REVIEWED BY EVIDENCE for ESSA
MEETS EVIDENCE STANDARDS

Curriculum Associates RESEARCH

Impact of *Magnetic Reading™ Foundations* on Overall Reading Scores in Grade 2

Molly K. Duncan, Ph.D.

Reading
Research Technical Report, October 2023

i-Ready Learning Magnetic Reading Foundations

Curriculum Associates RESEARCH

Impact of *Magnetic Reading Foundations* on Overall Reading Scores in Grade 2

Molly K. Duncan, Ph.D.

Reading

Research Technical Report, October 2023

Summary

This study examined the impact of *Magnetic Reading Foundations* on Grade 2 students' overall Reading scores, using a rigorous quasi-experimental design with a baseline-equivalent sample of students and hierarchical linear modeling to account for the nested nature of the data. In order to isolate the effects of the treatment, covariates such as pretest scores and exposure to other instruction were included in the outcomes model. The study included a large sample of students from 40 schools in four states during the 2022–2023 school year. Results show use of *Magnetic Reading Foundations* is associated with a nine-point increase in scores on the *i-Ready® Diagnostic* for Reading. The effect size of this difference in scores is .15, considered moderate to large when compared to similar interventions investigated with comparable methodologies. This study meets the requirements for Every Student Succeeds Act (ESSA) Level 2 evidence.

i-Ready Learning Magnetic Reading Foundations

Introduction

Given the importance of early reading proficiency for becoming a successful learner (Fiester, 2010) and recent reports of low reading proficiency among students in the United States (Curriculum Associates, 2023b; National Assessment of Educational Progress, 2022), it is critical for educators to find effective and efficient methods for teaching reading to students in the early elementary years. One such solution is *Magnetic Reading Foundations*, a foundational skills program designed for Grades K-2 that combines explicit, systematic instruction with rich and engaging texts for students (Curriculum Associates, 2023a). It is designed to be used as the foundational skills block of a full English language arts curriculum in non-intervention settings, and it provides supports for individualizing instruction for all learners, including multilingual learners and students who need additional instructional support. This study was designed to examine the impact of *Magnetic Reading Foundations* usage on overall Reading scores in Grade 2.

Magnetic Reading Foundations contains Unit Assessments educators can use to understand students' mastery of the specific concepts taught in each unit. However, for this study, overall reading skill was measured with the *i-Ready Diagnostic* for Reading because of the desire to measure growth in overall reading skill rather than skills related to the specific concepts taught in each unit of *Magnetic Reading Foundations*.

The *i-Ready Diagnostic* for Reading is a computer-adaptive assessment developed by Curriculum Associates that provides information about students' performance in relation to both grade-level standards and national norms. Although both the Diagnostic and *Magnetic Reading Foundations* are produced by Curriculum Associates, the *i-Ready Diagnostic* is not designed as a measure specific to *Magnetic Reading Foundations*. *Magnetic Reading Foundations* is designed primarily to support foundational reading skills, while the *i-Ready Diagnostic* is designed as a measure of general reading skills and captures information about Grade 2 students' performance in several Reading domains, including Phonics, Vocabulary, Comprehension: Literature, Comprehension: Informational Text, High-Frequency Words, and, for Grade 2 students who are striving learners in the Phonics domain, Phonological Awareness.

Research Methodology

This study addresses the following research question:

1. What was the impact of one year of use of *Magnetic Reading Foundations* on overall Reading scores for students in Grade 2?

Data

This study utilized data regularly collected by Curriculum Associates, including student demographic and *i-Ready Diagnostic* data. For this study, the only demographic variables used were student ethnicity and Hispanic origin. Districts can choose whether they provide student demographic data to Curriculum Associates, and many districts choose not to provide it or to provide it for only some students. Therefore, schools were excluded from this analysis if they reported Hispanic origin for fewer than 95% of the students in Grades K-2 who had taken an *i-Ready Diagnostic* in the fall. These criteria were applied at the school level because schools that report demographic data for only some students may be including data only for students who meet certain criteria, such as recording a value for Hispanic origin only if a student is Hispanic. This study also utilized an indicator of students' ethnicity, but because of small sample sizes, many ethnicity categories had to be combined. Therefore, each student has a value of "Black," "White," or "Other ethnicity or missing ethnicity" as well as a value of "Hispanic origin" or "non-Hispanic origin" for the Hispanic origin variable.

Data from students' fall and spring *i-Ready Diagnostics* for Reading were also used for this study. Scores on the *i-Ready Diagnostic* range from 100 to 800. For a student in Grade 2, a score of 489 is the minimum score in the Early On Grade Level range, indicating that a student has partially met grade-level expectations and would benefit from continued on-grade level instruction.

The data for this study were collected in two stages. The first stage took place in April 2023, before any of the included students had taken their spring *i-Ready Diagnostic*. At this point, all data except the spring *i-Ready Diagnostic* data were collected and cleaned, and student matching took place. The second stage of data collection took place in mid-June 2023, when all spring *i-Ready Diagnostics* had been completed. During this stage, the data from spring *i-Ready Diagnostics* were added to the dataset, and outcome analyses were completed. All data for this study are from the 2022–2023 school year.

Full Sample

Magnetic Reading Foundations is intended to be used by an entire classroom of students, rather than as an intervention for individual students, and as such it tends to be adopted by an entire district or school. The treatment group for this study is a convenience sample of the schools the researcher could identify as using *Magnetic Reading Foundations* and that met other selection criteria. *Magnetic Reading Foundations* usage or comparison group eligibility was determined through communication with the school's *i-Ready Partners*. Additional school-level criteria for study eligibility included the following: Each school was required to have fall *i-Ready Diagnostics* for Reading for at least 90% of the students rostered in Grade 2, report race and ethnicity data for at least 95% of the students who had taken an *i-Ready Diagnostic* for Reading in the fall, and not have administered their spring *i-Ready Diagnostics* at the time of the first stage of data collection.

In order to be included in the treatment group for this analysis, students in *Magnetic Reading Foundations* schools had to be rostered in Grade 2, have a non-missing value for Hispanic origin reported in Curriculum Associates' data system, have taken an eligible fall *i-Ready Diagnostic* that was completed within 45 calendar days of the first day of school (school calendar information was taken from online district calendars accessed in April 2023), and not have taken a spring *i-Ready Diagnostic* for Reading at the time of the first stage of data collection.[1]

[1]Although schools that had already begun administering their spring *i-Ready Diagnostics* to all students were not included in the sample, some schools occasionally administer the *i-Ready Diagnostic* to a few students at a time outside the school's standard testing window. There were only two students who met the other eligibility criteria and had to be dropped for having already taken their spring *i-Ready Diagnostic* for Reading.

The comparison group was selected by first identifying schools that had characteristics similar to one or more treatment school(s), then by using propensity score matching to select students who were similar. Comparison schools had to meet the same criteria as treatment schools regarding Diagnostic completion and data reporting in the *i-Ready* system. To select comparison schools from the schools that were eligible according to the aforementioned criteria, information from the National Center for Education Statistics' Common Core of Data was used to identify schools that were similar to at least one treatment school based on its location (that is, in the same state), having the same or a similar Title 1 status, and having similar percentages of students who were Black, Hispanic, eligible for free or reduced-price lunch or economically disadvantaged, and in a district with a similar percentage of English Learners. Because individual students were matched with a propensity score model, this step was only for the purpose of ensuring we had a large pool of similar students to choose from, and there were no explicit criteria for how big of a difference in percentage was allowed. Student-level inclusion criteria for the comparison group were the same as the student-level inclusion criteria for the treatment group.

Matching

All data processing and analyses for this report were completed in R version 4.1.3 (R Core Team, 2022). Matching was completed with the MatchIt package, version 4.5.0 (Ho et al., 2011). A single-level propensity score model was used to select students from the comparison group who were most similar to students in the treatment group on a vector of covariates. Matching methods such as propensity score matching can reduce bias and provide a better estimate of the treatment effect. Therefore, studies that use a matching model and include a pretest measure as a covariate in the outcome model are the preferred alternative when randomized controlled trials are not possible or desirable (Austin, 2011; Shadish et al., 2002; What Works Clearinghouse, 2022). Several one-to-one nearest neighbor matching models were compared to determine which yielded a balanced sample while retaining the most students. All matching models used a one-to-one nearest neighbor algorithm without replacement, and they all included the student's fall Diagnostic score, a binary variable indicating student's Hispanic origin, a vector of binary variables indicating the student's ethnicity, and a vector of binary variables indicating the state in which the student's school was located. Additionally, all models required an exact match on state. The models differed in the caliper used for matching, the order in which treatment students received a match, and the inclusion of an additional predictor, which was a continuous variable indicating the number of calendar days between the district's first day of school and the completion of the student's fall *i-Ready Diagnostic*. The final matching model used a .3 SD caliper, matched the treatment students in order of the size of the propensity score (smallest first), and did not include the variable indicating days between school start and Diagnostic completion. Because MatchIt version 4.5.0 breaks ties in propensity score or in distance based on the sort order of the data, the data were first sorted in a random but replicable order.[2]

[2]After setting a specific seed, the dataset was sorted by the student ID, then a random number was generated using the rnorm function. The dataset was then sorted by the random number.

The total matched sample included 872 *Magnetic Reading Foundations* students and 872 comparison students. Because matching took place before the outcome measure was collected, some students from the matched sample could not be included in the final analysis because they did not have a spring *i-Ready Diagnostic*. The treatment group lost 50 students (i.e., 5.7% of the group), and the comparison group lost 79 students (i.e., 9.1%). The final sample contains 11 schools from two states in the Northeast, six schools from one state in the Midwest, and 24 schools from one state in the West. Individual states and districts are not identified to preserve anonymity. Table 1 below summarizes sample size and demographic information about the original dataset (i.e., before matching and attrition) and the analysis dataset (i.e., after matching and attrition) by groups.

Table 1. Sample Size and Demographic Information

	Total Students	Total Schools	Percentage Black	Percentage White	Percentage Other Ethnicity or Missing Ethnicity	Percentage Hispanic
Original Dataset (i.e., Before Matching and Attrition)						
Magnetic Reading Foundations Group	987	12	3.55%	90.27%	6.18%	56.64%
Comparison Group	1,714	29	9.16%	68.96%	21.88%	43.00%
Analysis Dataset (i.e., After Matching and Attrition)						
Magnetic Reading Foundations Group	822	12	4.01%	89.42%	7.06%	51.46%
Comparison Group	793	28	4.04%	88.9%	6.57%	49.43%

Note: Many ethnicity categories had to be combined due to low sample sizes.

The groups had similar average values on the fall *i-Ready Diagnostic*. Information about average fall *i-Ready Diagnostic* scores is summarized in Table 2 below.

Table 2. Fall Diagnostic Score by Group

	Average Fall Diagnostic Score	Standard Deviation of Fall Diagnostic Score	Glass's Delta
Original Dataset (i.e., Before Matching and Attrition)			
Magnetic Reading Foundations Group	432.79	51.98	.22
Comparison Group	444.39	52.40	
Analysis Dataset (i.e., After Matching and Attrition)			
Magnetic Reading Foundations Group	444.34	44.62	.10
Comparison Group	439.28	51.09	

Outcome Model

To estimate the impact of *Magnetic Reading Foundations* usage after controlling for other variables, a multilevel model was specified using the package lme4, version 1.1-34 (Bates et al., 2015). Multilevel modeling is one method of properly accounting for the nested nature of the data and was preferable over other methods, such as cluster robust standard errors and generalized estimating equations, due to this dataset's overall sample size and cluster sizes (McNeish et al., 2017; Raudenbush & Bryk, 2002). Prior to fitting the outcome model, a random-intercepts-only model predicting the spring *i-Ready Diagnostic* score was fit to calculate the intraclass correlation coefficient (ICC). This is the proportion of total variability in spring *i-Ready Diagnostic* scores that can be attributed to clustering at the school level. In this dataset, about 21 percent of the variability of spring *i-Ready Diagnostic* scores is related to school membership. This high ICC reinforces the need for multilevel models to properly account for the clustered nature of the data.

To determine the final outcome model, a series of models were fit with maximum likelihood estimation and compared. The primary criterion for the comparison was a likelihood ratio test of nested models. The Akaike Information Criterion and Bayesian Information Criterion were also considered and were found to support the same conclusions as the likelihood ratio tests (Peugh, 2010; Whittaker & Furlow, 2009). Models differed in which fixed effects they included. The inclusion of random effects aside from the random intercept related to school membership was not tested due to sample size. The demographic variables used for matching were considered for inclusion, but the model fit indices suggested they did not improve the model fit. This may be partially because the students within each state were somewhat racially homogenous. The final outcome model was refit with restricted maximum likelihood estimation because it tends to produce less biased estimates (Raudenbush & Bryk, 2002), and the restricted maximum likelihood estimates are reported in the remainder of this paper. The model selected as the final outcome model was:

Level 1 (Student):

$$Y_{ij} = \beta_{0j} + \beta_{1j}(Centered\ Fall\ Diagnostic\ Score_{ij}) + \beta_{2j}(Centered\ Days\ between\ Diagnostics_{ij}) + \beta_{3j}(Centered\ Hours\ of\ Personalized\ Instruction_{ij}) + e_{ij}$$

Level 2 (School):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Magnetic\ Reading\ Foundations_j) + \Sigma\gamma_{02}(State_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

where Y_{ij} represents the expected score on the spring *i-Ready Diagnostic* for Reading for student *I* in school *j*; β_{0j} represents the school-level intercept, that is, the expected spring Diagnostic score for a student in school *j* for a student for whom all predictors are zero or the centered value (see Table 3); β_{1j} represents the difference in spring *i-Ready Diagnostic* score associated with a one-point increase in fall *i-Ready Diagnostic* score; β_{2j} represents the change in spring *i-Ready Diagnostic* score associated with one additional calendar day in between completion of the fall and spring *i-Ready Diagnostics*; β_{3j} represents the change in spring *i-Ready Diagnostic* score associated with one additional hour of *i-Ready Personalized Instruction* usage in between the fall and spring *i-Ready Diagnostic*; γ_{00} represents the overall grand mean spring score; γ_{01} represents the change in spring score associated with *Magnetic Reading Foundations* usage, that is, the treatment effect; and $\Sigma\gamma_{02}$ represents the sum of a vector of binary variables indicating the state in which school *j* is located. Table 3 presents information about the variables in this model and the value that 0 represented for each variable.

Table 3. Information about Outcome Model Covariates

Variable	Values	Rationale
Fall *i-Ready Diagnostic* Score	Continuous variable centered at 489	489 is the minimum score in the Early On Grade Level range for a student in Grade 2.
Days between *i-Ready Diagnostics*	Continuous variable centered at 252	252 is the approximate median for this variable in this sample.
Hours of *i-Ready Personalized Instruction*	Centered at 20	20 is the approximate median for this variable in this sample.
Magnetic Reading Foundations	0 for comparison group; 1 for *Magnetic Reading Foundations* group	Allows estimation of treatment effect
State	3 binary variables; all variables were 0 if school was in State 1	State 1 was chosen as the reference state because it had the most students in this analysis.

Estimation of the Treatment Effect

Glass's Delta is used for estimating a standardized treatment effect. As is true of most standardized effect size calculations, the basic calculation involves dividing the group differences by the standard deviation of the outcome. In the case of this study, Glass's Delta is calculated by dividing the fixed effect estimate associated with *Magnetic Reading Foundations* usage by the standard deviation of the spring *i-Ready Diagnostic* for Reading score in the control group. The fixed effect estimate for *Magnetic Reading Foundations* usage is used because it is the difference in scores between the treatment and control group once adjustments for all other covariates have been considered. The standard deviation for the control group is used as the denominator because interventions may sometimes be expected to affect the variance of the outcome measure, so using a standard deviation that has been unaffected by the intervention is preferable (Evidence for ESSA, 2023; Ferguson, 2009).

Results

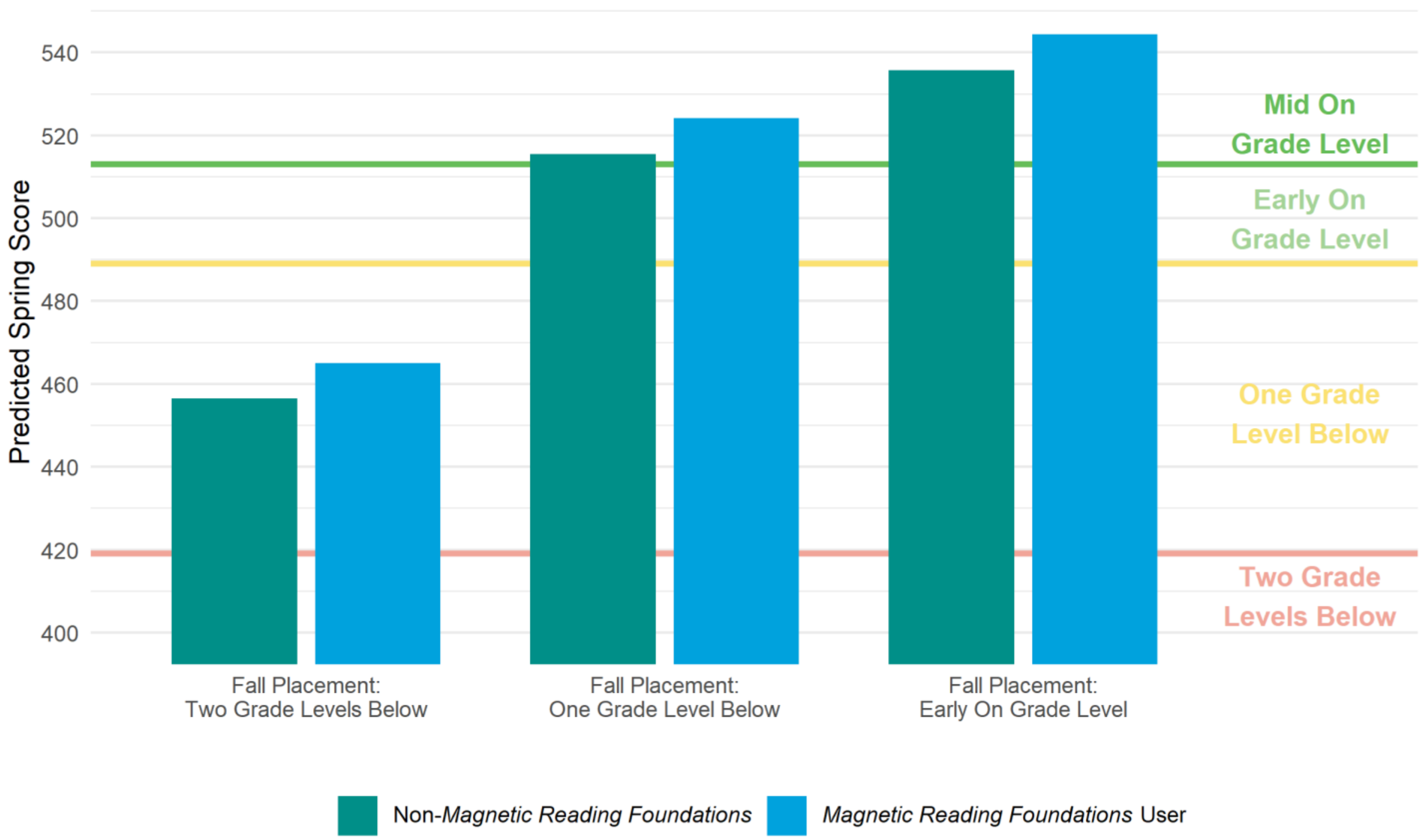
The use of *Magnetic Reading Foundations* had a positive and statistically significant effect on spring *i-Ready Diagnostic* scores after controlling for other important covariates. Students in schools that use *Magnetic Reading Foundations* scored about nine points higher (i.e., p = .02) than similar students in schools that did not use *Magnetic Reading Foundations*. This corresponds to an effect size of .15. Considering the features of this study and this intervention, this may be considered a moderate or large effect size (Cheung & Slavin, 2016; Kraft, 2020; Lipsey et al., 2012). Additional information about the results of the final model is available in the Appendix.

Conclusion

This study demonstrates that the use of *Magnetic Reading Foundations* in Grade 2 is related to higher overall Reading scores. Importantly, these differences were found in a matched sample of students after controlling for covariates that included the fall score on the same reading assessment used as the outcome, the number of days between assessments, and the amount of time spent on *i-Ready Personalized Instruction*. This was a rigorous quasi-experimental designed to meet ESSA Level 2 evidence standards. Importantly, the treatment sample was not restricted to students who received a certain "dosage" of the intervention, and the treatment schools carried out the curriculum in real-world conditions.

The nine-point increase in scores attributed to *Magnetic Reading Foundations* is practically significant for students and for schools. As shown in Figure 1, the use of *Magnetic Reading Foundations* moves students closer to or, in some cases, over the threshold to the next step of grade-level proficiency. For students, a difference of nine points translates to additional passages that can be understood, additional content that can be learned, additional tests that can be passed, and positive effects in reading skills and other areas for many years in the future (Fiester, 2010; Herbers et al., 2012; van Bergen et al., 2021). For schools, a difference of nine points means many additional students who are ready for grade-level work. For example, in the comparison sample from this study, if all students had scored nine points higher on the spring *i-Ready Diagnostic*, an additional 49 students (i.e., 6%) of the comparison group) would have placed on grade level.

Figure 1. Difference in Spring Scores Due to *Magnetic Reading Foundations* for Select Fall Scores



Note: The fall scores depicted here represent the highest score possible in the specified placement. The predicted spring scores assume a value of 0 after centering for all covariates aside from fall score and *Magnetic Reading Foundations* usage where indicated.

This study provided valuable information about the impact of *Magnetic Reading Foundations* usage for Grade 2 students under typical conditions. Future studies can examine the effects of *Magnetic Reading Foundations* in Grades K and 1. Because we expect different patterns of growth among students with particular educational needs, such as multilingual learners and students with disabilities, and among students with different fall placements, it will also be important to investigate whether *Magnetic Reading Foundations* has even more impact among particular student groups. Examination of the impact of *Magnetic Reading Foundations* on other assessments of reading and on assessments of reading domains (e.g., assessments of phonics knowledge) would also be informative. Additionally, because this study was undertaken in the first year *Magnetic Reading Foundations* was available, it was not possible to examine the effects of a more mature implementation. More experience with the program may lead to even more favorable outcomes. Notable limitations of this study include its quasi-experimental nature, the lack of information about how the program was implemented, and the lack of information about comparison conditions. Future research may utilize different methodologies or collect additional data to provide more information about the impact of this program.

References

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399-424. https://doi.org/10.1080/00273171.2011.568786

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5). https://doi.org/10.3102/0013189X16656615

Curriculum Associates. (2023a). *Magnetic Reading Foundations research base: Grades K-2.* https://www.curriculumassociates.com/research-and-efficacy/magnetic-research-base-grades-k-2

Curriculum Associates. (2023b). *State of student learning in 2023.* https://www.curriculumassociates.com/research-and-efficacy/annual-report-the-state-of-student-learning-in-2023

Evidence for ESSA. (2023). *Evidence for ESSA: Standards and procedures: Version 2.0.* https://evidenceforessa.org/wp-content/themes/evidenceforessa/uploads/FINAL_Standards_and_Proc%2005_18_23.pdf

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532-538. https://doi.org/10.1037/a0015808

Fiester, L. (2010). *Early warning! Why reading by the end of third grade matters.* Annie E. Casey Foundation. https://assets.aecf.org/m/resourcedoc/AECF-Early_Warning_Full_Report-2010.pdf

Herbers, J. E., Cutuli, J. J., Supkoff, L. M., Heistad, D., Chan, C., Hinz, E., & Masten, A. S. (2012). Early reading skills and academic achievement trajectories of students facing poverty, homelessness, and high residential mobility. *Educational Researcher, 41*(9), 366-374. https://doi.org/10.3102/0013189X12445320

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, *42*(8), 1–28. https://doi.org/10.18637/jss.v042.i08

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4). https://doi.org/10.3102/0013189X20912798

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. Institute of Education Sciences, National Center for Special Education Research. https://ies.ed.gov/ncser/pubs/20133000/pdf/20133000.pdf

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22*(1), 114-140. https://doi.org/10.1037/met0000078

National Assessment of Educational Progress. (2022). *NAEP report card: 2022 NAEP Reading Assessment.* https://www.nationsreportcard.gov/highlights/reading/2022

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*(1), 85-112. https://doi.org/10.1016/j.jsp.2009.09.002

R Core Team. (2022). *R: A language and environment for statistical computing*. R Project for Statistical Computing. https://www.R-project.org

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Wadsworth Cengage Learning.

van Bergen, E., Vasalampi, K., & Torppa, M. (2021). How are practice and performance related? Development of reading from age 5 to 15. *Reading Research Quarterly, 56*(3), 415-434. https://doi.org/10.1002/rrq.309

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0.* Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5.0-0-508.pdf

Whittaker, T. A., & Furlow, C. F. (2009). The comparison of model selection criteria when selecting among competing hierarchical linear models. *Journal of Modern Applied Statistical Methods, 8*(1), 173-193. https://doi.org/10.22237/jmasm/1241136840

Appendix

Table A.1. Parameter Estimates for the Outcome Model

Variable	Note	Unstandardized Estimate	Standard Error	*t*	*p*
Intercept		516.37	3.07	168.17	< .001
Hours of *i-Ready Personalized Instruction Usage*	Centered at 20	.22	.06	3.83	< .001
Fall Diagnostic Score	Centered at 489	.84	.02	51.81	< .001
Days between Diagnostics	Centered at 252	.37	.12	2.99	.003
State 2		25.15	6.51	3.87	< .001
State 3		18.09	5.17	3.50	.001
State 4		21.85	4.97	4.39	< .001
Magnetic Reading Foundations		8.59	3.51	2.44	.022