

**A STUDY OF THE DIFFERENTIAL EFFECTS
OF ELL TRAINING AND MATERIALS
(Study 2.1c)**

**OMB Clearance Package Supporting
Statement**

***Part B: Collection of Information
Employing Statistical Methods***

**Regional Educational Laboratory
for the
Central Region**

Contract #ED-06-CO-0023

Submitted to:

Institute of Education Sciences
U.S. Department of Education
555 New Jersey Ave., N.W.
Washington, DC 20208

Submitted by:

Mid-continent Research
for Education and Learning
4601 DTC Blvd., #500
Denver, CO 80237
Phone: 303-337-0990
Fax: 303-337-3005

Project Officer

Sandra Garcia, Ph.D.

Project Director:

Louis F. Cicchinelli, Ph.D.

Deliverable 2007-2.3/2.4

May, 2007

TABLE OF CONTENTS

B. Collection of Information Employing Statistical Methods.....	1
1. Respondent Universe and Sampling Methods.....	3
2. Statistical Methods for Sample Selection and Analysis.....	5
3. Methods to Maximize Response Rates and Deal with Non-response.....	7
4. Test of Procedures or Methods.....	8
5. Individuals Consulted on Statistical Aspects of the Design.....	9
References.....	8

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

Use of the *OWE* materials and participation in the *RISE* professional development are conceptualized as school-wide programs. As a result, the proposed study employs a randomized control design wherein schools, not teachers or their students, are the unit of assignment. Student-level data are therefore nested within classroom or school-level clusters, wherein teachers either participate in the *RISE* professional development or do not. To address the research questions, a randomized-control (experimental) design will be employed. In year one, schools (and teachers) in grades 1–5 will be assigned to the treatment condition or to a no-treatment control condition.¹ Treatment schools will be provided *RISE* through a facilitator trained by the developer in a train-the-trainer model and *OWE* materials for use in their classrooms. In control schools (year 1), teachers in grades 1–5 will use their existing strategies and materials for ELL students. Table 4 summarizes the study design.

Table 4. Overview of Study 2c Design and Sample

Groups	Program	Grades providing student data in year one	Grades providing student data in year two
Treatment	Two years of both <i>OWE</i> and <i>RISE</i>	1-4	2-5
Control	Two years of existing approaches to ELL instruction	1-4	2-5

Target Population & Eligibility Criteria. The target population for this study is elementary schools from a single state, serving a high percentage of Spanish-speaking ELL students in grades 1–5, and which meet the eligibility criteria detailed herein. To ensure an adequate final sample of committed schools from one state, only states with a total of at least 80 eligible schools will be targeted for recruitment. Specific eligibility criteria were developed to

¹ During year one, data will be collected in a cohort of students in grades 1–4; during year two data will be collected on the cohort now in grades 2–5. Grade 5 classrooms in year 1 and grade 1 classrooms in year 2 will implement the curriculum in the same way as teachers in other grades, however, they will not provide outcome data.

identify a list of proposed sites for this study. Eligible schools are public elementary schools serving grades 1–5 that have (1) at least 33% Spanish-speaking ELLs; (2) at least one classroom per grade with at least 10 ELL students per classroom; (3) a transience rate less than 50%; and (4) no previous use of Harcourt’s *OWE* or *RISE* programs. In addition, the following schools are not eligible to participate in this study: charter or language immersion schools, schools undergoing restructuring, and schools engaged in a competing intensive intervention or program, such as Success for All.

Sampling Strategy. The sample for this study of ELL-specific interventions will be drawn from a single state with a high percentage of Spanish-speaking elementary school students. Limiting the sample to one state is more cost effective and reduces the variation that naturally occurs among state requirements and policies for teaching and testing ELLs. Similarly, the geographic proximity of eligible schools within a state is a consideration with regard to the cost of conducting the study. A preliminary review of publicly available data led to the identification of seven states as likely candidates for this study: Arizona, California, Colorado, Nevada, New Mexico, and Texas. These states were identified as having the largest limited English proficiency (LEP) enrollment and/or significant increases in LEP enrollment in recent years. Additionally, their state ELL legislation does not conflict with study requirements (e.g., the state does not maintain that ELL students are only eligible for English as a Second Language classes for one year), and there are no changes in related legislation anticipated. A U.S. language population map was then used to identify counties that were likely to contain schools with high percentages of Spanish-speaking students based on 2000 census data.² An estimate of the number of schools within these predominately Spanish-speaking counties that had ELL populations of at least 33% was generated using the Common Core of Data³ and data from state departments of education. Further elimination of schools occurred by excluding those that had previously used or were currently using *OWE* and/or *RISE* based on data currently available from the publisher. After eliminating states where fewer than 80 eligible elementary schools remained, the list of eligible states was reduced to four: Texas, Florida, Colorado and California (in rank order by number of eligible schools).

Table 5. Proposed Sites, Accessible Schools, and Targeted Samples

Proposed Sites/States	Accessible Schools	Selected Sample	Actual Sample*
Texas	505	80	48
Florida	152	80	48

² MLA language map, http://www.mla.org/census_percentage&source=county, retrieved 2/13/07

³NCES Common Core of Data, <http://nces.ed.gov/ccd/>, retrieved 2/13/07.

Colorado	95	80	48
California	83	80	48

* The actual sample size derived from the power analysis.

School Recruitment

Based on expressed interest from districts in the four eligible states, one state will be identified for school recruitment. Researchers will contact eligible school administrators to describe the study in detail, explain the benefits of participating, and gain school support and approval. All procedures for conducting research in the schools, such as adhering to district policies regarding the collection of student information, will be followed. After an awareness presentation by a representative of the publisher, and an affirmative vote of at least 80 percent of the school staff to move forward with study participation, a school will be eligible for the study. Of those eligible schools, both schools and teachers must agree to the specifics of the study. Principals of participating schools will be asked to sign Memoranda of Understanding (MOUs) pledging a commitment to carry out the responsibilities of the study. Whenever feasible and appropriate, support from the school board will also be requested. A copy of the MOU appears in Exhibit C.

Sample. It is assumed that at least one teacher in each participating grade will participate. Following this assumption, it was determined that 38 schools would be required for a power of .80. To account for anticipated attrition, 25 percent was added to this estimate, indicating a necessary sample size of 48 schools. In schools where more than one teacher in the targeted grades is eligible for participation, s/he will also be asked to participate.

2. Statistical Methods for Sample Selection and Analysis

Outcomes at the student, teacher, and school level will be analyzed, hence power analyses were conducted for each set of outcomes. The number of schools needed was determined by the outcomes for which the largest number of schools was required to achieve the needed level of statistical precision. Thus, this analysis used a conservative estimate of the number of students per classroom based on the assumption that schools would have a minimum of 10 ELL students per grade level. This estimate would reflect a pull-out program; however, other approaches for teaching ELLs would likely contain more students. Moreover, we assumed that a minimum of four teachers (one in each grade 1-4 during year one, and 2-5 during year two) would be participating per school (in schools where multiple sections of

grades are eligible for participation, each of those sections would be invited to participate). The following paragraphs provide details on each of these power analyses.

For the student-level power analysis (individual growth), schools were considered to be the unit of assignment, and student achievement data were considered the main dependent variable. Prior research suggesting an appropriate effect size did not exist; thus, we proposed that an effect size of at least .35 was needed. An effect size of a smaller magnitude was deemed to not be meaningful. This value is a conservative estimate based on the literature on effective ELL interventions addressing English language acquisition and academic achievement (What Works Clearinghouse, 2006). Effect sizes from the literature on ELL programs vary according to the type of assessment intervention and the outcome measure. Gunn et al. (2000), in their analysis of the supplemental ELL instruction program, Reading Mastery, report that the effects on reading achievement average .72 at the end of the intervention and .60 one year later. A recent study of a reading, language arts, and English language development curriculum reports effects averaging .49 across outcomes measured (Vaughn, et al., 2006). An intra-class coefficient of 0.10 was selected based on the work of Raudenbush, Spybrook, Liu, and Congdon (2005), citing typical intra-class correlation coefficients for educational achievement to be between 0.05 and 0.15. The midpoint value (0.10) of this reported interval was assumed for purposes of the power analysis. Assuming a one-tailed test and $p < .05$, and desired power > 0.80 , it was determined that sample and cluster size would be adjusted to reach this goal. Additionally, it was assumed that pull-out classes (the smallest possible unit of class participating in the study) contain ten students. Power analyses were conducted using the smallest possible class size potentially participating in this study; should whole classes or mainstream classes of ELL students participate, the class sizes would obviously be greater than ten (indicating that the proposed ten students per class is a conservative estimate). Optimal Design software, version 1.55 (Liu, Spybrook, Congdon, & Raudenbush, 2005), calculated that 33 clusters were needed to reach the desired power of $>.80$. To account for anticipated attrition, 25 percent was added to this estimate, indicating a necessary sample size of 42.

Secondary research questions require the examination of teachers nested in schools. These questions address the effect of teachers/classrooms on outcomes. Power analyses were conducted to ensure that the above sample size for student-level analyses would also generate adequate power for the classroom-level analyses. Again, a high level of power (>0.80) was desired; a higher effect size of 0.50 was selected because teacher-level effects were expected to be larger than those expected for the student intervention. Based on previous research on the effects of teacher professional

development and classroom practices, we expect the effect size of the RISE professional development to be about .3 and the effect size of classroom practices using OWE to be approximately .2; this combined effect size is a far more conservative than that calculated by Wenglinsky (see <http://www.ncrel.org/gap/library/text/teachersmake.htm>). The same (mid-point) intra-class correlation coefficient was identified (0.10) and the proportion of post-intervention variance anticipated to be explained by pre-intervention survey data was set at $R^2 = .50$, which assumes a strong relationship between teacher practices on pre- and post-surveys. It was assumed that one teacher in each participating grade would participate. Using these values and Optimal Design software, it was determined that 38 schools would be required for a power of .80. To account for anticipated attrition, 25 percent was added to this estimate, indicating a necessary sample size of 48 schools.

The third level is a between-school model, allowing researchers to account for the variability at the school level. Power analyses were conducted to ensure that the above sample size for student-level and classroom-level analyses would also generate adequate power for the school-level analyses. Again, a high level of power (>0.80) was desired. An effect size of .30 was selected and it was assumed that 10 percent of the variation was between classrooms and 10 percent was between schools. Using the class size of 10, number of teachers in the school of 4, and school socio-economic status as a covariate (assuming $R^2 = .50$), Optimal Design determined that 35 schools would be required for a power of .80. To account for anticipated attrition, 25 percent was added to this estimate, indicating a necessary sample size of 44 schools.

Researchers will over-sample 25 percent to offset potential school attrition, yielding a sample of 48 schools. This will ensure adequate power (power $> .8$). It is anticipated that these 48 schools will be randomly selected from the pool of eligible and willing sites.

Exhibit A details the administration of the instruments as well as the purpose and use of each instrument. It is important to note that efforts will be made to minimize possible bias introduced with the administration or collection of data. For example, to minimize potential bias from teachers administering assessments to their own students, site coordinators in treatment and control schools administer the IPT (students enrolling in school later than the start of the study will also be tested). During the study orientation session, site coordinators will be trained on proper administration of the IPT, including training on the importance of their role in ensuring data integrity. Site coordinators will be provided with examples of how data integrity might be compromised, including how “coaching” students in completing test items

leads to invalid test scores. Although researchers are unable to strictly control the administration of this assessment and therefore cannot unequivocally assert that site coordinators will administer the IPT in a way consistent with appropriate test administration, other important factors must be considered when speculating on whether appropriate test administration guidelines will be followed. Although teacher logs do not require an inordinate amount of time to complete, follow-up emails will be sent to teachers who fail to complete their logs. If follow-up reminders do not result in completion of the logs, McREL will contact teachers' site coordinators to make additional attempts to ensure the completion of these data collection activities. Teacher responses in online logs will be verified during site observations (although the observations will only be conducted with a random sample of participating teachers, we believe that teacher log quality can be adequately assessed). In addition, site coordinator logs and teacher logs are expected to align. This provides another means of ensuring the quality of the data being submitted.

The full analysis plan is detailed in Part A of this OMB submission, item A16 and is therefore not included here.

3. Methods to Maximize Response Rates and Deal with Non-response

A separate level of accuracy will be needed for teacher logs than is needed for student data collection. Teacher response burden is to be minimized through the use of online administration of teacher logs. Online logs provide numerous advantages over traditional paper and pencil methods. These include allowing for designs that contain complex skip patterns; range and consistency checks that enhance data quality; availability of previous information, which reduces respondent burden; quick availability of data; and a decrease in the number of clerical errors that can occur during data-entry. Instrumentation developed for specific use in this study will be revised as necessary to ensure clarity in instructions, clarity of items, and time efficiency. In cases when it is required, modified instruments will be submitted for OMB review. Data collection using online instruments will be managed electronically. Reminders about upcoming and current data collection activities will regularly be sent to participating teachers via email. Two weeks before each data collection is due, teachers will receive an email message providing them with a link to the instrument and a requested timeline for completion. Non-responders will receive follow-up reminder emails. The full data collection schedule will be communicated to respondents at the onset of the study. The advance schedule, reminder, and response window structure will allow participants to plan and to incorporate the data collections into their schedules. Moreover, site coordinators will be

asked to intervene in cases where teachers are not completing study instruments. The use of site-based personnel has been beneficial in our previous experiences conducting randomized control trials. To calculate the teacher response rate, the number of returned, completed logs will be divided by the number of teachers who started the study. Based on previous experience with this approach to collecting teacher data, it is anticipated that this response rate will exceed 85%. To account for teacher attrition, 25% was added to the original sample.

Student response burden is to be minimized through the use of standardized, group-administered student assessments. As previously indicated, the required study sample will allow for 25% attrition. This will allow for some loss of student data. Student response rates will be calculated for the second and third IPT test administration. Response rates will be accomplished by dividing the total number of returned IPT assessments for each the second and third administration by the number of students providing data at the first administration. Response rates lower than 80% will be considered problematic; however researchers recognize that it is likely that attrition rates will be high with the ELL population under study. Because of this, researchers' have over-sampled by 25%.

4. Test of Procedures or Methods

Student Assessments. Since we will be using a standardized test and will be following precisely the procedures from the test publisher, we do not have a need to field test the student assessments. Variation from the predefined procedures would compromise the assessment.

Teacher Pedagogical Practices: Online Logs. We will field test the online logs used to assess pedagogical practices as reflected in teacher behaviors on a small sample (no more than nine respondents). We will ask teachers in these schools to complete the logs and to answer a few questions about it. We will ask (1) how long the log took to complete; (2) whether any log items were unclear; (3) whether teachers felt any important topics had been omitted; (4) whether any topics were covered in too much detail; and (5) whether teachers encountered any technical problems when completing the log. We will examine the log responses of the teachers to determine whether we are obtaining usable information (e.g., are responses in the appropriate ranges, are skip patterns and directions being followed). Based on comments provided by the teachers, we will make final revisions to the survey instruments.

5. Individuals Consulted on Statistical Aspects of the Design

The statistical aspects of the design have been reviewed thoroughly by staff at the Institute of Education Sciences, Mathematica, as well as by members of the study's expert panel listed in Section A.8. The following individuals have worked closely in developing the statistical procedures and will be responsible for data collection and data analysis:

Dr. Sheila Arens, Principal Investigator, (303) 632-5625
Dr. Edward Wiley, Chair, Research and Evaluation Methodology
Program
and Assistant Professor of Education, University of Colorado, Boulder,
(303) 492-5204

REFERENCES

Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and Non-Hispanic students in early elementary school. *The Journal of Special Education, 34*, 90-103.

Raudenbush, S.W., Spybrook, J., Liu, X., & Congdon, R. (2005). *Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software*. Retrieved September 8, 2005, from <http://sitemaker.umich.edu/group-based>

Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Cardenas-Hagan, E., Pollard-Durodola, S. D., Fletcher, J. M., & Francis, D. J. (2006). Effectiveness of a Spanish Intervention and an English Intervention for English Language Learners at Risk for Reading Problems. *American Educational Research Journal, 43*(3), 449-487.

What Works Clearing House. (2006). Interventions for Elementary School English Language Learners: Increasing English Language Acquisition and Academic Achievement. Retrieved December 5, 2006 from <http://www.whatworks.ed.gov/TopicAbstract.asp?tid=10>.