# Strengthening Adult Reading Instructional Practices
## 200807-1830-001

*Supporting Statement for*
*Request for OMB Approval of Data Collection*

*Part B: Collection of Information Employing Statistical Methods*

# TABLE OF CONTENTS

# LIST OF APPENDICES (attached as individual Word files)

Appendix A: Introduction to SARIP Study

Appendix B: Frequently Asked Questions from Learners

**Supporting Statement for Request for OMB Approval of Data Collection**

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

**Introduction**

This document presents Part B of the Supporting Statement for "Strengthening Adult Reading Instructional Practices" (SARIP), a study sponsored by the U.S. Department of Education's Office of Vocational and Adult Education (OVAE). JBL Associates, Inc. (JBLA) and Abt Associates Inc. (Abt) are the contractors for the study. The study is authorized under the Adult Education and Family Literacy Act, Title II of Public Law 105-220, Section 243, National Leadership Activities. Section 243 allows the Secretary of Education to establish and carry out a program of national leadership activities to enhance the quality of adult education and literacy programs nationwide.

**Overview of SARIP Study**

The SARIP study is an initial investigation of whether the Student Achievement in Reading (STAR) training and materials are effective in developing adult basic education (ABE) instructors' capacity to deliver evidence-based reading instruction and, consequently, in improving intermediate-level (4th-8.9th grade equivalence) adult learners' reading skills. The Office of Vocational and Adult Education began STAR in 2005 as a pilot project to build state capacity to implement research-based reading reform in adult education classrooms. The STAR toolkit, which contains information and resources to improve reading classroom instruction in ABE, was developed to be used in training and in providing technical assistance to local ABE administrators and ABE instructors. The information used in creating the toolkit was based on the body of knowledge on effective reading practices developed by the Partnership for Reading and summarized in Kruidenier (2002).

STAR began with a pilot phase, during which OVAE worked with state adult education administrators, professional developers, local ABE administrators, and classroom instructors in six states to test the STAR toolkit and training. Forty-four ABE programs and 144 instructors were involved in the STAR pilot project. As a result of the pilot project, STAR has grown to a national initiative with a National Technical Assistance Team, an online professional development system, and national dissemination activities. As additional states have received STAR training, there is a need to understand the effects of STAR on learners' reading development.

**Design and Sample**. The SARIP study will employ a quasi-experimental design to examine whether learners who are taught by ABE instructors that have been trained in the STAR methods and materials and have become proficient in these methods make greater gains in developing their reading skills compared to learners who have been taught by ABE instructors that have not participated in STAR. Criteria are being developed for determining high-implementing, STAR-trained instructors, and a sample of high-implementing, STAR-trained instructors will be selected based on these criteria. The adult learners in the reading classes taught by the sample of

high-implementing, STAR-trained instructors will constitute the treatment group for the SARIP study. The treatment learners will be compared to data from a matched sample of adult learners that have been taught by ABE instructors who have not participated in STAR training. This comparison group will be drawn from extant data that were collected in either of two previous studies conducted by the contractor (Abt Associates Inc.), which investigated intermediate-level learners' development of reading skills in ABE classes (*Study of Effective ABE Programs and Practices for First-Level Learners*[1] and *Building a Knowledge Base for Adult Decoding*[2]). The learner, instructor, class, and ABE program data in these two studies were collected using the same instruments that will be used in the SARIP study. Thus there will be comparable data for the treatment and comparison groups.

**Study Questions**. The SARIP study's primary research questions address learners' reading outcomes. These questions are the following.

1. What size reading gains do learners who are taught by high-implementing, STAR-trained instructors achieve?
    a. What is the relationship between learners' background characteristics and their reading gains?
    b. What is the relationship between learners' attendance and their reading gains?

2. Do the reading outcomes (i.e., alphabetics, fluency, vocabulary, and comprehension) of intermediate ABE learners who are taught by instructors that are high implementers of STAR differ from the reading outcomes of intermediate ABE learners who have participated in ABE reading classes taught by non-STAR trained instructors?

While the primary focus of the study is on learner outcomes, the contractor also will investigate differences between the types of reading instruction provided by the instructors in the treatment and comparison groups. This information will be useful in understanding the reasons for any differences in learner reading outcomes between the treatment and comparison groups. The study questions that address reading instruction are:

3. Is the instructional content provided in the reading classes taught by high-implementing STAR-trained instructors different from the content of the reading classes taught by non-STAR trained instructors?

4. Are the reading instructional strategies used by high-implementing, STAR-trained instructors different from the strategies used by non-STAR trained instructors?

---

[1]    The *Study of Effective ABE Programs and Practices for First-Level Learners* was conducted by Abt Associates Inc. and funded by the U.S. Department of Education/OVAE, and Office of Policy and Program Studies Services during 1995-2003. This was a descriptive study that investigated a range of reading instructional strategies for adult learners.

[2]    *Building a Knowledge Base for Adult Decoding* is being conducted by the University of Delaware and Abt Associates Inc. under a grant from NIH/NICHD in partnership with OVAE and the National Institute for Literacy. This is an experimental study that is investigating the impact of the use of a decoding curriculum for intermediate-level adult learners that was developed by the University of Delaware.

A third topic that will be addressed concerns the operational characteristics of the ABE programs in the treatment and comparison groups. Since the STAR training includes guidance about selected operational characteristics of ABE programs, such as the use of diagnostic reading assessments to identify strengths and weaknesses in learners' reading skills that can guide targeted reading instruction, the study will examine the key characteristics of ABE program operations to understand any differences between the operation of the ABE programs in the treatment and comparison groups. This information may be helpful in understanding whether there are any program-level factors that affect differences in reading outcomes between the treatment and comparison groups. The study question regarding ABE program operations is the following:

5.  Does the operation of ABE programs differ according to whether the program had instructors that are high implementers of STAR, or had only non-STAR trained instructors? Specifically, does the use of diagnostic reading assessments for organizing reading instruction for learners differ between the treatment and comparison programs?

**Analyses**. The statistical analyses that will be used to address study questions 1, 2, 3, and 4 are discussed under section B.2.b. of this document. To address question 5, the contractor will conduct case studies of each ABE program in the treatment group.

The SARIP study will produce descriptive information about programs, instructors, and learners involved in STAR instruction, as well as results from regression analyses within an HLM framework regarding differences in reading learning outcomes between learners who were taught by high-implementing, STAR-trained instructors and learners who were taught by instructors that did not receive STAR training. The U.S. Department of Education will use these results to determine whether a more rigorous evaluation of STAR is appropriate at this stage of STAR's implementation.

**B.1. Describe (including a numerical estimate) the potential respondent universe and any sampling or respondent selection methods to be used.**

This study is designed to yield data on important outcomes for a purposive sample of STAR-trained instructors, their learners, and a matched group of learners taught by non-STAR trained instructors. The contractor will use a systematic process to identify 20-26 high-implementing STAR-trained instructors for the study. While these instructors will be selected from the universe of 134 STAR-trained instructors from STAR's pilot phase of implementation, the selection process is not designed to identify a representative sample of those instructors—or a representative sample of the subset of instructors with certain characteristics. Therefore, the study is not designed to produce estimates that generalize to pilot phase instructors and their learners.

In contrast, the study sets up a test to determine whether the STAR program warrants further examination and policy consideration. If the study finds more favorable learner outcomes for the 20-26 high-implementing STAR-trained instructors than for learners taught by non-STAR trained instructors, the results may warrant a larger evaluation with a more representative sample to examine STAR's effects.

The remainder of this section describes (a) the sample design, including the sampling plan, sampling targets, and statistical power; and (b) details of the plan for selecting the "treatment" instructors—that is, the 20-26 high-implementing, STAR-trained instructors to be included in the study.

### B.1a. Sample Design.

The study's plan to select 20-26 high-implementing STAR-trained instructors is based on the study's cost constraints, which limits the study to 186 treatment students and 186 matched comparison learners. Data from a descriptive study of the pilot phase of STAR (Westchester Institute, 2007) suggest that some instructors will teach more than one class, and that 20-26 treatment teachers will yield data on approximately 36 classes of adult learners. Furthermore, these data also suggest that on average, each class will have 6-7 learners at enrollment, and that 5-6 learners will end the class with complete data, which assumes a response rate of 80 percent. Therefore, a sample of 20-26 high-implementing STAR-trained instructors should yield complete data on approximately 36 classes of 5-6 learners per class, for an expected total of approximately 186 learners with complete data in the treatment group. Unless the sample of 20-26 high-implementing STAR-trained instructors yields a larger than anticipated number of learners, the study will not need to sample learners. Rather, all of the adult learners taught by these teachers will be asked to participate in the study.

Using propensity score matching, the study will select one comparison learner for each treatment learner, for a total of 186 matched comparison learners with complete data. These matched comparison learners will be selected from participants in the two previous adult reading studies conducted by the contractor (Abt Associates): (1) *Building a Knowledge Base for Adult Decoding* and (2) the *Study of Effective ABE Programs and Practices for First-level ABE Learners*. In selecting matched comparison learners, the study will match on pre-test reading achievement levels and place of birth/education (born and educated outside U.S., or not). These matching variables were strong predictors of learners' reading gains in Abt Associates' previous adult reading studies.

The sample size targets for the study's treatment sample are presented in Exhibit 1.

### B.1.b. Study's Plan for Selecting High-Implementing, STAR-Trained Instructors.

The contractor will select the treatment sample by identifying 20-26 high-implementing STAR instructors whose implementation of the STAR program is consistent with the STAR model. To do this, the contractor has taken or will undertake the following steps:

1. **Identify the key features of the STAR program.** In order to develop a clear understanding of the goals and objectives of STAR and of the training and materials that STAR-trained instructors have received, the contractor has: (a) interviewed the STAR developer and lead trainers (three individuals), (b) reviewed STAR materials, (c) observed a STAR training workshop, and (d) reviewed the STAR evaluation report and data collected in the pilot phase of the STAR implementation.

| Exhibit 1 | | |
|---|---|---|
| Sample Sizes and Response Rates of Treatment Group | | |
| **Data Collection Instrument** | **Estimated Sample Size** | **Estimated Number of Respondents (Target Response Rates)** |
| ***A. Learner Data*** | | |
| 1. Tests of reading skills – baseline | *233 learners* | *233 learners (100%)* |
| 2. Tests of reading skills – post-test | *233 learners* | *186 learners (80%)* |
| 3. Learner background interview – baseline | *233 learners* | *233 learners (100%)* |
| 4. Learner background interview – post-test | *233 learners* | *186 learners (80%)* |
| 5. Learner class attendance from program records | *233 learners* | *233 learners (100%)* |
| ***B. Instructor/Class Data*** | | |
| 1. Class observation form | *36 classes* | *36 classes (100%)* |
| 2. Instructor background characteristics | *20-26 instructors* | *20-26 instructors (100%)* |
| 3. Instructor log (15 / teacher) | *20-26 instructors* | *20-26 instructors (100% of teachers, 90% of logs)* |
| 4. Instructor interview protocol | *20-26 instructors* | *20-26 instructors (100%)* |
| ***C. Program Operations Data*** | | |
| 1. ABE program protocol | *20 programs* | *20 programs (100%)* |

2. **Develop draft criteria for classifying STAR-trained instructors as high-implementing.** Based on the information that was gathered in Step 1 above, the contractor has developed draft criteria for classifying STAR-trained instructors as high-implementing. These draft criteria are the following:

**Diagnostic Assessment:**
- Uses STAR-recommended reading diagnostic assessment instruments in the four reading components (alphabetics, fluency, vocabulary, and comprehension) to place learners into reading instruction;
- Develops individual or class profile of learners' reading skills to guide instruction;

**Instruction:**
- Develops a lesson plan based on the profile;
- Provides differentiated reading instruction based on four reading components;
- Focuses on the reading components that are assessed as weak in the diagnostic assessments;
- Uses explicit or direct instruction;
- Models appropriate research-based instructional strategies for the reading components that are the focus of instruction; and

- Uses materials that are at the appropriate level and that relate to the instructional strategies that have been selected.

3. **Review draft criteria with STAR developer and lead trainers**. The contractor will review the draft criteria with the STAR developer and two lead trainers. The contractor will revise the criteria based on the review.

4. **Screen potentially eligible instructors against the criteria.** The pool of potentially eligible instructors includes 134 instructors from 44 programs in six states that participated in the STAR pilot project. The screening process will involve (a) pre-screening the instructors based on a review of the available data from the STAR pilot project (instructors' post-test scores on Knowledge of Teaching Adult Reading Skills assessment and instructors' responses on post survey regarding their perceived knowledge and skills in the use of STAR), (b) verifying with instructors that have high post-test scores on the reading assessment and perceived that they have strong skills and knowledge regarding STAR, their current use of STAR, and their planned teaching of reading during the 2008-2009 program year, and (c) identifying 20-26 instructors who meet the criteria for high-implementing developed in step 3 of the process.

**B.2. Describe the procedures for the collection of information.**

In the SARIP study, data will be collected from treatment learners, instructors, and ABE programs. For the collection of learner data, the study will recruit and train data collectors from the local communities in which the treatment programs are located. Local data collectors will be hired by the contractor as consultants to the study and will function as independent data collectors for the study. The contractor will collect information from treatment group instructors and ABE program staff. The measures and instruments that will be used in the SARIP study are described below.

**Learner Data Collection**. Three types of data will be collected about treatment learners: (a) reading skills using pre and post versions of standardized reading tests, (b) demographic and background information using a standardized interview protocol (pre and post versions) that was used in the contractor's (Abt Associates) two prior adult reading studies, and (c) learner attendance from ABE program files. Presented in Exhibit 2 are the reading constructs for the study, reading instruments, and data collection methods for the reading instruments.

**Instructor Data Collection**. The study will collect three types of data from treatment group instructors: (a) background characteristics of instructors, (b) documentation of instructors' reading teaching activities in the study's target class, and (c) information about instructors' use of the STAR methods and materials. The forms that will be used to collect these data were used in the contractor's (Abt Associates) previous reading studies. Sections of the Instructor Interview form and the Instructor Log have been customized for the SARIP study. Presented in Exhibit 3 are the measures and sources of data for the instructor data collection.

| Exhibit 2 Measures of Reading Skills, Instruments, and Data Collection Methods | | |
| --- | --- | --- |
| **Measure** | **Instrument** | **Data Collection** |
| Word recognition | Woodcock-Johnson-R: Letter-Word Identification | I (Individual Testing) |
| Word analysis | Woodcock-Johnson-R: Word Attack | I |
| Word recognition | WRAT-3: Word Reading | I |
| Fluency/word recognition | TOWRE: Sight Word Efficiency | I |
| Fluency/word analysis | TOWRE: Phonemic Decoding | I |
| Fluency | NAAL Passage Reading | I |
| Vocabulary | Nelson Reading: Word Meaning | G (Group Testing) |
| Reading comprehension | Nelson Reading: Reading Comprehension | G |
| Reading comprehension | Woodcock-Johnson-R: Passage Reading | I |

**ABE Program Data Collection**. Information about the operational characteristics of the ABE programs in the study's treatment group will be collected during the pre-test period for the study. The contractor will use an ABE Program Protocol to conduct face-to-face interviews with the program's director and two key program staff. This protocol was used in the contractor's (Abt Associates) previous two reading studies. The constructs measured in the instrument are listed in Exhibit 3.

| Exhibit 3 Instructor and Program Measures, Instruments, and Data Collection Methods | | |
| --- | --- | --- |
| **Measure** | **Instrument** | **Data Collection** |
| Instructor Characteristics | Instructor Background Characteristics Form | Interview |
| Instructional Approach in Reading Class | Class Observation Form | Direct Observation |
| Instructional Approach in Reading Class | Instructor Log | Form Completion |
| Instructional Approach in Reading Class | Instructor Interview | Interview |
| Program operations | ABE Program Protocol | Interview |

## B.2a. Statistical methodology for stratification and sample selection

The selection of the sample was described in the response to question 1.

## B.2b. Estimation procedure

The study's estimation procedures are described for study questions 1-4, which will involve statistical analyses.

**Study Question 1: Describe the Distribution of Gain Scores for Treatment Students.** To address Study Question 1, the contractor will analyze the gain score data in conjunction with data on learner attendance and learner background characteristics. These analyses will describe the amount of change in learners' reading skills from pre-test to post-test in each of the two groups. The data for the basic change analyses will be change scores obtained from reading test data collected before and after the instructional treatments. The description of the distribution of change scores will include calculation of statistics and production of plots. The statistics will include the mean, standard error, minimum, maximum, median, and the 5th, 25th, 75th and 95th percentiles. Note that if the study were to find, for example, that the mean, median, 25th and 75th percentiles were all greater than zero, this would be strong evidence that students are learning, and would provide a good description of how much they were learning. The study will examine these distributional statistics for both groups of learners: those in classes with high-implementing STAR instructors, and those in non-STAR classes.

The study also will test the statistical significance in learners' change scores. Paired t-tests will be conducted. If the average change is positive and the null hypothesis is rejected, this will be evidence that program participation has shifted the distribution of reading skills among the population of participants.

In addition, the study will examine the effect size of the pre-post changes in learners' scores on all of the reading measures. The effect size is a measure of the mean change expressed in standard deviation units. An advantage of expressing mean change as an effect size is that the mean change for each of the study's standardized reading tests can be directly compared to one another because they are each expressed in a common metric. Further, a commonly used rule-of-thumb for interpreting the magnitude of effect sizes allows one to discuss the effect sizes of 0.20 or lower as representing a "small" effect, an effect size of about 0.50 as "medium," and an effect size of 0.80 or greater as representing a "large" effect (Cohen, 1988). An example table shell for displaying gains expressed as effect sizes with 95% confidence intervals is shown in Exhibit 4.

**Study Questions 1a and 1b: Describe the Relationship Between Learner Gains and (a) Learners' Background Characteristics and (b) Learners' Attendance**. To address Study Questions 1a and 1b, the study will regress measures of test score gains on learners' demographics and learners' attendance in ABE training with high-implementing STAR trained instructors. These models will control for learners' pre-test reading scores, and will be modeled in a hierarchical linear modeling framework that will account for the clustering of learners within classes and programs. These models will be similar in form to the models that will be used to address Study Question 2, described below, but will be fit to a subset of data comprised only of the sample of learners with high-implementing STAR instructors, and will consequently have no term corresponding to the treatment/comparison group indicator. An example table shell for display of results of models that will be used to address Study Question 1a is shown in Exhibit 5. Model results corresponding to Question 1b will be displayed in a similar manner.

**Exhibit 4**
**Test Score Gains for Learners of High-implementing STAR-trained Instructors**

**(Study Question 1)**

| Subtest | STAR Mean Gain (95% CI) |
|---|---|
| WJ-R Word Attack | |
| WJ-R Letter Word ID | |
| WRAT3 Word Reading | |
| Towre Sight Word Efficiency | |
| Towre Phonemic Decoding | |
| Nelson Word Meaning | |
| Nelson Reading Comp | |
| WJ-R Passage Reading | |
| **Sample Size** | n = |

**Exhibit 5**
**Test Score Gains as a Function of Learner Characteristics**
**(Study Question 1a)**

| | Regression Coefficient | Standard Error | P-value |
|---|---|---|---|
| Pre-test score | | | |
| Non-native born | | | |
| Prior ABE | | | |
| Learning Problem | | | |
| Current Employment | | | |
| | | | |
| **Sample Size** | | | |

**Study Question 2:  Estimating the Impacts of High-Implementing STAR-Trained Instructors on Learners' Reading Skills**.  To address Study Question 2, the study will estimate

models to measure the regression-adjusted differences in learners' post-test scores between high-implementing STAR-trained instructors and the matched comparison group. As described and justified in Part B of the supporting statement, the estimates suggest that the study design will be sufficient to detect impacts on reading skills of .37 standard deviations, or an effect size of .37.

To estimate these differences, the study will use two-level or three-level hierarchical linear models (HLMs) where individual learners (level-1) are nested in classes (level-2) and classes are nested in programs (level-3). Since it is likely that there will be insufficient variation at level-3 to support the inclusion of a third level in the model, the analytical model will be specified as two-level HLM. If there is sufficient variance at level-3 to support a third level, the analytical model will be a straightforward generalization of the model shown below[3]. Models will be of the form:

$$Y_{\Delta,ij} = (Y_{Post,ij} - Y_{Pre,ij}) = \beta_0 + \alpha_{0j} + \beta_1(Trt_j) + \beta_2(Y_{Pre,ij}) + \sum_{m=1}^{M} \beta_{2+m}(X_{m,ij}) + \varepsilon_{ij} \text{ , [Eqn 1]}$$

where:

$Y_{\Delta,ij}$     is a pre-post change score on a reading assessment (e.g., WRAT Reading assessment) for the $i^{th}$ student in the $j^{th}$ class;

$Y_{Post,ij}$     is a post-treatment score on a reading assessment;

$Y_{Pre,ij}$     is a pre-treatment score on a reading assessment;

$Trt_j$     =1 if class $j$ is a treatment class, =0 if comparison class;

$X_{m,ij}$     is the $m^{th}$ of up to M covariates measured at pre-treatment for the $i^{th}$ student in the $j^{th}$ class;

$\beta_0$     is the grand mean intercept value;

$\alpha_{0j}$     is a random intercept term for the $j^{th}$ class, assumed to be normally distributed with mean = 0 and variance = $\tau^2$;

$\beta_1$     is the treatment effect, which is equal to the mean difference between treatment and comparison group in change scores, conditional on (controlling for) pre-treatment score and all other model covariates;

$\beta_2$     is the effect of the pre-test score on the change score;

$\beta_{2+m}$     is the effect of the $m^{th}$ covariate on the change score;

$\varepsilon_{ij}$     is the residual for the $i^{th}$ learner in the $j^{th}$ class, assumed to be normally distributed with mean = 0 and variance = $\sigma^2$.

We note that this model is equivalent to the following model:

---

[3]     An additional level of clustering is classes within instructors. Although there will be some instructors that teach more than one ABE class within a year, the planned analysis models will ignore the clustering of classes within instructors. This is because in the contractor's (Abt Associates) experience, data sets of the size proposed for the current study cannot support the partitioning of variance into the four levels that would be required to account for the clustering of classes within instructors. There is usually a greater correlation among measures from learners within a class, than there is among learners in different classes, but who share the same instructor. Thus, collapsing those two levels of clustering into one usually works well. The study will attempt to model the variation among classes within programs, but that level of clustering may also not be estimable for most outcomes, and may therefore have to be dropped from most or all of the models.

$$Y_{Post,ij} = \beta_0 + \alpha_{0j} + \beta_1(Trt_j) + \beta^*{}_2(Y_{Pre,ij}) + \sum_{m=1}^{M} \beta_{2+m}(X_{m,ij}) + \varepsilon_{ij} \qquad \text{[Eqn 2]}$$

where $\beta^*{}_2$ in Equation 2 is equal to $\beta_2 + 1$ in Equation 1. All other terms in the model result in estimates and standard errors that are identical (Allison, 1990). Thus, the treatment effect and its standard error are identical in these two models, but the model specified in Equation 1 is more convenient because it expresses the outcome measure in a metric that is of substantive interest – the change in reading ability between pre-treatment and post-treatment.

Potential model covariates will include learner demographic characteristics, (including age, gender, and an indicator for whether the learner was born and educated outside of the United States), disabilities, health, and general functioning, prior participation in adult basic education, and measures of goals and expectations. All potential model covariates will be exogenous variables created from items measured at pre-treatment. Covariates satisfying a $p<0.20$ criterion will be retained in the final analysis models. This criterion is a reliable indicator for whether a covariate either serves to control for confounding, or helps reduce residual variation, which will decrease the standard error of the treatment variable (Budtz-Jorgensen et. al, 2007; Maldonado & Greenland, 1993).

Estimated treatment effects will be converted to standardized effect sizes by dividing the treatment effect estimate by the pre-treatment pooled standard deviation of the treatment and comparison groups.

The results from the analysis will be presented in tables with different rows for each outcome measure. A draft table to present results from the analysis designed to answer Study Question 2 is presented in Exhibit 6.

**Study Questions 3 and 4: Estimating the Differences between High-Implementing, STAR-trained Instructors and Non-STAR trained Instructors in (a) Instructional Content and (b) Instructional Strategies.** To address Study Questions 3 and 4, the study will conduct analyses of the instructor-level data. Before addressing these questions directly, the study will produce descriptive statistics (means, frequency distributions) on the characteristics of instructors in the treatment and comparison groups. Instructor characteristics include gender, number of years of teaching adult basic education, highest degree completed, academic area of specialty, type of reading training completed, and use of formal lesson plans.

**Exhibit 6**
**Difference in Test Score Gains Between High-Implementing,**
**STAR-Trained Instructors and Non-STAR Trained Instructors**
**(Study Question 2)**

| | High Implementing STAR Mean Gain[a] | Non-STAR Mean Gain[b] | Difference[c] | P-value |
|---|---|---|---|---|
| WJ-R Word Attack | | | | |
| WJ-R Letter Word ID | | | | |
| WRAT3 Word Reading | | | | |
| Towre Sight Word Efficiency | | | | |

Towre Phonemic Decoding

Nelson Word Meaning

Nelson Reading Comp

WJ-R Passage Reading

**Sample Size**

[a] Unadjusted Mean Gain for learners taught by high-implementing STAR-trained instructors.

[b] Mean gain for non-STAR learners, adjusted for differences in the mean characteristics included in the models shown in Exhibit 5.

[c] Difference between mean gain of STAR and non-STAR learners, adjusted for differences in the mean characteristics included in the models shown in Exhibit 5.

To address Question 3, the study will estimate the differences in instructional content provided by high-implementing STAR and non-STAR instructors using two-level hierarchical linear models, where instructors (level-1) are nested in programs (level-2). These models will have measures of instructional content as outcome (dependent) measures, and an indicator for STAR vs. non-STAR on the right-hand side of the model (independent variable).

To address Question 4, the study team will estimate differences in instructional strategies used by high-implementing STAR and non-STAR instructors using two-level hierarchical linear models, where instructors (level-1) are nested in programs (level-2). These models will have measures of instructional strategies as outcome (dependent) measures, and an indicator for STAR vs. non-STAR on the right-hand side of the model (independent variable).

### *B.2.c. Degree of accuracy*

With the expected treatment sample size of 186 learners from classes with high-implementing STAR instructors, a power analysis indicates that the expected precision of an estimate of pre-post change on a reading outcome measure is such that the width of a 95 percent confidence interval around the change score will be plus or minus 0.10 standard deviation units. This means that if the estimated effect size of pre-post change is anything larger than 0.10 standard deviation units, then the 95 percent confidence interval around the estimate would not include zero, and there would be confidence that learning gains had occurred.

Power calculations based on the treatment and matched comparison learner samples suggest that the study will be able to detect differences in post-intervention test scores between the treatment and matched comparison groups of approximately 0.37 standard deviation units with 80 percent power.[4] In either of the rules-of-thumb due to Cohen (1988) or Lipsey (1990), effects of this size are in the range of small to medium. This Minimum Detectable Effect (MDE) is larger than in most randomized controlled trials (RCTs) where MDEs of 0.20 are more common.[5] However,

---

[4] The estimated MDEs were obtained using Optimal Design Software (Liu et al., 2006) and are dependent on assumptions about the form of the analytical model, power, hypothesis type, alpha level, sample size, and intraclass correlation. This computation is based on the following assumptions: (1) a two-level hierarchical model of students nested within classes, (2) two-sided hypothesis testing, (3) alpha level of 0.05, (4) an intraclass correlation of 0.13 (based on estimates from the previous two Abt Associates studies of adult learners), and (5) 372 students nested in 72 classes across both STAR and non-STAR conditions.

[5] It is noted that in order to achieve a minimum detectable effect size of 0.20, which is a common target for full sized "effectiveness studies," if the same set of assumptions were to be used as described above, the study would need to have 120 treatment classes and 120 comparison classes, and samples of 600 learners in each group.

there are two reasons to believe that a higher MDE is acceptable for this study. While it is probably unreasonable to expect that STAR instruction implemented at an *average* level of fidelity would produce effect sizes of greater than 0.20, this study will estimate the effects of STAR instruction implemented at a *high* level of fidelity, which could easily produce this effect size.

For comparisons between high implementing STAR and non-STAR instructors on measures of instructional strategies and content, the study is expected to have 80 percent power to detect differences that are equivalent to about 0.83 standard deviation units[6]. In a previous Abt study using the same measures of instructional strategies and content, differences between life skills and phonics-based classes on these measures were, in some cases, more than twice that size. Therefore, it is reasonable to expect to see large differences between high-implementing STAR and non-STAR instructors on some of these reading measures.

### B.2.d. Unusual problems requiring specialized sampling procedures

ED/OVAE does not anticipate any unusual problems requiring specialized sampling procedures.

### B.2.e. Use of periodic data collection cycles to reduce burden

No periodic data collection cycles will be undertaken. The SARIP study will collect data on learners, instructors, classes and programs within a one-year time frame, spanning the period from January 2009 through June 2009.

### B.3. Describe the methods to maximize response rates and to deal with issues of non-response.

To maximize the number of learners who volunteer to participate in the SARIP study, each treatment group instructor and the data collector for that instructor's class will meet with the reading class and describe the study to class participants. Each instructor will explain the purpose of the SARIP study, the data collection activities that will be conducted, and the confidentiality of the information that will be collected (see Appendix A, *Introduction to SARIP Study* and Appendix B, *Frequently Asked Questions by Learners*). The contractor's (Abt Associates) experience in conducting its two previous adult reading studies indicates that this is an effective process for recruiting adult learners to participate in the study, since the previous two studies had recruitment rates of 95 percent and 100 percent.

To maximize the number of learners who are retained from pre- to post-test, the contractor will send postcards to study participants four weeks and two weeks prior to the post-test to alert them about the upcoming post-test. The data collectors also will contact learners two weeks prior to the post-test to make an appointment to meet with each learner for the post-test interview and testing. The data collector will contact each learner the week of the scheduled appointment to confirm the time and date of the meeting.

---

[6] This computation is based on the following assumptions: (1) a two-level hierarchical model of teachers nested within programs, (2) two-sided hypothesis testing, (3) alpha level of 0.05, (4) an intraclass correlation of 0.10, and (5) 52 teachers nested in 40 programs across both STAR and non-STAR conditions.

To address nonresponse that might otherwise bias the study's estimates, the study plans to use standard weighting adjustments. Despite the study's plan for ensuring high response rates, study response rates will not be 100 percent, and the study will need to take appropriate steps for addressing nonresponse bias due to missing data. For example, among treatment learners—those who receive ABE from high-implementing STAR-trained instructors—it is anticipated that 20 percent of these learners will not take the battery of post-tests and thus the reading skills test score measures will be missing.

Therefore, the study will weight the data to account for differential nonresponse across different groups of learners. In particular, the study will use the characteristics of the learners in the sample, such as age, place of birth and education (US or non-US), and pre-test scores, to estimate a response propensity for each sample member. Learners will be stratified into a small number of groups based on their response propensities. Within each stratum, the study will compute a response rate, compute a weighting adjustment factor that equals the inverse of the response rate, and reweight the data by multiplying the initial weight by the adjustment factor. These weights will be used in the estimation described in our response to Study Question 2.

## B.4. Describe any tests of procedures or methods to be undertaken.

The data collection instruments that will be used for the SARIP study have been tested and used in the previously completed Abt Associates studies (1) *Building a Knowledge Base for Adult Decoding*, and (2) the *Study of Effective ABE Programs and Practices for First-level ABE Learners*. These instruments have been thoroughly tested on large samples with prior OMB clearance. Therefore, no additional tests of instruments or data collection procedures are planned.

## B.5. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

The statistical aspects of the design have been reviewed by staff at Abt Associates Inc. and by members of the Technical Work Group for the SARIP study.

The following individuals have worked closely in developing and reviewing the statistical aspects of the design.

| Name | Organization and Title | Telephone |
|------|------------------------|-----------|
| Stephen Bell | Abt Associates, Principal Associate | 301-634-1721 |
| Robert Olsen | Abt Associates, Senior Associate | 301-634-1716 |
| Cristofer Price | Abt Associates, Principal Scientist | 301-634-1852 |
| Elizabeth Stuart | Johns Hopkins University, Associate Professor | 410-502-6222 |
| John Sabatini | Educational Testing Service, Research Scientist | 609-921-9000 |

The contractors (JBL Associates, Inc. and Abt Associates Inc.) will be responsible for data collection and data analysis.

# References

Allison, P.D. (1990). Change scores as dependent variables in regression analysis, In *Sociological Methodology,* C. Clogg (ed.). Oxford: Basil Blackwell, 93-114.

Budtz-Jorgensen, E., Keiding, N., Grandjean, E.M., & Weihe, P. (2007). Confounder selection in environmental epidemiology: Assessment of health effects of prenatal mercury exposure. *Annals of Epidemiology,* 17, 27-35.

Bryck, A. & Raundenbush, S. *(1992)*. *Hierarchical linear models.* Newbury Park, CA: Sage Publications.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd edition. Hillsdale, NJ: Lawrence Erlbaum.

Hanna, G., Schell, L.M., & Schreiner, R. (1977). *The nelson reading skills test*. Itasaca, IL: Riverside.

Kruidenier, J. (2002). *Research-based principles for adult basic education:  Reading instruction*. Washington, DC: National Institute for Literacy.

Lipsey, M.W. (1990). *Design sensitivity: Statistical power for experimental researc*h. Newbury Park, CA: Sage Publications.

Liu, X., Spybrook, J., Congdon, R., Martinez, A., & Raudenbush, S. (2006). *Optimal design software: Optimal design for multi-level and longitudinal research.* Version 1.77, 2006 HLM Software. Ann Arbor, MI: Survey Research Center of the Institute of Social Research, University of Michigan.

Maldonado, G., & Greenland, S., (1993). Simulation study of confounder-selection strategies. *American Journal of Epidemiology,* 138(11), 923-936.

The Westchester Institute. (May 2007). *Student achievement in reading (STAR): Final evaluation report on pilot implementation*. White Plains, NJ: author.

Torgesen, J.T., Wagner R.K., & Rashotte C.A. (1999). *Test of word reading efficiency (TOWRE)*. Austin, TX: Pro-ed.

Wilkinson, G.S. (1993).  *Wide range achievement test – Revision 3*. Wilmington, DE: Jastak Associates, Inc.

Woodcock, R.W. & Johnson, M.B. (1989, 1990). *Woodcock-Johnson Psycho-educational Battery – Revised*.  Itasca, IL: Riverside.