



Reversing Downstream Consequences of School Hiatus on Reading in Disadvantaged, At-Risk Children

Eduardo Estrada¹ · Emilio Ferrer² · Bennett A. Shaywitz³ · John M. Holahan³ · Sally E. Shaywitz³

Received: 6 November 2021 / Revised: 6 January 2022 / Accepted: 10 January 2022 / Published online: 7 February 2022
© The Author(s) 2022

Abstract

The spread of COVID-19 has led to the disruption of K-12 education for about 90% of the world's student population. The effects on children's academic development are unknown. We examined how disruption in schooling over three consecutive summers in disadvantaged minority children affects reading and whether an intensive intervention can ameliorate these effects. Our data were collected prior to the COVID-19 pandemic. We applied Latent Change Score models to examine developmental trends in a longitudinal study of reading in 111 economically disadvantaged children, assessed biannually from grades 1 to 4, including 3 summers (for a total of 6 months of school hiatus). The students fell behind the normative population in their ability to understand written texts, a decrease in their relative percentile of 0.25 of a standard deviation each summer, and an effect 3–4 times greater than prior studies suggested. Compared to children in a comparison group, children who received an evidence-based intervention during the school year were better able to maintain their reading scores. These findings provide evidence that disruptions in schooling, for example, those implemented to slow the spread of COVID-19, may have a significant detrimental effect on the reading abilities of disadvantaged children and that children who received a reading intervention were better able to maintain their reading scores during the hiatus. It is critical that policy makers prioritize the allocation of necessary resources to minimize the negative effects on reading this pandemic has wrought on these most disadvantaged children.

Keywords Disadvantaged children · Dyslexia · Reading · School hiatus

In an attempt to slow the spread of COVID-19, most countries have interrupted formal schooling. This has led to approximately 90% of the world's student population having their schools closed during varying periods (UNESCO, 2021). The effects of such an unprecedented disruption in school on children's academic development are unknown. However, it is possible to extrapolate possible effects of the current and ongoing disruptions in school from studies which have examined the effects of the summer hiatus on academic performance. Previous studies have shown that the summer hiatus results in negative effects on children's educational achievement, with an average decrease of approximately one-tenth of a standard deviation over the summer period (Cooper et al., 1996).

Here we report findings on the effects of the summer hiatus in schooling in a contemporary sample of economically disadvantaged children in a longitudinal study over the first 4 years of school. In addition, we hypothesized and our data support significant ameliorative effects of an intensive evidence-based reading intervention provided during the school year on the summer hiatus. We note that our data were collected prior to the COVID-19 pandemic. These results have important implications for the estimated millions of children who have missed and are currently still missing school because of the coronavirus pandemic.

✉ Sally E. Shaywitz
sally.shaywitz@yale.edu

Bennett A. Shaywitz
bennett.shaywitz@yale.edu

¹ Department of Social Psychology and Methodology, Universidad Autónoma de Madrid, Madrid, Spain

² Department of Psychology, University of California, Davis, Davis, California, USA

³ Department of Pediatrics, Yale University School of Medicine, 129 York St. Suite 1P, New Haven, CT 06511, USA

Methods

Study Oversight

Parents or caretakers provided written consent for their children to participate in the study, and children also provided assent. The Institutional Review Board at the Yale University approved this study. This study was conducted in accordance with the ethical principles that have their origin in the Declaration of Helsinki and are consistent with good clinical practices and applicable laws and regulations.

Participants and Study Design

One hundred and eleven students from two elementary charter schools in New Haven participated in the study (58.6% were girls). Both schools were in districts with high proportions of economically disadvantaged and minority children; 79.0% of the children received free lunches at the beginning of the study; 96.6% were African American, Hispanic, or Latino. All the students in first grade in the two schools were eligible for the study. No students were excluded from the study.

All the participants were tested each year from grade 1 through grade 4, from 2015 through 2018. We used the *Phonological Awareness* measure of the Comprehensive Test of Phonological Processing (CTOPP-2) (Wagner et al., 2013) as a selection variable for deciding which children were at risk for reading failure. All students below the 25th percentile were considered at risk and were provided with the intervention ($n = 20$) (designated as *intervention*). The remaining students served as a comparison group. Among them we selected 18 children closely matched to the intervention cases in sex, age, and the next closest scores on *Phonological Awareness* (designated as *matched comparison group*). The remaining children served as a *remaining comparison group* ($n = 73$).

All students in the *intervention* and *matched comparison* groups were assessed at the beginning and end of each school year (fall and spring test periods). This allowed us to evaluate the effect of the interruption of schooling during the summer. The *remaining comparison* children were assessed in the fall only, which allowed us to study the overall trend. The mean age (and *SD*) in months for each group at the first test period were as follows: *intervention*, 84.2 ($SD = 6.8$); *matched comparison*, 78.2 ($SD = 3.7$); and *remaining comparison*, 76.4 ($SD = 4.3$). The sample size available at each wave is reported in Table 1. There was very little attrition in the study, and therefore the sample sizes for the intervention and matched comparison groups show little variation over time. The differences in sample size across time points in the remaining comparison group were due to measuring those students in the fall waves only.

Evidence-Based Intervention

The 20 students in the *intervention* group received, in addition to their regular reading program, an evidence-based intervention beginning in first grade and continuing in second, third, and fourth grades. The intervention used an Orton-Gillingham approach (Ritchey & Goeke, 2006) and was provided Mondays through Thursday for 90 min each day in small groups of 4 students by instructors trained to administer Orton-Gillingham instruction who worked together with the regular classroom teachers.

Outcome Measure

At every test period, we assessed participants reading comprehension using the *Passage Comprehension (PC)* component of the Woodcock-Johnson IV Tests of Achievement (WJ IV ACH) (Schrank et al., 2014). We selected reading comprehension as the outcome following long-standing federal guidelines (National Reading Panel, 2000) that reading comprehension represents "... the 'essence of reading' that sets the stage for children's later academic success..." (Gamse et al., 2008) (p. 10).

Because we were interested in the participants' position relative to their normative population at each test period, we used age-standardized z scores. In this metric, the population mean and standard deviation equal 0 and 1 points, respectively. Therefore, at every test period, a score of zero is exactly at the population's mean, and scores of -1 and 1 points are one *SD* below and above the mean, respectively. Note that a student whose score is zero at every test period would show typical development, which would be associated with actual *growth* in the absolute level of ability. This is because reading ability is expected to increase during the developmental period examined here.

Statistical Analysis

To characterize the longitudinal development of the various constructs, we applied a Latent Change Score (LCS) model (Ferrer & McArdle, 2010; McArdle, 2009; McArdle, 2001). This dynamic structural equation model allows capturing the key elements of change as a set of statistical parameters. A path diagram representing the model is depicted in Figure 1.

The LCS model includes a measurement structure that makes it possible to separate the variance in the relevant construct from the variance due to measurement error. This allows estimating the latent true trajectory. The measurement occasions (y) are modeled as an auto-regressive system: any given test period y_t is predicted by two components. The first component is the level at the previous test period y_{t-1} , and the second component Δy_t captures all the innovations, or changes, from y_{t-1} to y_t . The latent variable y_0 (with mean μ_0

Table 1 Means (standard deviations) and sample sizes for passage comprehension at all test periods. Age-normed *z* scores

Assessment	Mean (Sd)			Sample size		
	Interv.	Matched comparison	Remaining comparison	Interv.	Matched comparison	Remaining comparison
fall '14	-.85 (.91)	.59 (.50)	.58 (.64)	20	18	71
fall '15	-1.17 (.87)	.19 (.49)	.32 (.60)	17	18	70
spring '16	-.98 (.74)	.08 (.52)		17	18	2
fall '16	-1.25 (.77)	-.22 (.42)	-.19 (.53)	19	18	67
spring '17	-1.13 (.61)	-.37 (.54)		17	18	0
fall '17	-1.51 (.75)	-.54 (.54)	-.52 (.57)	17	17	61
spring '18	-1.34 (.68)	-.45 (.60)		14	17	0
Mean decrease wave 1→7	-.659	-1.126	-1.105			
Mean decrease wave 1→8	-.496	-1.036				

and variance σ^2_0) allows capturing the initial level (i.e., latent intercept). Another latent variable y_s (with mean μ_s and variance σ^2_s) characterizes the linear developmental change from each test period to the next (i.e., latent additive component). These two latent variables are allowed to be correlated through

the covariance σ_{0s} . Because the system is auto-regressive, any change introduced at one particular test period is propagated to later time points. We specified loadings from the latent additive component to the innovations with a fixed value of

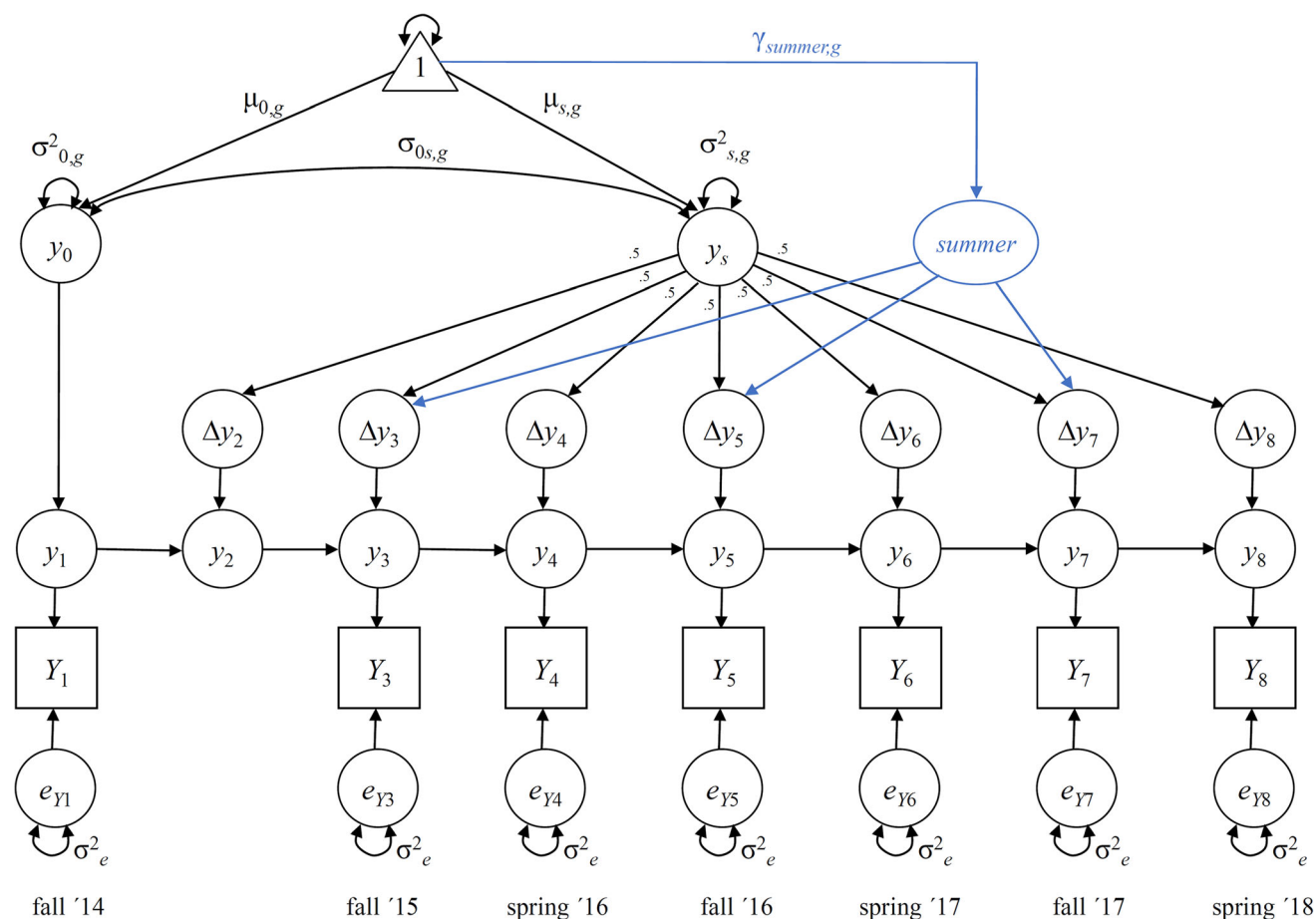


Fig. 1 Path diagram of the Latent Change Score model applied to characterize the developmental trajectories. Note: Sub index *g* indicates different estimates for each group. The square boxes represent observed variables. The circles represent latent variables. A triangle represents a

constant for the estimation of means. Single-ended arrows represent regression loadings. Double-ended arrows represent variances and covariances. Regression loadings without a value in the diagram were fixed to one.

0.5. This parameterization allows interpreting its mean μ_S as the mean yearly change.

We included a latent variable (*summer*), with mean γ_{summer} and no variance (i.e., a fixed effect). This latent variable had a specified effect on test periods 3, 5, and 7 only (i.e., the fall of grades 2 through 4, which were taken right after each summer). Because the model is auto-regressive, it allows any effect due to the summer introduced at a given test period to be carried over to later test periods.

To examine differences between the students in intervention and the comparison groups, we applied a two-group SEM including all participants and test periods. Because the same measurement tool was used for the participants in both groups, we constrained the measurement error variance to have the same value for both groups. The rest of the parameters were freely estimated. The *R* package *lavaan* was used for model estimation (Rosseel, 2012).

For estimating the model, we used full information maximum likelihood (FIML) (Arbuckle, 1996). This procedure was chosen to deal with unequal sample sizes at each wave. Previous literature has shown that FIML estimation is superior to other methods for dealing with incomplete data in the context of structural equation modeling (Enders & Bandalos, 2001; Schlomer et al., 2010). Given our relatively small sample size, we conducted a power analysis to assess the ability of our model to detect the summer effect in the intervention group. Our analysis indicates that, given our sample size and the effects found in our data, our model would be able to detect such effects at least 63% of the time with samples of size 20 (i.e., 63% power). A power of 80% would be achieved with a sample of size 30. In other words, our sample allows detecting summer effects of the size estimated here.

Results

Descriptive Analysis

Table 1 reports the means and standard deviations of the outcome variable (*Passage Comprehension, PC*) at each test period. A generalized decline in the *z* scores is readily apparent from fall '14 to spring '18. However, we found marked differences between the *intervention* group and the two comparison groups. Students in *intervention* started off with considerably lower mean levels: $-.85$ (percentile 20). In contrast, both comparison groups had higher initial scores: $.58$ and $.59$ (percentile 72). However, *intervention* students were better able to maintain their relative position from test period 1 to 7, declining by $.659$ points, whereas the *matched* and *remaining comparison* students lost 1.126 and 1.105 points, respectively. Because these are *z* scores, they can be readily interpreted as standard deviations and as Cohen's *d*-indices (Cohen, 1988).

These results are surprising because they show that those students who did not receive the intervention started off above the reference mean but fell behind their reference group from grade 1 to grade 4. They show also that both comparison groups were fairly similar. Note that, because these are *z* scores, a decrease from one test period to the next indicates a reduction in the group's standing relative to their reference population, not a decrease in the absolute level. In other words, reading scores in the comparison groups did not decline over time; rather, they did not improve at the rate of their reference population.

Effects of the Summer Interruption on Longitudinal Development

Table 2 includes parameter estimates and fit indices from the LCS models for the *intervention* and *matched comparison* groups.

An important feature of the LCS model is that it allows isolating the effect of the summer interruption on the longitudinal trajectories (and, consequently, it makes it possible to estimate what would be the yearly change when such effects are discounted). Figure 2 depicts the total yearly change in reading with and without the summer effects, for each group. A negative (or positive) effect implies that the group decreases (or increases) its mean percentile, relative to the reference population, every year. The gray bars (right side of each pair) depict the estimated yearly change, including the effect of the summer interruption. The black bars (left side of each pair) depict what would be the yearly change if the effect of summer is discounted (i.e., without the summer interruption). The net effect of summer is the difference between the two values and was negative for reading comprehension.

The effect of summer interruption was $d_{summer,Int} = -.32$ (95% CI $-.60$ to $-.04$) and $d_{summer,MC} = -.20$ (95% CI $-.47$ to $.07$) for *intervention* and *matched comparison* groups, respectively. In other words, the *intervention* group declined by approximately one-third of a standard deviation every summer, whereas the *matched comparison* group declined by one-fifth of a standard deviation. The total yearly change was negative for both groups. Note that without the summer disruption, the yearly change would be positive for the *intervention* students, $d_{year,AR} = .14$, and negative but much closer to null for the *matched comparison* group, $d_{year,MC} = -.16$. This implies that, without the summer interruption, the *intervention* students would gradually reduce their gap with the reference population, whereas the *matched comparison* would fall behind much more slowly.

Figure 3 depicts individual trajectories for reading in *intervention*, *matched comparison*, and *remaining comparison* groups, along with the means estimated by the LCS model. This figure indicates that, when compared to their reference

Table 2 Parameters (and standard errors) of the multigroup Latent Change Score model

Parameter	Intervention	Matched comparison
Mean total yearly change (w/summer interruption)	-.18	-.36
Mean yearly change w/o summer interruption	.14 (.14)	-.16 (.13)
Effect of summer	-.32 (.14)	-.20 (.10)
Initial mean	-.89 (.20)	.54 (.10)
Initial mean variance	.69 (.24)	.12 (.06)
Variance of yearly change w/o summer interruption	.03 (.02)	.01 (.01)
Covariance initial-yearly	-.11 (.05)	-.01 (.02)
Measurement error variance		.11 (.01)
LogLikelihood	-146.6	
χ^2 (57df)	90.0	$p = .003$
BIC	340.5	

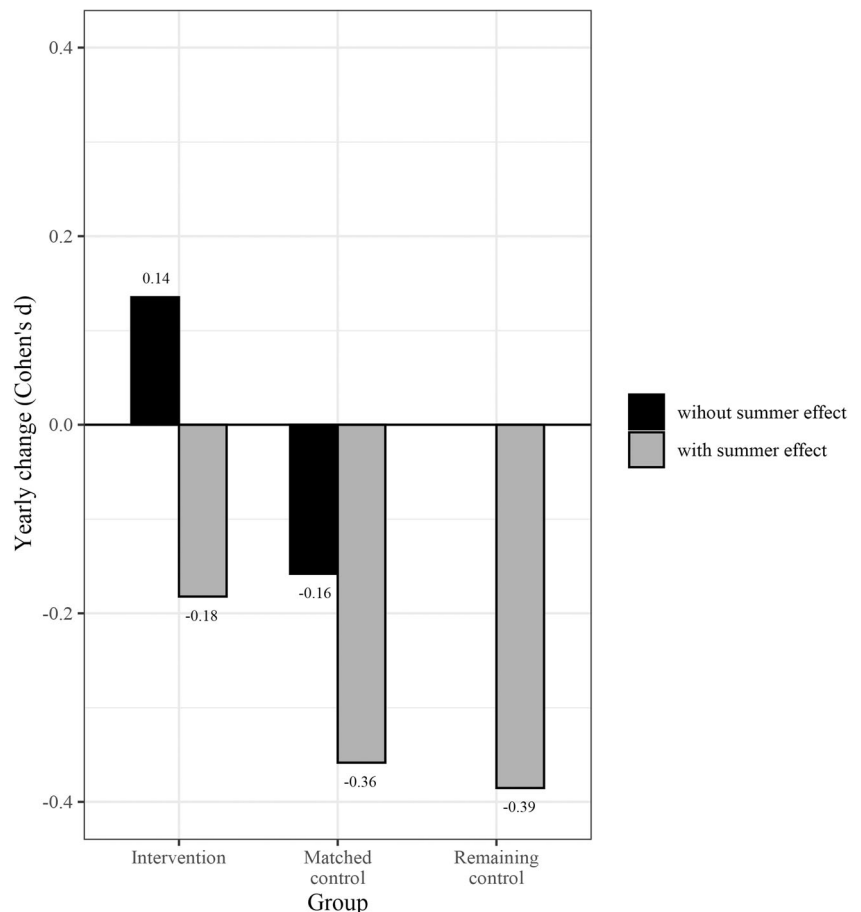
Note: Models fitted to age-normed Z scores. Negative changes imply declines in the groups' relative percentile compared to their normative population.

population and controlled for age, students in the *matched comparison* group generally fell behind from grade 1 to grade 4, whereas *intervention* students were better able to decelerate their decline and maintain their relative standing. The total decline was larger for the *matched comparison* students. In both cases, the decrease is accelerated during the summer periods.

Discussion

Our findings document a significant negative effect of a school hiatus in economically disadvantaged children in the first 4 years of school. This negative effect, here in a contemporary cohort, is 3–4 times greater than studies over two

Fig. 2 Yearly change in passage comprehension with and without the effect of the summer interruption. Note: Changes expressed in age-normed z scores. A z score of 0 implies that the group keeps its relative position with respect to the reference population from 1 year to the next. Negative and positive scores imply, respectively, that the group falls behind, or learn faster, than the reference population



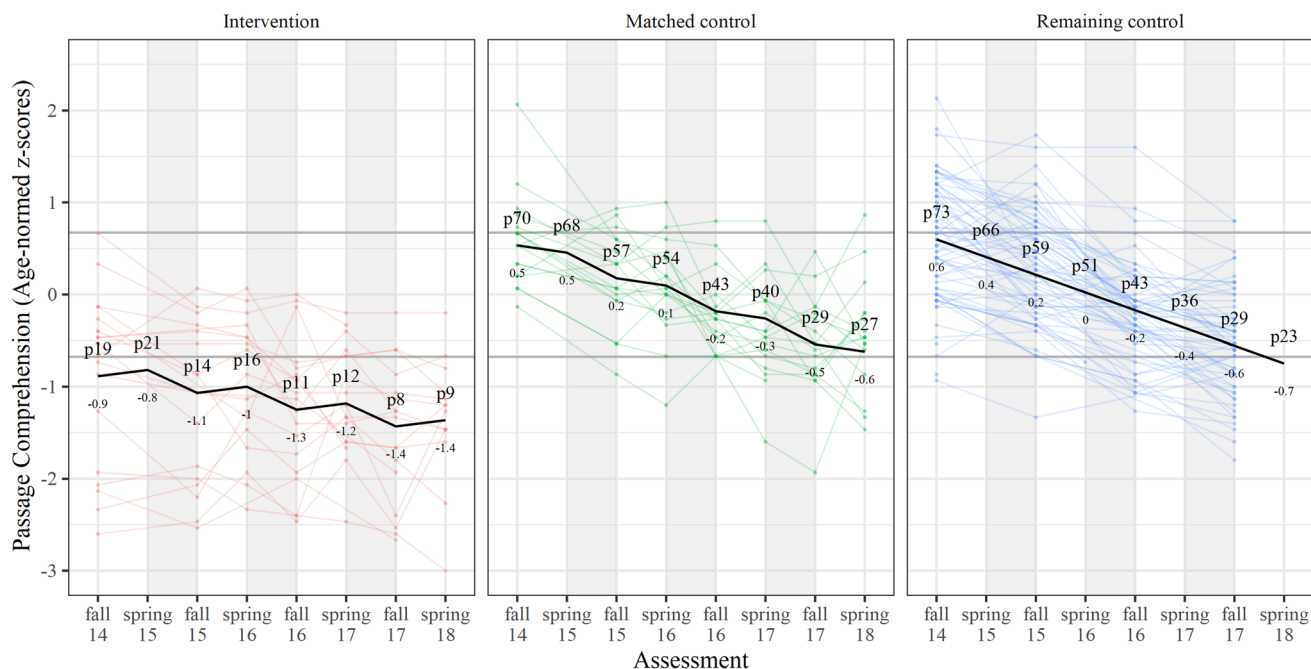


Fig. 3 Longitudinal development of passage comprehension for the intervention and comparison students (summer periods represented by shaded areas). Note: The black line represents the model-implied means.

The numbers above and below the line indicate mean percentiles and mean z scores, respectively. The horizontal gray lines indicate values containing the central 50% of the normative population.

decades ago had suggested and has serious implications for the millions of children who have missed and may currently still missing school because of the coronavirus pandemic. Of particular relevance, an intensive, evidence-based reading intervention provided to those children most at-risk for reading failure (intervention group) improved their reading during the time they were in school and modified though did not fully reverse the effects of the summer hiatus. On a positive note, even with the negative effects of the summer hiatus, those children who received the intensive reading intervention fared much better than children who did not, whose yearly decline was more than double that observed in the intervention group.

More specifically, our findings suggest that, without the summer interruption, the *intervention* group, those students at greatest risk (with lower initial scores: mean percentile = 19 in the fall of grade 1), would gradually improve in relation to the normative population. In other words, without the summer disruption, their yearly change in reading would have been positive. However, the interruption in schooling due to the summer hiatus disrupted this process and reversed the net yearly change from positive 0.14 to negative 0.18. This decrease of $d = -.32$ is more than three times as great as the effects of the summer hiatus reported in previous studies (Cooper et al., 1996). Furthermore, the students at greatest risk for reading failure, who received an evidence-based intervention during the school year, showed, overall, smaller declines in reading regardless of the effect of summer hiatus.

While the positive effects observed in the intervention group provide a degree of optimism, the generalizability of our findings is tempered by the resource intensive nature of the intervention. In particular, children received the reading intervention for 90 min 4 days/week in small groups provided by teachers who had considerable experience and training using the Orton-Gillingham method.

We note that our data were not collected during the COVID-19 pandemic. Two important differences between the schooling disruptions examined in this study and the disruption due to the COVID pandemics are as follows: (a) each of the summer breaks considered here is shorter than the COVID disruption, and (b) summer breaks are an expected part of a normal school year, whereas the COVID disruption is not. Nevertheless, our findings are very useful to inform decisions in the current situation. First, the total disruption time adding the three consecutive summers is approximately 6 months. This period is similar to that in which many school districts have been closed due to the COVID pandemics. Second, even if these types of disruptions are not equivalent, both situations involve children missing school for months. Our results provide support to indicate that the current disruptions in schooling implemented to slow the spread of COVID-19 are likely having a significant detrimental effect on the reading abilities of those students at greatest risk—economically disadvantaged children.

In addition our findings make it imperative that schools put in place programs to compensate for the declines during the period without regular classes. It is critical that vulnerable children are given an opportunity to begin to narrow their yawning achievement gap in reading. These data further suggest that the longer these disruptions persist, the greater will be the negative effects on reading for the most disadvantaged children. They mandate that educational policy makers must both prioritize and implement specific measures to minimize as much as possible these serious consequences. In particular, they must ensure that these most vulnerable students are a major focus of effective, evidence-based reading interventions throughout the school year as well as over the summer. These efforts must be especially proactive in low-income African American and Hispanic populations where identification of academic deficits may be more likely to go unrecognized (Magnusson et al., 2017).

Providing such interventions presents a real challenge since currently, even in school settings with relatively small numbers of students in a class, teachers may have very limited time to help any individual student. Interventions in the current study were provided for 90 min 4 days/week to students in groups of four students or less. In the interests of equity, it is critical that policy makers prioritize the allocation of necessary resources to minimize the negative effects on reading this pandemic has wrought on these most disadvantaged children.

Acknowledgements Support for the work reported in this article was provided by: The Seedlings Foundation. EE was supported by the Ministry of Science and Innovation of Spain (ref. PID2019-107570GA-I00 / AEI / doi: <https://doi.org/10.13039/501100011033>).

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arbuckle, J.L. (1996). Full information estimation in the presence of incomplete data. In Marcoulides GA & Schumacker RE (Eds.), *Advanced Structural Equation Modeling: Issues and Techniques* (Vol. 243, pp. 243-277): Erlbaum.
- Cohen, J. . (1988). *Statistical power analysis for the behavioral sciences* (2 ed.): L. Erlbaum Associates.
- Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66(3), 227–268.
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5
- Ferrer, E., & McArdle, J. J. (2010). Longitudinal modeling of developmental changes in psychological research. *Current Directions in Psychological Science*, 19(3), 149–154.
- Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B., & Unlu, F. (2008). *Reading First Impact Study Final Report (NCEE 2009-4038)*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences.
- Magnusson, D. M., Minkovitz, C. S., Kuhlthau, K. A., Caballero, T. M., & Mistry, K. B. (2017). *Beliefs regarding development and early intervention among low-income African American and Hispanic mothers Pediatrics*, 140(5), e20172059. <https://doi.org/10.1542/peds.2017-2059>
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60(1), 577–605.
- McArdle, J. J. . (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling, present and future: A festschrift in honor of Karl Jöreskog* (pp. 7-46): Scientific Software International.
- National Reading Panel. (2000). *Report of the national reading panel, teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. US Government Printing Office.
- Ritchey, K. D., & Goeke, J. L. (2006). Orton-Gillingham and Orton-Gillingham—based reading instruction: A review of the literature. *The Journal of Special Education*, 40(3), 171–183.
- Rosseel, Y. . (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2).
- Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology*, 57(1), 1–10. <https://doi.org/10.1037/a0018082>
- Schrank, K. S., Mather, N., & Woodcock, R. (2014). *Woodcock-Johnson-IV*. Riverside Publishing.
- UNESCO. (2021). COVID-19 educational disruption and response: UNESCO.
- Wagner, R, Torgesen, J, Rashotte, C, & Pearson, N. (2013). *CTOPP-2: Comprehensive test of phonological processing* (2nd ed Ed.). : PRO-ED.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.