

Title: Effects of Intervention Duration on Outcomes of Experiments in Education

Authors and Affiliations: Marta Pellegrini, University of Florence, Italy
(marta.pellegrini@unifi.it)

Amanda Inns, Johns Hopkins University
(ainns1@jhu.edu)

Cynthia Lake, Johns Hopkins University
(clake5@jhu.edu)

Robert E. Slavin, Johns Hopkins University
(rslavin@jhu.edu)

Background / Context

In 2015, the U.S. Congress passed the Every Student Succeeds Act (ESSA), which for the first time defines in law what it means for educational programs to have evidence of effectiveness. It distinguishes between three levels of evidence: “strong”, which requires at least one randomized study with at least one positive outcome; “moderate”, which requires at least one quasi-experimental or matched study with a positive outcome; and “promising”, which requires at least one correlational study with a positive outcome. This policy change makes it particularly important that researchers discover methodological factors that may affect outcomes of experimental evaluations.

Cheung and Slavin (2016), Pellegrini, Inns, Lake, and Slavin (2019), de Boer, Donker, and van der Werf (2014), Wolf and Slavin (2018), and other researchers have reported evidence on the effects of design issues on conclusions of meta-analyses, such as research design (randomized vs. quasi-experiment) and outcome measure (independent vs. researcher-made). One important design element evaluated by de Boer et al. (2014) is the intervention duration. The study examined 58 studies, 95 interventions and 180 outcomes included in a previous meta-analysis on the effects of learning strategy instruction on student achievement (Donker et al., 2014). Meta-regression analysis reported a small effect of intervention duration indicating that briefer interventions have higher effect sizes than longer interventions. Previous studies have found similar findings showing a small effect of duration on student mathematics performance (Dignath & Büttner, 2008) or no significant effects (Bangert-Drowns et al., 2004; Chiu, 1998).

Purpose / Objective / Research Question

The purpose of this study is to determine the effect of intervention duration on outcomes of experiments using a large sample of studies included in the What Works Clearinghouse reviews on reading and mathematics (K-12). We controlled for other variables such as design features and subject areas.

Data Collection

The data were obtained from the WWC Individual Study Database file, a database that contains information on all studies included in WWC reviews. Only studies eligible for WWC standards (WWC, 2017) in the areas of elementary and secondary reading/literacy and mathematics (K–12)¹ were included in the sample. We further excluded subgroup analyses, such as ELL students or students with disabilities. Five studies had at least one outcome that was missing, and these were dropped from the sample. The final sample consisted of 671 outcomes from 171 studies.

The WWC database includes many variables, collected and reported by WWC reviewers. Of relevance to this analysis were the type of outcome measure (independent measure or one created by the program developer or researchers), the research design (either randomized or quasi-experimental), and the effect size. Additional factors were coded, including the level of assignment (student or cluster), publication date (either before or after 2004), and subject (mathematics or reading/literacy). Intervention duration was coded in weeks, taking the information from the What Works Clearinghouse intervention reports or, if needed, from the original studies.

Data analysis

We used a meta-regression model with robust variance estimation to conduct the meta-analysis (Hedges, Tipton, & Johnson, 2010), because in the 171 studies included in this analysis, there were an average of 3.92 outcomes per study ($SD = 3.64$). The use of robust variance estimation (RVE) allows the inclusion of multiple effect sizes per study, which differs from a traditional meta-analysis where there can only be a single effect size included for each study (Hedges, Tipton, & Johnson, 2010; Tipton, 2015). The packages *metafor* and *ClubSandwich* were used to estimate all random-effects models with RVE (Pustejovsky & Tipton, 2018; Viechtbauer, 2010).

¹ The WWC data were extracted in December 2017 and updated in January 2019. These data included studies in the elementary school math, middle school math, high school math, primary math, secondary math, beginning reading, and adolescent literacy protocols.

The first step was to estimate the variance for each outcome, adjusting for cluster-assigned studies as described by Hedges (2007). Those variances were used to weight the studies using an inverse variance procedure. The meta-regression model to estimate the effect of duration controlled for research design, type of measure, subject area, publication date, and level of assignment.

Preliminary Findings

The study included 671 findings from 171 studies (see Table 1). Intervention duration varied between 15 days and 5 years, with few very brief studies (34 studies with less than 1 month of intervention) and with many studies with more than two years of intervention ($k = 104$).

The analysis revealed no effect of intervention duration ($\beta = -.00, p = .44$) when controlling for the other variables (Table 2). Further analysis found a quite strong correlation between duration and sample size ($r = -.42$), level of assignment ($r = -.43$), and research design ($r = -.40$). These results showed that briefer studies usually involved a small sample, and used student level assignments and quasi-experimental designs more often than longer studies. These correlations could explain why it an effect of duration was not detected when controlling for these variables.

Discussion

Methodological features are important elements to be studied to know which factors affect effect sizes in meta-analyses in education. This study's findings revealed that intervention duration has no effect controlling for other variables that affect effect sizes, such as measure type and research design. However, we found strong correlations between intervention duration and other factors, which make it hard to separate the effect of duration from the effect of measure type, research design, and level of assignment. Despite these results, intervention duration is an important attribute in program evaluations because briefer studies provide results hard to replicate in real school settings, where interventions are used over many months or years. Longer interventions, in addition to using higher methodological standards, as shown from the correlation coefficients, are more likely to be replicated in school over extended periods.

References

- Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, 74(4), 29–58. doi:10.3102/00346543074001029
- Cheung, A., & Slavin, R. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292. <https://doi.org/10.3102/0013189X16656615>
- Chiu, C. W. T. (1998, April). *Synthesizing metacognitive interventions: What training characteristics can improve reading performance?* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students: A meta-analysis on intervention studies at primary and secondary school level. *Metacognition Learning*, 3, 231–264. doi:10.1007/s11409-008-9029-x
- Donker, A. S., de Boer, H., Kostons, D., Dignath-van Ewijk, C. C., & van der Werf, M. P. C. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review*, 11, 1–26. doi:10.1016/j.edurev.2013.11.002
- de Boer, H., Donker, A. S., & van der Werf, M. P. (2014). Effects of the attributes of educational interventions on students' academic performance: A meta-analysis. *Review of Educational Research*, 84(4), 509-545. doi:10.3102/0034654314540006
- Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, 32(4), 341–370. doi:10.3102/1076998606298043
- 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.
- Pellegrini, M., Inns, A., Lake, C., & Slavin, R. E. (2019). *Effects of researcher-made vs. independent measures on outcomes of experiments in education*. Paper presented at the Annual Meeting of the Society for Research on Educational Effectiveness (SREE), Washington, DC.
- Pustejovsky, J. E., & Tipton, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, 36(4), 672-683.
- Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, 20(3), 375-393.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48.
- Wolf, R., & Slavin, R. E. (2018). *Do developer-involved evaluations inflate effect sizes among studies accepted by the What Works Clearinghouse?* Manuscript submitted for publication.

Tables

Table 1. Study characteristics.

<i>Characteristics</i>		
Total Studies		171
Total Findings		671
<i>Continuous Variables</i>		<i>Mean (SD)</i>
Findings per study		3.92 (3.64)
Sample size		746.65 (1666.37)
<i>Categorical Variables</i>		<i>N (%)</i>
Subject		
	Math	45 (26.3)
	Reading/literacy	126 (73.7)
Date of Publication		
	2005-2016	100 (58.5)
	1984-2004	71 (41.5)
Research Design		
	QED	62 (36.3)
	RCT	109 (63.7)
Level of Assignment		
	Cluster-Assigned	86 (50.3)
	Student-Assigned	85 (49.7)
Measure Type		
	Independent	549
	Researcher-made	122

Table 2. Results of meta-regression.

Coefficient	beta	SE	tstat	df	p_Satt
Intercept	0.344	0.061	5.641	54.742	0.000
Duration	-0.000	0.000	-0.797	15.244	0.438
Researcher measures	-0.198	0.053	-3.752	32.766	0.001
Reading	0.055	0.033	1.655	47.826	0.104
Pre-2005	0.100	0.037	2.689	68.579	0.009
Matched	-0.013	0.031	-0.426	57.586	0.671
Student level analysis	0.003	0.038	0.083	15.569	0.935