

# Empirical Application

To illustrate the new beta-function selection model in an empirical application, we reanalyzed the data from a meta-analysis conducted by BSCS Science Learning. BSCS Science Learning conducted a meta-analysis of science interventions effects on K–12 student science achievement (Taylor et al. 2018). The meta-analytic sample included findings from 96 studies with 292 effects. Approximately half of the studies contributed multiple effects per study, due to either multiple samples or multiple outcomes, leading to dependent effect sizes. Effect sizes were measured as standardized mean differences representing the effects of the interventions on student achievement outcomes. We conducted the analyses using R Version 4.4.3 (R Core Team 2023).

We used the `rma.uni()` function from the `metafor` package to fit a standard random-effects model based on an independent effects working model with random effects for each study (Viechtbauer 2010). To adjust for selective reporting, we used the `selection_model()` function from the `metaselection` package to fit the new beta-function selection model along with our recently developed step-function selection model (Pustejovsky and Joshi 2025). Both selection models are based on an independent effects working model with random effects for each study and account for dependent effects using RVE. For the step-function model, we estimated both a single-step model with a threshold at  $\alpha_1 = 0.025$  and a two-step model with thresholds at  $\alpha_1 = 0.025$  and  $\alpha_2 = 0.5$ .

The overall estimate of the average effect using the unadjusted model is 0.44, 95% CI [0.384, 0.496], which is significantly different from zero ( $p = 0$ ). The estimates from the selection models are also all positive and significant, but they vary in magnitude and thus their adjustment for selective reporting. Compared to the unadjusted model, the average adjusted effect is 25% small when using the beta-function selection model ( $ES = 0.33$ ), 32% larger when using the single-step model ( $ES = 0.582$ ), and 12% larger when using the two-step selection model ( $ES = 0.492$ ).

A key advantage of selection models over other methods for diagnosing and adjusting for selective reporting bias is their ability to include both discrete and continuous moderators. This allows meta-analysts to distinguish selective reporting bias from systematic variation in effect sizes that can be explained by primary study characteristics. The BSCS Science Learning meta-analytic data includes a number of potential moderators of effect size. Among these is an indicator for assessment author type, defined as 1 for assessment developer is the author or researcher and 0 otherwise. We reran all four models using assessment author type as a moderator (Taylor et al. 2018).

When the assessment developer is the author or researcher, the unadjusted estimate is 0.087, 95% CI [-0.039, 0.214], which is not significantly different from zero ( $p = 0.177$ ). The estimates from the selection models are also all nonsignificant and ranged from 0.096 to 0.14. However, when the assessment developer is not the author or researcher, the models lead to different conclusions. The estimate from the unadjusted model is 0.376 and significant (95% CI [0.267, 0.485],  $p = 0$ ). The estimate from the beta-function selection model is 0.21 and nonsignificant ( $p = 0.387$ ). The estimates from the step-function selection models are both significant, with estimates of 0.511 for the single-step selection model and 0.401 for the two-step selection model.

The model estimates reveal some key differences between estimation methods. In this dataset, the beta-function selection model adjusts the estimates downward—so much so in the moderator analysis that neither estimate remains statistically significant. The step-function selection model, on the other hand, adjusts the estimates upward, though the adjustment is minimal when using the two-step model. These differences suggest that the models may behave differently in different conditions. To investigate this, we conducted simulations across a wide range of conditions. The simulations will allow us to draw firmer conclusions about the comparative performance of these methods, including their robustness to misspecification of the selection function.

- Pustejovsky, James, and Megha Joshi. 2025. *Metaselection: Meta-Analytic Selection Models with Cluster-Robust and Cluster-Bootstrap Standard Errors for Dependent Effect Size Estimates*. <https://github.com/jepusto/metaselection>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Taylor, Joseph A., Susan M. Kowalski, Joshua R. Polanin, Karen Askinas, Molly A. M. Stuhlsatz, Christopher D. Wilson, Elizabeth Tipton, and Sandra J. Wilson. 2018. “Investigating Science Education Effect Sizes: Implications for Power Analyses and Programmatic Decisions.” *AERA Open* 4 (3): 1–19. <https://doi.org/10.1177/2332858418791991>.
- Viechtbauer, Wolfgang. 2010. “Conducting meta-analyses in R with the metafor package.” *Journal of Statistical Software* 36 (3): 1–48.