

Effect Sizes Larger in Developer-Commissioned Studies than in Independent Studies

Rebecca Wolf, Jennifer Morrison, Amanda Inns, Robert Slavin, and Kelsey Risman

Rigorous evidence of program effectiveness has become increasingly important with the 2015 passage of the Every Student Succeeds Act (ESSA). One question that has not yet been addressed is whether findings from program evaluations carried out or commissioned by developers are as trustworthy as those identified in studies by independent third parties. Using study data from the What Works Clearinghouse, we found evidence of a “developer effect,” where program evaluations carried out or commissioned by developers produced average effect sizes that were substantially larger than those identified in evaluations conducted by independent parties.

Why is it important to accurately determine the effect sizes of an educational program?

Policymakers may select educational programs based on research that the program “works” and results in meaningful improvements in student outcomes. Because decisions to select one program over another are relative, to the extent that studies overstate the true effect sizes, a less effective program may be selected instead of a more effective one.

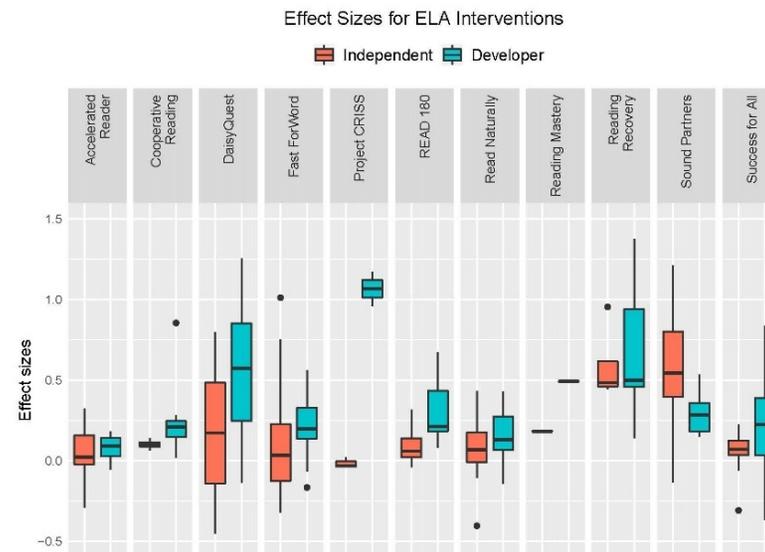
Did developer-commissioned studies produce larger effect sizes than independent studies?

When looking within the same program, developer-commissioned studies—developers conducted or funded the study—produced average effect sizes that were 1.7 times greater than those in independent studies. This is a big difference. The study used advanced meta-analytic techniques and also controlled for study and program factors that could influence the effect sizes.

Why would there be a difference?

The evidence is not clear about why developer-funded studies might be more susceptible to bias than others, but we found some evidence that developers are more likely than independent researchers to “bury” disappointing findings. If an independent researcher conducts a study, it is likely to be funded by a government or foundation

funder, who will expect a report, even if the results are disappointing. In contrast, a developer who commissions a study with disappointing results may choose to not release the report. Developers could also influence whether only the most positive findings are included in the report, and the null or negative ones withheld.



Note: Box plots of (unadjusted) effect sizes for English language arts (ELA) educational programs by independent versus developer study

How should we proceed?

These findings beg the question of whether we should trust results from “developer” studies to the same extent we trust results from independent studies. We encourage policymakers, practitioners, and educational researchers to pay more attention to contextual factors that may influence effect sizes, such as who conducted or paid for the evaluation. We also advocate for preregistration of program evaluations in education, which may mitigate any bias resulting from selective reporting of only the best outcomes.

Full Article Citation:

Wolf, R., Morrison, J., Inns, A., Slavin, R., & Risman, K. (2020). [Average effect sizes in developer-commissioned and independent evaluations](#). *Journal of Research on Educational Effectiveness*, 13(2), 428–447.