

**Effectiveness of McGraw-Hill's  
*Jamestown Reading Navigator*  
in Grades 9–10:**

A Study of Intensive Reading Classes  
in Miami-Dade High Schools

Valeriy Lazarev

Sara Atienza

Garrett Lai

Boya Ma

Laurel Sterling

Empirical Education Inc.

October 1, 2010



## Executive Summary

**Background.** A quasi-experimental study designed to provide evidence of the effectiveness of the *Jamestown Reading Navigator (JRN)* program for students in grades 9–10 was conducted in the Miami Dade Public School District (MDCSD). *Jamestown Reading Navigator* is a reading intervention program developed by McGraw-Hill Education to raise reading competencies of struggling middle school and high school students. The publishers state that the program “motivate[s] reluctant readers with highly engaging online and print-based content written exclusively for adolescents” (*Jamestown Reading Navigator*, 2009). *JRN* was implemented in MDCSD during the 2008–2009 school year. It is used in Intensive Reading (IR) classes enrolling students who are reading two or more levels below their grade.

The current study tests two research questions:

- Do students participating in the *Jamestown Reading Navigator* program attain higher reading achievement scores than comparable students who are not participating in *JRN*?
- Are there discernible differences in the size of impact on children of different gender, ethnicity, disability, English learner status, and pretest score?

**Study Design.** The study uses a quasi-experimental comparison group design to test the effectiveness of *JRN* by comparing outcomes for students who used *JRN* to those who did not use *JRN*, adjusting for differences between the *JRN* and comparison groups on baseline characteristics.

**Source Data.** The primary data for this study come from two sources: 1) student records from the 2007–2008 and the 2008–2009 school years, which were provided by the Miami-Dade school district, 2) a list of teachers who had used *JRN* during the 2008–2009 school year, provided by McGraw-Hill. The student records include demographic information such as gender, ethnicity, English proficiency, disability, and socio-economic status as measured by participation in the subsidized school lunch program. Two years of reading test scores and the English/language arts teacher’s name were also included in the MDCSD student records.

**Sample Selection.** To improve its quality, the sample was reduced in the following ways: At the school level, the sample was limited to the 40 (out of 70) schools that had both *JRN* and non-*JRN* IR classes. At the class level, the sample was limited to remedial reading classes termed Intensive Reading (IR) by MDCSD. At the individual student level, only records for the students enrolled in an IR class for both semesters of 2008-09 were included in the analysis. Student records with missing data elements were removed. The size of the resulting sample, by grade level, is shown in the table above.

|                 | Number of students |             |
|-----------------|--------------------|-------------|
|                 | <i>JRN</i>         | Comparison  |
| <b>Grade 9</b>  | 1769               | 862         |
| <b>Grade 10</b> | 3391               | 1475        |
| <b>Total</b>    | <b>5160</b>        | <b>2337</b> |

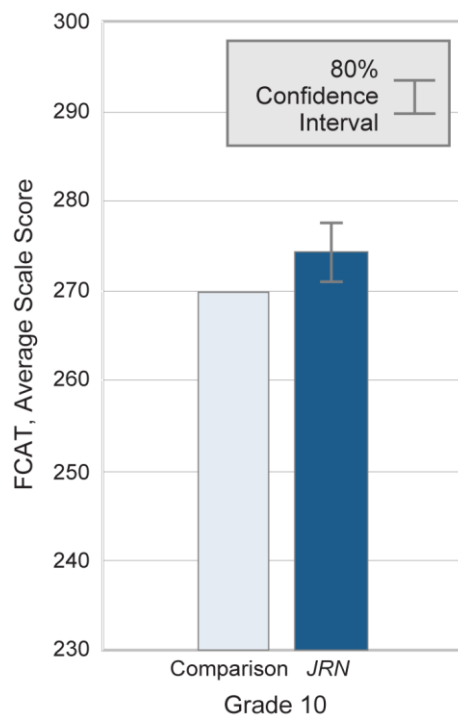
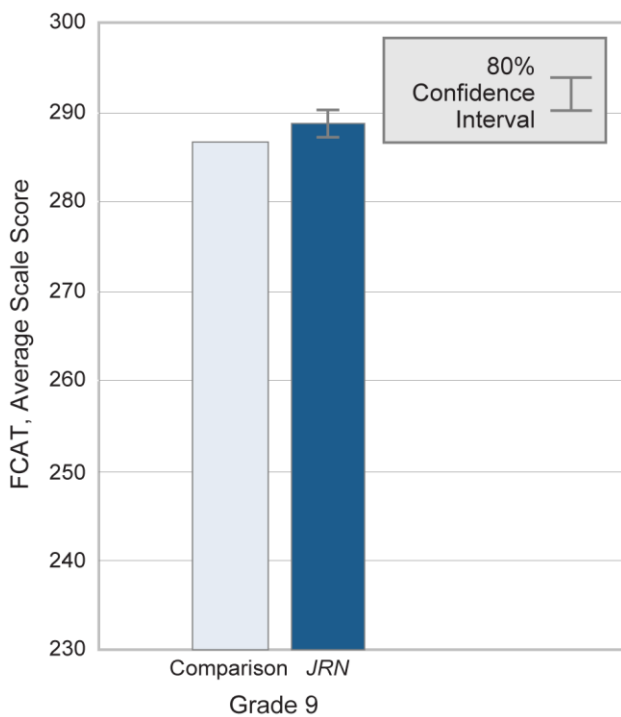
**Outcome Measure.** The impact of *JRN* was evaluated using student reading scores from the Florida Comprehensive Assessment Test (FCAT). Spring 2009 FCAT reading scores were used as the outcome measure, while Spring 2008 FCAT reading scores were used as a pretest measure to control for possible differences in the background between *JRN* and comparison students.

**Analytical Methods.** The data analysis was performed using a multi-level linear mixed model, producing the estimates of *JRN* impact adjusted for the differences in student-level covariates (gender, ethnicity, disability, English learner and socio-economic status, and pretest scores) and for clustering of students by teacher and school. In addition, moderator analyses were performed to explore subgroup differences in the effectiveness of *JRN*. All estimates were performed for 9<sup>th</sup> and 10<sup>th</sup> graders separately to accommodate for differences in the test content and potential differences in factors determining learning patterns and selection into IR classes between the two grade levels.

**Results.** The study found that *JRN* had a positive impact on IR student reading scores both in grade 9 and grade 10. The following table gives the size of the difference between control and *JRN* expressed in units of the standard deviation. Both estimates are given a *p* value, which is a measure of the probability that we would see a difference with an absolute value as large or larger than the one observed with the sample in this study if in actuality there was no difference.

|                 | Effect size | <i>p</i> value |
|-----------------|-------------|----------------|
| <b>Grade 9</b>  | 0.06        | .092           |
| <b>Grade 10</b> | 0.11        | <.001          |

A *p* value under .05 gives a high level of confidence in the result, while a *p* value above .05 but under .2 gives us some confidence in the result. We therefore have a high level of confidence in the result estimated for the 10<sup>th</sup> grade (*p* value less than .001) and have some confidence in the 9<sup>th</sup> grade result (*p* value equals .092).



Moderator analysis revealed that *JRN* has a positive differential impact on 10<sup>th</sup> graders with disabilities with effect size of 0.1, which implies that the effect size for students with disabilities using *JRN* was about twice the size as for otherwise comparable students without disabilities. We have limited confidence in this result (*p* value = .18). No significant moderating effects of other student characteristics or pretest were established, and no significant differential impact on 9<sup>th</sup> graders with disabilities was identified.

**Conclusion.** This study demonstrates that *JRN* had a positive impact on student achievement in reading for students in the Miami Dade Public School District’s Intensive Reading classes. It also indicates a potential for *JRN* to benefit students with disabilities. Data availability limited the impact evaluation to the first year of program implementation and precludes assessing *JRN*’s long term benefits.

# Table of Contents

|  |          |
|--|----------|
| <b>BACKGROUND</b> .....  | <b>1</b> |
| <b>METHODS</b> .....   | <b>1</b> |
| <b>STUDY DESIGN</b> .....  | <b>1</b> |
| <b>DATA</b> .....  | <b>2</b> |
| <b>ANALYTICAL SAMPLE SIZE AND CHARACTERISTICS</b> .....                                  | <b>2</b> |
| <i>Table 1. Size of the Analytical Sample</i> .....                                      | <i>3</i> |
| <i>Table 2. Characteristics of Students Included in the Analytical Sample</i> .....      | <i>4</i> |
| <b>ANALYTICAL METHODS</b> .....  | <b>4</b> |
| <b>RESULTS</b> .....   | <b>5</b> |
| <b>AVERAGE PROGRAM EFFECT</b> .....  | <b>5</b> |
| <i>Table 3. Estimated Effect of JRN in Grades 9 and 10</i> .....                         | <i>5</i> |
| <i>Figure 1. FCAT Reading Assessment for Grade 9 IR Students</i> .....                   | <i>6</i> |
| <i>Figure 2. FCAT Reading Assessment for Grade 10 IR Students</i> .....                  | <i>6</i> |
| <b>MODERATOR ANALYSIS</b> .....  | <b>6</b> |
| <i>Figure 3. Estimated Effect of Disability Status for 10<sup>th</sup> Graders</i> ..... | <i>7</i> |
| <b>CONCLUSION</b> .....  | <b>7</b> |









## Background

The goal of this study was to determine the effectiveness of the *Jamestown Reading Navigator (JRN)* Program for students in grades 9 and 10 in the Miami Dade Public School District (MDCSD). *JRN* was implemented in MDCSD during the 2008–2009 school year. It is used in Intensive Reading (IR) classes enrolling students who are reading two or more levels below their grade. For each study school, performance of students in IR classes who used *JRN* was compared to the performance of students in IR classes who did not use *JRN*. Florida Comprehensive Assessment Test (FCAT) Reading scores are the outcome measures.

*Jamestown Reading Navigator (JRN)* is a reading intervention program designed specifically for students in grades 6–12 who are reading two or more levels below their grade level. The publishers state that the program “motivate[s] reluctant readers with highly engaging online and print-based content written exclusively for adolescents” (*Jamestown Reading Navigator*, 2009). The online component directly instructs, assesses, and provides practice for students on vocabulary, reading comprehension, writing, and oral fluency. Each level in the online component is created to provide scaffolded instruction that moves students from a highly structured environment with explicit instruction toward more independent reading. The *Jamestown* print-based component provides teachers with resources to construct lessons that reinforce and reteach the skills that students receive through the online program. The print-based component consists of a set of leveled inClass Reader textbooks, inTime student magazines, a set of *Jamestown* lesson plans that are linked to *Jamestown* blackline masters, and a set of *Jamestown* teacher support binders.

The study tests two research questions:

- Do students participating in the *Jamestown Reading Navigator* program attain higher reading achievement scores than comparable students who are not participating in *JRN* ?
- Are there discernible differences in the size of impact on children of different gender, ethnicity, disability, English learner status, and pretest score?

## Methods

### Study Design

This study uses a quasi-experimental comparison group design, where the effectiveness of the *JRN* reading program is estimated by comparing the reading achievement of students enrolled in Intensive Reading (IR) classes who used the *JRN* reading program to the reading achievement of students who did not use *JRN*, adjusting for differences between the two student groups on baseline characteristics. The study focuses on reading comprehension performance of students in grades 9–10 as measured by their FCAT Reading scores. The unit of analysis in this study is the student. The level of assignment was assumed to be the class. No information reflecting the actual use of *JRN* materials by individual students was available.

Quasi-experimental studies are used to estimate the impact of a program in situations where the program has already been implemented and the data pertaining to its impact may already be available for analysis or when randomized assignment is impossible or undesirable. The major challenge in a quasi-experimental study is that program and comparison groups may systematically differ in terms of background characteristics that affect performance. Unless the estimate of the program effect is adjusted for the effects of these covariates, the result will be inaccurate or biased.

To minimize the bias in a quasi-experimental study, a sample selection procedure was developed and implemented to ensure that the program and comparison groups possess statistically similar properties. Analysis of the available data for the 2008–09 school year showed that some of the high schools in the Miami-Dade school district had not adopted *JRN*. Researchers were concerned that bias might occur due to unobservable characteristics that had led these schools to not adopt the

program. Because the schools that adopted *JRN* contained both *JRN* and comparison group teachers, we decided to use data from only these schools in the analysis.

An analytical strategy based on the use of linear mixed models is employed that further reduces the potential for bias by controlling for the effects of important covariates, including the pretest, and adjusting for clustering of students in classes and schools and differences in achievement across schools.

## Data

The primary data for this study come from two sources: 1) student records from the 2007–2008 and the 2008–2009 school years, which were provided by the Miami-Dade school district, and 2) a list of teachers who had used *JRN* during the 2008–2009 school year, which was provided by McGraw-Hill. The student records include student characteristics, two years of reading test scores, and the English/language arts teacher's name.

The following student characteristics were provided by MDCSD.

- Gender
- Ethnicity (African American, Hispanic, White/Non-Hispanic, Asian, Native American, or Mixed)
- English proficiency (English Language Learner status)
- Disability status
- Enrollment in the National School Lunch Program (proxy for socio-economic status)

All these variables were included in the analysis. With respect to ethnicity, a poolability analysis was performed, which showed that only 'African American' had a statistically significant effect on student achievement. The categories of the ethnicity variable were therefore aggregated into a binary indicator "is African American" for the purposes of the analysis.

Identification of program participation was performed using nominal record linkage of student data and usage logs obtained from the *JRN* program. For the lack of actual classroom information, class was defined as all students taught by one teacher, and teachers were identified on the basis of their recorded names and school affiliation. Intensive Reading teacher names from the student records were matched to the names on the *JRN* usage logs. Teachers identified as a result were marked as *JRN* teachers and all students associated with such teachers were marked as *JRN* students, while the rest were marked as comparison group students.

The impact of *JRN* was evaluated using student reading scores from the FCAT. Spring 2009 FCAT reading scores were used as the outcome measure, while Spring 2008 FCAT reading scores were used as a pretest measure. Because FCAT scores vary widely across grade levels, schools, and over time, the scores were standardized by subtracting school averages for the given year and grade level and dividing by the standard deviation. Score standardization also eliminated significant differences in the average pretest between the *JRN* and comparison groups.

## Analytical Sample Size and Characteristics

Program impact analysis in quasi-experimental studies requires that cases are selected to ensure the highest possible reliability of the data and balance between the comparison and program groups. To create maximally favorable conditions for accurate program impact analysis, three decisions were made concerning reducing the sample size, while increasing its quality.

At the school level, the sample was limited to the 40 (out of 70) schools that had both *JRN* and non-*JRN* Intensive Reading classes. In the remaining 30 schools, either all classes were using *JRN* or *JRN* was not used at all. This decision was aimed at reducing the potential bias from self-selection of schools into program participation.

At the class level, the sample was limited to Intensive Reading classes based on information from the district that *JRN* was implemented in only IR classes (email communication with the Instructional Supervisor at MDCPS, 5/5/2010.) In addition, unusually small classes (with fewer than 12 students)—most of which were special education classes (classes consisting of only students with disabilities)—were excluded. Exclusion of students from classes other than IR allowed for a more transparent study design and limited the data collection requirements. Excluding atypical IR classes reduced excess variability in the data and limited the impact of anomalous cases on the estimates.

At the individual student level, only records for the students enrolled in an IR class for both semesters of 2008-09 were included in the analysis. This eliminated ambiguity in terms of whether a student belonged in the program or comparison condition. Student records with missing data elements were removed. Limiting the sample this way reduces the measurement error that creates a potential downward bias in the program impact estimate.

These additional requirements, in conjunction with limitations in data availability, reduced the sample used in the analysis to 23 schools with 72 *JRN* and 56 comparison teachers and a total of 7,497 students (5,160 in grade 9 and 2,337 in grade 10, respectively). The breakdown of students and teachers in the analytical sample by grade level and condition appears in Table 1.

**Table 1. Size of the Analytical Sample**

|                 | Number of students |             | Number of teachers    |                       |
|-----------------|--------------------|-------------|-----------------------|-----------------------|
|                 | <i>JRN</i>         | Comparison  | <i>JRN</i>            | Comparison            |
| <b>Grade 9</b>  | 1769               | 862         | 30                    | 24                    |
| <b>Grade 10</b> | 3391               | 1475        | 52                    | 38                    |
| <b>Total</b>    | <b>5160</b>        | <b>2337</b> | <b>72<sup>a</sup></b> | <b>56<sup>a</sup></b> |

<sup>a</sup> The sums of grade level numbers exceed the respective column totals because some teachers teach at both grade levels.

This analytical sample is characterized by the baseline equivalence of the student characteristics using the threshold level of 0.25 standard deviations as set by the guidelines for the federal What Works Clearinghouse (Institute of Educational Studies, 2008, p.15). The differences between the *JRN* and comparison groups were under 0.25 of their standard deviations in all cases except one: the proportion of 9<sup>th</sup> grade African American students differed by more than 0.25 standard deviations (see Table 2). The differences in the mean pretest scores, although substantial, were due in part to the differences in test performance across schools in the sample, and the imbalance was reduced by the standardization of scores by school, as previously described.

**Table 2. Characteristics of Students Included in the Analytical Sample**

|   | <i>JRN</i> | Comparison | Pooled standard deviation | Difference between means in units of standard deviation |
|---|------------|------------|---------------------------|---|
| <b>Grade 9</b>  |            |            |                           |   |
| Male (%)  | 51.61      | 51.50      | 49.98                     | <0.01   |
| African American (%)  | 32.05      | 54.64      | 48.88                     | 0.46  |
| Students with disabilities (%)                              | 7.97       | 12.06      | 29.06                     | 0.14  |
| English proficient (%)                                      | 99.89      | 99.30      | 5.51                      | 0.11  |
| Low SES (enrolled in the National School Lunch Program) (%) | 67.10      | 67.05      | 46.99                     | <0.01   |
| Pretest, FCAT reading scale score (mean)                    | 277.59     | 264.45     | 29.62                     | 0.44  |
| <b>Grade 10</b>   |            |            |                           |   |
| Male (%)  | 49.45      | 54.37      | 49.99                     | 0.10  |
| African American (%)  | 40.22      | 46.10      | 49.36                     | 0.12  |
| Students with disabilities (%)                              | 9.88       | 16.68      | 32.43                     | 0.21  |
| English proficient (%)                                      | 99.91      | 99.80      | 3.51                      | 0.03  |
| Low SES (enrolled in the National School Lunch Program) (%) | 62.99      | 60.27      | 48.50                     | 0.06  |
| Pretest, FCAT reading scale score (mean)                    | 282.50     | 266.19     | 32.56                     | 0.50  |

## Analytical Methods

The data analysis was performed using a multi-level linear mixed model to estimate the effect of the *JRN* impact, adjusting for differences between the program and comparison groups on baseline covariates and accounting for clustering of students in classes and schools. Variation in baseline characteristics across schools and classes was modeled using a random effects approach. Program impact estimates were obtained using the implementation of linear mixed models in the *R*-language package *lme4* (Bates, 2010).

All estimates were performed for 9<sup>th</sup> and 10<sup>th</sup> graders separately to accommodate differences in the test content and potential differences in factors which determine learning patterns and selection into IR classes between the two grade levels.

In addition to the estimation of the average program effect, moderator analyses were performed to explore potential subgroup differences in the effectiveness of *JRN*. All student-level covariates and

school-level averages of student characteristics were tested as potential moderator variables by including a corresponding interaction term. Each moderator analysis involved only one interaction term. Due to relatively small cell sizes, especially in the comparison group, no higher-order interactions (such as a joint moderating effect of gender and ethnicity on the program effect) were tested.

## Results

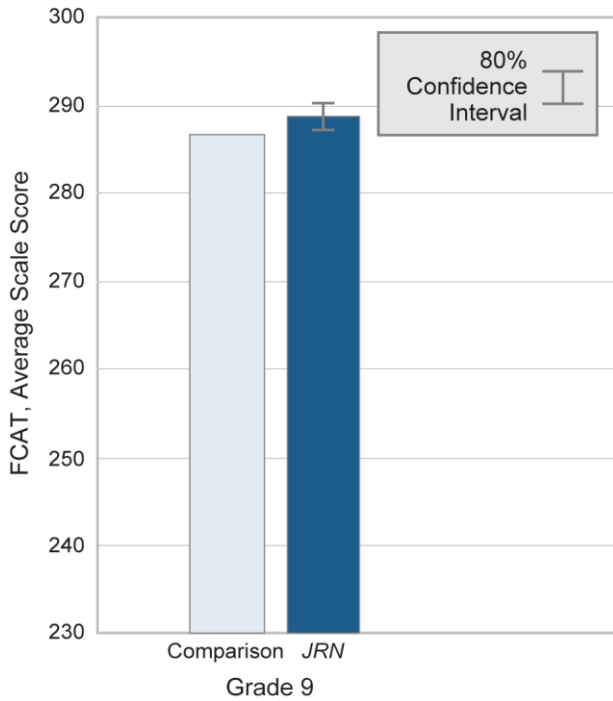
### Average Program Effect

Table 3 shows that the study found a positive impact of *JRN* on IR students' reading scores both in grade 9 and grade 10. The effect size (in units of standard deviation) was higher for the 10<sup>th</sup> grade (0.11) than for the 9<sup>th</sup> grade (0.06). Test scores of 9th graders enrolled in *JRN* classes were 2.5 percentiles higher than those of the comparable students enrolled in comparison IR classes. For the 10<sup>th</sup> graders, test scores were equivalent to 4.3 percentiles of the score distribution. With a *p* value of less than .001, we have strong confidence that the result for 10<sup>th</sup> grade is not a matter of chance. For 9<sup>th</sup> grade, with a higher *p* value, we have some confidence that the result is not a matter of chance factors.

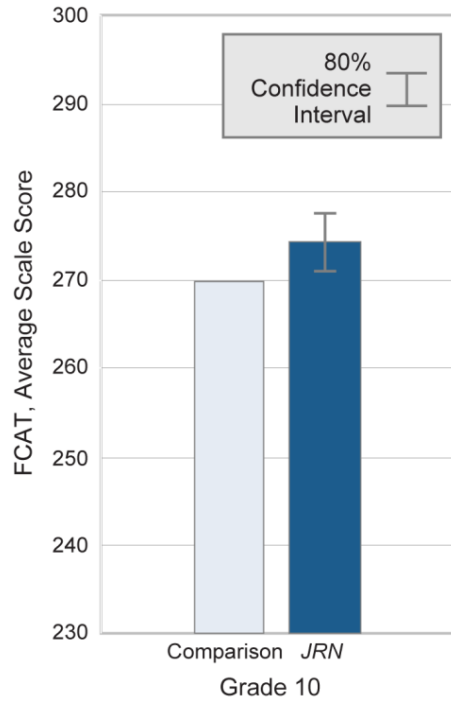
**Table 3. Estimated Effect of *JRN* in Grades 9 and 10**

|                 | Impact on reading (FCAT) | Effect size | <i>p</i> value |
|-----------------|--------------------------|-------------|----------------|
| <b>Grade 9</b>  | 2.06                     | 0.06        | .092           |
| <b>Grade 10</b> | 4.32                     | 0.11        | <.001          |

The bar graphs in Figure 1 and Figure 2 present the effect of *JRN* on IR students in grades 9 and 10 respectively in terms of FCAT scale scores. The left (light blue) bars in each of the two diagrams show the actual average FCAT scale scores for the control group and the right (darker) bars show average FCAT scale scores for the *JRN* group adjusted for the differences in student characteristics between the two groups. The height of the bars corresponding to *JRN* students can be interpreted as the estimated average FCAT score that would have been achieved by control group students if they had been enrolled in *JRN* classes in 2008-2009. Brackets on top of the bars corresponding to the *JRN* group represent the 80% confidence intervals for the program effect estimates.



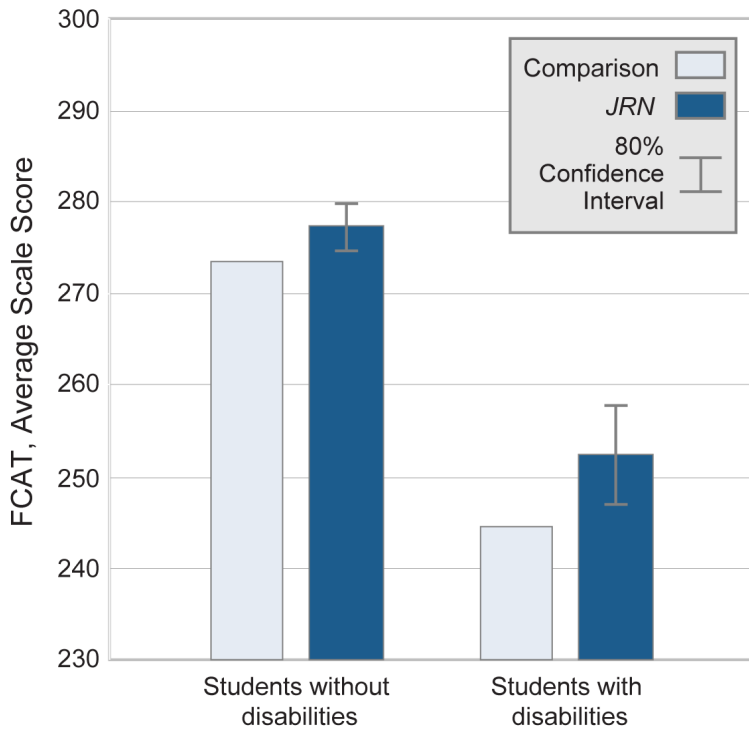
**Figure 1. FCAT Reading Assessment for Grade 9 IR Students**



**Figure 2. FCAT Reading Assessment for Grade 10 IR Students**

## Moderator Analysis

A moderator analysis suggested that *JRN* may have had a positive differential impact on the reading scores of 10<sup>th</sup> graders with disabilities, with an effect size of 0.1. This means that program effects among *JRN* students with disabilities may be twice as high as program effects among *JRN* students without disabilities. We have a limited level of confidence in this result with the *p* value equal to .18. Figure 3 presents the results of the moderator analysis involving the disability status by separating the average program effect on students with and without disability.



**Figure 3. Estimated Effect of *JRN* by Disability Status for 10<sup>th</sup> Graders**

No significant moderating effects of other student characteristics or pretest scores were established, and no significant differential impact on 9<sup>th</sup> graders with disabilities was identified. The lack of a statistically significant estimate of the moderating effect of disability in the 9<sup>th</sup> grade could be a consequence of the smaller number of students with disabilities in the 9<sup>th</sup> grade compared to the 10<sup>th</sup> grade.

## Conclusion

This study shows that *JRN* has a positive impact on reading achievement in the Intensive Reading classes in MDCSD’s high schools. It also suggests a potential for *JRN* to benefit students with disabilities. While the effect for 10<sup>th</sup> grade appeared stronger than for 9<sup>th</sup> grade, these differences should be taken with caution because the number of 9<sup>th</sup> graders in the analytical sample was substantially lower than the number of 10<sup>th</sup> graders. This caution also applies to the appearance of a moderating effect of disability in 10<sup>th</sup> but not 9<sup>th</sup> grade.

An important limitation of the study was that the data were available for the first year of the program implementation only. The study does not reflect the potentially greater impact of longer term usage for teachers or the potential cumulative effect of multiple years of exposure to the *JRN* program for the students. Another limitation of this study is related to the lack of teacher characteristics that may have influenced the choice to adopt *JRN* by some teachers but not others.

While this study provides the Miami-Dade schools with evidence of the benefits of *JRN*, we must be cautious in generalizing these results to other districts, especially where there are different populations. Additional quasi-experimental studies of *JRN* implementation across a variety of school districts with a greater geographical coverage, a more balanced representation of grade levels, and a larger number of years after the program adoption would have the potential to produce more accurate estimates of the impact of *JRN* on the reading performance of students reading two or more levels below grade level.

## References

- Bates, D. (2010, August). *Linear mixed model implementation in lme4*. University of Wisconsin-Madison. Retrieved June 2010 from <http://cran.r-project.org/web/packages/lme4/vignettes/Implementation.pdf>
- Florida Department of Education. (2009). *Florida Comprehensive Assessment Test (FCAT): Sunshine State Standards, State Report of District Results: Grade 9, Reading, 2009* [Excel data file]. Retrieved August 2010 from [http://fcats.fldoe.org/xls/2009/F09\\_GR09\\_RDIS.XLS](http://fcats.fldoe.org/xls/2009/F09_GR09_RDIS.XLS)
- Florida Department of Education. (2009). *Florida Comprehensive Assessment Test (FCAT): Sunshine State Standards, State Report of District Results: Grade 10, Reading, 2009* [Excel data file]. Retrieved August 2010 from [http://fcats.fldoe.org/xls/2009/F09\\_GR10\\_RDIS.XLS](http://fcats.fldoe.org/xls/2009/F09_GR10_RDIS.XLS)
- Institute of Educational Studies. (2008, May). *What Works Clearinghouse Evidence Standards for Reviewing Studies, Version 2.0*, (p 15). Retrieved June 2010 from [http://ies.ed.gov/ncee/wwc/pdf/wwc\\_procedures\\_v2\\_standards\\_handbook.pdf](http://ies.ed.gov/ncee/wwc/pdf/wwc_procedures_v2_standards_handbook.pdf)
- Jamestown Reading Navigator. (2009). Retrieved August 2010 from [http://www.readingnavigator.com/mkt/po/po\\_main.html](http://www.readingnavigator.com/mkt/po/po_main.html)