Published in final edited form as:
Rev Educ Res. 2014 September 1; 84(3): 328–364. doi:10.3102/0034654313500826.

Meta-Analysis With Complex Research Designs: Dealing With Dependence From Multiple Measures and Multiple Group Comparisons

Nancy Scammacca,
Nancy Scammacca, PhD, is a research associate at the Meadows Center for Preventing Educational Risk in the College of Education at the University of Texas at Austin, 1 University Station D4900, Austin, TX 78712

Greg Roberts, and
Greg Roberts, PhD, is the director of the Vaughn Gross Center for Reading and Language Arts and the associate director of the Meadows Center for Preventing Educational Risk at the University of Texas at Austin

Karla K. Stuebing
Karla K. Stuebing, PhD, is a research professor at the Texas Institute for Measurement, Evaluation, and Statistics at the University of Houston

Nancy Scammacca: nancyscammacca@austin.utexas.edu

Abstract

Previous research has shown that treating dependent effect sizes as independent inflates the variance of the mean effect size and introduces bias by giving studies with more effect sizes more weight in the meta-analysis. This article summarizes the different approaches to handling dependence that have been advocated by methodologists, some of which are more feasible to implement with education research studies than others. A case study using effect sizes from a recent meta-analysis of reading interventions is presented to compare the results obtained from different approaches to dealing with dependence. Overall, mean effect sizes and variance estimates were found to be similar, but estimates of indexes of heterogeneity varied. Meta-analysts are advised to explore the effect of the method of handling dependence on the heterogeneity estimates before conducting moderator analyses and to choose the approach to dependence that is best suited to their research question and their data set.

Keywords

meta-analysis; statistical dependence; heterogeneity analysis

---

The inclusion of statistically dependent effect sizes in a meta-analysis can present a serious threat to the validity of the meta-analytic results. Dependence can arise in a number of ways. One common way that dependence presents itself occurs when a study included in a meta-analysis uses more than one outcome measure, such as a reading intervention study that measures both reading fluency and reading comprehension. The resulting effect sizes are dependent because the same participants were measured more than once. Dependence also

NIH-PA Author Manuscript

NIH-PA Author Manuscript

commonly occurs when a study's research design includes two treatment groups compared with the same control group. Because the same control group participants are included in each treatment/control comparison, the resulting effect sizes are statistically dependent. Failure to resolve or model dependence results in artificially reduced estimates of variance, which in turn inflates Type I error (Borenstein, Hedges, Higgins, & Rothstein, 2009a). Treating dependent effect sizes as if they were independent also gives more weight in the meta-analysis to studies that have multiple measures or more than two groups. Statistical dependence must be resolved in a way that allows each study to contribute a single independent effect size to the meta-analysis or modeled using methodological techniques designed to handle dependence to avoid these threats to the validity of the meta-analytic results.

Prevalence of the Problem of Dependence in Meta-Analyses in Education Research

Education research studies commonly yield a set of dependent effect sizes. For example, Edmonds et al. (2009) extracted 78 effect sizes from 21 studies of interventions for struggling readers, an average of nearly four per study across multiple measures and multiple dependent comparisons. Tran, Sanchez, Arellano, and Swanson (2011) calculated 107 effect sizes from multiple measures across 13 response-to-instruction studies, meaning that an average of eight outcome measures had been used in these studies. In their meta-analysis on the effectiveness of Reading Recovery, D'Agostino and Murphy (2004) calculated 1,379 effect sizes across the multiple outcomes, group comparisons, and testing occasions in the 36 studies that met their inclusion criteria, for an average of approximately 38 effect sizes per study. In a review of education meta-analyses published since 2000, Ahn, Ames, and Myers (2012) found that 37.5% of the 56 meta-analyses included in their report averaged three or more effect sizes per study. The average number of effect sizes per study across all 56 meta-analyses was 3.71. Just 7 of the 56 meta-analytic reports stated that dependence of effect sizes was not an issue in their data set.

Statistical Methods for Handling Dependence From Multiple Outcomes

Much has been written by prominent researchers about how to resolve dependence of effect sizes in a meta-analysis when faced with multiple outcomes. Some methods are more complex and challenging to implement with education research studies than others. On the less complex end of the spectrum, Card (2012) recommended choosing between two straightforward methods of resolving dependence. The first is to select a single outcome to include based on the focus of the meta-analysis. He cautioned that this approach is appropriate only when the meta-analyst can make a strong case for including one outcome over others. A second option, and one that is frequently implemented in education meta-analyses, is to aggregate all measures by computing an average effect size. Although computing an average effect across measures within a study is easy to do, the result may not be the best measure of the effect of the study. This approach effectively punishes studies for attempting to measure the impact of their treatment across a broad array of measures. For example, researchers testing a reading fluency intervention might be interested in knowing if their intervention has any effect on reading comprehension. Such a study conceivably could result in a large effect of 0.80 on a measure of reading fluency and a small effect of 0.20 on a measure of reading comprehension. If these measures are averaged for inclusion in a meta-analysis that is focused broadly on the effect of reading interventions on reading skills, the resulting effect size of 0.50 would not accurately represent the effectiveness of this study's intervention.

Reflecting on this problem, Marín-Martínez and Sánchez-Meca (1999) cautioned meta-analysts to consider whether or not effect sizes within a study are homogenous before averaging them to resolve dependence. If effects within studies are not homogenous, another approach to resolving dependence should be implemented. Cooper (1998) suggested a variation on simply averaging all outcomes. In his shifting-unit-of-analysis approach, effect sizes within studies are combined based on the variables of interest in the meta-analysis to provide a single estimate of the overall effect to include in the meta-analysis. Cooper stated that this approach minimizes violations of the assumption of independence of the effect sizes while preserving as much of the data as possible. However, using this approach can result in running multiple meta-analyses for each outcome type, with some analyses having a small number of studies and little power as a result.

More complex approaches to dealing with dependence from multiple outcomes involve accounting for the correlation between measures when computing a summary effect size across multiple dependent outcomes. As Borenstein et al. (2009b) pointed out, averaging effect sizes across measures makes an implicit assumption that the correlation between measures is 1.0—meaning that each outcome essentially duplicates the information provided by other outcomes. When meta-analysts ignore dependence and include effect sizes from all measures as if the effects were independent, the assumed correlation between measures is 0—meaning that each outcome contributes information that is unrelated to any other outcome. According to Borenstein et al., when making either of these assumptions about the correlation between measures, the result is an incorrect estimate of the variance of the composite effect size that the study contributes to the meta-analysis. Assuming a correlation of 1.0 results in an overestimate of the variance of the composite effect size because all the information provided by the outcomes is redundant. Assuming a correlation of 0 results in an underestimate of the variance for the composite effect size because each effect size is seen as contributing independent information. A larger estimate of the variance results in a larger confidence interval around the effect size and an increased likelihood of finding that the effect size is not significantly different from zero (a Type II error). The opposite is true when an inaccurately small estimate of the variance is calculated, resulting in an inflation of the Type I error rate.

When the correlation between outcomes is known, the dependence can be accounted for mathematically when computing a mean effect for a study. Rosenthal and Rubin (1986); Raudenbush, Becker, and Kalaian (1988); Gleser and Olkin (1994); and Borenstein et al. (2009b) provided equations for calculating an effect size for a study with multiple outcomes that include the correlations between the outcomes. More complex approaches incorporate the correlation between measures into multivariate models for conducting meta-analysis. Kalaian and Raudenbush (1996) described and illustrated the use of multivariate multilevel modeling to conduct meta-analysis in a way that models dependency in effects within studies. In their example, they meta-analyzed studies of the impact of coaching on performance on the Scholastic Aptitude Test (SAT) math and verbal subtests. Given that the correlation between these subtests has been reported by the developers of the SAT, Kalaian and Raudenbush were able to compute the covari-ance matrix needed for implementing their modeling technique. The structural equation modeling (SEM) approach to meta-analysis proposed by Cheung (2010) also requires that the correlations between multiple measures within a study are known.

In her discussion of multivariate meta-analysis, Becker (2000) acknowledged that in many cases the meta-analyst does not know the correlations between multiple measures used in a particular study. She suggested consulting previous studies or manuals from test publishers to impute a correlation. Theoretically, such an approach makes sense. However, it is often impractical or impossible for a meta-analyst working with education research studies to implement any of these suggestions. Researcher-designed measures are commonly used in education research, and the correlations between such measures are not routinely reported. When a study measures outcomes using standardized tests, the correlations between them might be available from test publishers or in the research literature, but the extent to which these correlations generalize beyond the normative sample to a special population (such as students with learning disabilities) is rarely documented.

When it is not possible to locate the correlation from these sources, Becker (2000) and Borenstein et al. (2009b) suggested conducting sensitivity analyses to determine a possible range of correlations between measures. Conducting sensitivity analyses can be a workable solution when a small number of measures are involved and only a few studies use multiple measures. However, when more than two or three measures are used in multiple studies to be included in the meta-analysis, conducting sensitivity analyses for every pair of outcomes quickly become so laborious and time-consuming that it is not feasible, especially because computer programs to conduct sensitivity analysis are not available. In these instances, averaging outcomes with an assumed correlation of 1.0 and inflating Type II error is considered the more conservative approach.

Statistical Methods for Handling Dependence From Multiple Group Comparisons

Many of the same researchers who have suggested methods for dealing with dependence when including studies with multiple outcomes also have described methods for dealing with dependence from multiple group comparisons within studies. Gleser and Olkin (1994) provided equations for a matrix of effect sizes that come from a set of studies where multiple treatments are compared with a no-treatment control group. They assumed that the corpus of studies that the meta-analyst has gathered includes a common and defined set of treatments (such as several types of diet or exercise routines), with some studies including perhaps two of these treatments compared with a no-treatment control group and others including three or four or more. In this scenario, regression models can be fit that account for the dependence in the group comparisons within studies. This approach works well in fields where treatments are standardized or come from a common set of treatments, such as medicine. Within education research, it is rare that the same treatments are present across studies, making it impossible to construct the type of matrix needed to implement Gleser and Olkin's approach.

Borenstein et al. (2009c) proposed a way of dealing with the dependence inherent in multiple group comparisons that is more easily applied to education research. First, they advised meta-analysts to consider if their interest is in comparing the effects of two specific treatments or in computing a combined overall effect of treatment compared with the control group. If one's interest is in comparing treatments, and two treatment groups are compared with a single control group in a given study, an effect size can be computed from the information provided for the two treatments that indicates the benefit of one treatment over the other. In this case, effect sizes from treatment–control comparisons are not included in the meta-analysis, eliminating the dependence from the shared control group. This approach makes sense only if the two treatments are present in a similar enough form across the corpus of studies to allow for similar contrasts across the meta-analysis.

If one's interest is in the overall effect of different types of treatment compared with a control group, calculating a combined effect size and its variance for studies in which multiple treatments are compared with the same control group is a straightforward process as long as the number of participants in each treatment group and the control group is known. The correlation between the effect size for the first treatment group versus the control group and the effect size for the second treatment group versus the control group can be calculated based on the number of participants in each group. A combined weighted mean effect size can be computed that gives more weight to an effect from a treatment with a larger sample size than to another treatment in the same study with a smaller sample size. The variance of this combined effect can be computed in a manner that takes into account the proportion of all study participants that are shared members of the control group. For example, if 50 participants are in one treatment group, 50 participants are in a second treatment group, and 50 participants are in the control group, the proportion of shared participants in the comparison of the each treatment group with the control group is 0.50 because 50% of the participants in each comparison are the same. More simply, in cases where means, standard deviations, and sample sizes are available for all treatment groups and the control group, the meta-analyst can create a combined mean simply by calculating a weighted mean and standard deviation for a study with all treatment conditions combined and using this mean and standard deviation with the mean and standard deviation of the control group to calculate a standardized mean difference effect size.

Borenstein et al.'s (2009c) approach to computing a combined, weighted mean effect is easier to apply to the types of research methodologies typically found in education research reports than Gleser and Olkin's (1994) approach. It is a sound means of preserving the statistical independence of effect sizes in a meta-analysis. However, independence comes at the cost of losing information about the unique effect of each treatment. Averaging the effects of treatment may not represent the intent of a study's researchers when they designed a multiple treatment versus control study. Additionally, when there are vast differences in the effectiveness of the treatments, this approach handicaps the most effective treatment in a study by averaging it with less effective treatments. When there are many studies with multiple dependent comparisons in a meta-analysis, the overall mean effect will be reduced by the presence of weaker and stronger treatments homogenized into a middling studywise effect size.

New Approaches to Dealing With Dependence From Multiple Outcomes and Comparisons

Robust Variance Estimation

Hedges, Tipton, and Johnson (2010) proposed a new approach to dealing with dependence that can be applied no matter the source or sources of dependence in a data set of effect sizes. Known as robust variance estimation (RVE), it overcomes the need to include the known correlations between measures in order to include all effect sizes from all measures and all group comparisons in the meta-analysis. Instead of modeling dependence as is done in multivariate approaches to meta-analysis that require known correlations, RVE mathematically adjusts the standard errors of the effect sizes to account for the dependence (Hedges et al., 2010; Tanner-Smith & Tipton, 2013). An intraclass correlation (ρ) that represents the within-study correlation between effects must be specified when implementing RVE to estimate the effect size weights, but because RVE is not affected very much by the choice of weights, it does not matter if the correlation is precise (Hedges et al., 2010; Tanner-Smith & Tipton, 2013). Because the same ρ is applied to all dependent effect sizes within each study in the meta-analysis, sensitivity analysis with a range of values for ρ can be conducted quite easily to determine how the correlation that is chosen affects the resulting estimates of the mean effect and its variance. Dependence from multiple sources, including multiple measures and multiple group comparisons, can be accommodated simultaneously (Tanner-Smith & Tipton, 2013). RVE is reasonably easy to implement with syntax for several popular statistical software packages provided by Tanner-Smith and Tipton and available from the Peabody Research Institute (n.d.).

There are some important limitations to consider when implementing RVE. Because the math involved in RVE relies on the central limit theorem, simulation studies have shown that a minimum of 10 independent studies are needed to estimate a reliable main effect and a minimum of 40 independent students are needed to estimate a meta-regression coefficient (Hedges et al., 2010; Tanner-Smith & Tipton, 2013). RVE can be used only in meta-regression. If a meta-analysis involves categorical moderators with more than two levels, the dummy-coding of variables required to analyze all pairwise comparisons can be cumbersome to implement in currently available statistical software. Additionally, because the degrees of freedom used to test the statistical significance of the meta-regression coefficients is equal to the number of independent studies minus the number of parameters estimated, meta-analyses with a small number of studies will be restricted in the number of covariates that can be included (Tanner-Smith & Tipton, 2013). Tanner-Smith and Tipton's simulation studies indicated that a minimum of 40 studies with an average of at least five effect sizes per study are needed to estimate a meta-regression coefficient. When fewer studies are included, they found that the confidence interval for the coefficient tends to be too narrow, meaning that the *p* value for the estimate will be inaccurate. Nevertheless, RVE is a mathematically sound method for modeling dependence that should be strongly considered by education meta-analysts when their data sets meet its requirements.

Three-Level Meta-Analysis

Konstantopoulos (2011) proposed three-level meta-analysis as an extension of the use of two-level random-effects models in meta-analysis. In two-level models, Level 2 variance represents between-study differences in effect size estimates, with the assumption that all studies are contributing an independent effect size. Three-level meta-analysis allows for clustering of dependent effect sizes within studies at Level 2; between-study effects are then estimated at Level 3. Cheung (2013) described how three-level meta-analysis can be used to pool dependent effect sizes within each study, modeling the within-study dependence at Level 2 and the between-study mean effect size and variance at Level 3. This approach to dependence can be applied when the correlations between the dependent effect sizes are not known, as is usually the case when multiple measures are used in a study. Unlike in RVE, three-level meta-analysis provides estimates of both the Level 2 (within study) and Level 3 (between study) variance so that meta-analysts can determine where the variation in effects is the greatest. Covariates can be included in the three-level model at both Level 2 and Level 3 to attempt to explain the variance present at each level.

Cheung (2013) described how to use SEM to conduct a three-level meta-analysis. Some advantages of the SEM approach include its ability to handle missing data on covariates and to provide a means for empirical comparison of the two-level and three-level models to determine which model best fits the data. Cheung provided syntax and a package for running three-level meta-analysis in R, making it easier for other meta-analysts to implement his approach. Like RVE, three-level meta-analysis is a promising solution to the problem of dependence in meta-analysis. However, as Cheung noted, additional studies are needed to demonstrate the strengths and potential limitations of both approaches to dependence because neither technique has been used widely in published research.

How Education Researchers Handle Dependence in Meta-Analysis

Drawing from the methods described above, education researchers have implemented a variety of means of handling dependence from multiple measures and/or multiple group comparisons when conducting a meta-analysis. In their meta-analysis of the effect of writing instruction on reading, Graham and Hebert (2011) resolved the dependence from multiple measures using Cooper's (1998) shifting-unit-of-analysis approach. They separated measures by construct (e.g., reading comprehension, reading fluency) and meta-analyzed effect sizes for each construct separately. When studies included multiple measures of a single construct, they included the average of the effects in their meta-analysis. Graham and Hebert's approach yielded multiple sets of independent effects that they meta-analyzed separately. This approach also can be implemented when studies provide multiple treatment comparisons by conducting separate meta-analyses for each type of treatment.

The advantage of this approach is that it allows the meta-analyst to retain all of the information from each study while preserving statistical independence. However, to do so the meta-analyst must run multiple analyses and cannot draw conclusions about the overall effect from the corpus of studies. Additionally, dividing the corpus of studies into groups by measure type and/or treatment type can result in a significant reduction in power. Nevertheless, this approach remains popular with meta-analysts and has been implemented in a number of other recent meta-analyses in education (e.g., Flynn, Zheng, & Swanson, 2012; Gersten et al., 2009; Tran et al., 2011). In their review of 56 education meta-analyses, Ahn et al. (2012) found that 26.8% of the meta-analyses in their data set used the shifting-unit-of-analysis approach to resolve dependence.

Another common approach to handling dependence in meta-analysis is to select a single measure and/or group comparison that seems to best represent the study's primary research question. Graham and Hebert (2011) took this approach to resolving the statistical dependence in studies that had multiple group comparisons, and Chambers (2004) used it in a meta-analysis of the effects of computers in classrooms. In their meta-analysis of reading comprehension instruction for students with learning disabilities, Berkeley, Scruggs, and Mastropieri (2010) implemented a hybrid of this approach and the approach described above, selecting a single outcome measure from each study that best represented the research question while conducting separate meta-analyses for different types of measures and for measures of treatment effect, maintenance effect, and generalization effect. This approach was used in 14.3% of the 56 education meta-analyses reviewed by Ahn et al. (2012).

The main advantage of this method of resolving dependence is that it contributes the effect size that conveys the central finding of the study to the meta-analysis. When meta-analysts select a single outcome or group comparison for the meta-analysis, studies that include additional outcomes or comparisons in an attempt to measure the effects of their intervention more broadly or compare it with other types of treatment do not have the effect size of their primary outcome or comparison of interest reduced by averaging it with smaller effects from tertiary outcomes or weaker treatments. However, in large-scale or multicomponent interventions, researchers often expect to see effects of treatment on multiple types of measures or are interested in determining which of several treatments is most effective. In these cases, it can be difficult for the meta-analyst to pick a single measure or group comparison that will best represent the study in the meta-analysis, especially if the study's authors are not clear in describing the outcome or comparison they view as most central to the purpose of their study.

Ahn et al. (2012) documented the use of other approaches to dealing with dependence in the 56 education meta-analyses they reviewed. The approach most commonly used in these meta-analyses was averaging or weighted averaging of the dependent effect sizes within studies. This approach was implemented in 42.9% of the meta-analyses. They also found that a multivariate approach was used in 7.1% of the meta-analyses. A combination of approaches was used in 12.5% of the meta-analyses. In 32.2% of the meta-analyses, researchers either failed to mention whether dependence was an issue in their data set or mentioned it but did not report how they handled it.

Because Hedges et al.'s (2010) RVE approach is a relatively new technique for dealing with dependence, published examples of its use are few in number. Wilson, Tanner-Smith, Lipsey, Steinka-Fry, and Morrison (2011) used RVE to account for dependence in their meta-analysis of high school dropout prevention programs that included 504 effect sizes from 317 independent samples and 152 studies. Uttal et al. (2013) implemented RVE in a meta-analysis that included 1,038 effect sizes from 206 studies that assessed the effect of training programs on spatial skills. Outside of educational research, RVE has been implemented in meta-analyses on the effectiveness of outpatient substance abuse treatment for adolescents (Tanner-Smith, Wilson, & Lipsey, 2013), the relationship between social goals and aggressive behavior in youth (Samson, Ojanen, & Hollo, 2012), and the effect of mindfulness-based stress reduction on physical and mental health in adults (de Vibe, Bjørndal, Tipton, Hammerstrøm, & Kowalski, 2012). No published examples of the use of three-level meta-analysis to handle dependence were found in the educational research literature. Both Konstantopoulus (2011) and Cheung (2013) illustrated the use of three-level meta-analysis with extant data sets. Van den Noortgate, López-López, Marín-Martínez, and Sánchez-Meca (2013) used simulated data sets in their exploration of three-level meta-analysis as a method for handling dependence.

A Case Study in Methods of Dealing With Dependence

To better understand the impact of the choices education meta-analysts face when dealing with multiple measures and multiple group comparisons within studies, different methods of handling dependence were implemented using a set of effect sizes from a meta-analytic study by Scammacca, Roberts, Vaughn, and Stuebing (in press) of reading interventions for struggling readers in Grades 4 to 12. Researchers chose to use an extant set of effect sizes from a recent meta-analysis rather than a simulated data set because we believed that a real-world data set can better emulate the types and nature of dependence that typically exist in studies that education researchers struggle to meta-analyze. In doing so, we acknowledge that simulation studies make an important contribution to the knowledge base and are a necessary next step to the work we present here.

The Scammacca et al. (in press) report involved separate and combined analyses of effect sizes from research published between 1980 and 2004 and between 2005 and 2011. For this report, only effect sizes from the 2005 to 2011 group of 50 studies were used. This more recent group contained many more instances of studies with more than two groups (k = 17) and with multiple measures (k = 43) than the earlier group of studies. The proportion of these more complex research designs within the set of 50 is more representative than the older set of the sets of studies that would cause a meta-analyst to confront the issues addressed here. See the appendix for the effect size data used in this case study.

This case study sought to answer the following research questions:

Research Question 1: How do different approaches to dealing with dependence in data from multiple outcomes within studies affect meta-analytic estimates of mean effect size, variance, and indexes of heterogeneity?

Research Question 2: How do different approaches to dealing with dependence in data from multiple group comparisons within studies affect meta-analytic estimates of mean effect size, variance, and indexes of heterogeneity?

The approaches to handling dependence in this case study include those implemented in other meta-analytic studies that involved education data and others chosen to illustrate alternative means of estimating the overall effect from a study with multiple dependent effects. Additionally, meta-analyses were attempted with all outcomes and all groups as independent for comparison purposes.

Method

Procurement of Corpus of Studies

The studies used in Scammacca et al. (in press) were located through a computer search of ERIC and PsycINFO using descriptors related to reading, learning difficulties/disabilities, and reading intervention; a search of abstracts from other published research syntheses and meta-analyses and reference lists in seminal studies; and a hand search of major journals in which previous intervention studies were published. Studies were included in the meta-analysis if (a) participants were English-speaking struggling readers in Grades 4 to 12 (age 9–21), (b) the study's research design used a multiple-group experimental or quasi-experimental treatment-comparison or multiple-treatment comparison designs, (c) the intervention provided any type of reading instruction, (d) data were reported for at least one dependent measure that assessed one or more reading constructs, and (e) sufficient data for calculating effect sizes and standard errors were provided.

Studies that met criteria were coded using a code sheet that included elements specified in the What Works Clearinghouse Design and Implementation Assessment Device (Institute of Education Sciences, 2008) and used in previous research (Scammacca et al., 2007). Researchers with doctorate degrees and doctoral students with experience coding studies for other meta-analyses and research syntheses completed the code sheets. All coders had completed training on how to complete the code sheet and had reached a high level of reliability with others coding the same article independently. Every study was independently coded by two raters. When discrepancies were found between coders, they reviewed the article together and discussed the coding until consensus was reached.

Effect Size Calculation

Effect sizes were calculated using the Hedges (1981) procedure for unbiased effect sizes for Cohen's *d* (this statistic is also known as Hedges's *g*). Hedges's *g* was calculated using the posttest means and standard deviations for treatment and comparison (or multiple treatment) groups when such data were provided. In some cases, Cohen's *d* effect sizes were reported and means and standard deviations were not available. For these effects, Cohen's *d* for posttest mean differences between groups and the treatment and comparison group sample sizes was used to calculate Hedges's *g*. For each effect, estimates of Hedges's *g* were weighted by the inverse of the variance to account for variations in precision based on sample size in the studies. All effects were computed using the Comprehensive Meta Analysis (Version 2.2.064) software (Borenstein, Hedges, Higgins, & Rothstein, 2011). Effects were coded for all measures and pairwise group comparisons between treatment and control groups or different treatment groups when no control group was included in the study. The 36 research reports yielded 50 independent studies with a total of 366 effect sizes, an average of about 7 effect sizes per study. At this point, researchers in the original study were faced with the dilemma of how to combine multiple effect sizes from multiple measures and multiple dependent group comparisons to best estimate the mean effect of reading intervention.

Calculating Mean Effects From Studies With Multiple Measures

Nearly all studies provided data on multiple outcome measures. Scammacca et al. (in press) averaged the effect sizes from multiple measures within each pairwise group comparison using the procedure recommended by Card (2012), and included the average effect size and the average of its standard error in the meta-analysis. Five other approaches to computing a mean effect across multiple measures within a single independent group comparison were conducted for the present report:

1. The measure that yielded the highest effect size was selected for each independent group comparison.
2. A measure was selected at random using a random number generator for each independent group comparison.
3. A measure was selected for each independent group comparison that seemed to best represent the primary focus of the study's intervention.
4. Measures were analyzed separately based on the type of reading skill measured (fluency, vocabulary, spelling, reading comprehension, word, and word fluency) for each independent group comparison and mean effects were calculated for each skill.
5. All measures were treated as independent estimates of effects for each independent group comparison.

All five approaches were used in meta-analyses to calculate a mean effect and its standard error across all studies for all measures included in the research reports and for norm-referenced, standardized measures only. Because researcher-designed measures tend to have lower reliability than standardized measures, repeating the meta-analyses with only standardized measures allows researchers to investigate the effects of different approaches to dealing with multiple measures while constraining some of the influence of measurement error.

Calculating Mean Effects From Studies With Multiple Dependent Groups

Seventeen of the research reports contained more than one dependent treatment-control or multiple-treatment group comparison. In Scammacca et al. (in press), the procedure recommended by Borenstein et al. (2009c) was implemented for comparisons that involved dependent groups. This procedure involves computing a combined weighted mean effect size and its standard error in a manner that reflects the degree of dependence in the data. Four other approaches to computing a mean effect across multiple dependent group comparisons were completed for this report:

1. The group comparison that yielded the highest mean effect size across measures included in the study was selected and included in the meta-analysis.

2. A group comparison was selected at random using a random number generator and its mean effect size across measures was included in the meta-analysis.
3. A group comparison was selected that seemed to best represent the primary focus of the study's intervention and its mean effect size across measures was included in the meta-analysis.
4. All group comparisons were treated as independent and each mean effect size across measures was included in the meta-analysis.

Meta-analyses were then conducted on the resulting data using all types of measures and using standardized measures only, for the reason stated above. In each of the different analyses for the multiple dependent group comparisons, the average of the effect sizes for all measures involving the group comparison of interest was used to hold constant the effect of multiple measures while examining different approaches to the problem of multiple dependent group comparisons. In a similar way, the effect of multiple dependent group comparisons was held constant in the analyses involving multiple measures. In these analyses, the Borenstein et al. (2009c) method of combining multiple dependent comparisons was implemented. Results for the RVE approach and three-level meta-analysis are reported separately.

Meta-Analytic Procedures

For all the methods of dealing with multiple measures and multiple dependent group comparisons, a random-effects model was used to analyze effect sizes. This model allows for generalizations to be made beyond the studies included in the analysis to the population of studies from which they come. Mean effect size statistics and their standard errors were computed and heterogeneity of variance was evaluated using the Q statistic, the I^2 statistic, and the tau-squared statistic. For all but the RVE and three-level meta-analysis approaches, the meta-analyses were conducted in Comprehensive Meta Analysis (Version 2.2.064) software (Borenstein et al., 2011). For the RVE approach, unrestricted, intercept-only meta-regression models were run in SPSS using a macro provided by Tanner-Smith and Tipton (2013) and Peabody Research Institute (n.d.). Sensitivity analysis with a range of values for ρ was conducted to determine the effect of varying intraclass correlations on estimates of the mean effect size, the Q statistic, and the tau-squared statistic. For three-level meta-analysis, Cheung's (2013) R syntax for the metaSEM package he authored was used. Finally, meta-regression was conducted using number of measures and number of groups as a predictor of effect size in a mixed-effects model using unrestricted maximum likelihood estimation.

Results

Approaches to Handling Dependence From Multiple Measures

The meta-analyses that implemented different methods of resolving the dependence resulting from having multiple measures within a study produced some points of similarity and some differences across the methods used when considering all types of measures. See Table 1 for results for all types of measures. The mean effect size and variance when using the mean of measures method was nearly identical to the mean effect size when all measures were treated as independent and when a measure was selected based on the primary research question. Using the highest effect size produced a much larger mean effect, as would be expected, and a slightly larger variance. Random selection of an effect size also produced a somewhat larger estimate of the mean effect and a slightly larger variance. Estimates of heterogeneity varied widely depending on the method used to resolve dependence from multiple measures. Treating all measures as independent, using the highest effect size, and randomly selecting an effect size resulted in the largest values across all three indexes of heterogeneity.

When all types of measures were meta-analyzed by the type of reading skill tested (the shifting-unit-of-analysis approach), reading comprehension measures, word measures, and word fluency measures produced similar mean effect sizes and variances to the mean of measures, all measure independent, and select by research question measures. Results differed for all measures of fluency, vocabulary, and spelling skills, perhaps due to the number of small number of studies that included vocabulary and spelling measures or due to true differences in the effectiveness of reading interventions on these reading skills. The results for indexes of heterogeneity also differed depending on the domain of reading skills analyzed, with fluency measures showing the most heterogeneity and spelling measures the least.

Results of the meta-analyses that included standardized measures only are shown in Table 2. As with the analyses of all measures, the mean of measures approach resulted in a similar mean effect size and variance as was obtained when all measures were treated as independent or when measures were selected based on the study's primary research question. Random selection of a measure resulted in a similar mean effect size and variance as well, whereas choosing the measure with the highest effect size resulted in the largest mean effect but a similar variance to other approaches. The Q and I^2 measures of heterogeneity again had large values for the independent approach, the random approach, the highest effect size approach, and the select-by-research-question approach. In the analyses by reading skill, fluency and vocabulary measures again showed much smaller effects than other domains. The I^2 index of heterogeneity had large values for reading comprehension and small values for other domains, likely due to the large number of studies that included a standardized measure of reading comprehension.

Approaches to Handling Dependence From Multiple Group Comparisons

Results from the five approaches used to deal with the dependence from multiple group comparisons are shown in Tables 3 (all types of measures) and 4 (standardized measures only). Mean effect sizes and variances were very similar across all approaches for both sets of analyses. Selecting the group comparison with the highest effect size resulted in the largest mean effect size for both all types of measures and standardized measures only, but not by much. The difference was especially small in the analysis of standardized measures. Interestingly, the variance did not increase when all group comparisons were treated as independent. However, treating all group comparisons as independent did result in very large estimates on the Q index of heterogeneity. Tau-squared and I^2 estimates were less affected.

The Robust Variance Estimation Approach to Handling Dependence

Results from the meta-analyses that implemented RVE are shown in Tables 5 (all types of measures) and 6 (standardized measures only). Results are reported to the fifth decimal place to show that very little change occurred in the mean effect size and measures of heterogeneity based on varying the intraclass correlation ρ. Compared with the results presented above for other approaches to dealing with dependence, the RVE approach produced estimates of the mean effect size and standard error that were very similar to those found when using the mean of measures approach and the weighted mean for group comparisons approach when looking both at the meta-analysis of all types of measures and only at standardized measures. The *Q* statistic for the meta-analysis of standardized measures was just slightly larger using the RVE approach than in other methods used to handle multiple dependent group comparisons, but the increase was enough to indicate the presence of statistically significant heterogeneity. The estimates of heterogeneity generally were larger than those obtained with other methods of dealing with dependence but less than those obtained when dependence was ignored.

Handling Dependence With Three-Level Meta-Analysis

Results from the three-level meta-analysis using all types of outcome measures were similar to those obtained using RVE. The estimate of the mean effect was 0.27 with a standard error of 0.05 (95% confidence interval = 0.18, 0.37). The tau-squared estimate of variance was 0.10 (*SE* = 0.01) at Level 2 (within studies) and 0.07 (*SE* = 0.02) at Level 3 (between studies), meaning that more within-study than between-study variation was present. In three-level meta-analysis, I^2 is calculated based on the *Q* statistic; thus, it is on a different scale and is interpreted differently than the I^2 statistics that have been presented previously in this article. The Level 2 I^2 and Level 3 I^2 values were 0.48 and 0.35, respectively, meaning that 48% of the variation in effect sizes was due to within-study factors and 35% of the variation was due to between-study factors. These finding suggest that Level 2 covari-ates should be included in the model to account for the within-study variation before between-study covariates are considered.

A three-level meta-analysis using effect sizes from standardized measures only was attempted; it failed to converge on an optimal solution. When restricted maximum likelihood estimation was used to examine the variance components, a solution was reached that estimated tau-squared at Level 2 (within studies) as 1×10^{-10} and at Level 3 (between studies) as 0.02. Therefore, it seems likely that the very small value for tau-squared at Level 2 caused the model to fail to converge on an optimal solution.

Meta-Regression on Number of Measures and Number of Groups

To evaluate the relationship between effect size and number of measures and number of groups in a study, meta-regression was conducted using these variables as a predictor of effect size in a mixed-effects model using unrestricted maximum likelihood estimation. Meta-regression was run predicting the effect size using all types of measures and standardized measures only. The number of measures used in a study was not a statistically significant predictor of effect size when considering all types of outcome measures (β = 0.00, *SE* = 0.01, *Q-model* = 0.06, *df* = 1, *p* = .805, T^2 = 0.03) or only standardized outcome measures ($\beta = -0.01$, $SE = 0.01$, $Q\text{-}model = 0.43$, $df = 1$, $p = .513$, $T^2 = 0.00$). See Figures 1 and 2 for scatterplots of effect sizes by number of measures.

The number of groups used in a study was not a statistically significant predictor of effect size when considering all types of outcome measures ($\beta = 0.00$, $SE = 0.05$, $Q\text{-}model = 0.00$, $df = 1$, $p = .999$, $T^2 = 0.03$). However, number of groups was a statistically significant predictor of effect size when considering only standardized measures ($\beta = -0.06$, $SE = 0.03$, $Q\text{-}model = 5.04$, $df = 1$, $p = .024$, $T^2 = 0.00$), with effect sizes from standardized measures decreasing as the number of groups increased. See Figures 3 and 4 for scatterplots of effect sizes by number of measures.

Discussion

The case study presented here was conducted to demonstrate the effects of different methods of dealing with dependence from multiple measures and multiple group comparisons within studies on meta-analytic results. Results indicated that most approaches to handling dependence produced similar estimates of the mean effect and variance for this set of effect sizes. The mean effect and variance were especially similar when only standardized measures were included in the analyses. The expected increase in the variance of the mean effect was not observed in the case study data when all measures or all group comparisons were included in the analysis as if they were independent effect sizes.

These findings are not what would be expected based on previous research. In their simulation study of dependence from multiple group comparisons, Kim and Becker (2010) found that the variance of the mean effect increased as the proportion of studies in the meta-analysis that contained dependent comparisons increased. They found that the variance estimate was at least somewhat inflated when as few as 20% of the studies in the meta-analysis included dependent comparisons. In the present case study, 34% of the studies had dependent group comparisons. However, Kim and Becker also noted that variance estimates were most inflated when treatment groups were larger than control groups, which was not generally the case in the studies included in the present case study. Additionally, Kim and Becker's simulations involved a set of 10 studies with 12 and 15 effect sizes representing 20% and 50% dependence. In the present case study, 50 studies with 92 effect sizes were included in the analysis with multiple dependent group comparisons. It may be that the larger number of effect sizes in the case study contributed to the difference in findings. Additional simulation studies with a larger set of effect sizes and additional scenarios of dependence are needed to determine under what circumstances and to what extent dependence inflates variance estimates.

Based on a single set of effects from reading intervention studies, the case study presented here cannot provide definitive guidance on the best way to resolve dependence resulting from multiple measures or multiple group comparisons within studies. Indeed, there may not be only one best way to resolve dependence, given that data sets of effect sizes can differ widely in the degree and nature of the dependence present in them. Additionally, the choice of method in dealing with dependence must take into account the overall purpose and research questions behind the meta-analysis. Despite being unable to offer definitive guidance, the case study presented here raises some important issues for meta-analysts of education research to consider as they deal with dependence. Furthermore, it draws attention to ways in which primary researchers can assist meta-analysts by providing the information needed to make the best decision about how to handle the multiple dependent effects from their studies.

Implications for Education Meta-Analysts

Consider the Effect of Your Method of Dealing With Dependence When Conducting Moderator Analyses—The greatest differences between the various methods of dealing with dependence in the case study were seen in the indexes of heterogeneity. For the methods of handling multiple measures in the meta-analysis of all measures, *Q* values varied widely, ranging as high as 363.04. A good deal of variation also was seen in the *Q* values in the meta-analysis for standardized measures only, though the range was smaller. *Q* values were especially large when all group comparisons were treated as independent, providing another reason why this approach to dealing with dependence should be avoided.

Meta-analysts who find large *Q* values likely will want to find meaningful moderator variables within their set of studies that explain the heterogeneity. If a moderator variable was confounded with the approach taken to deal with dependence, the moderator analysis could show significant differences falsely based on a moderator variable that is a characteristic of the studies in the meta-analysis when in fact the heterogeneity being explained is due to the method used to deal with dependence. This false finding could occur, for example, if grade level is used as a moderator variable and multiple measures were more commonly administered to students in upper grades than students in lower grades. Conversely, when the indexes of heterogeneity are increased as a result of the method chosen to deal with dependence and that method is not confounded with any moderator variable, meta-analysts may not be able to find moderators that explain the heterogeneity and not know why it remains unexplained, while failing to realize that the actual source of the heterogeneity is the method used to deal with dependence.

Therefore, it is critical for meta-analysts to consider and account for the impact of their method of dealing with dependence when their meta-analysis results in a large *Q* statistic. If possible, running the meta-analysis using only standardized, norm-referenced measures instead of researcher-developed measures also can be helpful in detecting whether the size of the *Q* statistic is due to variance in measurement rather than meaningful differences between studies. Additionally, looking beyond the often-reported *Q* statistic and evaluating I^2 and tau-squared as measures of heterogeneity is important. Simulation studies are needed to model the impact of different approaches to dealing with dependence on estimates of heterogeneity and determine under what circumstances these estimates are artificially inflated or constrained.

Match Your Method of Dealing With Dependence to Your Research Question and Your Data Set—When the correlations between measures are known or can be obtained, a mul-tivariate approach using multilevel modeling or SEM is the best approach for handling dependence in meta-analytic data sets. Because this is rarely the case in education meta-analyses, different approaches to dependence when correlations are not known have been recommended by different research methodologists. Given the lack of guidance currently available on a single optimal way to deal with dependence from multiple measures or multiple group comparisons when correlations are not available, meta-analysts of education research would do best to choose an approach that is suited to the data available from their set of studies and the questions they hope to answer through their meta-analyses.

When an overall estimate of the effect of treatment is more central to the purpose of a meta-analysis and/or treatments and measures are sufficiently similar to one another, the RVE, three-level meta-analysis, and mean of measures approaches are the best options. The RVE and three-level meta-analysis approaches are particularly well suited to meta-analyses with a large numbers of studies and when continuous moderators or dichotomous categorical moderators are of interest. Because three-level meta-analysis provides variance estimates at both the within-study and between-study levels and allows for covariates to be introduced at both levels, it is ideal for meta-analyses where researchers are interested in exploring sources of systematic variance at the within-study level. However, given the newness of RVE and three-level meta-analysis and the small number of published meta-analyses that have used these techniques, more research is needed to explore their benefits and limitations as solutions to the problem of dependence in education meta-analyses before they can be considered the optimal methods.

When the meta-analyst's research questions are addressed to the mean effect of particular domains of treatment or measurement and obtaining an overall estimate of effect across domains is not important, the shifting-unit-of-analysis approach works well if domains of treatment or measurement can be sorted cleanly into categories and enough independent effect sizes are available in each domain to allow for sufficient power. When the research question driving the meta-analysis is clearly and at least somewhat narrowly defined, selecting a single measure and/or group comparison that is best aligned with the purpose of the meta-analysis is a reasonable approach to dependence. Intentional selection of a single measure or comparison also is warranted when the authors of the studies in the meta-analysis define causal models in a way that makes clear which measures their treatments should affect most directly or which treatment in a multiple-comparison study is most central to their hypothesis. In data sets where a great deal of dependence is present, multiple approaches to resolving dependence might be attempted and the range of the mean effect, variance, and indexes of heterogeneity for each approach reported.

Practice Full Disclosure—The American Psychological Association's Meta-Analysis Reporting Standards (American Psychological Association, 2008) recommended that meta-analysts describe the method used to arrive at a single independent effect size for studies that contain multiple dependent effect sizes. However, it seems that this recommendation is not routinely followed. Ahn et al. (2012) reported that 32.2% of the education meta-analyses they reviewed failed to disclose any information about the dependence present in their corpus of studies or how it was handled. Similarly, in a review of a random sample of 100 meta-analyses in psychology and related disciplines, Dieckmann, Malle, and Bodner (2009) found that information on dependence was missing from 34% of the reports they reviewed.

Given the importance of maintaining statistical independence in a meta-analysis, failure to report the extent and type of dependence present in one's corpus of studies and how that dependence was resolved is inexcusable and raises questions about the validity of the meta-analytic results. Additionally, meta-analysts should describe briefly why a particular approach to resolving dependence was chosen and what attempts were made to determine how the mean effect size, variance, and estimates of heterogeneity were affected by the chosen approach.

Implications for Primary Researchers

Clearly Specify All Aspects of Your Causal Model—Given the complexity present in studies where dependence of effect sizes occurs, primary researchers can assist meta-analysts who will be working with these effect sizes by clearly stating the way in which the measures and/or multiple treatment groups are related in their theoretical conceptualization of the effect of their treatment. Readers and future meta-analysts will benefit from knowing which measures a study's researchers view as a primary indicator of the effectiveness of the treatment and which are secondary or tertiary indicators. This information helps meta-analysts who choose to deal with dependence by focusing only on primary indicators to know which measure to include.

When a study introduces dependence from multiple group comparisons, researchers can help meta-analysts by carefully describing the treatment provided to all groups (including details on what, if any, treatment the control group receives, especially if control group members are receiving a business-as-usual treatment provided by their school). Complete descriptions of all groups allow meta-analysts to select group comparisons that align with their research questions and to choose moderator variables to use in attempting to explain heterogeneity. Additionally, primary researchers who include multiple treatment groups should explain why different variations of treatments are being provided, how the treatments differ, and which outcomes are considered primary for each treatment. This information is helpful to meta-analysts who are aggregating independent effects across studies based on similar treatment characteristics or who are interested in including independent effects from certain types of treatment only. Finally, primary researchers might consider whether their causal model would be best represented in future meta-analyses if separate control groups were provided for each treatment group. When researchers are interested in determining the distinct effect of two or more different treatments compared with a control condition, the cost of creating separate control groups would be warranted. Doing so preserves independence while allowing the effect size for each treatment–control comparison to be included rather than averaged.

Provide All Relevant Statistics Needed to Deal With Dependence—Another way that primary researchers can assist meta-analysts in dealing with dependence is by providing all the necessary statistical information needed to allow meta-analysts to implement multivariate meta-analytic methods that model dependence. Researchers should provide the correlations between all measures used so that meta-analysts can create the covariance matrices needed for meta-analytic multilevel modeling and multivariate SEM. A simple table of measures and their correlations based on all participants in the study's sample is all that is needed. Additionally, primary researchers with multiple dependent group comparisons should be sure to include the initial sample size and the sample size after attrition for all treatment and comparison groups so that meta-analysts can use this information to calculate a sample-weighted effect size for all treatments versus the shared control group.

Conclusion

With primary researchers increasingly designing more complex, large-scale studies at the request of grant providers, statistical dependence of the resulting effect sizes has become a significant issue for meta-analysts in education research. All the approaches available to meta-analysts to deal with dependence that were described in this report and demonstrated in the case study have benefits and limitations. At the present time, selecting a method for dealing with dependence is one of many choices a researcher must make when conducting a meta-analysis. Further research is needed to test these approaches with simulated and nonsimulated data to determine the conditions under which each approach is best implemented and to provide better guidance in selecting the best approach for a given set of dependent effect sizes. While waiting for this guidance to become available, the best way forward for education meta-analysts is to weigh carefully the advantages and disadvantages of each approach and to provide as much information as possible on the chosen approach so that readers can consider this information when interpreting the meta-analytic results.

Acknowledgments

This research was supported by Grant P50 HD052117 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and by the Institute of Education Sciences, U.S. Department of Education, through Grant R305F100013 to The University of Texas at Austin as part of the Reading for Understanding Research Initiative. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the National Institutes of Health, the Institute of Education Sciences, or the U.S. Department of Education.

Appendix

Effect Size Data Used in this Case Study

Study	Independent subgroup	Comparison[a]	Outcome[b]	Hedges's *g*	*SE*
1		T1 vs. C	A	1.39	0.35
			B	1.08	0.33
			C	0.13	0.31
		T2 vs. C	A	0.92	0.33
			B	1.01	0.33
			C	−0.16	0.31
2		T vs. C	A	0.85	0.30
3	A	T1 vs. T2	A	0.06	0.32
			B	1.09	0.34
			C	0.05	0.32
		T1 vs. T3	A	0.41	0.32

Study	Independent subgroup	Comparison[a]	Outcome[b]	Hedges's *g*	*SE*
			B	2.53	0.43
			C	0.28	0.32
		T1 vs. T4	A	0.15	0.32
			B	0.87	0.33
			C	0.16	0.32
		T1 vs. T5	A	0.20	0.32
			B	0.06	0.32
			C	0.29	0.32
		T2 vs. T3	A	0.34	0.32
			B	3.86	0.54
			C	0.26	0.32
		T2 vs. T4	A	0.21	0.32
			B	2.06	0.40
			C	0.13	0.32
		T2 vs. T5	A	0.24	0.32
			B	1.16	0.34
			C	0.36	0.32
		T3 vs. T4	A	0.52	0.32
			B	1.87	0.38
			C	0.14	0.32
		T3 vs. T5	A	0.51	0.32
			B	2.45	0.42
			C	0.54	0.32
		T4 vs. T5	A	0.07	0.32
			B	0.79	0.33
			C	0.45	0.32
	B	T1 vs. T2	A	0.14	0.20
			B	0.02	0.20
			C	0.30	0.20
		T1 vs. T3	A	0.02	0.20
			B	3.36	0.31
			C	0.18	0.20
		T1 vs. T4	A	0.23	0.20
			B	0.26	0.20
			C	0.74	0.20
		T1 vs. T5	A	0.07	0.20
			B	0.61	0.20
			C	0.67	0.20
		T2 vs. T3	A	0.15	0.20
			B	3.09	0.29
			C	0.11	0.20
		T2 vs. T4	A	0.37	0.20

Study	Independent subgroup	Comparison[a]	Outcome[b]	Hedges's *g*	*SE*
			B	0.23	0.20
			C	0.94	0.21
		T2 vs. T5	A	0.18	0.20
			B	0.54	0.20
			C	0.88	0.21
		T3 vs. T4	A	0.20	0.20
			B	2.77	0.28
			C	0.84	0.21
		T3 vs. T5	A	0.05	0.20
			B	2.77	0.28
			C	0.78	0.20
		T4 vs. T5	A	0.12	0.20
			B	0.21	0.20
			C	0.07	0.20
4		T1 vs. C	A	0.71	0.33
		T2 vs. C	A	1.07	0.34
5		T vs. C	A	0.08	0.32
			B	0.96	0.34
			C	0.84	0.33
			D	1.02	0.34
6		T1 vs. T2	A	0.36	0.26
			B	0.15	0.26
			C	0.19	0.26
			D	0.43	0.26
			E	0.17	0.26
			F	0.32	0.26
		T1 vs. T3	A	0.19	0.26
			B	0.42	0.26
			C	0.03	0.26
			D	0.32	0.26
			E	0.25	0.26
			F	0.24	0.26
		T2 vs. T3	A	0.17	0.25
			B	0.58	0.26
			C	0.14	0.25
			D	0.79	0.26
			E	0.45	0.26
			F	0.61	0.26
7	A	T vs. C	A	0.21	0.12
	B	T vs. C	A	0.10	0.11
8		T1 vs. C	A	0.55	0.23
			B	1.09	0.24

Study	Independent subgroup	Comparison[a]	Outcome[b]	Hedges's *g*	*SE*
			C	0.04	0.23
			D	0.37	0.23
			E	0.36	0.23
		T2 vs. C	A	−0.01	0.22
			B	0.18	0.22
			C	0.11	0.22
			D	0.09	0.22
			E	0.37	0.23
		T3 vs. C	A	0.09	0.22
			B	1.65	0.26
			C	0.13	0.22
			D	0.23	0.22
			E	0.45	0.23
9		T vs. C	A	−0.21	0.23
			B	0.24	0.23
			C	0.17	0.23
			D	0.87	0.24
10	A	T1 vs. T2	A	0.45	0.37
			B	0.14	0.35
			C	0.95	0.38
		T1 vs. T3	A	0.19	0.42
			B	0.02	0.41
			C	0.56	0.43
		T2 vs. T3	A	0.34	0.40
			B	0.15	0.39
			C	0.54	0.40
	B	T1 vs. T2	A	0.11	0.34
			B	0.10	0.34
			C	0.06	0.34
		T1 vs. T3	A	1.06	0.44
			B	0.78	0.43
			C	0.31	0.41
		T2 vs. T3	A	1.01	0.42
			B	0.89	0.41
			C	0.45	0.40
11		T1 vs. C	A	−0.42	0.39
			B	−0.20	0.39
			C	−0.12	0.39
			D	0.22	0.39
			E	−0.25	0.39
			F	0.39	0.39
			G	−0.07	0.39

Study	**Independent subgroup**	**Comparison[a]**	**Outcome[b]**	**Hedges's *g***	***SE***
			H	−0.71	0.40
		T2 vs. C	A	0.05	0.37
			B	0.13	0.37
			C	−0.11	0.37
			D	0.20	0.37
			E	−0.46	0.38
			F	0.67	0.38
			G	0.41	0.38
			H	0.38	0.38
		T3 vs. C	A	−0.59	0.38
			B	0.17	0.37
			C	0.16	0.37
			D	0.37	0.37
			E	0.43	0.37
			F	0.41	0.37
			G	0.15	0.37
			H	−0.58	0.38
		T4 vs. C	A	−0.32	0.40
			B	0.47	0.40
			C	−0.19	0.40
			D	0.20	0.40
			E	−0.17	0.40
			F	0.76	0.41
			G	0.13	0.40
			H	0.01	0.40
12	A	T vs. C	A	−0.38	0.28
			B	0.11	0.26
			C	−0.43	0.27
	B	T vs. C	A	−0.12	0.30
			B	0.20	0.28
			C	−0.53	0.30
			D	−0.39	0.29
13		T vs. C	A	1.36	0.29
			B	1.21	0.28
			C	0.93	0.27
			D	0.48	0.26
			E	1.03	0.28
14		T1 vs. C	A	1.32	0.54
			B	2.12	0.62
		T2 vs. C	C	0.26	0.54
			D	2.12	0.69
15		T vs. C	A	0.10	0.12

Study	Independent subgroup	Comparison[a]	Outcome[b]	Hedges's *g*	*SE*
			B	0.02	0.12
			C	0.00	0.12
			D	0.00	0.12
			E	0.00	0.12
			F	−0.12	0.12
			G	−0.04	0.12
			H	−0.09	0.12
16		T vs. C	A	0.93	0.35
			B	0.84	0.35
			C	0.84	0.35
			D	0.73	0.35
			E	0.41	0.34
17	A	T1 vs. C	A	−0.20	0.14
		T2 vs. C	A	−0.11	0.14
		T3 vs. C	A	−0.09	0.14
	B	T1 vs. C	A	0.26	0.10
		T2 vs. C	A	0.01	0.10
		T3 vs. C	A	0.25	0.10
18		T vs. C	A	0.57	0.13
			B	1.90	0.14
			C	0.33	0.13
			D	0.45	0.13
			E	0.15	0.13
19		T vs. C	A	−0.66	0.33
			B	−0.46	0.32
			C	0.12	0.32
			D	−0.36	0.32
			E	−0.51	0.32
			F	−0.48	0.32
			G	0.27	0.32
20		T1 vs. T2	A	0.88	0.45
			B	0.47	0.44
			C	0.95	0.46
			D	0.99	0.46
			E	0.26	0.43
			F	0.44	0.44
			G	0.60	0.44
21		T1 vs. C	A	−0.05	0.13
		T2 vs. C	A	0.22	0.13
22		T1 vs. T2	A	0.15	0.30
			B	0.04	0.30
			C	0.33	0.30

Study	Independent subgroup	Comparison[a]	Outcome[b]	Hedges's *g*	*SE*
			D	0.31	0.30
23	A	T vs. C	A	−0.41	0.27
			B	0.08	0.40
	B	T vs. C	A	0.13	0.28
			B	0.63	0.24
	C	T vs. C	A	0.15	0.25
			B	0.27	0.40
	D	T vs. C	A	−0.01	0.27
			B	0.51	0.32
	E	T vs. C	A	0.08	0.21
			B	0.01	0.31
	F	T vs. C	A	0.11	0.23
			B	−0.10	0.48
	G	T vs. C	A	−0.07	0.23
			B	0.44	0.31
24		T1 vs. C1	A	−0.39	0.37
			B	−0.61	0.37
			C	−0.60	0.37
			D	−0.30	0.37
		T2 vs. C2	A	−0.83	0.40
			B	−0.48	0.39
			C	0.56	0.39
			D	0.53	0.39
25		T1 vs. C1	A	0.12	0.04
			B	0.00	0.04
		T2 vs. C2	A	0.05	0.04
			B	0.03	0.04
26		T vs. C	A	0.45	0.26
			B	0.36	0.26
			C	0.14	0.26
			D	0.09	0.26
			E	0.49	0.26
			F	0.33	0.26
27		T vs. C	A	0.81	0.31
			B	0.25	0.29
			C	0.56	0.30
			D	1.59	0.29
			E	0.49	0.26
			F	1.38	0.28
			G	1.39	0.31
			H	0.33	0.26
			I	0.98	0.28

Study	**Independent subgroup**	**Comparison**[a]	**Outcome**[b]	**Hedges's *g***	***SE***
28		T vs. C	A	0.04	0.36
			B	0.07	0.36
29	A	T vs. C	A	−0.01	0.18
			B	−0.11	0.18
			C	0.07	0.18
			D	0.11	0.18
			E	0.02	018
			F	−0.06	0.18
			G	−0.04	0.18
	B	T vs. C	A	0.08	0.20
			B	−0.05	0.20
			C	0.28	0.20
			D	0.14	0.20
			E	0.04	0.20
			F	0.35	0.20
			G	0.01	0.20
	C	T vs. C	A	0.00	0.21
			B	0.09	0.21
			C	−0.10	0.21
			D	−0.03	0.21
			E	0.17	0.21
			F	0.29	0.21
			G	0.14	0.21
	D	T vs. C	A	0.10	0.22
			B	0.02	0.22
			C	0.13	0.22
			D	0.15	0.22
			E	0.12	0.22
			F	0.12	0.22
			G	0.02	0.22
30		T vs. C	A	0.77	0.19
			B	0.99	0.19
			C	0.47	0.19
			D	0.22	0.18
			E	0.45	0.19
			F	0.12	0.18
31		T vs. C	A	0.09	0.12
			B	0.07	0.12
			C	0.24	0.12
			D	0.18	0.12
			E	0.13	0.12
			F	0.19	0.12

Study	Independent subgroup	Comparison[a]	Outcome[b]	Hedges's *g*	*SE*
			G	0.30	0.12
			H	0.15	0.12
			I	0.19	0.12
			J	0.22	0.12
			K	0.15	0.12
			L	0.14	0.12
32		T vs. C	A	0.04	0.20
			B	0.36	0.21
			C	0.03	0.20
			D	0.16	0.20
33		T1 vs. C	A	−0.06	0.10
			B	−0.04	0.10
			C	0.10	0.10
			D	−0.23	0.11
			E	−0.11	0.10
			F	0.06	0.10
			G	0.07	0.10
			H	−0.01	0.10
			I	0.01	0.10
			J	0.03	0.10
			K	−0.06	010
			L	0.06	0.10
		T2 vs. C	A	−0.16	0.15
			B	−0.03	0.15
			C	0.35	0.16
			D	−0.13	0.17
			E	−0.03	0.15
			F	0.19	0.16
			G	0.24	0.16
			H	0.08	0.16
			I	0.23	0.16
			J	0.09	0.16
			K	0.16	0.16
			L	0.15	0.16
34		T1 vs. C	A	0.09	0.19
			B	0.24	0.20
			C	0.65	0.20
			D	0.17	0.19
			E	0.33	0.20
			F	0.31	0.20
		T2 vs. C	A	0.03	0.19
			B	0.23	0.19

Study	Independent subgroup	Comparison[a]	Outcome[b]	Hedges's *g*	*SE*
			C	0.70	0.20
			D	0.39	0.20
			E	0.05	0.19
			F	0.29	0.19
35		T vs. C	A	0.36	0.18
			B	0.48	0.18
			C	0.05	0.18
			D	0.21	0.18
			E	−0.02	0.18
36		T1 vs. C	F	0.11	0.25
			G	0.04	0.25
			H	−0.31	0.25
			I	−0.05	0.25
			J	−0.10	0.25
		T2 vs. C	A	−0.17	0.25
			B	−0.23	0.27
			C	−0.38	0.25
			D	−0.11	0.25
			E	−0.18	0.25

Note. T = treatment; C = comparison.

[a]In studies with more than one treatment and/or comparison groups, groups are numbered.

[b]Letters represent the different outcome measures within studies.

References

Ahn S, Ames AJ, Myers ND. A review of meta-analyses in education: Methodological strengths and weaknesses. Review of Educational Research. 2012; 82:436–476.10.3102/0034654312458162

American Psychological Association. Reporting standards for research in psychology: Why do we need them? What might they be? American Psychologist. 2008; 63:839–851.10.1037/0003-066X.63.9.839 [PubMed: 19086746]

Becker, BJ. Multivariate meta-analysis. In: Tinsley, HA.; Brown, SD., editors. Handbook of applied multivariate statistics and mathematical modeling. San Diego, CA: Academic Press; 2000. p. 499-525.

Berkeley S, Scruggs TE, Mastropieri MA. Reading comprehension instruction for students with learning disabilities, 1995–2006: A meta-analysis. Remedial and Special Education. 2010; 31:423–436.10.1177/0741932509355988

Borenstein, M.; Hedges, LV.; Higgins, JPT.; Rothstein, HR. Complex data structures: Overview. In: Borenstein, M.; Hedges, LV.; Higgins, JPT.; Rothstein, HR., editors. Introduction to meta-analysis. Chichester, England: John Wiley; 2009a. p. 215-216.

Borenstein, M.; Hedges, LV.; Higgins, JPT.; Rothstein, HR. Multiple outcomes or time-points within a study. In: Borenstein, M.; Hedges, LV.; Higgins, JPT.; Rothstein, HR., editors. Introduction to meta-analysis. Chichester, England: John Wiley; 2009b. p. 225-238.

Borenstein, M.; Hedges, LV.; Higgins, JPT.; Rothstein, HR. Multiple comparisons within a study. In: Borenstein, M.; Hedges, LV.; Higgins, JPT.; Rothstein, HR., editors. Introduction to meta-analysis. Chichester, England: John Wiley; 2009c. p. 239-242.

Borenstein, M.; Hedges, LV.; Higgins, JPT.; Rothstein, HR. Comprehensive meta analysis. Englewood, NJ: Biostat; 2011. Version 2.2.064

Card, N. Applied meta-analysis for social science research. New York, NY: Guilford; 2012.

Chambers EA. An introduction to meta-analysis with articles from the *Journal of Educational Research* (1992–2002). Journal of Educational Research. 2004; 98:35–44.

Cheung MWL. Fixed-effects meta-analyses as multiple-group structural equation models. Structural Equation Modeling. 2010; 17:481–509.10.1080/10705511.2010.489367

Cheung MWL. Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. Psychological Methods. 2013 Advance online publication. 10.1037/a0032968

Cooper, H. Synthesizing research: A guide for literature reviews. 3rd. Thousand Oaks, CA: Sage; 1998.

D'Agostino J, Murphy J. A meta-analysis of Reading Recovery in United States schools. Educational Evaluation and Policy Analysis. 2004; 26:23–38.10.3102/01623737026001023

de Vibe M, Bjørndal A, Tipton E, Hammerstrøm K, Kowalski K. Mindfulness based stress reduction (MBSR) for improving health, quality of life, and social functioning in results. Campbell Systematic Reviews. 2012; 310.4073/csr.2012.3

Dieckmann NF, Malle BF, Bodner TE. An empirical assessment of meta-analytic practice. Review of General Psychology. 2009; 13:101–115.10.1037/a0015107

Edmonds MS, Vaughn S, Wexler J, Reutebuch CK, Cable A, Tackett KK. A synthesis of reading interventions and effects on reading outcomes for older struggling readers. Review of Educational Research. 2009; 79:262–300.10.3102/0034654308325998 [PubMed: 20072704]

Flynn LJ, Zheng X, Swanson HL. Instructing struggling older readers: A selective meta-analysis of intervention research. Learning Disabilities Research & Practice. 2012; 27:21–32.10.1111/j.1540-5826.2011.00347.x

Gersten R, Chard DJ, Jayanthi M, Baker SK, Morphy P, Flojo J. Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. Review of Educational Research. 2009; 79:1202–1242.10.3102/0034654309334431

Gleser, LJ.; Olkin, I. Stochastically dependent effect sizes. In: Cooper, H.; Hedges, LV., editors. The handbook of research synthesis. New York, NY: Russell Sage Foundation; 1994. p. 339-355.

Graham S, Hebert M. Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. Harvard Educational Review. 2011; 81:710–744.

Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. Journal of Education Statistics. 1981; 6:107–128.

Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. Research Synthesis Methods. 2010; 1:39–65.10.1002/jrsm.5

Institute of Education Sciences. What works clearinghouse study review standards. 2008. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_version1_standards.pdf

Kalaian HA, Raudenbush SW. A multivariate mixed linear model for meta-analysis. Psychological Methods. 1996; 1:227–235.10.1037/1082-989X.1.3.227

Kim R, Becker B. The degree of dependence between multiple-treatment effect sizes. Multivariate Behavioral Research. 2010; 45:213–238.10.1080/00273171003680104

Konstantopoulos S. Fixed effects and variance components estimation in three-level meta-analysis. Research Synthesis Methods. 2011; 2:61–76.10.1002/jrsm.35

Marín-Martínez F, Sánchez-Meca J. Averaging dependent effect sizes in meta-analysis: A cautionary note about procedures. Spanish Journal of Psychology. 1999; 2:32–38. [PubMed: 11757259]

Peabody Research Institute. (n.d.). Methods resources. Retrieved from http://peabody.vanderbilt.edu/research/pri/methods_resources.php

Raudenbush SW, Becker BJ, Kalaian H. Modeling multivariate effect sizes. Psychological Bulletin. 1988; 103:111–120.10.1037/0033-2909.103.1.111

Rosenthal R, Rubin DB. Meta-analytic procedures for combining studies with multiple effect sizes. Psychological Bulletin. 1986; 99:400–406.10.1037/0033-2909.99.3.400

Samson JE, Ojanen T, Hollo A. Social goals and youth aggression: Meta-analysis of prosocial and antisocial goals. Social Development. 2012; 21:645–666.10.1111/j.1467-9507.2012.00658.x

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Scammacca, N.; Roberts, G.; Vaughn, S.; Edmonds, M.; Wexler, J.; Reutebuch, CK.; Torgesen, JK. Reading interventions for adolescent struggling readers: A meta-analysis with implications for practice. Portsmouth, NH: RMC Research Corporation, Center on Instruction; 2007.

Scammacca N, Roberts G, Vaughn S, Stuebing K. A meta-analysis of interventions for struggling readers in Grades 4–12: 1980–2011. Journal of Learning Disabilities. in press.

Tanner-Smith, EE.; Tipton, E. Research Synthesis Methods. 2013. Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. Advance online publication

Tanner-Smith EE, Wilson SJ, Lipsey MW. The comparative effectiveness of outpatient treatment for adolescent substance abuse: A meta-analysis. Journal of Substance Abuse Treatment. 2013; 44:145–158.10.1016/j.jsat.2012.05.006 [PubMed: 22763198]

Tran L, Sanchez T, Arellano B, Swanson HL. A meta-analysis of the RTI literature for children at risk for reading disabilities. Journal of Learning Disabilities. 2011; 44:283–295.10.1177/0022219410378447 [PubMed: 21521870]

Uttal DH, Meadow NG, Tipton E, Hand LL, Alden AR, Warren C, Newcombe NS. The malleability of spatial skills: A meta-analysis of training studies. Psychological Bulletin. 2013; 139:352–402.10.1037/a0028446 [PubMed: 22663761]

Van den Noortgate W, López-López JA, Marín-Martínez F, Sánchez-Meca J. Three-level meta-analysis of dependent effect sizes. Behavior Research Methods. 2013; 45:576–594.10.3758/s13428-012-0261-6 [PubMed: 23055166]

Wilson SJ, Tanner-Smith EE, Lipsey MW, Steinka-Fry K, Morrison J. Dropout prevention and intervention programs: Effects on school completion and dropout among school-aged children and youth. Campbell Systematic Reviews. 2011; 810.4073/csr.2011.8

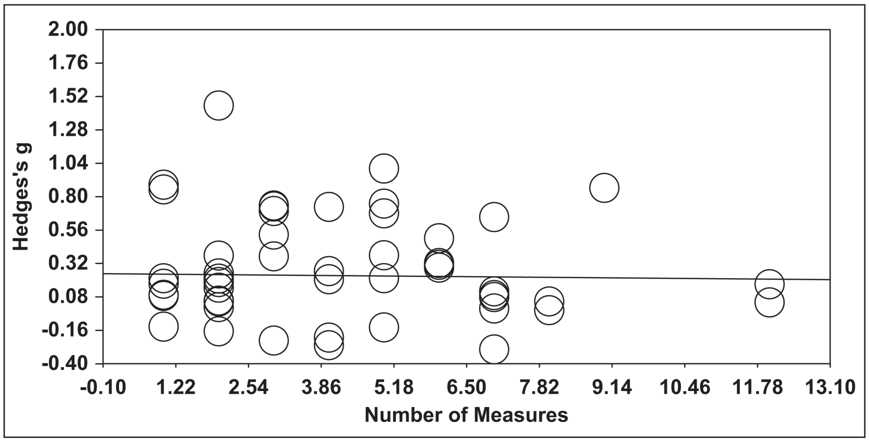

Figure 1.

Scatterplot of effect size by number of measures used in a study.

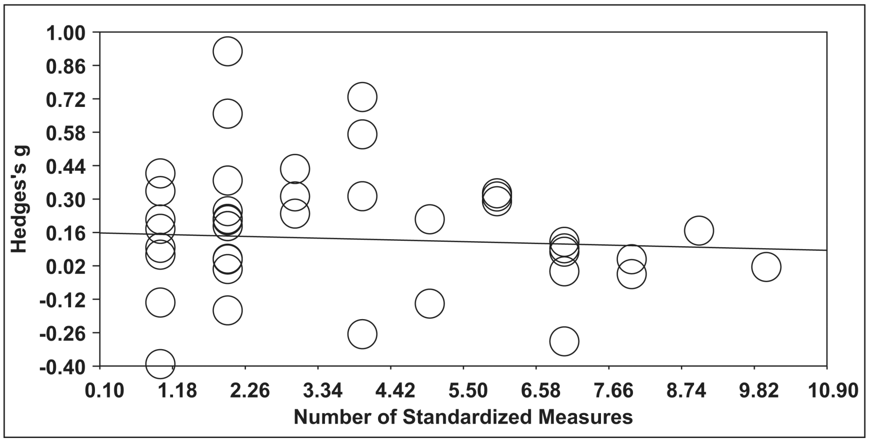

Figure 2.

Scatterplot of effect size by number of standardized measures used in a study.

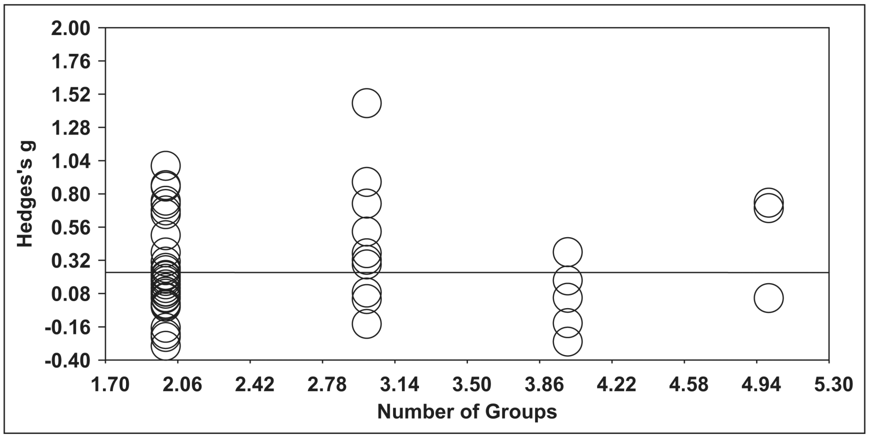

Figure 3.

Scatterplot of effect sizes by number of groups in a study using all types of measures.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

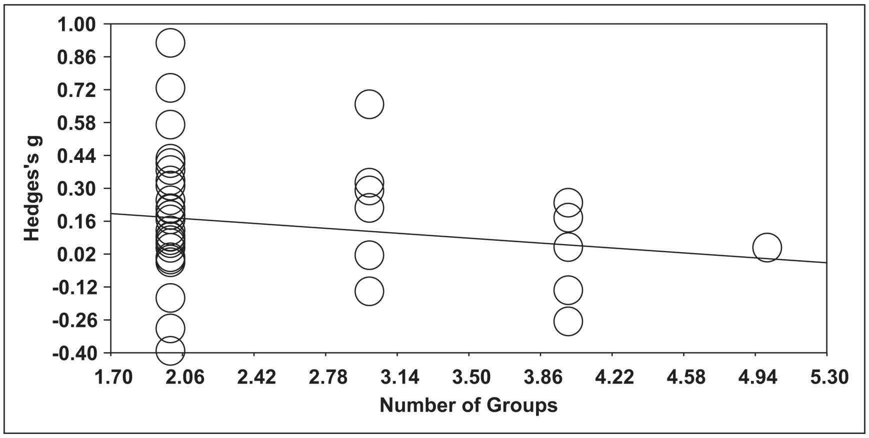

Figure 4.
Scatterplot of effect sizes by number of groups in a study using standardized measures only.

Table 1

Meta-analytic results for various methods of summarizing multiple measures within a study for all types of measures

Type	*k*	Mean ES	*SE*	Variance	95% CI	*P*	*Q*	*df*	*p* of *Q*	I^2	Tau^2
Mean of measures	50	0.23	0.04	0.002	0.15,0.31	<.001	98.53	49	<.001	50.27	0.03
Highest ES selected	50	0.53	0.07	0.005	0.39, 0.67	<.001	325.46	49	<.001	84.94	0.18
Random selection	50	0.37	0.07	0.004	0.24, 0.49	<.001	270.31	49	<.001	81.87	0.18
Select by research question	50	0.26	0.05	0.002	0.17,0.35	<.001	121.46	49	<.001	59.66	0.05
By skill: Fluency	22	0.21	0.13	0.017	−0.05, 0.47	.107	182.20	21	<.001	88.47	0.32
By skill: Reading comprehension	47	0.24	0.04	0.002	0.16,0.33	<.001	96.62	46	<.001	52.39	0.03
By skill: Vocabulary	8	0.06	0.13	0.020	−0.20, 0.32	.638	30.28	7	<.001	76.89	0.09
By skill: Word	19	0.22	0.06	0.004	0.10,0.34	<.001	27.62	18	.068	34.83	0.02
By skill: Word fluency	16	0.21	0.07	0.004	0.08, 0.33	.002	28.27	15	.02	46.94	0.03
By skill: Spelling	6	0.14	0.09	0.007	−0.03,0.31	.098	5.93	5	.314	15.61	0.01
All measures independent	50	0.24	0.04	0.001	0.17,0.32	<.001	363.04	49	<.001	86.50	0.05

Note. ES = effect size; CI = confidence interval.

Table 2

Meta-analytic results for various methods of summarizing multiple measures within a study for standardized measures

Type	*k*	Mean ES	*SE*	Variance	95% CI	*P*	*Q*	*df*	*p* of *Q*	I^2	Tau^2
Mean of measures	41	0.13	0.03	0.001	0.07,0.18	<.001	42.03	40	.383	4.82	0.00
Highest ES selected	41	0.30	0.04	0.002	0.22, 0.38	<.001	76.10	40	<.001	47.44	0.03
Random selection	41	0.19	0.04	0.002	0.11,0.28	<.001	71.71	40	.001	47.44	0.02
Select by research question	41	0.18	0.04	0.002	0.10,0.26	<.001	67.66	40	.004	40.88	0.02
By skill: Fluency	14	0.07	0.06	0.004	−0.05,0.19	.262	13.34	13	.422	2.58	0.00
By skill: Reading comprehension	39	0.19	0.04	0.002	0.11,0.27	<.001	66.26	38	.003	42.65	0.02
By skill: Vocabulary	4	0.01	0.06	0.004	−0.11,0.13	.885	3.63	3	.304	17.33	0.00
By skill: Word	18	0.19	0.05	0.003	0.10,0.29	<.001	19.01	17	.328	10.58	0.01
By skill: Word fluency	14	0.13	0.05	0.002	0.03, 0.22	.008	9.23	13	.755	0.00	0.00
By skill: Spelling	6	0.14	0.09	0.007	−0.03,0.31	.098	5.93	5	.314	15.61	0.01
All measures independent	41	0.15	0.03	0.001	0.10,0.21	<.001	126.43	40	<.001	68.36	0.02

Note. ES = effect size; CI = confidence interval.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 3

Meta-analytic results for various methods of summarizing multiple comparisons within a study for all types of measures

Type	*k*	Mean ES	*SE*	Variance	95% CI	*P*	*Q*	*df*	*p* of *Q*	I^2	Tau^2
Mean	50	0.23	0.04	0.002	0.15, 0.31	<.001	98.53	49	<.001	50.27	0.03
Highest ES selected	50	0.29	0.05	0.002	0.20, 0.38	<.001	126.78	49	<.001	61.35	0.05
Random selection	50	0.22	0.04	0.002	0.14, 0.30	<.001	107.52	49	<.001	54.43	0.04
Select by research question	50	0.20	0.04	0.001	0.13, 0.28	<.001	99.35	49	<.001	50.68	0.02
All comparisons independent	92	0.31	0.04	0.001	0.24, 0.38	<.001	265.09	91	<.001	65.67	0.06

Note. ES = effect size; CI = confidence interval.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 4

Meta-analytic results for various methods of summarizing multiple comparisons within a study for standardized measures

Type	*k*	Mean ES	*SE*	Variance	95% CI	*P*	*Q*	*df*	*p* of *Q*	I^2	*Pau*^{2}
Mean	41	0.13	0.03	0.001	0.07,0.18	<.001	42.03	40	.383	4.82	0.00
Highest ES selected	41	0.15	0.03	0.001	0.09,0.21	<.001	43.67	40	.381	8.40	0.00
Random selection	41	0.13	0.03	0.001	0.06,0.19	<.001	47.01	40	.207	14.92	0.01
Select by research question	41	0.14	0.03	0.001	0.08, 0.20	<.001	44.63	40	.283	10.38	0.00
All comparisons independent	61	0.12	0.02	0.001	0.07,0.17	<.001	69.13	60	.196	13.21	0.00

Note. ES = effect size; CI = confidence interval.

Table 5

Meta-analytic results for summarizing multiple outcomes and multiple comparisons within a study using robust variance estimation for all types of measures at different values of ρ

ρ	*k*	Mean ES	*SE*	Variance	95% CI	*P*	*Q*	*df*	*p* of *Q*	*Tau*^{2}
0.0	50	0.25172	0.04410	0.00194	0.16308,0.34035	<.001	187.71	49	<.001	0.09038
0.2	50	0.25173	0.04410	0.00194	0.16310,0.34037	<.001	187.71	49	<.001	0.09047
0.4	50	0.25175	0.04411	0.00194	0.16311,0.34039	<.001	187.71	49	<.001	0.09056
0.6	50	0.25177	0.04411	0.00194	0.16312,0.34041	<.001	187.71	49	<.001	0.09065
0.8	50	0.25178	0.04411	0.00194	0.16313,0.34044	<.001	187.71	49	<.001	0.09074
1.0	50	0.25180	0.04411	0.00194	0.16314,0.34046	<.001	187.71	49	<.001	0.09083

Note. ES = effect size; CI = confidence interval.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 6

Meta-analytic results for summarizing multiple outcomes and multiple comparisons within a study using robust variance estimation for standardized measures at different values of ρ

ρ	*k*	Mean ES	*SE*	Variance	95% CI	*P*	*Q*	*df*	*p* of *Q*	Tau^2
0.0	41	0.12706	0.02816	0.00079	0.07013,0.18399	<.001	58.39	40	.030	0.01264
0.2	41	0.12710	0.02817	0.00079	0.07016,0.18403	<.001	58.39	40	.030	0.01271
0.4	41	0.12713	0.02817	0.00079	0.07019,0.18408	<.001	58.39	40	.030	0.01279
0.6	41	0.12717	0.02817	0.00079	0.07021,0.18412	<.001	58.39	40	.030	0.01286
0.8	41	0.12720	0.02818	0.00079	0.07024, 0.18416	<.001	58.39	40	.030	0.01293
1.0	41	0.12724	0.02818	0.00079	0.07027, 0.18421	<.001	58.39	40	.030	0.01300

Note. ES = effect size; CI = confidence interval.