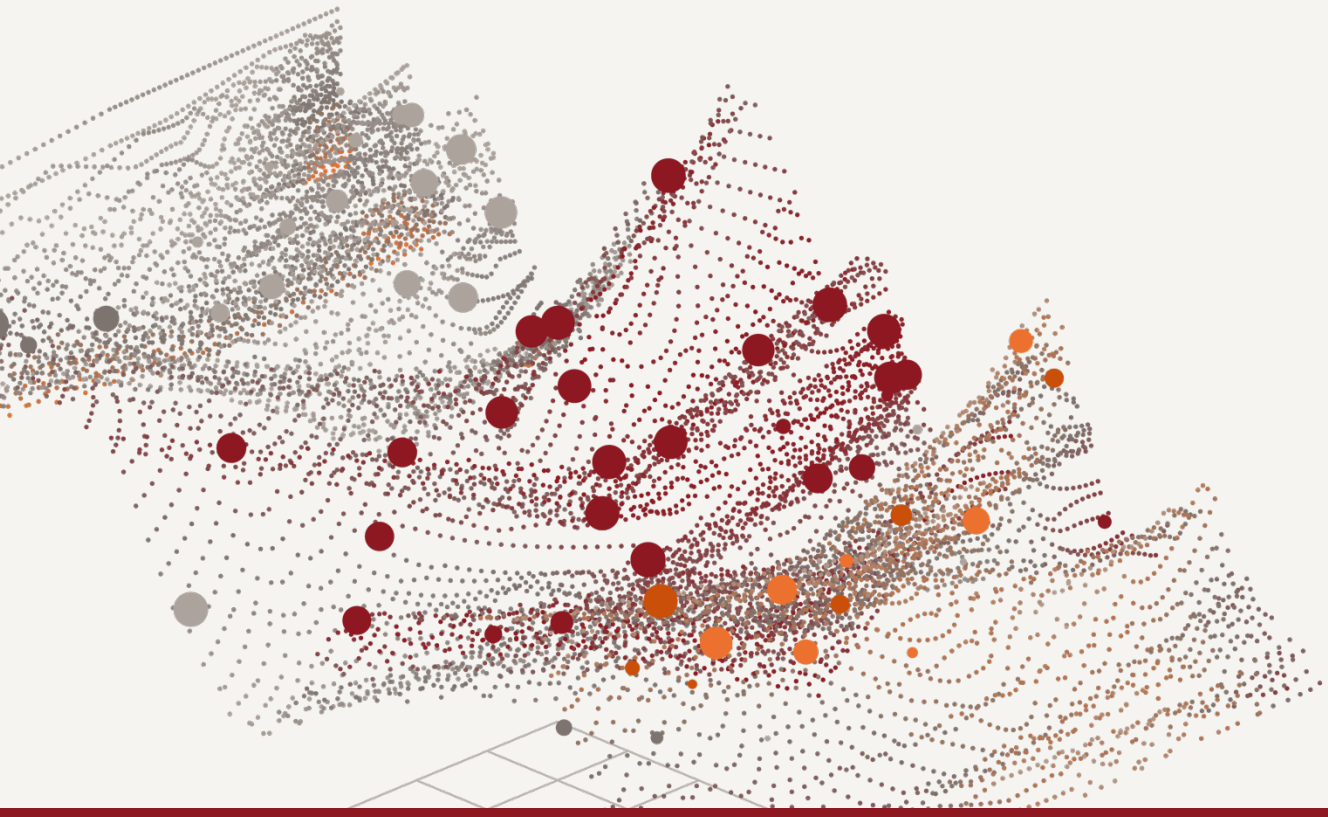
NORC Education

FINAL REPORT
MAY 2024

Magnetic Reading:
Effects on Reading Achievement for Grades 3-5

Presented by:
Neil Seftor,
NORC at the University of Chicago

Presented to:
Curriculum Associates



Abstract

Magnetic Reading is an evidence-based reading comprehension program for students in grades 3 through 5 built on four pedagogical pillars: knowledge-rich learning, culturally and linguistically responsive pedagogy, scaffolds to support learner variability, and data to inform instruction. This study uses a prospective matched comparison group quasi-experimental design to examine the impact of *Magnetic Reading*. At the beginning of the 2022-23 school year, we used coarsened exact matching to match students in schools that used *Magnetic Reading* to students in schools that did not using student demographic characteristics and a fall achievement test score. We subsequently conducted an analysis of spring state achievement test scores on the post-attrition sample after establishing that the groups were similar on baseline characteristics. Our HLM analysis found a statistically significant impact of 8.4 points—an effect size of 0.22 standard deviations—on reading scores.

Introduction

Magnetic Reading is an evidence-based reading comprehension program for students in grades 3 through 5. Students are intended to use the teacher-led reading program for 30-45 minutes daily as a component of their literacy instruction. It is built on four pedagogical pillars: knowledge-rich learning, culturally and linguistically responsive pedagogy, scaffolds to support learner variability, and data to inform instruction (Curriculum Associates, 2022).

- **Knowledge-rich learning**. *Magnetic Reading* explicitly builds knowledge thorough a curated series of coherent grade-level texts, adding to a student's stored background knowledge and vocabulary that they can use as they encounter new texts.
- **Culturally and linguistically responsive pedagogy**. *Magnetic Reading* encourages students to feel comfortable and excited to learn by highlighting the strengths of diverse students and engaging students through validating and affirming their diverse cultural backgrounds.
- **Scaffolds to support learner variability**. *Magnetic Reading* uses an asset-based pedagogical approach that employs frameworks for teachers to draw on to strategically suit the strengths and needs of their learners.
- **Data to inform instruction**. Using the *i-Ready Diagnostic* in conjunction with *Magnetic Reading* allows teachers to identify where students are so they can plan reading and standards-based instructional scaffolds to address students' individual needs.

Having only launched recently in fall 2021, the research base on *Magnetic Reading* is limited. One previous study found *Magnetic Reading* had a positive, statistically significant effect of 0.23 standard deviations on test scores for grade 3 through 5 students in schools that used the program compared to similar students in schools that did not (Holzman & Duncan, 2023).

The current study builds on these positive findings. It tests the effectiveness of *Magnetic Reading* in typical practice with varying usage and implementation, using a larger sample and a prospective matching design that includes student demographic characteristics. The primary research question to be addressed was: To what extent does *Magnetic Reading* impact the reading achievement of students in grades 3 through 5?

Study Design

This study uses a prospective matched comparison group quasi-experimental design to examine the impact of *Magnetic Reading*. At the beginning of the 2022-23 school year, we matched students in schools that used *Magnetic Reading* (treatment students) to students in schools that did not (comparison students) using student demographic characteristics and a fall achievement test score.[1]

[1] A prospective quasi-experimental design satisfies the requirements for review by Evidence for ESSA (Evidence for ESSA, 2024).

We subsequently conducted the analysis of spring achievement test scores on the post-attrition sample after establishing that the groups were similar on baseline characteristics.[2]

Data

We collected student data in two stages. First, we collected student demographic data, such as race, ethnicity, and economic disadvantage status from school administrative records. We linked these data to *i-Ready Diagnostic* scale scores from the beginning of the school year. We used fall *i-Ready Diagnostic* scale scores as the pre-intervention achievement measure for this study. All students completed the fall *i-Ready Diagnostic* during the standard fall testing window.

Second, we collected Iowa Statewide Assessment of Student Progress (ISASP) English Language Arts (ELA) scale scores from the end of the school year. We used spring ISASP ELA scale scores as the post-intervention achievement measure for this study. Fall *i-Ready Diagnostic* and spring ISASP ELA scores had a correlation of 0.84, suggesting that the fall *i-Ready Diagnostic* scores were an acceptable pre-intervention proxy for the analysis.

Assessments

The *i-Ready Diagnostic* is an adaptive, online assessment designed to place students in relation to grade-level standards and national norms. Scale scores range from 100-800.

The ISASP assessment is the statewide summative, standardized assessment completed by all Iowa students in the spring of each school year. Scale scores range from 345-590.

Sample

The study was conducted on grade 3 through 5 students in four treatment and 30 comparison elementary schools across Iowa.[3] We assigned students to either the treatment or the comparison group based on school usage of the intervention. All students assigned to the treatment group are analyzed as if they received the intervention, regardless of whether, or how much, they used *Magnetic Reading*—an intent-to-treat (ITT) estimate. By analyzing participants based on their assigned treatment group, irrespective of their adherence to it, we obtain an unbiased estimate of the effect of the curriculum. Table 1 presents characteristics of the two groups prior to matching.

[2] Demonstrating equivalence on baseline characteristics for the analytic sample satisfies the requirements for both Evidence for ESSA and the What Works Clearinghouse (What Works Clearinghouse, 2022).

[3] The large pool of comparison schools allows matches to be identified for nearly all treatment students in the sample. At the same time, the study was conducted in one state (Iowa) which limits the generalizability of the findings beyond schools and students in similar environments.

Table 1. Characteristics of Unmatched Study Sample, by Condition

Characteristic	Comparison	Treatment	Difference	Effect Size	Significance
Male	51.6%	50.2%	1.4%	0.03	
White	59.1%	75.3%	-16.2%	0.33	***
Hispanic	9.2%	22.6%	-13.4%	0.44	***
Economic disadvantage	21.4%	28.7%	-7.3%	0.18	***
Grade 4	33.1%	36.0%	-3.9%	-0.06	
Grade 5	32.9%	33.1%	-0.1%	0.00	
Fall *i-Ready* Diagnostic score	509.01	523.58	-14.57	0.23	***
Number of students	4,016	478			

Note: *** represents $p < 0.001$.

Matching

After student characteristics and fall achievement were obtained, but before ISASP ELA scores were available, we used coarsened exact matching (CEM) to create analysis groups that were balanced on pre-intervention achievement and other important characteristics. Like other matching methods, CEM is designed to create treatment and comparison groups that are similar on important, observed characteristics to reduce the confounding influence of those measures on the outcome.

We selected CEM for this study because it could ensure a close match on the pre-intervention measure while allowing matching on other characteristics, it is superior to propensity score and other matching methods in reducing imbalance and model dependence, and it could retain a large proportion of students in the initial sample. Using CEM, continuous variables are coarsened into categorical variables and then all possible combinations of the coarsened measures are used to stratify the sample. Students in strata that contain both treatment and comparison students are assigned weights according to the ratio of students in the stratum. In the analysis, the continuous variables are used along with the weight obtained through CEM. We matched students on the following characteristics which have been shown to be predictive of student achievement: fall *i-Ready Diagnostic* score; grade; male; white; Hispanic; and economic disadvantage.

Attrition and Equivalence

After the collection of spring ISASP ELA scores, we evaluated the matched sample for attrition—students who were matched using CEM but subsequently did not have a spring test score. Table 2 presents the loss of sample and corresponding rates of attrition.

Table 2. Attrition Due to Missing Spring Test Score, by Condition

Sample	Comparison	Treatment	Total
Matched treatment and comparison	2,628	473	3,101
Missing spring test score	31	3	34
Attrition	**1.2%**	**0.6%**	**1.1%**

Table 3 presents the baseline characteristics and pre-intervention achievement of students in the matched, post-attrition analysis sample. There are no significant differences between the two groups and the largest difference is 0.01 standard deviations. By creating groups that are similar on these observed characteristics at baseline (at the beginning of the school year), we are better able to attribute differences in outcomes (at the end of the school year) to the curriculum.[4]

Table 3. Characteristics of Post-Attrition Analysis Sample, by Condition

Characteristic	Comparison	Treatment	Difference	Effect Size	Significance
Male	51.7%	51.7%	0.0%	0.00	
White	80.4%	80.4%	0.0%	0.00	
Hispanic	9.8%	9.8%	0.0%	0.00	
Economic disadvantage	17.8%	17.8%	0.0%	0.00	
Grade 4	32.7%	32.7%	0.0%	0.00	
Grade 5	31.4%	31.4%	0.0%	0.00	
Fall *i-Ready* Diagnostic score	521.73	521.01	0.72	0.01	
Number of students	2,597	470			

Note: No differences are statistically significant at the $p < 0.05$ level.

[4] By its nature, a quasi-experimental design is limited to accounted for observed differences between treatment and comparison schools and students and cannot account for unobserved differences that may affect the outcome of interest.

Analysis

We conducted the analysis using a two-level hierarchical linear model (HLM) in which students were nested within schools. Within the two-level HLM framework, the ITT effect of *Magnetic Reading* can be estimated by:

$$\text{Level 1 (student):} \quad Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \beta_2 G_{ij} + \varepsilon_{ij}$$
$$\text{Level 2 (school):} \quad \beta_{0j} = \gamma_{00} + \gamma_{01}T + \gamma_{02}W_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$

where Y_{ij} is the outcome of interest for student *i* in school *j*; *T* is an indicator for the treatment; X_{ij} includes background characteristics of students and their prior academic outcomes; G_{ij} are grade indicators; W_j are school indicators; and ε_{ij} and u_{0j} are normally-distributed errors with mean 0. The parameter estimate, γ_{01}, provides unbiased covariate-adjusted estimates of the effect of *T*. We calculate standardized effect sizes by dividing the impact estimate by the pooled standard deviation (SD) derived from the unadjusted sample SDs for the outcome pooled across the groups (Hedges's *g*) and by the unadjusted sample SD of the comparison group only (Glass's delta).

Results

Magnetic Reading was found to have a strong statistically significant impact on reading scores, as measured by the state assessment. Accounting for student characteristics and prior achievement, students attending a school with *Magnetic Reading* scored, on average, 8.4 points higher ($p < 0.01$) on the ISASP ELA test compared to similar students in schools where *Magnetic Reading* was not used.[5] Table 4 presents the analysis findings, demonstrating the positive, statistically significant effect of *Magnetic Reading* on the spring ISASP ELA test.

Table 4. Impact of *Magnetic Reading* on Spring Reading Achievement

Covariate	Coefficient	Std Error	*p*-value
Treatment	**8.38**	**2.70**	**0.002**
Male	-2.71	1.01	0.007
White	2.96	2.27	0.192
Hispanic	0.20	2.04	0.921
Economic disadvantage	-3.18	1.90	0.094

[5] The unadjusted mean scores on the ISASP ELA spring test were 442.62 for the treatment group and 431.52 for the comparison group, a difference of 11.1 points.

Covariate	Coefficient	Std Error	*p*-value
Grade 4	7.83	1.89	0.000
Grade 5	15.72	1.59	0.000
Fall *i-Ready* Diagnostic score	0.53	0.02	0.000
Number of students	3,067		

The finding translates to an effect size of 0.22 for both Glass's delta (comparison standard deviation = 37.43) and Hedges's *g* (pooled standard deviation = 37.86).

Discussion

This study adds to the evidence base on the effectiveness of *Magnetic Reading*. Replicating and expanding on previous research (Holzman & Duncan, 2023), we used a prospective matched comparison group quasi-experimental design to estimate the impact of *Magnetic Reading* on the reading achievement of students in grades 3 through 5. We found a positive, statistically significant effect of 0.22 standard deviations, consistent with previous research (Holzman & Duncan, 2023). Next steps in expanding the evidence base further would include conducting a similar study in a different setting or using a more rigorous design, such as a randomized controlled trial.

References

Curriculum Associates. (2022). *Magnetic Reading Research Base.* Curriculum Associates.

Evidence for ESSA. (2024). *Evidence for ESSA: Standards and Procedures, Version 2.2.* Evidence for ESSA.

Holzman M. & Duncan. M. (2023). *Impact of Magnetic Reading in Iowa Schools: Evidence from the 2021-2022 School Year.* Curriculum Associates.

What Works Clearinghouse. (2022). *Procedures and Standards Handbook, Version 5.0.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.