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:

NEW YORK STATE ENGLISH LANGUAGE ARTS EXAMINATION

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

Striving Readers Project

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

NEW YORK STATE ENGLISH LANGUAGE ARTS EXAMINATION

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 5
Data Collection Plan 5
Summary of Analytic Approach 5
3. IMPACTS ON STUDENT ACHIEVEMENT 5
Measures of Student Outcomes/Dependent Variables 5
Independent Variables 6
Covariates 6
Impact Analyses 6
Impacts on Reading Achievement 7
ELA Scale Score 8
ELA Performance Level 9
ELA Pass/Fail Rating 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 New York State English Language Arts examination (NYS ELA).

2. IMPACT EVALUATION DESIGN

Study Design

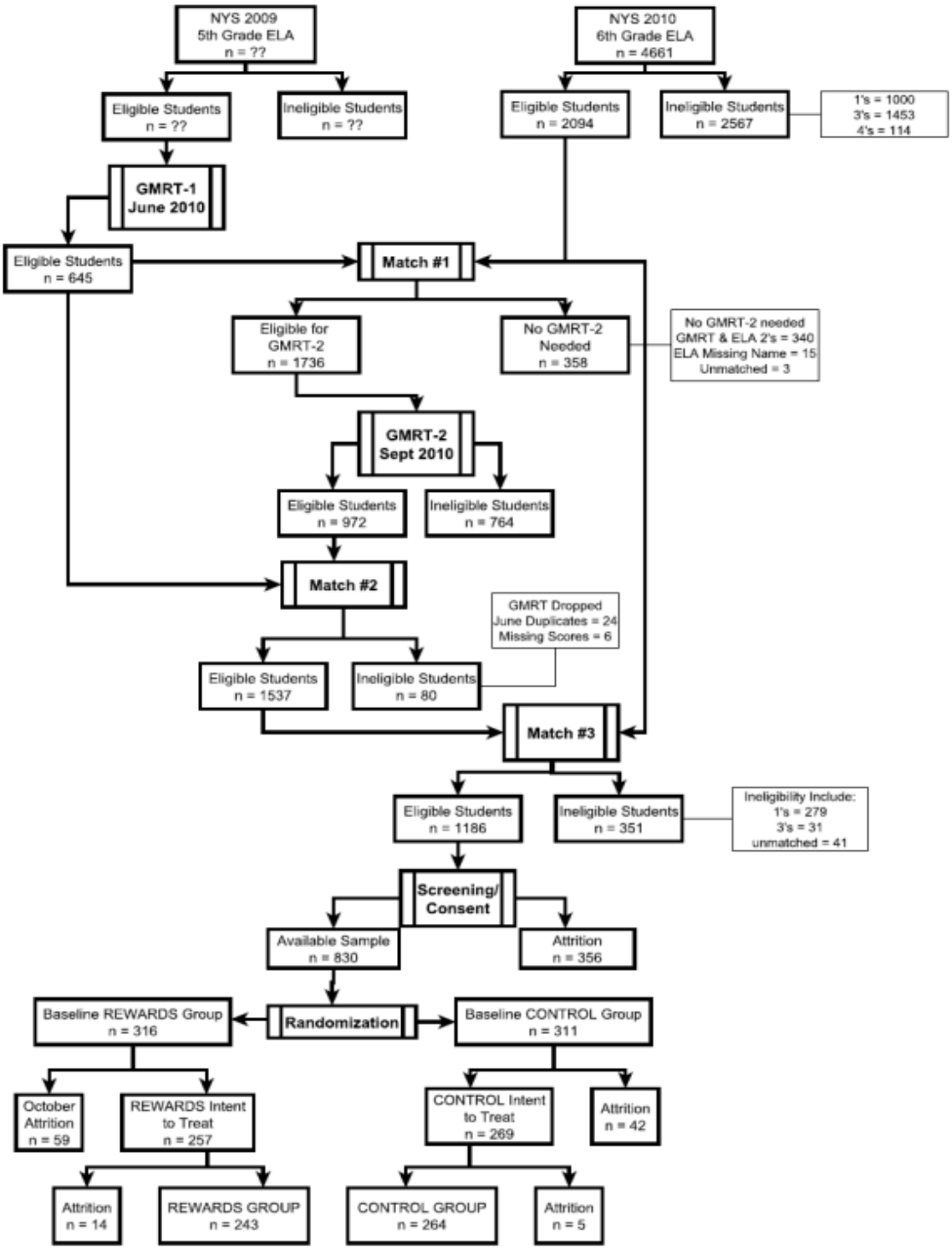
The *Striving Readers Project* focused on increasing reading achievement in 7th grade students who struggled in reading. The methodology employed in the NYS project was an experimental pre-post control group design with random assignment.

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*. After attrition, the final sample consisted of 507 students from 11 school buildings (treatment group *n*=243, control group *n*=264). This report includes NYS ELA results for 517 students (treatment group *n*=253, control group *n*=264; data were available for students who moved within the district during the school year). 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.

Figure 1. **Sampling Plan Consort Diagram**



Sample Size and Power. A-priori statistical power analyses were conducted to determine the probability of detecting treatment effects using *Power in Two-Level 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 = 11
- 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 NYS ELA test performance. Data were collected pre- and post-intervention on the state-mandated English Language Arts examination. Pre-intervention testing occurred April 26-28, 2010, and post-intervention testing occurred May 3-6, 2011. These measures were administered by NYCDOE staff.

The NYS English Language Arts Exam Grade 7 is the required New York State test for students in grades 3-8. Psychometrics are established yearly by New York State Education Department (Cronbach's alpha reliability=.92). The grade 7 test consists of a section containing multiple-choice and short-response questions based on reading selections and a section containing multiple-choice and short-response questions based on a listening selection, as well as an editing task. Raw scores are converted to Scale Scores (2011 *Mean* = 663.71, *Standard Deviation* = 19.60), and Performance Levels (i.e., 1, 2, 3, or 4) which are established annually based on the Scale Scores.

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 NYS ELA consisted for 3 dependent variables: Scale Scores, Performance Level (1, 2, 3, or 4), and Pass/Fail outcome. 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 (NYS ELA 2010 grade 6), the following scores from the NYS ELA 2011 were used as dependent variables in data analyses:

1.a. NYS ELA Scale Scores (2011 *Mean* = 663.71, *Standard Deviation* = 19.60)

1.b. NYS ELA Performance Level (NYSED has established four State-designated levels of performance: 1=Below Standard [not meeting standards], 2=Meets Basic Standard [not fully meeting standards], 3=Meets Proficiency Standard, 4=Exceeds Proficiency Standard); categorical score most often reported and used by the schools

1.c. NYS ELA rating Pass/Fail (*Levels 1 & 2 coded 0, Levels 3 & 4 coded 1*); categorical 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 11 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 (2 of the variables were constant at the 2010 pretest (e.g., all 2s or Fail rating). There were no Level 2 covariates at the school level in the data set. 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 NYS ELA consisted for 3 dependent variables: Scale Scores, Performance Level (1, 2, 3, or 4), and Pass/Fail rating. 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 11 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 NYS ELA data of this research. Both random effects and fixed effects models were fitted to the NYS ELA 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 NYS ELA 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 NYS ELA 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).

NYS ELA Scale Score. The NYS ELA Scale Score was modeled as a fixed effects linear model with an intercept, a pretest covariate (NYS ELA 2010 grade 6 Scale Score), and a treatment effect. The REWARDS-Control mean difference (656.70-655.29 = 1.41) was not significantly different from zero (p = .057). Specifically, the analysis revealed no significant intervention effect (refer to Tables 1.a, 1.b, 1.c); the obtained effect size of .15 was below the .16 criterion identified in the power analysis reported above. These findings are exhibited graphically in Figure 1 which illustrates the similarity in NYS ELA 2011 grade 7 test performance across the 2 groups.

Table 1.a
Pre-Screening for Choosing Random versus Fixed Effects Model

Random Effects (from unconditional null model)

Level	Variance Component	Variance	ICC	Wald Test	*p*
School	Level 2	5.83	.062	1.63	.104
Student	Level 1	88.27			

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
NYS ELA Scale Score

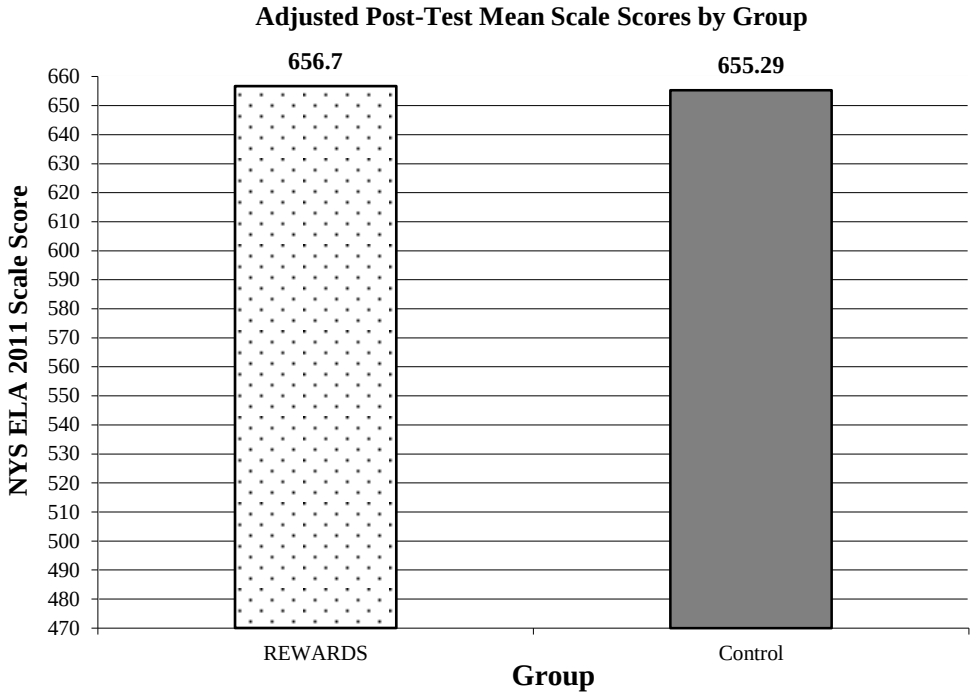
	Control Group		Treatment Group				
Subtest	Mean	SD	Model – Adjusted Mean	SD	Estimated Impact	Effect Size	*p*-value
NYS ELA SS11	655.29	9.69	656.70	9.60	1.41	.15	.057

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 NYS ELA SCALE SCORES
Fixed Effects Coefficients

Level	Effect	Impact(β)	S.E.	df	*t*	*p*
Student	Intercept	-1.11	51.59	514	-.02	.983
	Treatment	1.41	.74	514	1.91	.057
	Pre-test	1.00	.08	514	12.72	<.001

***Figure 1*. NYS ELA Scale Score Means by Group**



NYS ELA Performance Level. For the NYS ELA Performance Level (a categorical variable) there was no covariate in the model as the NYS ELA 2010 grade 6 performance level was a constant 2.0. Hence, the linear model involved an intercept and a treatment term for the REWARDS/Control contrast. As noted in Tables 2.a, 2.b, and 2.c, the treatment effect was not significant (p = .087), with a mean difference between REWARDS and Control of 2.15 – 2.08 = .07. That is, the students in the REWARDS and Control groups performed similarly on the ELA 2011 grade 7 exam. Moreover, the effect size of .152 was quite small by conventional standards (Cohen, 1988).

Table 2.a
Pre-Screening for Choosing Random versus Fixed Effects Model

Random Effects (from unconditional null model)

Level	Variance Component	Variance	ICC	Wald Test	*p*
School	Level 2	.012368	.056	1.56	.118
Student	Level 1	.207889			

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
NYS ELA Performance Level

	Control Group		Treatment Group				
Subtest	**Mean**	**SD**	**Model – Adjusted Mean**	**SD**	**Estimated Impact**	**Effect Size**	***p*-value**
NYS ELA P11	2.08	.46	2.15	.47	.07	.15	.087

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 NYS ELA PERFORMANCE LEVELS
Fixed Effects Coefficients

Level	Effect	Impact(β)	S.E.	df	*t*	*p*
Student	Intercept	2.08	.03	515	72.30	<.001
	Treatment	.07	.04	515	1.72	.087

NYS ELA Pass/Fail Rating. The Pass/Fail rating of the NYS ELA (a categorical variable) was fitted as a linear probability model (OLS regression fitted to a 1-0 dependent variable). The fixed effects linear model included an intercept and the treatment effect; no covariate was included in the model as the NYS ELA 2010 grade 6 rating was a constant 2.0/Fail. The REWARDS-Control effect revealed a mean difference of .19 - .15 = .04, which was not statistically different from zero ($p = .165$). Furthermore, the resulting effect size of .139 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

Random Effects (from unconditional null model)

Level	Variance Component	Variance	ICC	Wald Test	*p*
School	Level 2	.008	.060	1.66	.097
Student	Level 1	.133			

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
NYS ELA Pass/Fail Rating

	Control Group		Treatment Group				
Subtest	Mean	SD	Model – Adjusted Mean	SD	Estimated Impact	Effect Size	*p*-value
NYS ELA Pass Fail	.15	.36	.194	.40	.05	.13	.165

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 NYS ELA PASS/FAIL RATING
Fixed Effects Coefficients

Level	Effect	Impact(β)	S.E.	df	*t*	*p*
Student	Intercept	.15	.02	515	6.39	<.001
	Treatment	.05	.03	515	1.39	.165

4. CONCLUSIONS

Multilevel analyses consistently revealed no detectable overall impacts of the REWARDS intervention on student reading achievement as measured by the NYS ELA examination. More specifically, across all post-intervention scores examined (Scale Score, Performance Level, and Pass/Fail status) 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 the state-mandated test. Moreover, the effect sizes in the present investigation (.14-.15) 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).