

**IMPROVING ADOLESCENT LITERACY ACROSS THE  
CURRICULUM IN HIGH SCHOOLS: AN EVALUATION  
OF THE STRATEGIC INSTRUCTION MODEL'S  
CONTENT LITERACY CONTINUUM**

---

**OMB CLEARANCE REQUEST  
Supporting Statement Part B**

**November 2007**

Prepared for:  
Institute of Educational Sciences  
United States Department of Education  
Contract No. ED-06-CO-0019

Prepared by:  
MDRC  
16 East 34<sup>th</sup> Street  
New York, NY 10016

Learning Point Associates  
1120 East Diehl Road, Ste. 200  
Naperville, IL 60563

**SUPPORTING STATEMENT**  
**FOR PAPERWORK REDUCTION ACT SUBMISSION**

**B. Collection of Information Employing Statistical Methods**

**1. Respondent Universe and Sampling Methods**

The study design calls for high schools from diverse types of school districts across at least two of the states in the Midwest region (Illinois, Indiana, Iowa, Michigan, Minnesota, Ohio, and Wisconsin). We will seek 50 high-need high schools serving Grades 9-12 where at least one third of the students come from low-income families and at least 50 percent of the students are struggling readers (e.g., reading at least two grades below grade level or are “below basic” or “below proficient” on the eighth-grade state or district assessments). These schools will likely come from about 10 to 12 school districts. To capture the diversity of school districts and high schools within the region, we, ideally, seek to include:

- 1) at least three large urban school districts (with six to eight or more high schools);
- 2) at least three midsize school districts (with two to four high schools);
- 3) and at least one rural school district or consortium of districts (with two to four high schools).

Because this study is an efficacy study, the study team is not seeking a sample of schools that allows for generalization to the entire region. Even though such a sample (e.g., a random selection of districts and schools from across the region) is not sought, the study team will attempt to be inclusive of at least some of the variety of districts served by the REL Midwest. Thus, the study team plans to recruit schools that reflect at least some of that diversity of the Midwest region.

**2. Information Collection Procedures**

**a. Statistical methodology and stratification**

We will use a cluster random assignment design with schools being the key unit of treatment and the unit of random assignment. The CLC program is dependent upon collaboration and activities across the school such that a student will experience CLC throughout the school day. Thus the school must be the unit of assignment and analysis. Random assignment will be blocked by district (i.e., conducted within each district) with equal proportions of schools assigned to treatment and control status. This will yield 25 treatment schools and 25 control schools across the sample as a whole. Including both treatment and control schools within each district in the study sample is necessary to eliminate treatment-control differences by district as a possible causal factor in

explaining impact results. MDRC will conduct the computerized random assignment of schools.

## **b. Estimation procedures / Analysis methods**

The key research question in this study is: What is the impact of a literacy across the curriculum intervention on student outcomes? We propose to address this question by comparing average student outcomes in a set of treatment schools randomly assigned to receive the intervention to average student outcomes in a set of control schools randomly assigned not to participate in the intervention. As noted earlier, we plan to recruit a sample of 50 schools across 12 districts or consortia of districts. Random assignment will be blocked by district (i.e., conducted within each district) with equal proportions of schools assigned to treatment and control status. This will yield 25 treatment schools and 25 control schools across the sample as a whole.

The covariates in our analyses will include information about prior academic achievement. Where available we will include student-level and school-level performance on prior state and district achievement tests. The covariates will also include student-level background demographic data, such as race/ethnicity, free/reduced-price lunch status, and gender. The use of covariates will contribute to increased precision in our estimates, particularly the covariates measuring prior academic achievement. They will also allow us to account for any spurious differences that may be observed between treatment and control schools following random assignment.

Our approach to estimating the effects of CLC has the following core features:

- 1) A focus on impacts based directly on the experimental design.
- 2) Estimation of impacts in ways that account for the randomization of schools.
- 3) Use of student- and school-level baseline covariates to increase precision.
- 4) Estimation of impacts separately for each follow-up year and for each grade in question.

The basic logic of our analysis strategy is to compare the schools that are randomly assigned to receive the treatment to those that are not. As random assignment occurs at the school level, schools are the primary unit of analysis. However, the data for this evaluation can be thought of as nested, as individual students are nested within schools. Individual student observations tend to vary as a group rather than being independent of each other. For example, student outcomes are likely to be correlated with the “clusters” (e.g., schools) within which each student outcome is associated. If such clustering exists and is not accounted for, standard errors will be underestimated and statistical significance will be overstated. Since observations within the same group are not statistically independent of one another, the most appropriate way to estimate the effect of the intervention and correctly estimate statistical precision is to apply a multilevel model (HLM) that estimates separate equations at the student and school levels. Specifically, impacts would be estimated using a two-level model as follows:

**Level 1: Students-Within-Schools.** Our system of equations begins at the student level. Equation 1 describes the relationship between student achievement, individual background characteristics, and random variation among the students in each school.

$$Y_{ik} = \pi_{0k} + \pi_{1k} X_{ik}^* + e_{ik} \quad (1)$$

In this model,

- $Y_{ik}$  = achievement of student  $i$ , at school  $k$ ; and
- $X_{ik}^*$  = individual student characteristics (e.g., race/ethnicity, free and reduced-price lunch status, prior academic achievement) of student  $i$ , at school  $k$  (centered around the grand mean across the sample).

Therefore,

- $\pi_{0k}$  = average achievement at school  $j$ , for students with average characteristics and prior achievement;
- $\pi_{1k}$  = the relationship between individual student characteristics and student achievement at school  $k$ ; and
- $e_{ik}$  = the difference between the achievement of student  $i$ , and average achievement in classroom  $j$  at school  $k$  (adjusted for student background characteristics).

**Level 2: Schools.** Given that random assignment occurs at the school level, program impacts are estimated at the following level of the system of equations:

$$\pi_{0k} = \beta_{00} + \beta_{01} Treatment_k + \beta_{02} \bar{X}_k^* + \mu_{0k} \quad (2)$$

$$\pi_{1k} = \beta_{10} + \mu_{1k}, \quad (3)$$

Where,

- $Treatment_k$  = 1 if school  $k$  is in the treatment group, 0 otherwise;
- $\beta_{01}$  = the difference between average achievement at schools randomly assigned to the treatment group versus schools assigned to the control condition, i.e., the effect of the intervention on student achievement.

This two-level system of equations will be estimated separately within each district and translated into an effect-size metric (i.e., the impact estimate divided by the standard deviation). The average effect will then be estimated by taking a simple average of the impact estimates across all the districts in question. Though we can explore variation across districts, we are likely to lack the statistical precision to discern whether or not the observed variation represents systematic differences in program effects.

**Fixed vs. Random Effects.** Discussion of fixed and random effects focuses on the district level, with schools being our unit of random assignment within districts (as well as our unit of analysis). Fundamentally, the “fixed-effect” approach addresses the question “what is the average effect for districts in the study sample?” and the “random-effect” approach addresses the question “what is the average effect for the population of

districts that is represented by the study sample?” Hence the fixed-effect model restricts its inferences to the sample in the study whereas the random-effect model attempts to infer to a broader population. To date, given the typically small number of sites (districts) for most social experiments, it has been common practice to use fixed effect models for pooling experimental findings (Schochet, 2005). This is because the small number of sites (districts) does not provide enough information about how true impacts vary across sites to support generalizations with adequate precision. Moreover, in this study districts are not selected to be a random sample of a larger population. Instead, they are selected because they match particular criteria. Therefore, our model for this study is a district-level fixed-effect model.

Randomization gives us confidence in the similarity of the units of analysis in the treatment and control conditions. Also, random assignment of schools within districts or consortia of districts helps account for potential district differences between units, and the inclusion of school-level and student-level covariates helps us account for other possible sources of variation (dissimilarity) in our analyses.

### **c. Degree of accuracy needed**

The prevailing standard of precision for randomized controlled trials funded by the U.S. Department of Education (ED) is a minimum detectable effect size of approximately 0.20 standard deviations. Existing empirical research for estimating minimum detectable effect sizes based on the magnitudes of student-level and school-level variance components of outcomes and the predictive power of baseline covariates with respect to these outcomes is based almost exclusively on data for elementary and middle schools (Bloom, Richburg-Hayes, Michalopoulos, & Black, 2005; Bloom, Bos, & Lee 1999). MDRC calculations show that approximately 60 elementary schools are required for randomization to attain the current standard of 0.20 effect sizes at the third- and fifth-grade levels. These analyses also show that only about 30 to 40 middle schools are needed for randomization to achieve the same level of precision on eighth-grade test scores. The difference is due to the much higher school-level predictive power (R-square) of baseline pretests for middle schools than for elementary schools, a phenomenon observed in two large urban districts for which MDRC has analyzed data for minimum detectable effects parameters. Furthermore, findings for tenth grade test scores in these districts suggest even higher school-level predictive power for pretests and thus even greater precision. Bloom, Richburg-Hayes, & Black (2005) present evidence that with prior school mean achievement test scores for the same grade of students used as covariates in impact analyses, minimum detectable effects (MDEs) of approximately 0.11 can be expected for a study of 40 randomized schools (20 treatment and 20 control). At a minimum we expect to obtain these data for participating schools to use in our analyses. If we also succeed in obtaining prior achievement test scores for individual students, smaller MDEs may be possible. As explained in Bloom, Richburg-Hayes, and Black (2005), the inclusion of covariates reduces unexplained variances and consequently reduces the minimum detectable effect size. School-level covariates can only reduce random variation between schools because their values are constant for all students in a school. Student-level covariates can reduce random variation between schools and across

students within schools because their individual values can vary across students within schools and their mean values can vary between schools. Nonetheless, some school-level covariates can reduce minimum detectable effect sizes by as much as or more than student-level covariates.

This analysis assumes that both individual student prior achievement and average school-level prior achievement for each of the previous two years are available for the analysis. At a bare minimum, absent changes in test administration, average school performance on achievement tests is generally available to the public and will still have an appreciable influence on minimum detectable effect sizes.

The study team’s preliminary assessment is that the evaluation should be designed to detect effect sizes as small as .10 to .15 standard deviations. This minimum detectable effect size (MDES) corresponds to improvements of 2 to 3 normal curve equivalent (NCE) points or moving students from the 40<sup>th</sup> percentile to approximately the 45<sup>th</sup> percentile on a norm referenced standardized test. This is a critical design parameter for the evaluation that we have reviewed with IES.

Table B1 (below) illustrates the MDES estimates and sample size requirements for the school-level random assignment design. The table assumes that half the schools would implement the CLC program and half the schools in the study sample would continue with “business as usual” (i.e., the control condition). That is, the schools will be randomly assigned within their districts at a 1:1 ratio to each experimental condition: CLC or non-CLC. The table also shows how the MDES estimates change based on the number of students that are in the sample from each school. These numbers will vary based on the size of the schools that participate, but as the table shows, there is very little change in the MDES estimates as the numbers of students change. These numbers also provide a gauge for the power of the analyses for subgroups of students.

There are three key parameters in the calculation of MDES:  $\rho$  (rho: the intra-class correlation),  $R^2_c$  (the proportion of the random variance between schools that is reduced by the covariate(s)), and  $R^2_i$  (the proportion of the random variance within schools that is reduced by the covariate(s)). The work of Bloom, Richburg-Hayes, and Black (2005) provides us with estimates for  $\rho$  (.22),  $R^2_c$  (.945), and  $R^2_i$  (.56). These estimates are based on tenth grade reading test data from two school districts (from an average of 229 students at each of 12 schools in one district, and from an average of 265 students at each of 32 schools in the other). This work indicates that for high school reading test scores, it is reasonable to assume that a pretest covariate can account for more than 90 percent of the variation in the posttest outcome measure. Given that our primary outcome is students’ reading achievement and we plan to measure it with a reading assessment in common across all sites, eighth grade reading scores for the participating students will be an important covariate. All states in the Midwest region test their eighth graders in reading, and we expect to obtain access to these scores. As previously discussed in the design report, our model assumes fixed effects at the district level. We will include a dummy variable for each district (i.e., a cluster-level covariate) in our model.

We have calculated the MDES presented in the table below using the following equation:

$$MDES = M_{J-K} \sqrt{\frac{\rho(1-R_c^2)}{P(1-P)J} + \frac{(1-\rho)(1-R_i^2)}{P(1-P)Jn}}$$

Where:

$M_{J-K}$  = a multiple of the standard error of the estimator (the “degrees of freedom multiplier”)

$J$  = the total number of schools randomized

$K$  = the number of cluster-level covariates used (1 for the treatment/control variable, 1 pretest covariate, and assumed 1 district covariate for every 4 schools – i.e.,  $J/4$ ; e.g., 40 schools would represent 10 districts)

$\rho$  = the intra-class correlation

$R_c^2$  = the proportion of the random variance between schools that is reduced by the covariate(s) – i.e., their school-level explanatory power

$R_i^2$  = the proportion of the random variance within schools that is reduced by the covariate(s) – i.e., their individual-level explanatory power

$P$  = the proportion of schools randomized to treatment (assumed to be .5)

$n$  = the number of students per school in the grade(s) of interest

**Table B1**  
**Minimum Detectable Effect Sizes by Numbers of Schools and Students**

MDES		Number of students per school				
		50	100	200	400	800
Number of schools	10	0.30	0.28	0.26	0.25	0.25
	20	0.19	0.17	0.16	0.15	0.15
	30	0.15	0.13	0.13	0.12	0.12
	40	0.13	0.11	0.11	0.10	0.10
	50	0.11	0.10	0.10	0.09	0.09
	60	0.10	0.09	0.09	0.08	0.08

In short, for this school-level random assignment design, our preliminary assessment suggests that the study sample should include at least 40 schools (20 randomly assigned to the treatment group and 20 randomly assigned to the control group), preferably with at least 100 students per grade. With a reading pre-test covariate, we would be able to detect impacts as small as a .11 effect size. If the school sample is evenly split across two states and we wanted to look at the results separately for each state (as mentioned in one of our responses above), we would be able to detect impacts as small as a .17 effect size. Detecting impacts with smaller subsamples of schools would only be possible if there were relatively large impacts (e.g., .28 effect size for 10 schools). We will also have the power to look at impacts for subgroups of students. As the table shows, with 40 schools in the sample, we could detect impacts as small as a .13 effect size for subgroups of 50 students per school. However, to protect against some attrition in the sample and the possibility that our estimation of  $R^2_c$  is not conservative enough, we seek to recruit 50 schools for this study.<sup>1</sup>

The calculations in the MDES table assume that statistical significance is determined at the  $p = .05$  level. However, there is currently discussion at ED about whether to adjust determinations of statistical significance based on multiple hypothesis tests, and if so, how. We recognize the potential problems associated with conducting multiple hypothesis tests and running the risk of drawing conclusions about program effectiveness on the basis of falsely rejecting a true null hypothesis. Consequently, we seek to keep the number of outcome measures in the study to a minimum. Yet, even with a limited number of outcomes and only a few subgroups, we can quickly accumulate a large number of hypothesis tests and increase the risk of basing conclusions on false positive results. The analyses will adjust for multiple hypothesis testing in line with What Works Clearinghouse standards. For precision, REL RCTs are powered conservatively and in consideration of prior research, to account for the fact that non-RCT designs often find larger impacts and also to protect against attrition and non-response.

### 3. Methods to maximize response rates

<sup>1</sup> If  $R^2_c = .75$  (a more conservative estimate), a sample of 50 schools would allow the study to detect impacts as small as a .20 effect size. For the same  $R^2_c$ , a sample of 40 schools would allow the study to detect impacts as small as a .22 effect size.



The research team will collaborate with KU-CRL (the program developer) and the participating districts as the program unfolds in order to monitor the potential attrition of schools from the CLC intervention. In addition, through field research the study team will be able to monitor whether staff in the intervention schools are not participating as needed in the CLC professional development activities. The combination of communication and monitoring will produce strong implementation of the random assignment design, and ongoing implementation effort from the schools in the treatment group. If the study team determines that a treatment school has decided to drop the CLC intervention, we will work with district, school, and KU-CRL staff to determine the source of this decision and see if the school can be persuaded or supported to continue with the program. If a school refuses to continue with the CLC program, any remaining funding support would be withdrawn. The school would remain in the study and would continue to be counted as part of the treatment group in the impact analysis.

As part of the project, the REL Midwest is covering most of the costs of the intervention, serving as encouragement to schools to participate in the intervention. The participation of the control schools in data collection activities will be supported at the district level. Data collection activities that will require district support include the compilation of electronic student data records as well as interactions with the research team to determine logistics surrounding data collection activities (such as coordinating site visits). Additionally, control schools will receive compensation ~~from the control schools~~ for their study participation. (See Supporting Statement A, item 9.)

Given that the value of the intervention is estimated at \$200,000 per school over two years of implementation and being supported by the REL Midwest, this participation fee should not prove off-putting to treatment schools. Additionally, the compensation provided to the control schools should keep them interested in the study. A similar process being used in the 2007 *Study of the impact on student achievement of teacher professional development designed to enhance teacher content knowledge and pedagogical content knowledge in mathematics*, has thus far proven to be appreciated by the control schools and helped keep them engaged in the project.

For the collection of information from building and district administrators through interviews, the study team intends to follow two main principles.

- 1) *Justification: Providing respondents with sufficient information about why their participation is important.* District and building administrators will be given information about the context of the study, the importance of their participation, and advance notice of site visits. Additionally, school and district leadership will have committed to participation in this data collection effort when they sign a Role and Responsibilities document that indicates their agreement to participate in the study and specifies what types of involvement is needed from them.
- 2) *Accommodation: Working with the respondents' schedules.* Field researchers will be flexible in scheduling interviews with administrators and will make efforts to complete interviews at the respondent's convenience on site. However, if this is not possible, interviewers will seek to complete the

interviews over the phone.

The ability to communicate the importance of the respondents' participation in the study and the ability of the study team to be flexible in seeking these interview data are expected to result in high response rates; the study team anticipates that there will be at least a 95% response rate by district and building interviewees.

#### **4. Tests of procedures to be undertaken**

The items in the building level administrator interviews are taken almost directly from the interview protocols designed and validated by LPA for their Striving Readers program evaluation. The items in the district level administrator interviews are drawn the LPA Striving Readers protocols as well as from instruments used in prior MDRC studies of high school reform. Thus, the items in both instruments have been used successfully in the past.

The GRADE is a widely used, group-administered paper-and-pencil test. Because subtests can be administered separately, the test can be divided across two or more sessions to accommodate schools' class schedules if necessary. The directions for administering the GRADE are easy to follow, so school staff and the research team can readily accomplish this task. Tests can be easily scored by hand or using software by the developer.

#### **5. Individuals consulted on statistical aspects of design**

James Heckman, University of Chicago  
Larry Hedges, Northwestern University  
Rebecca Maynard, University of Pennsylvania  
James Kemple, MDRC  
Howard Bloom, MDRC  
William Corrin, MDRC

## REFERENCES PART B

- Bloom, H.S., Bos, J.M., and Lee, S. (1999). "Using Cluster Random Assignment to Measure Program Impacts: Statistical Implications for the Evaluation of Education Programs." *Evaluation Review* 23(4): 445-69.
- Bloom, H.S., Richburg-Hayes, L., Michalopoulos, C. and Black, A. (2005). *Using Covariates to Improve the Precision of Studies that Randomize Schools to Measure Intervention Effects on Student Achievement*. New York: MDRC. Retrieved April 17, 2007 from <http://epa.sagepub.com/cgi/reprint/29/1/30>
- Bloom, H.S., Richburg-Hayes, L., and Black, A. (2005). *Using covariates to improve precision: Empirical guidance for studies that randomize schools to measure the impacts of educational interventions*. New York: MDRC. Retrieved April 17, 2007 from <http://www.mdrc.org/publications/417/full.pdf>
- Schochet, P.A. (2005). *Statistical power for random assignment evaluations of education programs*. Princeton, NJ: Mathematica Policy Research. Retrieved April 17 from <http://www.mathematica-mpr.com/publications/PDFs/statisticalpower.pdf>