Striving Readers Project
New York State Department of Education/New York City Department of Education

REPORT OF INTENT TO TREAT ESTIMATES OF PROGRAM IMPACTS ON STUDENT ACHIEVEMENT:

GATES-MACGINITIE READING TESTS

June 2012

Project Staff*:*
Dianna L. Newman, PhD, Principal Investigator; Deborah K. Kundert, PhD, Co-Principal Investigator

Assisted by:
Richard F. Haase, PhD and Tierney A. Gifford, MS

The Evaluation Consortium
University at Albany, State University of New York
Dutch Quadrangle Beverwyck B-7
1400 Washington Avenue
Albany, New York 12222
(518) 442-5027

REPORT OF INTENT TO TREAT ESTIMATES OF PROGRAM IMPACTS ON STUDENT ACHIEVEMENT:

GATES-MACGINITIE READING TESTS

New York State Education Department
Evaluation Consortium University at Albany
State University of New York

1. BRIEF OVERVIEW OF PROJECT 3
2. IMPACT EVALUATION DESIGN 3
Study Design 3
Sampling Plan 3
Sample Size and Power 3
Data Collection Plan 5
Summary of Analytic Approach 5
3. IMPACTS ON STUDENT ACHIEVEMENT 6
Measures of Student Outcomes/Dependent Variables 6
Independent Variables 6
Covariates 6
Impact Analyses 6
Impacts on Reading Achievement 7
GMRT TOTAL Score 8
GMRT Vocabulary Score 9
GMRT Comprehension Score 10
4. CONCLUSIONS 11

1. BRIEF OVERVIEW OF PROJECT

In October 2009, The New York State Education Department (NYSED), in partnership with the New York City Department of Education (NYCDOE), was granted funding as part of the *Striving Readers Project* to address the literacy needs of adolescent struggling readers early in middle school. The goal of the project was to implement and examine the impact of a one-year, comprehensive supplemental literacy intervention that was provided to seventh grade students across 11 New York City middle schools. The supplemental literacy intervention used in this study was the REWARDS Program (REWARDS Secondary-Multisyllabic Word Reading Strategies; REWARDS Plus; REWARDS Writing). The REWARDS Program provides comprehensive instruction in word analysis, fluency, vocabulary, reading comprehension and writing, and uses content-related text and extended discussion of text meaning and interpretation to enhance student motivation and engagement in literacy learning. The three components in the REWARDS Program were taught in an integrated sequence with careful attention to fidelity, by specially trained teachers who were assisted throughout the year with skilled coaching and expert support.

This report summarizes the examination of the impact of the REWARDS reading intervention on student achievement. Specifically, this evaluation examined differences between the treatment and control groups on reading achievement as measured on the Gates-MacGinitie Reading Tests (GMRT).

2. IMPACT EVALUATION DESIGN

Study Design

The *Striving Readers Project* focused on increasing reading achievement in 7th grade students who struggled in reading. The methodology employed in the NYS project was an experimental pre-post control group design with random assignment.

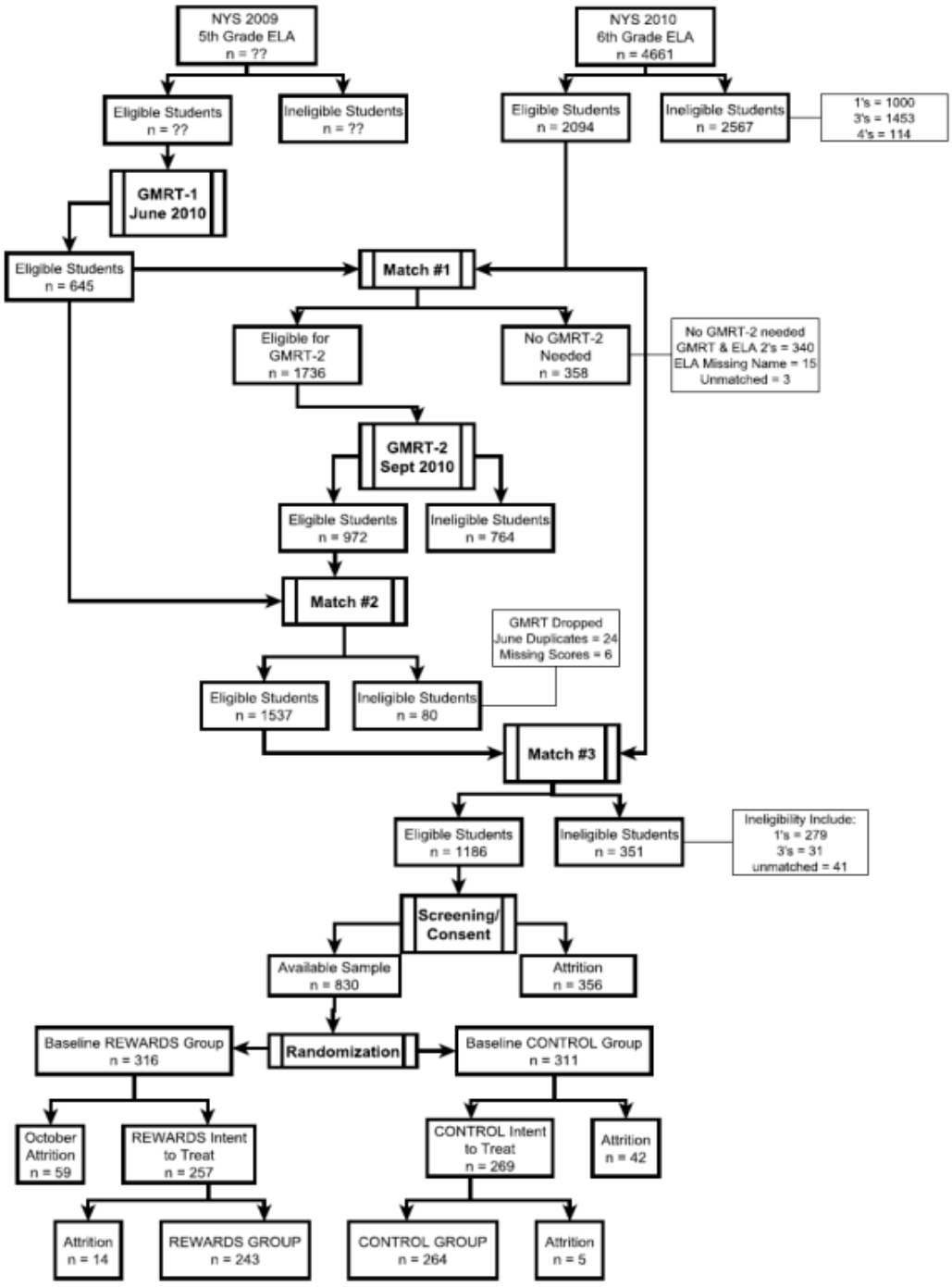
Sampling Plan. As required to participate in the *Striving Readers* grant, schools had to meet the following criteria:

- Be Title I eligible
- Have a minimum of 75 students in the grades to be served by the supplemental literacy intervention were struggling readers.
- Not currently using the REWARDS program

The implementation of the sampling plan is detailed in *Figure 1*. The final sample after attrition consisted of 507 students (treatment group *n*=243, control group *n*=264). This report includes results for 469 students across 10 school buildings (treatment group *n*=232, control group *n*=237). Comprehensive discussion of the random assignment process and sample descriptive characteristics is presented in the Random Assignment Report 2011 and the ITT Descriptive Analyses Report 2012.

Sample Size and Power. A-priori statistical power analyses were conducted to determine the probability of detecting treatment effects using *Power in Two-Level*

Figure 1. **Sampling Plan Consort Diagram**



Designs Software (*PinT* v. 2.12; Bosker & Snijders, 2007). The specific design used was person randomized trials at multisite trials. The minimal detectable effect calculated was .16. This estimate was based on the following assumptions:

- Two-level HLM model (student and school)
- Type I error rate (alpha) = .05
- Intra-class correlation (rho) = .05
- Number of sites = 10
- Average number of students/site = 47
- Minimum power level = 80%

This analysis indicates that there is sufficient statistical power to detect an intervention effect of less than one-fifth standard deviation in the project as planned.

Data Collection Plan: Included in this report are the analyses of the REWARDS program intervention impact on *student achievement* as represented by GMRT performance. Data were collected pre- and post-intervention on the GMRT. Pre-intervention testing occurred May 24-26, 2010 and September 13-16, 2010, and post-intervention testing occurred June 6-9, 2011. The tests were administered by trained NYCDOE staff.

The Gates-MacGinitie Reading Tests (GMRT) is a group- or individually-administered, norm-referenced survey measure of reading achievement for students from kindergarten-adult; group administration was used in this project. Two alternate forms are available for use. Vocabulary and reading comprehension are assessed via multiple-choice questions. Five types of scores are available: normal curve equivalent (NCE), percentile rank, stanine, grade equivalent, and extended scale score. These scores are available for each subtest and for total reading at each level. This project used grade-based NCE scores in data analyses. Test reviewers (Johnson, 2000; McCabe, 2000) noted that compelling evidence for reliability based on three comparisons is reported. Alternate form correlations for total scores ranged from .81 to .95. Internal consistency reliability for total scores ranged from .93 to .97. Evidence for validity for the Fourth Edition is based on: (a) the high score correlation with the Third Edition (for total scores on the Third and Fourth Editions ranged from .82 to .93), (b) strong validity indicators on the Third Edition, (c) piloting, and (d) the careful procedures in developing the Fourth Edition, including input from teachers. Test users are referred to the technical report of the Third Edition for some of the validity information, a somewhat inconvenient task (McCabe).

Summary of Analytic Approach

To estimate the impact of the REWARDS program intervention on student achievement, Hierarchical Linear Models (HLMs) were used. The data from the GMRT consisted of 3 dependent variables: GMRT TOTAL NCE score, GMRT Vocabulary NCE score, and GMRT Comprehension NCE score. These analyses focused on the intent-to-treat samples that are detailed in Intent to Treat Descriptive Variable Analyses Report. A

two-level model was employed, with student and school as the levels. For the variables analyzed and included in this report, there were few or no missing data. In the event there were missing data, they were deleted listwise by the SPSS mixed model analysis.

3. IMPACTS ON STUDENT ACHIEVEMENT

Measures of Student Outcomes/Dependent Variables

Controlling for pre-test scores (applicable GMRT score), the following scores from the GMRT 2011 were used as dependent variables in data analyses:

1.a. GMRT TOTAL NCE Score

1.b. GMRT NCE Vocabulary Score

1.c. GMRT NCE Comprehension Score

Independent variables

Two independent variables were included in the impact analyses: access to program and school. Access to program was coded as "yes" (1) or "no" (0). Each of the 10 schools included in the data analyses was numbered sequentially.

Covariates

The only covariates that were included in the analyses were the pretest scores on any of the variables for which these were requested, and only if the variable had some variability. Because no random effect of schools was found for any of the variables, there was no need to consider any covariate at the school level.

Impact analyses

Based on information provided at the March 2011 grant meeting in Washington, DC, both random effects and fixed effects models with covariates were explored to determine which more efficiently met the needs of the district under study. To make this determination, the analyses were completed in 2 stages. The data from the GMRT consisted for 3 dependent variables: GMRT TOTAL NCE score, GMRT Vocabulary NCE score, and GMRT Comprehension NCE score. All data were organized as an hierarchical linear model with Level 1 of the data consisting of students and the variable of interest at the student level being the REWARDS treatment or control group to which the students were randomly assigned. The students of the study were nested within 10 schools that constituted the Level 2 of the hierarchical linear model.

The first stage consisted of fitting a random effects, intercepts only, null model (Heck, Thomas, & Tabata, 2010) to the data in order to partition the variance components (σ^2) into two sources due to students (Level 1; σ_w^2) and schools (Level 2; σ_b^2). The linear model of a dependent variable, Y_{ij}, whose variability is predicted to be a function of a mean of the observations of the *i* students nested within the *j* schools is given as,

	$Y_{ij} = \beta_{0j} + r_{ij},$	(1)

The regression coefficient β_{0j} with subscripts $0j$ implies that the j intercepts (intercept denoted by β_0) are fitted separately within each of the j schools. It is possible to postulate that these intercepts (means within schools) also vary across schools and that this variability could be estimated. Letting the intercepts be predicted by a grand mean (i.e., γ_{00}) plus the deviation of each of the school means from that grand mean (i.e., $\mu_{0j} = \beta_{0j} - \gamma_{00}$), we can write,

	$\beta_{0j} = \gamma_{00} + \mu_{0j},$	(2)

A single reduced form equation can be constructed by substituting Equation 2 into Equation 1,

	$Y_{ij} = \gamma_{00} + \mu_{0j} + r_{ij},$	(3)

Equation 3 is fitted to the data and in the process the student variances at Level 1 (σ_w^2) and the school variances at Level 2 (σ_b^2), which are the additive parts of the total variance of *Y*, are estimated.

At stage 1 of the HLM analysis the purpose was to assess the proportion of the total variance that is attributable to the school effect. The intraclass correlation (ICC) is defined as this proportion,

	$ICC = \frac{\sigma_b^2}{\sigma_w^2 + \sigma_b^2}$	(4)

Most authors recommend that an ICC of less than .05 (less than 5% of the variance accounted for by Level 2) is typically too small a proportion to add any useful information beyond a fixed effects regression/linear model. Additionally, most commercial software for hierarchical linear model analysis computes a Wald test of significance of the ICC. Conventionally any ICC that is not statistically significant at $p <$ .05 would not be pursued in an hierarchical random effects model.

Stage 2 of an HLM analysis of a random effects intercept + slope model based on both school and student observations, would be pursued further ***only if*** the ICC > .05 and $p <$ .05. If these criteria are not met, Stage 2 reverts to fitting a theoretically interpretable, but more simple, fixed effects linear model to the Level 1 data.

Impact on Reading Achievement

The results of the impact analyses of the REWARDS intervention on student reading achievement are presented in this section. Two aspects of the results are discussed: whether the results were statistically significant at the .05 level, and whether any of the results reached an effect size threshold of .16 (based on the power analysis reported above). Effect sizes were calculated using Cohen's *d*.

The 2-stage process described above was implemented in the analyses of the GMRT data from this research. Both random effects and fixed effects models were fitted to the GMRT variables (as requested). Also as requested, following examination of the random or fixed effect models only one type of model was reported. For all of the GMRT variables in this study, **no** ICC evaluated by an intercept only random effects model was statistically significant or of substantial magnitude; consequently fixed effect linear models were fitted and presented in the tables in this section.

The random intercepts, null model was fitted to each of the three GMRT variables of this study. None of the ICCs exceeded .05, nor were any of the ICCs significantly different from zero (See the Table summaries for the pre-screening tests for each variable).

GMRT TOTAL NCE Score. The GMRT TOTAL NCE Score was modeled as a fixed effects linear model with an intercept, a pretest covariate (GMRT TOTAL NCE 2010), and a treatment effect. The REWARDS-Control mean difference of 41.70-41.42 = .28 was not significantly different from zero (p = .726). Specifically, the analysis revealed no significant intervention effect (refer to Tables 1.a, 1.b, 1.c); the obtained effect size of .02 was below the .16 criterion identified in the power analysis reported above. These findings are exhibited graphically in Figure 2 which illustrates the similarity in GMRT TOTAL NCE performance across the 2 groups.

Table 1.a
Pre-Screening for Choosing Random versus Fixed Effects Model
GMRT TOTAL NCE 2011

Random Effects (from unconditional null model)

Level	Variance Component	Variance	ICC	Wald Test	*p*
School	Level 2	8.04	.052	1.35	.178
Student	Level 1	147.03			

The unconditional model is a two-level model with students (level-1) nested in schools (level-2) and only an intercept term on the right hand side of the model. A non-significant ($p > .05$) Intraclass Correlation leads to the decision to fit only the fixed effects model to the data as summarized in Tables 1.b and 1.c

Table 1.b
Fixed Effects Model
GMRT TOTAL NCE 2011

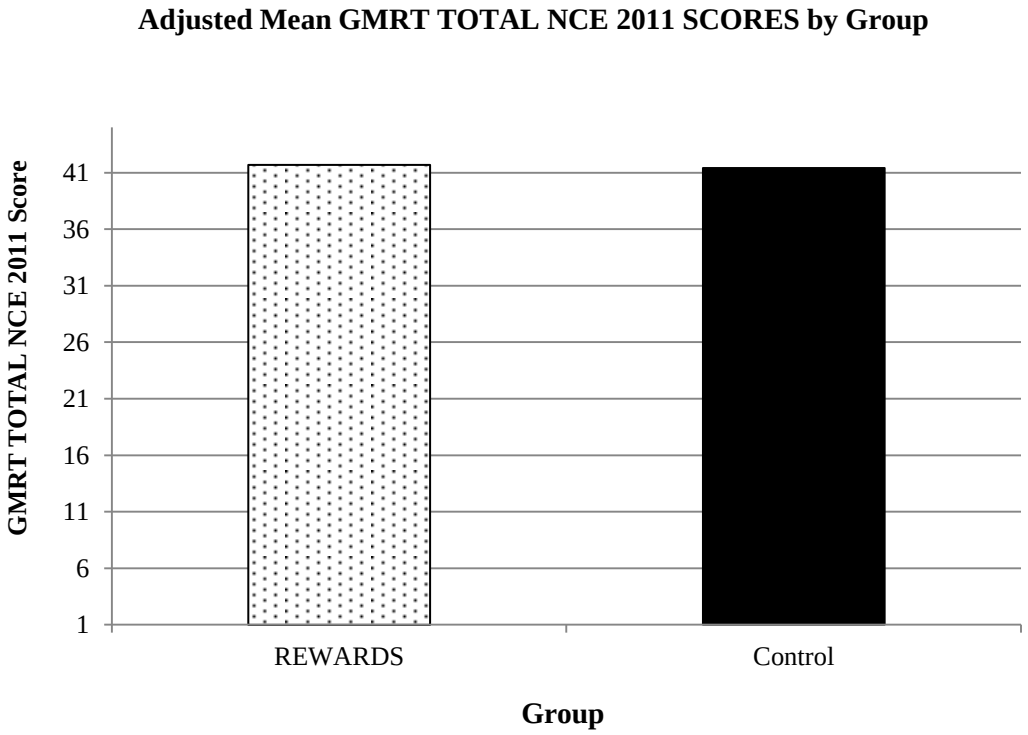
	Control Group		Treatment Group				
Subtest	Mean	SD	Model – Adjusted Mean	SD	Estimated Impact	Effect Size	*p*-value
GMRT TOTAL NCE 2011	41.42	12.25	41.70	12.52	.28	.02	.726

Effect size = Estimated Impact (β) / control group standard deviation
Model adjusted treatment group mean = control group mean + estimated impact

Table 1.c
Analysis Detail Table of GMRT TOTAL NCE 2011 Scores
Fixed Effects Coefficients

Level	Effect	Impact(β)	S.E.	df	*t*	*p*
Student	Intercept	14.67	1.35	460	10.84	<.001
	Treatment	.28	.81	460	.35	.726
	Pre-test	.71	.03	460	21.77	<.001

Figure 2. **GMRT TOTAL NCE Means by Group**



GMRT Vocabulary NCE Score. The analysis for this variable was a fixed effects linear model with an intercept, a pretest covariate (GMRT Vocabulary NCE 2010), and a treatment effect. As noted in Tables 2.a, 2.b, and 2.c, the treatment effect was not significant (p = .224), with a mean difference between REWARDS and Control of 39.17-38.04 = 1.13. That is, the students in the REWARDS and Control groups scored similarly on the GMRT Vocabulary NCE 2011 exam. Moreover, the effect size of .08 was quite small by conventional standards (Cohen, 1988).

Table 2.a
Pre-Screening for Choosing Random versus Fixed Effects Model
GMRT Vocabulary NCE 2011

Random Effects (from unconditional null model)

Level	Variance Component	Variance	ICC	Wald Test	*p*
School	Level 2	6.54	.035	1.29	.198
Student	Level 1	180.00			

The unconditional model is a two-level model with students (level-1) nested in schools (level-2) and only an intercept term on the right hand side of the model. A non-significant ($p > .05$) Intraclass Correlation leads to the decision to fit only the fixed effects model to the data as summarized in Tables 2.b and 2.c.

Table 2.b
Fixed Effects Model
GMRT Vocabulary NCE 2011

	Control Group		Treatment Group				
Subtest	Mean	SD	Model – Adjusted Mean	SD	Estimated Impact	Effect Size	*p*-value
GMRT Vocabulary NCE 2011	38.04	13.62	39.17	13.64	1.13	.08	.224

Effect size = Estimated Impact (β) / control group standard deviation
Model adjusted treatment group mean = control group mean + estimated impact

Table 2.c
Analysis Detail Table of GMRT Vocabulary NCE 2011 Scores
Fixed Effects Coefficients

Level	Effect	Impact(β)	S.E.	df	*t*	*p*
Student	Intercept	15.57	1.32	466	11.82	<.001
	Treatment	1.13	.93	466	1.22	.224
	Pre-test	.65	.03	466	19.65	<.001

GMRT Comprehension NCE Score. The GMRT Comprehension NCE score also was fitted as a fixed effects linear model with an intercept, a pretest covariate (GMRT Comprehension NCE 2010), and a treatment effect. The REWARDS-Control effect revealed a mean difference of 42.71-42.81 = -.10, which was not statistically different from zero ($p = .923$). Furthermore, the resulting effect size of -.008 was less than the necessary minimally detectable effect criterion of .16 based on the power analysis. Again, no significant difference was observed between the REWARDS and control groups on this variable.

Table 3.a
Pre-Screening for Choosing Random versus Fixed Effects Model
GMRT Comprehension NCE 2011

Random Effects (from unconditional null model)

Level	Variance Component	Variance	ICC	Wald Test	*p*
School	Level 2	8.04	.052	1.35	.178
Student	Level 1	147.03			

The unconditional model is a two-level model with students (level-1) nested in schools (level-2) and only an intercept term on the right hand side of the model. A non-significant ($p > .05$) Intraclass Correlation leads to the decision to fit only the fixed effects model to the data as summarized in Tables 3.b and 3.c.

Table 3.b
Fixed Effects Model
NCE Comprehension 2011

	Control Group		Treatment Group				
Subtest	Mean	SD	Model – Adjusted Mean	SD	Estimated Impact	Effect Size	*p*-value
GMRT Comprehension NCE 2011	42.81	12.78	42.71	13.45	-.10	-.008	.923

Effect size = Estimated Impact (β) / control group standard deviation
Model adjusted treatment group mean = control group mean + estimated impact

Table 3.c
Analysis Detail Table of GMRT Comprehension NCE 2011 Scores
Fixed Effects Coefficients

Level	Effect	Impact(β)	S.E.	df	*t*	*p*
Student	Intercept	17.37	1.78	460	9.74	<.001
	Treatment	-.10	.99	460	-.10	.923
	Pre-test	.64	.04	460	15.49	<.001

4. CONCLUSIONS

Multilevel analyses consistently revealed no detectable overall impacts of the REWARDS intervention on student reading achievement as measured by the GMRT. More specifically, across all post-intervention scores examined (GMRT TOTAL, Vocabulary, and Comprehension) the achievement level of the REWARDS group was similar to that of the control group. Based on examination of both statistical significance and effect size results in this study, it was noted that participation in the REWARDS reading intervention did not result in a significant increase on achievement scores on a nationally-normed reading test. Moreover, the effect sizes in the present investigation (-.008-.08) are lower than those reported in the available literature on academic interventions (.20-.30; Hill, Bloom, Black, & Lipsey, 2008). It is important to consider these results within the context of the larger study, including the program implementation fidelity and test administration fidelity (see previous reports for this information).