0.696	0.2201	0.729	0.2817	-.492	-.653	-.654
84	24	909	37.88	0.657	0.2081	0.653	0.1829	0.712	0.3457	-.046	-.379	.377
85	25	3614	144.56	0.488	0.1304	0.487	0.1290	0.486	0.1066	-.487	-.527	-.528
86	27	1360	50.37	0.975	0.2244	0.972	0.2120	1.007	0.2292	-.434	-.600	-.601
87	30	1981	66.03	0.773	0.3861	0.753	0.2646	0.824	0.5664	-.314	-.658	-.650
88	33	4422	134.00	0.373	0.0621	0.372	0.0591	0.379	0.0638	-.560	-.591	-.588
89	36	1631	45.31	0.859	0.0494	0.858	0.0454	0.880	0.0415	-.247	-.365	-.362
90	45	4735	105.22	0.007	0.0744	0.008	0.0641	0.003	0.0859	.040	.067	.072
91	53	5384	101.58	0.629	0.3142	0.626	0.2324	0.644	0.3321	.070	-.121	-.130
92	88	14069	159.88	0.504	0.0944	0.509	0.1379	0.503	0.0398	-.146	-.216	-.224

Note. MM = Mixture Model; DL = DerSimonian & Laird estimator; REML = Restricted Maximum Likelihood

Table S2 includes the 32 MAs analyzed in the *experimental field* ordered by the number of primary studies (k). The average number of primary studies in this field is 34.34, with an average of 49.68 participants per study. The first 10 MAs are composed of less than 10 primary studies. Excluding these MAs the average number of primary studies per MA increases to 47.09 and the average number of participants to 57.97.

Table S2
Estimations of μ_δ and τ^2 and correlations (Cor_) g – weights by procedure for each meta-analysis (MA) in the experimental field.

MA	k	Total Sample	Average Sample	μ REML	τ^2 REML	μ DL	τ^2 DL	μ MM	τ^2 MM	Cor_MM	Cor_REML	Cor_DL
1	2	68	34.00	0.054	0.0000	0.054	0.0000	0.055	0.0000	1.000	1.000	1.000
2	3	49	16.33	0.943	1.2905	0.941	1.2168	0.993	1.3421	.347	-.613	-.612
3	6	236	39.33	0.643	0.0000	0.643	0.0000	0.651	0.0000	.186	-.083	-.083
4	6	194	32.33	0.179	0.0000	0.179	0.0000	0.182	0.0000	.026	-.044	-.044
5	6	125	20.83	0.966	0.0000	0.966	0.0000	0.977	0.0000	.672	.579	.579
6	6	153	25.50	2.149	1.9213	2.120	1.5045	2.302	2.5810	-.165	-.903	-.903
7	8	445	55.63	0.374	0.0000	0.374	0.0003	0.384	0.0000	.102	-.013	-.012
8	8	107	13.38	0.477	0.0000	0.477	0.0000	0.717	0.7601	-.633	-.694	-.694
9	9	559	62.11	0.466	0.0000	0.466	0.0000	0.474	0.0000	.292	.192	.192
10	9	135	15.00	0.515	0.0000	0.515	0.0000	0.521	0.0000	-.337	-.438	-.438
11	12	591	49.25	0.504	0.0940	0.503	0.0954	0.506	0.1310	.347	.277	.277
12	12	532	44.33	0.754	0.0000	0.754	0.0000	0.761	0.0000	.136	.016	.016
13	14	305	21.79	0.259	0.0000	0.259	0.0000	0.266	0.0000	.158	.105	.105
14	15	858	57.20	0.350	0.0279	0.351	0.0258	0.360	0.0199	.136	.048	.049
15	16	961	60.06	0.620	0.0000	0.620	0.0000	0.628	0.0000	.164	.079	.079
16	18	1467	81.50	0.998	0.2242	0.994	0.1888	1.020	0.2586	-.218	-.563	-.561
17	19	1522	80.11	0.591	0.2269	0.589	0.1926	0.599	0.2510	-.245	-.377	-.375
18	19	925	48.68	0.592	0.0000	0.592	0.0000	0.606	0.0000	-.095	-.247	-.247
19	20	1061	53.05	0.048	0.4762	0.034	0.2859	0.082	0.7496	-.356	-.499	-.458
20	22	676	30.73	2.615	4.1480	2.558	2.5190	2.812	4.3564	.214	-.546	-.559
21	23	1123	48.83	1.446	0.5936	1.435	0.4451	1.497	0.7114	.014	-.557	-.551
22	31	1210	39.03	0.187	0.0369	0.180	0.0255	0.161	0.0000	-.575	-.622	-.618
23	41	3217	78.46	0.468	0.2656	0.466	0.2445	0.500	0.3951	-.245	-.496	-.488
24	42	3279	78.07	0.973	0.2233	0.973	0.2089	0.998	0.2485	-.001	-.293	-.290
25	54	2520	46.67	0.308	0.2220	0.308	0.1961	0.324	0.2485	.105	-.050	-.053
26	62	3148	50.77	0.514	0.1788	0.512	0.1635	0.535	0.2248	-.211	-.373	-.371
27	62	4397	70.92	0.180	0.0000	0.180	0.0000	0.181	0.0000	-.078	-.084	-.084
28	64	4474	69.91	0.895	0.3099	0.876	0.2041	0.971	0.6649	-.260	-.698	-.658
29	103	6667	64.73	1.046	0.0519	1.046	0.0525	1.057	0.0392	-.006	-.079	-.078
30	121	8239	68.09	0.632	0.4775	0.632	0.4119	0.638	0.6232	.208	.116	.109
31	128	8488	66.31	0.169	0.1577	0.165	0.1390	0.238	0.5518	-.327	-.575	-.560
32	138	9225	66.85	0.116	0.0565	0.119	0.0694	0.168	0.2971	-.301	-.376	-.397

Note. MM = Mixture Model; DL = DerSimonian & Laird estimator; REML = Restricted Maximum Likelihood

Table S3 includes the 54 MAs analyzed in the *educational field* ordered by the number of primary studies (k). The average number of primary studies in this field is 44.78, with an average of 508.92 participants per study. The first 10 MAs are composed of less than 10 primary studies. Excluding these MAs the average number of primary studies per MA increases to 53.64 and the average number of participants to 589.88.

Table S3

Estimations of μ_δ and τ^2 and correlations (Cor_) g – weights by procedure for each meta-analysis (MA) in the educational field.

MA	k	Total Sample	Average Sample	μ REML	τ2 REML	μ DL	τ2 DL	μ MM	τ2 MM	Cor_MM	Cor_REML	Cor_DL
1	2	53	26.50	0.239	0.0000	0.239	0.0000	0.242	0.0000	-1.000	-1.000	-1.000
2	3	192	64.00	0.130	0.0588	0.130	0.0575	0.139	0.0444	.992	.995	.995
3	3	486	162.00	1.424	0.1050	1.422	0.1205	1.432	0.0882	.656	.532	.529
4	5	2121	424.20	0.198	0.0000	0.198	0.0000	0.199	0.0000	-.875	-.878	-.878
5	5	126	25.20	0.764	0.6904	0.710	0.4898	1.052	2.0234	-.955	-.986	-.986
6	7	1281	183.00	0.738	0.0000	0.738	0.0000	0.722	0.0080	.490	.451	.451
7	7	845	120.71	0.184	0.0000	0.184	0.0000	0.186	0.0000	-.444	-.457	-.457
8	8	520	65.00	0.461	1.4200	0.450	1.1780	0.479	1.4696	-.631	-.755	-.754
9	9	2097	233.00	0.268	0.1128	0.237	0.0646	0.312	0.2341	-.375	-.430	-.420
10	9	2007	223.00	0.059	0.0000	0.059	0.0000	0.059	0.0000	-.473	-.475	-.475
11	10	1981	198.10	0.429	0.1733	0.423	0.1546	0.441	0.1883	-.663	-.713	-.714
12	11	8951	813.73	0.230	0.0184	0.236	0.0231	0.259	0.0563	-.581	-.572	-.581
13	12	805	67.08	0.910	1.3439	0.820	0.5507	1.014	2.0634	-.061	-.605	-.589
14	12	1078	89.83	0.466	0.7514	0.432	0.3761	0.506	0.9215	-.207	-.506	-.482
15	13	3009	231.46	0.178	0.0056	0.211	0.0278	0.328	0.3442	-.533	-.325	-.441
16	12	3277	273.08	0.111	0.0036	0.119	0.0135	0.140	0.0392	-.539	-.355	-.479
17	15	2982	198.80	1.058	1.0766	1.060	0.9279	1.073	1.0058	.420	.306	.307
18	16	1575	98.44	0.515	0.7814	0.510	0.5567	0.517	0.7794	-.208	-.249	-.251
19	20	7984	399.20	0.147	0.0320	0.142	0.0224	0.136	0.0142	-.292	-.258	-.281
20	21	3408	162.29	1.938	4.6248	1.926	2.6296	1.974	4.6019	.171	-.255	-.253
21	25	258373	10334.92	0.089	0.0209	0.068	0.0090	0.109	0.0390	-.493	-.483	-.433
22	26	90191	3468.88	0.259	0.0270	0.239	0.0119	0.070	0.0000	-.426	-.329	-.381
23	28	807	28.82	0.718	0.1680	0.718	0.1518	0.762	0.0773	.118	-.060	-.062
24	32	5462	170.69	0.549	0.9116	0.536	0.3211	0.565	0.9384	.117	-.106	-.133
25	34	1683	49.50	0.018	0.0665	0.019	0.0611	0.020	0.0484	.216	.237	.234
26	36	3195	88.75	0.754	0.5612	0.731	0.3198	0.798	0.7104	-.344	-.542	-.539
27	36	3026	84.06	0.302	0.0537	0.302	0.0542	0.321	0.0994	-.340	-.386	-.386
28	42	14069	334.98	0.103	0.0932	0.103	0.0978	0.103	0.0786	-.032	-.034	-.035
29	43	18096	420.84	0.287	0.0426	0.283	0.0331	0.293	0.0566	-.245	-.283	-.299
30	44	4428	100.64	0.030	0.1131	0.031	0.1043	0.024	0.1420	.251	.273	.272
31	46	51966	1129.70	0.159	0.0035	0.152	0.0018	0.132	0.0000	-.212	-.324	-.288
32	49	2399	48.96	0.225	0.1680	0.221	0.1440	0.246	0.2232	-.297	-.407	-.406
33	49	20272	413.71	0.531	0.1205	0.507	0.0792	0.578	0.2188	-.548	-.630	-.604
34	52	4489	86.33	0.610	0.2462	0.610	0.2469	0.627	0.2555	-.094	-.274	-.274
35	54	4617	85.50	0.636	0.4435	0.624	0.2792	0.668	0.5942	-.112	-.403	-.392
36	54	2446	45.30	0.456	0.0909	0.459	0.0995	0.462	0.0727	-.494	-.582	-.582
37	56	6257	111.73	0.316	0.1087	0.314	0.0874	0.328	0.1641	-.085	-.163	-.158
38	63	7133	113.22	1.545	0.4135	1.542	0.3813	1.576	0.4251	-.134	-.442	-.441
39	63	10566	167.71	0.464	0.3339	0.454	0.2113	0.474	0.3474	-.197	-.324	-.323
40	69	7407	107.35	0.429	0.2901	0.426	0.2196	0.439	0.3000	-.009	-.155	-.160
41	72	10639	147.76	0.434	0.2335	0.431	0.1738	0.444	0.2796	.026	-.110	-.122
42	72	151838	2108.86	0.872	0.2831	0.808	0.0311	0.923	0.4255	-.059	-.348	-.181
43	77	152059	1974.79	2.762	1.4122	2.509	0.1514	2.854	1.4767	-.003	-.588	-.487
44	79	11235	142.22	0.878	0.1853	0.879	0.1991	0.898	0.2088	-.046	-.222	-.226
45	82	5104	62.24	0.653	0.2001	0.648	0.1783	0.681	0.2173	-.241	-.425	-.426
46	85	6015	70.76	0.451	0.3103	0.416	0.1918	0.600	0.9822	-.656	-.787	-.749
47	87	19604	225.33	0.699	0.5003	0.700	0.6788	0.713	0.5192	.070	-.152	-.168
48	88	38138	433.39	0.552	0.2142	0.543	0.1374	0.596	0.4940	-.375	-.579	-.515
49	88	6622	75.25	0.042	0.0000	0.042	0.0000	0.043	0.0000	-.054	-.057	-.057
50	93	8487	91.26	0.460	0.2979	0.454	0.1919	0.475	0.3491	-.012	-.209	-.215
51	105	14001	133.34	0.673	0.1051	0.672	0.0885	0.688	0.1185	-.009	-.131	-.127
52	106	35185	331.93	0.464	0.4696	0.434	0.1360	0.481	0.5096	-.075	-.262	-.263
53	118	16518	139.98	0.607	0.1241	0.606	0.1159	0.627	0.1874	-.080	-.235	-.232
54	165	15524	94.08	0.542	0.1289	0.541	0.1209	0.555	0.1441	-.186	-.272	-.270

Note. MM = Mixture Model; DL = DerSimonian & Laird estimator; REML = Restricted Maximum Likelihood