Adult ESL Literacy Impact Study

Statement for Paperwork Reduction Act Submission

Part B: Collection of Information Employing Statistical Methods

Contract No. ED-01-CO-0026/0025

April 17, 2007

PREPARED FOR:
Institute of Education Sciences
United States Department of Education

PREPARED BY:
American Institutes for Research®

Table of Contents

Part B. Collection of Information Employing Statistical Methods

B.1 Respondent Universe and Sampling Method

The ESL Literacy Impact Study will test the effectiveness of the *Sam and Pat* curriculum in improving the English reading and speaking skills of low-literate adult ESL. These students are attending federally-funded adult education programs. There are a total of approximately 3,500 such programs, serving about 2.7 million adult students annually, of which about 1.1 million are ESL students. The exact number of adult ESL teachers working in these 3,500 local programs is not known, but the total is estimated to be over 25,000.

Through preliminary information gathered from existing sources, the evaluation team identified approximately 65 programs out of this total universe that potentially met criteria, and screening interviews were conducted with those programs as part of our first clearance (1850-0811). Programs were considered eligible for site visits if they met the study's criteria:

- Have a managed enrollment policy or have an enrollment policy where a majority of learners enter during the beginning of a course;
- Have site enrollments of adult ESL literacy learners large enough to support the evaluation (e.g., able to enroll approximately 90 students from the target population per semester, on average);
- Have a sufficient number of adult ESL literacy instructors to support the evaluation's requirements (e.g., at least 5 instructors per program in the target classes);
- Have common programmatic features (e.g. have classes that are of similar duration both in terms of total class hours and class hours per day/week); and
- Do not currently offer instruction based on *Sam and Pat* or similar instructional materials..

From a careful screening of these programs, approximately 20 programs appeared to meet eligibility requirements and were pursued for participation the study.

B.2 Statistical Methods for Sample Selection and Degree of Accuracy Needed

Stratification and Sample Selection

Approximately ten programs and 40 adult ESL teachers within those programs will participate in the study. The teachers recruited for the study will be the teachers of classes identified by their programs as ESL literacy classes.

The students to be recruited for the study are those students who would normally be enrolled in the ESL literacy classes selected by the program and study staff for inclusion in the study. Based on past experiences and information about programs learned during recruitment, we expect to recruit approximately 900 students per term (1800 total) from the study programs in fall 2008 and winter 2009.

Within each participating program, random assignment will be done at both the teacher/class and student levels. Teachers and students will have an equal chance of being assigned to either the treatment (*Sam and Pat*) or control condition. The teacher-level random assignment will be conducted in summer 2008, to identify teachers that should attend the *Sam and Pat* training in late summer/early fall 2008. The student-level random assignment will be conducted twice with two separate cohorts; first in fall 2008, and secondly in winter/spring 2009, when students attend their first day of class.

Exhibit B-1 provides a summary of the study design and anticipated sample sizes.

Exhibit B.1. Study Design Summary

Program	Treatment Group	Number of Teachers (unit of randomization)	Number of Students Across Two Terms (unit of randomization)
Program 1	Treatment	2	90
	Control	2	90
Program 2	Treatment	2	90
	Control	2	90
Program 3	Treatment	2	90
	Control	2	90
Program 4	Treatment	2	90
	Control	2	90
Program 5	Treatment	2	90
	Control	2	90
Program 6	Treatment	2	90
	Control	2	90
Program 7	Treatment	2	90
	Control	2	90
Program 8	Treatment	2	90
	Control	2	90
Program 9	Treatment	2	90
	Control	2	90
Program 10	Treatment	2	90
	Control	2	90
Total by Group	Treatment	20	900
	Control	20	900
TOTAL		40	1,800

Estimation Procedures

The basic analytic strategy for assessing the impact of *Sam and Pat* will be to compare outcomes for students that were randomly assigned to either the treatment (*Sam and Pat*) or the control condition. (More information on the multi-level statistical models to be used in the impact analyses is provided in section A.16 of this clearance package.) The impact analyses will focus on two types of student outcomes —English reading skills and English speaking skills.

The impact analyses will use an "intent-to-treat" approach and include students regardless of whether their enrollment status changes during the term. The estimates of effect will therefore reflect the impact of *Sam and Pat* on the intended sample.

As discussed in Section A.16, student and teacher or class-level covariates will be included in the model to increase the precision of the impact estimates. Missing values on covariates (e.g., student pre-test scores) will be replaced with the mean value for the participants in the student's program. Students with

missing data on the outcome variables, however, will be dropped from the impact analysis for which they lack data.

Degree of Accuracy Needed

For the feasibility and success of randomized experiments these studies should be designed with sufficient statistical power to detect meaningful treatment effects if they occur. One method of determining whether an effect size is meaningful is to look to the particular field in which the research was conducted. However, randomized trials are fairly recent in the education literature and no such distribution of effects is available. Given that some education studies have concluded that effect sizes as low as 0.20 are meaningful (e.g., the Tennessee STAR experiment), we have adopted that standard as our desired minimum detectable effect size (MDES).

Exhibit B.2. Summary of Planned Outcomes to be Tested

Domain	Outcome	Probable Data Source
English Reading	Decoding	WJ III
(4 outcomes)	Word Identification	WJ III
	Reading Fluency	WJ III
	Passage Comprehension	WJ III
English Speaking	Oral Vocabulary	WJ III
(3 outcomes)	Aural Vocabulary	ROWPVT
	Listening Skills	OWLS

In preparing for the implementation of our study we have access to program data that are a good approximation of the study sample and outcome data we will likely encounter once we recruit sites, enroll students, and conduct random assignment. Using these data it is possible to assess the extent of clustering of outcome data across individual students, teachers, and programs and it is possible to closely approximate the expected level of statistical power for a range of different impact analyses.

The data assembled for the early What Works Study (WWS) for Adult ESL Literacy Students are very similar to the data that we will collect in our study in that they cover students in multiple classrooms, who are taught by multiple, identified, teachers, who teach in a range of different programs in different locations. The data feature similar kinds of student background variables as will be available in our study and have test results for each student at three different time points. This enables us to control for baseline test scores in examining the variability of student outcomes at two "post-program" time points (in the case of the WWS data, these time points are set at three and nine months after program entry; here we only present outcomes at three months, which most closely resembles the timing of our outcome measurement).

The only major difference between the WWS data and our future data is that no specific treatment was being tested in the WWS. So there are no real treatment effects to measure, either in terms of size or in terms of statistical precision. However, under certain assumptions it is possible to simulate a program by creating a random program variable that does not represent an actual program assignment. When included in an "impact regression", the coefficient on such a meaningless random treatment dummy has the same statistical properties as a real treatment variable would have, except for the potential treatment effect on

the variance of the outcome variable. In other words, if the treatment would either narrow or widen the distribution of achievement in the treatment group, power estimates based on a hypothetical treatment variable would be too conservative or too optimistic, respectively. For the purpose of the power analyses presented here, we have to assume that the program either does not significantly affect the variance of the outcome variable or that the effect of such a change in variance on the study's statistical power is small relative to the other determinants of statistical power.

We used SAS, STATA, and EXCEL to conduct the following analyses:

1. We conducted impact regressions to determine the standard errors associated with the (zero) coefficient on a 50/50 random assignment dummy variable, with random assignment conducted at the individual student level, but by teacher. We used three outcomes, as detailed in the table below. For each of these impact regressions we calculated regular OLS standard errors and Huber-White standard errors obtained after correcting for clustering. We ran all analyses with and without a pretest to show the effect of including such a pretest on the statistical power.

2. We used the standard errors and other outcome statistics in formulas developed by Bloom (1995), Bloom, Bos and Lee (1997), and Bloom, Bos, and Lee (1999) to calculate minimum detectable effect sizes (MDES) for all the outcomes and analyses described above.

3. We scaled down the MDES estimates to represent a larger sample (800 in our case, for a large subgroup or a cluster of sites in our study). The analyses on the Condelli et al. data were conducted on a sample of approximately 150 students in 10 programs.[1]

Exhibit B.3 presents the results of these analyses.

Exhibit B.3. Adult ESL Impact Study Power Analyses

	Minimum Detectable Effect Size		
Outcome	**Estimated**	**Actual**	**Without Pretest**
Woodcock Johnson Basic Reading	0.08	0.12	0.23
Woodcock Johnson Reading Comprehension	0.09	0.07	0.12
BEST Total Score	0.10	0.14	0.15
N = 800[2]			

The analyses presented in Exhibit B.3 reflect the effects of conducting random assignment individually and including a strongly predictive pretest as a covariate in the impact analyses. Without such a pretest, the statistical power is much more limited and the effect of clustering of observations by teacher would be much more severe. However, since a pre-test will be included in the impact analyses, the MDES for large subgroups is likely to be below .20 for each outcome. MDES's for the full sample would be smaller. The estimated power is thefore more than adequate to find meaningful and statistically significant differences between treatment and control groups, as long as the pre-test is included.

[1] We excluded programs with only one teacher in the Condelli et al. data, because it was not possible to randomly assign teachers to program or control groups in those programs.
[2] This sample size represents the anticipated size of large subgroups in the sample that may be a focus of analysis (e.g., Spanish-speakers).

B.3 Methods to Maximize Response Rates

During the study, the anticipated response rate for the teacher data forms, program intake form, student interviews, and daily attendance is approximately 100 percent. These estimates are based on the previous experience of study staff in conducting similar studies, as well as our understanding of the typical program intake process. Several procedures will be used to ensure high response rates:

- Obtaining high response rates depends in part on the quality of the instruments. All instruments will be pre-tested to ensure that the questions are clear and as simple as possible for respondents to complete.
- The study will offer a social incentive to respondents by stressing the importance of the data collections as part of a study that will provide much needed information to programs.
- Respondents will be visited in their classrooms multiple times during the study, which will provide us with in-person follow-up opportunities.
- For those respondents who leave the program, we will attempt to contact the respondent at home or through local connections, such as the program staff.

B.4 Test of Procedures

All of the instruments included in this package underwent internal review as well as external review by project consultants (see exhibit B.4). The forms and procedures were then revised to ensure that the questions are clear and as simple as possible for respondents to complete. The project consultants also gave input on the selection of the standardized assessments (outcome measures) to pilot for the study. The assessments to be piloted prior to the test selection for the study include the Woodcock-Johnson III® Tests of Achievement (WJ III®), the Oral and Written Language Scales (OWLS) listening subtest, the Receptive One-Word Picture Vocabulary Test (ROWPVT), and the ETS SARA reading subtests.

B.5 Individuals Consulted on Statistical Aspects of Design

This project is being conducted under contract to the Department of Education by AIR, the Lewin Group, Berkeley Policy Associates (BPA), Mathematica Policy Research (MPR), Educational Testing Service (ETS) and World Education. In collaboration with other consultant groups, Hans Bos of BPA is responsible for the overall statistical aspects of the design. Larry Condelli and Stephanie Cronen at AIR, Mike Fishman at the Lewin Group, John Sabatini at ETS, and Susan Sprachman at MPR also provided input on both the design and the data collection plans. Contact information for the individuals involved is provided in Exhibit B.4.

Exhibit B.4. Contact Information for Individuals Involved in Study

Name	Organization	E-mail Address	Phone Number
Melanie Ali	Institute of Education Sciences	melanie.ali@ed.gov	202-208-7082
Hans Bos	BPA	hans@bpacal.com	510-465-7884
Larry Condelli	AIR	lcondelli@air.org	202-403-5330
Stephanie Cronen	AIR	scronen@air.org	202-403-5229
Mike Fishman	Lewin Group	mike.fishman@lewin.com	703-269-5655
John Sabatini	ETS	jsabatini@ets.org	609-921-9000
Susan Sprachman	MPR	Ssprachman@mathematica-mpr.com	609-275-2333