Abstract Title Page

Not included in page count.

Title: Meta-analyzing a complex correlational dataset: A case study using correlations that measure the relationship between parental involvement and academic achievement

Authors and Affiliations:

Joshua R. Polanin
Sandra Jo Wilson

Vanderbilt University

Abstract Body

Problem / Background / Context:
Practitioners and policymakers often rely on meta-analyses to inform decision-making around the allocation of resources and provision of interventions to students. (Petitti, 2000; Pigott, 2012). Indeed, high profile results from current reviews, such as Durlak et al.'s (2011) review of socio-emotional learning to support students' academic and emotional behavior, play a major role in the current zeitgeist of academic research.

Despite this encouraging utilization, meta-analytic results often remain simplistic and lack the statistical rigor and virtue that primary research analyses provide. A clear example of this phenomenon is the synthesis of correlations. Meta-analysis of correlation coefficients has the potential to provide strong evidence in support of theory required to begin a complex intervention. Often, however, the synthesis of correlation coefficients is confined to the analysis of bivariate relationships, independently synthesized without the consideration of covariance among other constructs or the inherent clustering of effect sizes within studies.

In order to deliver clear and appropriate theoretical understanding, the synthesis of correlations must advance beyond simple bivariate analysis; fortunately, recent advances in meta-analytic methodology make it possible. One relatively new procedure is to model the covariance structure through the use of meta-analytic structural equation modeling (MASEM; Cheung & Chan, 2005; Cheung, 2008). MASEM has all the benefits of traditional structural equation modeling, including the modeling of latent and observed variables simultaneously. The flexible analytical structure of MASEM can also handle the "general linear model, path analysis, and confirmatory factor analysis" as well (pg. 182). As a result, MASEM has the potential to answer much more complex questions relative to traditional synthesis of bivariate correlations (Hunter & Schmidt, 2004).

Another means to move beyond the relatively simplistic meta-analytic analyses is to embrace and extract as much information as possible from primary studies. Primary study authors often report, for example, multiple measures of constructs, across multiple samples, throughout multiple timeframes. It has been suggested in the past, however, to use only one effect size per study (Borenstein, Hedges, Rothstein, & Sutton, 2010), thus limiting the information available for synthesis. Indeed, it is no longer necessary to extract merely one effect size per study out of fear of the assumption of independence. Meta-analysts, instead, may utilize the analysis of multi-leveling modeling (Viechtbauer, 2010) or robust variance estimation (Hedges, Tipton, & Johnson, 2010) to handle the dependency among effect sizes within studies.

Despite these advances, practical and problematic questions continue to arise due to the inherent complexity of analyzing both multivariate and multilevel effect size data. Myriad problems exist, from missing data to dependent effect sizes, when attempting to model complex meta-analytic data. The purpose of this project, therefore, is to demonstrate the practical methods we have developed to utilize such a dataset.

The context for this project is a large-scale meta-analytic review of the predictors of academic achievement (Blinded for review, ****). We have collected over 700 primary studies that measured and reported longitudinal correlations of behavior, attitude, and school readiness predicting academic achievement. More than 150,000 bivariate correlations have been extracted.

A portion of these studies focus on the relationship between parental academic involvement and school performance (Hill & Taylor, 2004; Stevenson & Baker, 1987). Over the past few decades, conflicting research has shown that parental involvement is both positively (Catsambis, 2001) and negatively (Bronstein, Ginsberg, & Herrera, 2005) related to academic achievement. A widely-cited (see Epstein & Sanders, 2002) existing framework is to consider separately the influences of school-based involvement (i.e., communication between parents and teacher, parent-teacher relationships, etc.) or home-based involvement (i.e., parents helping students with homework, etc.). Developmental literature posits that parental involvement's influence on academic achievement may vary as a function of the child's age (Sing et al., 1995) or socioeconomic status (Duncan, 1994; Lent et al., 2000). Despite the burgeoning extant literature, questions remain as to the relationship between parental involvement and school performance. Figure 1 provides a graphical representation of our hypotheses.

Purpose / Objective / Research Question / Focus of Research:
This project is guided by three primary research questions:

1) How do we meta-analyze the heterogeneous and complex, multivariate, multilevel, and longitudinal correlational literature on parental involvement and academic achievement?
2) What are the associations between parental involvement and academic achievement, controlling for socioeconomic status and age?
3) Are there differential impacts on academic achievement as a function of the type of parental involvement (i.e., school-based versus home-based)?

Improvement Initiative / Intervention / Program / Practice:
The innovation of this project is the way in which we analyzed a complex meta-analytic dataset. The correlational literature on parental involvement and academic achievement is both heterogeneous and complex. The complexity derives from the myriad ways in which primary authors measure both parental involvement and then, subsequently, report the correlations. It is common practice, for example, for researchers to report multiple correlations for the same predictor using multiple outcomes. Moreover, the covariance structure among correlational effect sizes should be accounted for by the analysis technique. It is common practice, for example, to synthesize the bivariate correlations independently. Hill and Tyson's (2009) meta-analysis on the relationship between parental involvement and academic achievement synthesized 50 studies, for example, but the authors extracted only one effect size per study and then synthesized different constructs independently.

To move the field forward, both substantively and methodologically, we utilized methods of multilevel and structural equation meta-analysis. First, multilevel meta-analysis was utilized to handle the dependent nature of the correlations. Second, given a synthesized correlation *matrix* (as opposed to synthesized bivariate relationships), we utilized structural equation modelling for meta-analysis to analyze the relationships simultaneously, controlling for theoretically related constructs (i.e., socioeconomic status and age). We believe our technique to be the first of its kind to analyze multilevel and multivariate meta-analytic models.

Population / Participants / Subjects:
The population of interest, given that this project is a meta-analysis, is primary studies. The population of studies of interest for this project is part of an ongoing, larger meta-analysis (Blinded for review, ****).

Briefly, we included studies that:
1) Sampled school-aged students, from birth to 19 years old
2) Used a longitudinal, panel design, where the time between measurements was at least 6 months (unless the measurements were taken across subsequent semesters)
3) Measured parental involvement at wave 1 and school performance at wave 2

No restriction was placed on the type of student, school, or location. However, we excluded studies that used comparison designs. This included designs where authors create groups based on their membership in some outcome category (i.e., high achievement vs. low achievement) and compare the groups on earlier predictor status. In addition, intervention studies were excluded unless the control group students were identified and a longitudinal correlation was measured and extracted.

Research Design:
The research design of this project is a multilevel, multivariate meta-analysis. The primary studies were longitudinal panel designs, measuring students' parental involvement at wave 1 and school performance at wave 2. The following section will detail how we analyze and account for the complexity of these data.

Data Collection and Analysis:
Data Collection
Data collection has occurred over the course of the last several years. We searched and screened traditional and grey literature databases, including PsycInfo, ERIC, Digital Dissertations, as well as several conference websites and national study repositories. Experts in this field were also contacted.

The collected citations' abstracts and titles were independently screened for inclusion by several research assistants. The full text was then evaluated for inclusion. Trained research assistants coded included abstracts. All information was coded into a Filemaker Database (Apple Inc., 2013). From the larger database (described above), we selected studies that included a measure of parental involvement and academic achievement.

Analysis
The effect size of interest is the correlation coefficient (r). The majority of studies reported the bivariate correlation; however, a few studies reported means and standard deviations or cell frequencies. This information was used to estimate a bivariate correlation following the procedures outlined in Borenstein et al. (2009). The correlations were corrected for measurement problems such as dichotomization; large effect sizes were winsorized (Lipsey & Wilson, 2002).

In order to conduct our final analysis, using a structural equation model, we must first decide how to handle the complexities inherent in the dataset. Mainly, we have several levels of clustering to account for to ensure accurate effect sizes. In the past, meta-analysts had a few options available to model correlations in a structural equation. The first was to meta-analyze each correlation cell independently, then combining all cells into a correlation matrix. The problem with this approach is that it does not take into account the dependency among the correlations (the covariance of the correlations), thus overestimating the average effect sizes.

The second way to analyze correlation matrices is a relatively new phenomenon (Cheung, 2008). This process aggregates the correlation matrices simultaneously, accounting for the dependency among the correlations using the weighted least squares estimation. The problem with this approach, however, is how to handle primary studies that report more than one effect size per correlation constructs (i.e., two parental involvement measures correlated with reading achievement). Moreover, rarely will studies report the same correlation matrix structure (i.e., the same three predictors and same outcome) across multiple studies, thus engendering pervasive missing data.

To account for the inherent clustering, we first employed the use of a 4-level, random-effects model. Level-1, consistent with traditional meta-analysis, is the variance of the effect size estimate, assumed to be known in meta-analysis. The second level is the effect size level, where, for instance, multiple measures of home-based parental involvement and reading achievement are reported. The third level accounts for the correlational structure inherent within the study. Finally, the fourth level is the study. Table 1 provides an example of the dataset structure.

The results of this analysis provide a synthesized correlation matrix. To account for differences across the effect sizes and studies, we next run a meta-regression to control differences in attrition, age, and race. The meta-regression supplies an adjusted set of correlations, which we use to conduct the structural equation model. Because we are using a structural equation model, we can account for the covariance between the correlations, thus providing an estimate of the relationship between the different parental involvement measures and academic achievement while controlling for socioeconomic status.

Findings / Outcomes:
The results of this project are ongoing. We have 1,559 longitudinal correlation estimates (k = 85 studies) representing the relationship between parental involvement and academic achievement. Across all correlations, the simple weighted-mean effect size is positive and moderate ($\bar{r}$ = .21, 95% CI = .16, .26). We plan to report, however, a much more sophisticated and nuanced estimate of the effect sizes.

Conclusions:
Despite the prolific usage and reliability of meta-analytic results in policy decisions, the results from meta-analyses often lack sophistication. We plan to utilize a complex analysis to answer nuanced questions within the context of parental involvement and academic achievement. We believe this project has the potential to contribute substantially to the methodological literature on meta-analysis, and, moreover, to address plaguing issues in education research.

Appendices

Appendix A. References

Apple, Inc. (2013). Filemaker Pro (Version 13) [Software]. Santa Clara, CA: Apple, Inc.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. West Sussex, England: Wiley & Sons.

Bronstein, P., Ginsburg, G. S., & Herrera, I. S. (2005). Parental predictors of motivational orientation in early adolescence: A longitudinal study. *Journal of Youth and Adolescence*, *34*(6), 559-575.

Catsambis, S. (2001). Expanding knowledge of parental involvement in children's secondary education: Connections with high school seniors' academic success. *Social Psychology of Education*, *5*(2), 149-177.

Cheung, M. W. L., & Chan, W. (2005). Meta-analytic structural equation modeling: a two-stage approach. *Psychological Methods*, *10*(1), 40.

Cheung, M. W. L. (2008). A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. *Psychological methods*, *13*(3), 182.

Duncan, G. J. (1994). Families and neighbors as sources of disadvantage in the schooling decisions of white and black adolescents. *American Journal of Education*, 20-53.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405-432.

Epstein, JL., & Sanders, MG. (2002). Family, school, and community partnerships. In Bornstein, MH. (Eds.), *Handbook of parenting: Practical issues in parenting* (Volume 5) (p. 507-437). Mahwah, NJ: Erlbaum.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39-65.

Hill, N. E., & Taylor, L. C. (2004). Parental school involvement and children's academic achievement pragmatics and issues. *Current Directions in Psychological Science*, *13*(4), 161-164.

Hill, N. E., & Tyson, D. F. (2009). Parental involvement in middle school: a meta-analytic assessment of the strategies that promote achievement. *Developmental Psychology*, *45*(3), 740.

Hunter, H.E., & F. L. Schmidt. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Lent, R. W., Brown, S. D., & Hackett, G. (2000). Contextual supports and barriers to career choice: A social cognitive analysis. *Journal of Counseling Psychology*, *47*(1), 36.

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Petitti, D.B. (2000). *Meta-analysis, decision analysis, and cost effectiveness analysis.* New York, NY: Oxford University Press.

Pigott, T.D. (2012). *Advances in meta-analysis.* New York, NY: Springer.

Singh, K., Bickley, P. G., Trivette, P., & Keith, T. Z. (1995). The effects of four components of parental involvement on eighth-grade student achievement: Structural analysis of NELS-88 data. *School Psychology Review*, 24, 299-317.

Stevenson, D. L., & Baker, D. P. (1987). The family-school relation and the child's school performance. *Child Development*, 1348-1357.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1-48. URL http://www.jstatsoft.org/v36/i03/.

Appendix B. Tables and Figures

Table 1. Example of the Complex Meta-Analytic Dataset

Effect Size ID	Correlation ID	Study ID	Predictor	Correlation (Variance)
1	1	1	Home-based Parental Involvement	.12 (.03)
2	1	1	Home-based Parental Involvement	.22 (.05)
3	2	1	School-based Parental Involvement	.28 (.06)
4	2	1	School-based Parental Involvement	-.07 (.02)
5	1	2	Home-based Parental Involvement	.39 (.10)
6	2	2	School-based Parental Involvement	-.12 (.04)
7	2	2	School-based Parental Involvement	-.11 (.06)

Notes: Data are for presentation purposes only; Level-1 = Variance component; Level-2 = Effect size; Level-3 = Correlation pair; Level-4 = Study.

Figure 1. Meta-Analytic Path Coefficient Model