TECHNICAL RESPONSE

EDUCATION

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

Talia Berkowitz, Marjorie W. Schaeffer, Christopher S. Rozek, Erin A. Maloney, Susan C. Levine,* Sian L. Beilock*

Frank presents an alternative interpretation of our data, yet reports largely similar results to those in our original Report. A critical difference centers on how to interpret and test interaction effects. Frank finds no mistakes in our analyses. We stand by our original conclusions of meaningful effects of the Bedtime Learning Together (BLT) math app on children's math achievement.

The Bedtime Learning Together (BLT) math app provides a structured way for parents to increase math engagement with their children. Families were randomized at the classroom level to receive either a math or reading app. We hypothesized that the math app would increase children's math learning over the school year. Logically, we only predicted that the math app would improve students' math achievement if used. Based on our previous research showing that (i) high-math-anxious parents and teachers negatively affect children's math achievement (1, 2), (ii) people with high anxiety avoid math (3), and (iii) talking about math with children enhances math achievement (4–8), we also hypothesized that the math app would be most effective for children with high-math-anxious parents. The data support our hypotheses.

Frank raises issues with our data analysis (9). He argues that some of our analyses are "data-dependent"—that is, driven by our obtained results rather than by theory. This claim is unfounded because our previous work in this area directly leads to the a priori hypotheses that we posited.

Frank then goes on to reanalyze our data using tests that are less statistically powerful, and in our view less appropriate, than in the original Report and finds weaker yet similar results. We have a philosophical disagreement with Frank about the most appropriate analyses to employ. As we lay out below and in the original Report, we view our analyses as most appropriate for informing our hypotheses.

In our analyses, we wanted to assess whether the math app was specifically impactful for children of high-math-anxious parents. Thus, we split children into groups based on their parents' math anxiety levels. Our intent-to-treat (ITT) analysis

confirms a significant effect of app group, with children in the math group outperforming those in the reading group when parents are high-math-anxious [$\hat{\beta}_{21} = 5.25, t = 1.99, P = 0.048$] but not when parents are low-math-anxious [$\hat{\beta}_{31} = -0.61, t = -0.27, P = 0.79$]. Even though comparing the coefficients of intent-to-treat effects for children of high- and low-math-anxious parents (i.e., the interaction between parent-math-anxiety and app condition) is a low-powered test, especially in field trials (10, 11), we see a marginally significant

difference between the coefficients estimating the effects of the math app for these two groups of children at $P = 0.06$ (significant at $P = 0.03$ using a one-tailed test given our a priori hypothesis). In field trials like ours, not testing for pre-planned contrasts because of a nonsignificant interaction term can lead to unnecessarily missing heterogeneous treatment effects and, as a result, to misleading conclusions, both practically and theoretically (10, 11).

The ITT analysis underestimates the effect of using the app because it does not take app use into account (12). We can obtain an unbiased estimate of the effect of app use by using randomization as an instrumental variable (IV) to identify the effect of actual app use (12). We conducted an IV analysis on end-of-year math achievement (controlling for beginning-of-year math achievement), which estimates the effect of dosage on those whose dosage was induced by randomization. As in the ITT analysis, we obtained a significant effect of math app use on children of high-math-anxious parents (13) (Model S6). This analysis suggests that there is a causal effect of math app usage on end-of-year math achievement for the children of high-math-anxious parents and that this effect is negligibly influenced by selection bias, in contrast to Frank's speculation.

In our other main analysis, we show that the more times parents and their children used the app, the higher children's math achievement at school-year's end (controlling for beginning-of-year math achievement), but only for children in the math group—a group by use interaction
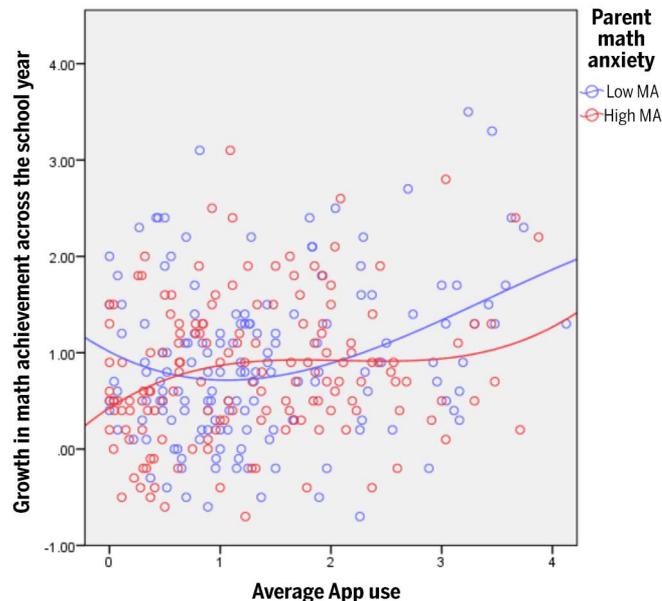
Fig. 1. The nonlinear relation between app usage and growth in math achievement across the school year (grade-equivalent change) in the math app group. As stated in the original Report, we use grade-equivalent change scores for illustrative purposes only; all analyses are conducted on W scores (a Rasch-scale score with equal intervals and thus the most appropriate outcome for our analyses). Beyond 3.29 average app uses per week, one is looking at 16 out of 310 families that fall more than 2 SD above the mean of app use. One should be cautious about conclusions regarding group differences in growth in math achievement across the school year based on this small number of families.

University of Chicago, Chicago, IL, USA.
*Corresponding author. E-mail: s-levine@uchicago.edu (S.C.L.); beilock@uchicago.edu (S.L.B.)

$[\hat{\beta}_{21} = 4.03, t = 2.83, P = 0.005]$ (*13*) (Model S1). This interaction is for children matched within schools, which is most appropriate for comparing the math and reading groups. However, when using the entire sample, the interaction holds $[\hat{\beta}_{21} = 2.55, t = 2.25, P = 0.03]$.

Conducting the above analysis using change scores, as Frank does, is a weaker-powered and less appropriate test than using end-of-year achievement as an outcome, controlling for beginning-of-year achievement as we do in our model (*14*). Further, we view the choice to examine only children who attend a school with classrooms in both conditions as the most conservative test of our hypothesis, given the possibility of school-level differences. For example, if students in the math group came from schools with stronger math instruction than students in the reading group, we might see a large advantage for the math group in terms of math growth over the school year, due to a selection bias instead of randomization to condition. Using a combination of lower-powered tests that do not control for school-level effects contributed to the marginally significant effect that Frank obtained. The statistical tests we report, which we consider more appropriate, show a significant effect of app usage in the math group but not in the reading group. We note that, although app use analyses do not equivocally rule out selection bias, showing strong effects of app use for the math but not the reading group does control for at least one family of endogeneity effects (e.g., parents' propensity to use an app with their children).

Finally, by binning app usage as a secondary analysis, we explored how usage amounts that were understandable in real-word terms related to math growth (e.g., once-a-week app use versus very little). As seen in Berkowitz *et al.*'s figure 2 (*13*) and Fig. 1 below, once-a-week app usage helps close the gap in math achievement growth between children with high- and low-math-anxious parents (although achievement is higher among children of low-math-anxious parents when app usage is highest). We note that the cubic pattern, the best fit for our data (Fig. 1), looks quite similar to the pattern highlighted in our binned data. For reference, there is a significant app use (cubic term of the continuous variable) by parent-math-anxiety interaction on children's end-of-year math achievement (controlling for beginning-of-year math achievement) $[\hat{\beta} = -1.21, t = 3.75, P < 0.001]$.

In sum, we view our data-analytic strategy as appropriate and as providing support for the efficacy of the BLT math app for promoting children's math achievement. We welcome debate about data analysis and hope that this discussion benefits the scientific community.

REFERENCES

1. E. A. Maloney, G. Ramirez, E. A. Gunderson, S. C. Levine, S. L. Beilock, *Psychol. Sci.* **26**, 1480–1488 (2015).
2. S. L. Beilock, E. A. Gunderson, G. Ramirez, S. C. Levine, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1860–1863 (2010).
3. R. Hembree, *J. Res. Math. Educ.* **21**, 33–46 (1990).
4. S. C. Levine, L. W. Suriyakham, M. L. Rowe, J. Huttenlocher, E. A. Gunderson, *Dev. Psychol.* **46**, 1309–1319 (2010).
5. E. A. Gunderson, S. C. Levine, *Dev. Sci.* **14**, 1021–1032 (2011).
6. S. C. Levine, K. R. Ratliff, J. Huttenlocher, J. Cannon, *J. Dev. Psychol.* **48**, 530–542 (2012).
7. S. M. Pruden, S. C. Levine, J. Huttenlocher, *Dev. Sci.* **14**, 1417–1430 (2011).
8. B. N. Verdine, C. M. Irwin, R. M. Golinkoff, K. Hirsh-Pasek *J. Exp. Child Psychol.* **126**, 37–51 (2014).
9. M. C. Frank, *Science* **351**, 1161 (2016).
10. C. F. Manski, *Social Choice with Partial Knowledge of Treatment Response* (Princeton University Press, 2005).
11. G. H. McClelland, C. M. Judd, *Psychol. Bull.* **114**, 376–390 (1993).
12. J. D. Angrist, G. W. Imbens, D. B. Rubin, *J. Am. Stat. Assoc.* **91**, 444–455 (1996).
13. T. Berkowitz *et al.*, *Science* **350**, 196–198 (2015).
14. G. J. P. Van Breukelen, *J. Clin. Epidemiol.* **59**, 920–925 (2006).

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

Talia Berkowitz *et al.*
Science **351**, 1161 (2016);
DOI: 10.1126/science.aad8555