Do Developer-Commissioned Evaluations Inflate Effect Sizes?

Wolf, R., Morrison, J., Slavin, R., & Risman, K.
Center for Research and Reform in Education, Johns Hopkins School of Education

Introduction

Rigorous evidence of program effectiveness has become increasingly important with the recent passage of the Every Student Succeeds Act (ESSA).

One question that has not yet been fully explored is whether program evaluations carried out or commissioned by developers produce larger effect sizes than studies by independent third parties.

Using study data from the What Works Clearinghouse (WWC), we find evidence of a "developer effect," where studies carried out or commissioned by developers produced average effect sizes that were substantially larger than those identified in independent studies. We explore reasons for this phenomenon.

Methods

DATA

- We used study data from the WWC database in the areas of K–12 mathematics and literacy. Studies met WWC standards.
- Subgroup analyses were excluded.
- Sample included 755 findings in 169 studies.

VARIABLES

"Developer" studies are those for which the program developer either authored or funded the study. Otherwise, studies were coded as "independent."

We controlled for factors thought to influence effect sizes:

- Outcome measure type (e.g., researcher-made v. independent)
- Sample size
- Non-experimental v. experimental design
- Program characteristics
- Grade levels
- Academic subjects
- Study year

ANALYSIS

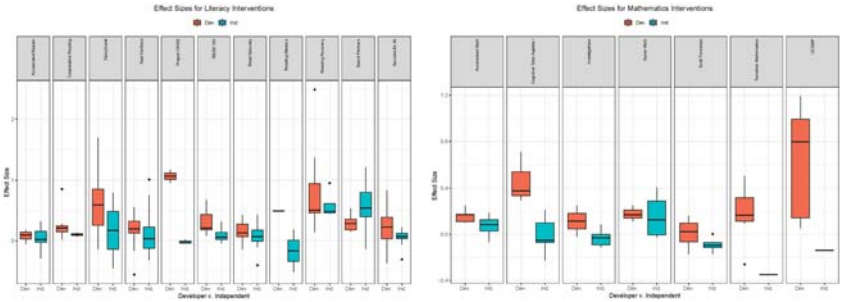
- Meta-regression model with robust variance estimation
- Used *robumeta* package in R
- Assumed findings within the same study were correlated with one another
- Controlled for factors known to influence effect sizes (see above) and intervention fixed effects

Results

	Model 1: Baseline + grade level, subject, study year	Model 2: Model 1 + developer effect	Model 3: Model 2 + factors known to influence effect sizes	Model 4: Model 3 + intervention fixed effects
Intercept	0.239*** (0.021)	0.173*** (0.023)	0.212*** (0.025)	0.194*** (0.039)
"Developer" effect		0.134*** (0.036)	0.117** (0.034)	0.156* (0.060)

FINDING 1: Systematic differences across developer and independent studies explained some of the developer effect, yet a developer effect of +0.117 persisted even when controlling for these factors.

FINDING 2: Even when adding fixed effects for common interventions to the model, we still found a developer effect of +0.156. The graphs to the right descriptively show the range of effect sizes for these interventions across developer and independent studies.



FINDING 3: The developer effect is likely attributable to both publication bias and decisions made in data cleaning and analysis. We found greater evidence of publication bias in developer studies.

	Developer (%)	Independent (%)
Quasi-experimental design	51%	15%
Researcher-made measure	29%	8%
Grades K–2	52%	40%
Grades 3–5	27%	40%
	Mean	**Mean**
Student N	392	659
Cluster N	12	26
Study year	2003	2005

Note. There were no statistical differences in program characteristics, academic subjects, or other grade levels.

Implications

- Effect sizes in developer-commissioned studies are inexplicably greater than those in independent studies, on average.
- We need to pay attention to who conducts and funds the research. Research funded by non-developer sources may be more informative.
- A national registry of studies may help reduce publication bias.

Questions?

Correspondence for this poster should be addressed to Betsy Wolf at BetsyWolf@jhu.edu.