

## Appendix

### Intervention Research Grant

#### **l) True experiments involving random assignment to well specified instructional conditions**

If properly designed and executed, this type of study is most powerful for providing information about the relative effectiveness of two or more interventions, and it should also provide information about the amount of growth students achieve in each condition. This type of study has a number of important characteristics: They are:

- **Random assignment.** Students must be randomly assigned to each instructional condition, so that every student participating in the study has an equal probability of being assigned to each of the interventions, or to a control group.
  - Approximately equal numbers of students must be assigned to each of the interventions within each of the schools participating in the study. In other words, if the study compares Intervention A to Intervention B, then roughly equal numbers of students must be randomly assigned to each of these interventions within each school. If this does not happen (i.e. Intervention A is implemented in Schools 1, 2, and 3, and Intervention B is implemented in schools 4,5, and 6, then true random assignment is not present, and it will be difficult to determine if the effects are due to the interventions themselves, or if they are due to the schools in which the interventions were implemented.
  - More than one teacher must implement each intervention. If only one teacher implements intervention A, and another teacher implements intervention B, then differences in impact might easily be due to differences between the two teachers in personality or general teaching effectiveness, rather than differences in effectiveness of the particular instructional approaches being studied.
- **Sufficient sample size.** It is very difficult to determine whether two interventions produce reliably different impacts if there are too few students participating in the study. Although we do not specify a minimum sample size in this RFP, projects should seek to include as many students and instructional groups within each experimental condition as possible. If students are taught in groups, the unit of analysis is the group, and a minimum sample size would probably involve 4-6 instructional groups within each condition. If only one or two instructional groups within each condition can be implemented, then the proposal should plan for a descriptive study rather than one that has sufficient power to have a reasonable chance of determining whether one intervention is more effective than another.
- **Adequate descriptions of the interventions.** The general instructional strategies used in each of the interventions should be clearly described. The amount of training provided to teachers, as well as the level of ongoing support for implementation should also be clearly described. The total hours of intervention in each condition should also be well documented, and the instructional group size should be specified. In general, it is desirable to include all information

necessary for the reader to understand the conditions under which each intervention was implemented.

- Observations of fidelity of implementation. In order to establish whether differences in reading growth produced by two different interventions are due to the type of intervention provided and not to differences in the quality with which the two interventions were implemented, some kind of systematic observations of instructional fidelity must be provided. Ideally, these observations will provide a quantitative estimate of the extent to which teachers in each condition implemented the instructional protocol in the way it was designed to be implemented.
- Reliable and valid measures of reading growth. It is critical to have measures of reading skill both before and after implementation of the intervention. As a minimum, all studies of this type should report FCAT Sunshine State Standards (SSS) and Developmental Scale Scores (DSS) scores in the year prior to the intervention as a pretest, and FCAT SSS and DSS scores in the year following (or during) the intervention as a post test. The pretest scores are necessary to determine that random assignment has produced experimental groups that were equivalent to one another in reading skill before the interventions began. In addition to FCAT reading scores, it is desirable to include other relevant reading measures. Two possible candidates are the Oral Reading Fluency passages and the Maze Tests developed by the Florida Center for Reading Research (FCRR), and available to all districts free of charge. The more measures that can be taken of reading growth, the richer will be the description of the range of effects of the interventions being studied.
- Adequate characterization of sample. A thorough description of the students participating in the study should be provided. At the minimum, the percent of students qualifying for free or reduced price lunch, percent minorities (African American, Hispanic), and percent of students who are English language learners (ELL) should be provided separately for students in each of the experimental groups.
- Appropriate statistical analysis. Most experiments of interventions in middle and high school will involve interventions delivered in groups of students. Data analysis procedures should take account of the nested, hierarchical structure of this type of experiment (students are nested within instructional groups and instructional groups are nested within schools), in order to form appropriate estimates of standard errors. It is also desirable to determine if interventions are equally effective for students entering with different levels of pre-intervention reading skills. Designs with sufficient power to test reading level x intervention interactions are especially encouraged. At the very least, some attempts should be made to determine if interventions are differentially effective depending on students' entering level of reading ability.

## **II) Quasi-experiments involving comparison of non-randomly assigned groups**

- If it is not possible to conduct a true experiment, a well designed quasi-experimental study can also provide useful information about both the amount of growth students experience under well specified conditions, and whether that growth is greater under some conditions than others. Quasi-experiments differ from true experiments primarily in the fact that the students are not randomly assigned to different interventions in the study, or to the experimental vs. the control group. The basic strategy in a quasi-experimental design involving a treatment and a control group is to identify a group of students as similar to the students in the experimental group as possible to use as the control group. Ideally, these students should be attending the same school, should have the same beginning reading skills, and should receive the same types of non intervention instruction as students in the experimental groups. Sometimes contrast groups are taken from different schools that have the same general level of academic achievement outcomes as the school in which the intervention is implemented. In other cases, “historical control groups” are used in which students who receive the intervention are compared to a group of students who attended the same school the year before the intervention being studied became available. Historical control groups are feasible in Florida schools because of the availability of FCAT reading data over multiple years in the same schools. In constructing these quasi-experimental contrast groups, the main goal is to identify a group whose only known difference from the students receiving the intervention is that they did not receive the intervention.
- The same considerations in terms of sample size, description of interventions, fidelity observations, measures of reading growth, and characterization of the sample apply to quasi-experimental designs as apply to true experiments. Data analysis strategies might differ slightly because the control group, who did not receive the intervention, will have a different group structure than the students who did receive the intervention.

## **III) Studies that evaluate the impact of a single intervention without a control group**

- The purpose of this type of study is to carefully establish the amount of growth in specific reading skills that occurred when students received a given amount of instruction following a well described instructional plan. The goal of this type of study is to provide district and school decision makers with an estimate of the amount of growth in reading skills they can expect from a specific type of intervention if it is implemented under similar conditions as those described in the study. This type of study should be proposed if the district does not have the capacity to conduct a study involving random assignment to treatment and control groups, or if formation of a quasi-experimental comparison group is not feasible. Districts can use the methodology described here to more fully document the effectiveness of an intervention they are currently using, or are investigating for broad use, than would otherwise be possible. The important conditions for this type of study include:

- Adequate descriptions of the intervention. The instructional strategies or intervention program that is being studied should be clearly and fully described. The amount of training provided to teachers prior to implementation of the intervention, as well as the level of ongoing support for implementation of the intervention, should also be clearly described. The total hours of intervention received by students in the study should be documented, and the instructional group size should be specified. In general, it is desirable to include all information necessary for the reader to understand the conditions under which the intervention was implemented.
- Observations of fidelity of implementation. In any descriptive evaluation of instruction, it is useful to understand the extent to which the teachers implementing the intervention followed the instructional plan of the intervention. Information should be provided about the extent to which teachers actually followed the scope and sequence of the intervention, whether they implemented the instructional strategies specified by the intervention, and whether they followed general principals of effective instruction. This information is critical in knowing whether the intervention was actually implemented as described in the study.
- Reliable and valid measures of reading growth. It is critical to have measures of reading skill both before and after implementation of the intervention. As a minimum, all studies of this type should report FCAT SSS and DSS scores in the year prior to the intervention as a pretest, and FCAT SSS and DSS scores in the year following (or during) the intervention as a post test. In addition to FCAT reading scores, it is desirable to include other relevant reading measures. Two possible candidates are the Oral Reading Fluency passages and the Maze Tests developed by FCRR, and available to all districts free of charge. If large numbers of students in the intervention struggle with basic reading accuracy, some measure of that skill should be provided so that the impact of the intervention in this area can be documented. The more measures that can be taken of reading growth, the richer will be the description of the range of effects of the interventions being studied.
- Adequate characterization of sample. A thorough description of the students participating in the study should be provided. At the least, the percent of students qualifying for free or reduced price lunch, percent minorities (African American, Hispanic), and percent of students who are English language learners (ELL) should be provided, along with pretest scores on the reading measures.
- Appropriate statistical analysis. The most important data from this type of study is an adequate quantitative estimate of the amount of growth, or change, in reading skills that occurred in the students receiving the intervention. The analysis should also attempt to determine whether the intervention was equally effective for students with different levels of reading skill on the pretest, or with different student characteristics (ELL vs. non ELL students).