TECHNICAL COMMENT

EDUCATION

# Comment on "Math at home adds up to achievement in school"

Michael C. Frank*

**Berkowitz *et al*. (Reports, 9 October 2015, p. 196) described a randomized field experiment testing whether a math app designed to increase parent-child interaction could also bring academic benefits. A reanalysis of the data suggests that this well-designed trial failed to find strong evidence for the efficacy of the intervention. In particular, there was no significant effect of the intervention on math performance.**

Can electronic apps increase parent-child interaction around academic subjects like math and in turn help improve children's school outcomes? Berkowitz *et al.* (*1*) reported a randomized field experiment testing this hypothesis. Children were randomly assigned to math and reading app groups, and their learning outcomes were reassessed at the end of the school year. The study had a strong design, including a large sample size, objective measures of app usage, standardized outcome measures, and a well-matched control group. Unfortunately, a reanalysis of Berkowitz *et al.*'s data—which they provided as part of their Report, in a commendable show of open practices—suggests that their results provide limited support for the effectiveness of the intervention.

First, the intervention resulted in no significant improvement in math performance for the experimental group compared with the control group (Fig. 1). A longitudinal mixed-effects regression predicting math performance as a function of condition, time, and their interaction (including random intercepts for each student and classroom and random slopes for each classroom) (*2*) showed no significant condition-by-time interaction for either grade-level equivalent scores or raw scores ($\beta_{GE}$ = 0.07, SE = 0.10, $t$ = 0.73, $P$ = 0.47; $\beta_{raw\text{-}W}$ = 1.00, SE = 1.74, $t$ = 0.58, $P$ = 0.56). Thus, the intervention was not successful overall.

Second, although greater use of the math app by families was related to greater growth in children's math performance, causal interpretation of this result is difficult because of endogeneity issues. Parents who used the app more might also have greater interest in math learning more generally, and both children's math gains and their app usage could plausibly reflect those underlying differences (rather than gains being a causal effect of random assignment). Further, this relationship did not reach statistical significance in all analyses. For example, when app usage is added to the longitudinal analysis described above, the three-way app usage by condition-by-time interaction was significant when computed on grade-level equivalent scores but only marginal using raw scores ($\beta_{GE}$ = 0.14, SE = 0.06, $t$ = 2.19, $P$ = 0.03; $\beta_{raw\text{-}W}$ = 1.87, SE = 1.11, $t$ = 1.68, $P$ = 0.09).

Third, the authors report that the math app intervention was especially effective for those children

Department of Psychology, Stanford University, Stanford, CA, USA.
*Corresponding author. E-mail: mcfrank@stanford.edu

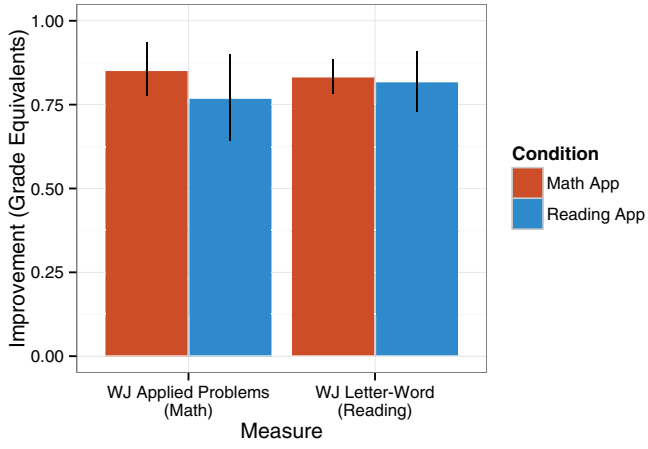


**Fig. 1. Student improvement (grade-level equivalents), plotted by intervention condition and measure.** Error bars show 95% confidence intervals computed by nonparametric bootstrap. The groups did not differ significantly on either measure in simple pairwise tests: $t(498)$ = 1.06, $P$ = 0.29 for math; $t(507)$ = 0.27, $P$ = 0.79 for reading.

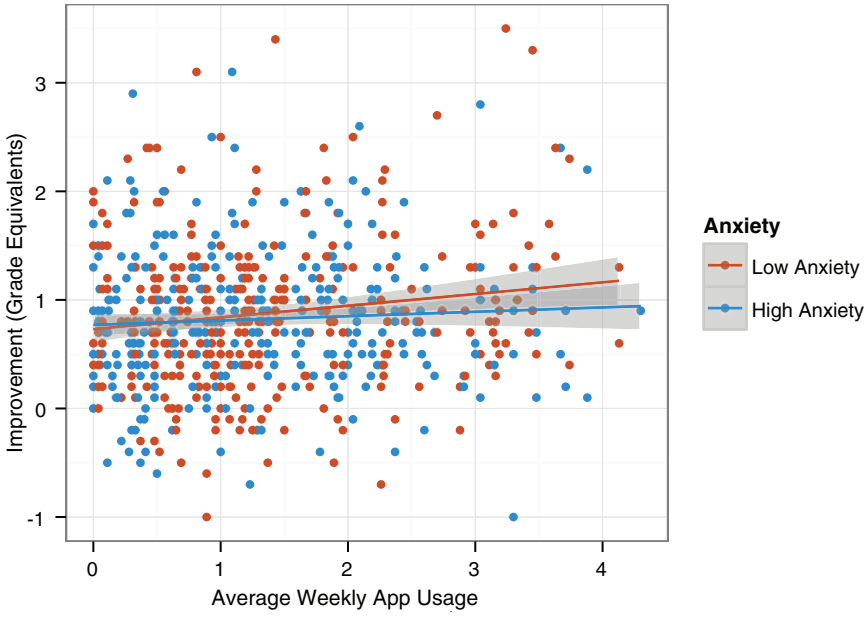


**Fig. 2. Student math improvement (grade-level equivalents) for children in the math intervention condition, plotted by average weekly app usage and split by parent math anxiety.** Dots show individual students, lines show the best fitting linear trend, and shaded areas show 95% confidence intervals.

whose parents were anxious about math, and especially for moderate app users. These descriptive claims are not supported by the data, however (Fig. 2). On average, children of highly math-anxious parents tended numerically to learn slightly less in both the intervention and control conditions. In addition, a statistical test of the claim that an intervention is especially effective for a subgroup requires a test for an interaction rather than a comparison of separate significance tests, the analysis that was performed in the original report (*3*, *4*). But when math anxiety is added to the basic longitudinal model described above, the three-way interaction (time by condition by math anxiety) was not significant for either measure ($t$s ≤ 1.28, $P$s ≥ 0.20). There were additionally no four-way interactions (time by condition by math anxiety by app usage) when app usage was added ($t$s ≤ 0.82, $P$s ≥ 0.41).

Finally, although some analyses of the original data do show statistically significant support for aspects of the intervention, these analyses rely on a variety of decisions that were not specified a priori. Hence, the findings run a heightened risk of being false positives (*5*, *6*). These decisions include (i) the discretization of continuous variables into two (math anxiety) or three (app usage) categorical bins, a practice that is also known to reduce statistical power (*7*), and (ii) the specification of primary analytic models to subsets of the data rather than the full data set (e.g., only the high-anxiety group or only a matched subset of families).

Although the authors may not have chosen to report statistical tests on the basis of the tests having produced significant results, their analysis strategy was nevertheless data-dependent. Consider the scenario in which the intervention as a whole had yielded a positive effect; in that case, the simple analysis in Fig. 1 would almost certainly have been a centerpiece of the report. This analytic situation, known as the "garden of forking paths," leads to an inflation of type 1 error, just as if analyses were actively selected (*6*). Preregistration of analytic hypotheses before data collection is currently the strongest method for protecting against this problem.

In sum, Berkowitz *et al*. report a well-designed study that shows at best weak support for an app-based intervention. This result, although disappointing, is nevertheless extremely informative for parents and policy-makers interested in the potential of app-based interventions.

## REFERENCES AND NOTES

1. T. Berkowitz *et al*., *Science* **350**, 196–198 (2015).
2. D. J. Barr, R. Levy, C. Scheepers, H. J. Tily, *J. Mem. Lang.* **68**, 255–278 (2013).
3. S. Nieuwenhuis, B. U. Forstmann, E. J. Wagenmakers, *Nat. Neurosci.* **14**, 1105–1107 (2011).
4. A. Gelman, H. Stern, *Am. Stat.* **60**, 328–331 (2006).
5. E. J. Wagenmakers, R. Wetzels, D. Borsboom, H. L. J. van der Maas, R. A. Kievit, *Pers. Psychol. Sci.* **7**, 632–638 (2012).
6. A. Gelman, E. Loken, *Am. Sci.* **102**, 460 (2014).
7. P. Royston, D. G. Altman, W. Sauerbrei, *Stat. Med.* **25**, 127–141 (2006).

## ACKNOWLEDGMENTS

The supplementary materials contain additional data and computer code. Thanks to Berkowitz *et al*. for posting raw data and providing feedback on a draft of this reanalysis. Thanks also to J. Haushofer and the members of the Language and Cognition laboratory at Stanford for valuable feedback.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/351/6278/aad8008/suppl/DC1
Methods
Data File

3 November 2015; accepted 21 January 2016
10.1126/science.aad8008

# Comment on "Math at home adds up to achievement in school"

Michael C. Frank

*Science* **351**, 1161 (2016);
DOI: 10.1126/science.aad8008