Contract No.: ED-06-CO-0029
Lab Internal Ref No.: 02-01

Effects of Odyssey® Math Software on the Mathematics Achievement of Selected Fourth Grade Students in the Mid-Atlantic Region: A Multi-Site Cluster Randomized Trial

Supporting Statement for Request for OMB Approval of Data Collection Instruments

Section B: Collection of Information Employing Statistical Methods

Submitted: January 7, 2007
Revision Submitted: March 21, 2007

Kay Wijekumar, Ph.D.
John Hitchcock, Ph.D.

Submitted to:

Institute of Education Sciences
555 New Jersey Ave, NW
Room 504E
Washington, DC 20208-5500
Telephone: (202) 208-3951

Project Officer:
Ok-Choon Park

Submitted by:

REL Mid-Atlantic
108 Rackley Building
University Park, PA 16802
Telephone:814-863-9319
Facsimile: 814-865-0555

Project Director:
Kyle Peck

TABLE OF CONTENTS

Page

LIST OF EXHIBITS

LIST OF APPENDICIES

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Sampling Procedures

As stated earlier, for the planned study, the Lab will recruit a volunteer sample of schools. This study does not aim to form a statistically representative sample of a national population. Thus, no statistical sampling procedures will be used. To be clear, after OMB approval of the Lab's clearance package, the Lab will send a letter to *all* Districts and schools in Pennsylvania, Maryland, New Jersey, Delaware, and Washington, D.C. to begin recruiting the volunteer sample of schools.

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

a. Stratification and Sample Selection

Historically, recruitment of schools into randomized experiments has been extremely challenging. A primary reason for the challenge is that 50 percent of the teachers who enroll in the trial will be assigned to the control condition and, consequently, will not use the treatment during the study. The Lab also recognizes that recruitment of enough teachers to meet the power requirements, discussed in the next section, is critical to the success of the proposed CRT. The lack of power has been the Achilles heel of previous randomized trials conducted in the social, behavioral, and education sciences in detecting statistically significant effects (Boruch, 1997; Orr, 1998; Shadish, Cook, and Campbell, 2002).

During the recruitment process, the LES will describe the study and provide a fact sheet (see Exhibit K) that the district administrator and school principal can share with teachers. This fact sheet will describe the purpose of the study as well as the requirements and incentives for participation. School staff that want to participate in the study will contact the LES (LESs will be asked to complete at least one follow-up phone call if they do not hear a response). In order to participate in the CRT, each school must meet the following criteria:

1) Is not currently using Odyssey® Math in the third and fourth grades but would like to do so;
2) Have access to the technologies needed to support four fourth grade classrooms' use of the Odyssey® Math products (see Appendix A for details on computer requirements);
3) Is willing to be randomly assigned to either teach using Odyssey® Math or the curriculum they usually use to teach mathematics.

Ideally schools should have four teachers who are willing to accept the conditions listed above, although the study will be sufficiently powered to allow for a few schools to have fewer than four. This will probably limit the number of small schools that would limit the external validity of findings. However, the analytic benefit of having at least two classrooms in each condition is this avoids problems with lack of degrees of freedoms within schools, and hence improves the statistical conclusion validity of study.

b. Estimation Procedures: Statistical Power and Intended Sample Sizes

The lack of internal validity of the existing empirical studies on Odyssey® Math makes it difficult to form an empirical basis for a hypothesized effect size to be used in power calculations. As Bloom (2005) notes, Jacob Cohen suggested that a small effect size is approximately .20 standard deviations, .40 is medium, and .80 is large. Lipsey and Wilson (2001) have generated empirical support for this suggestion. More recently, Agodino, Dynarski, Honey, and Levin (2003) present empirical evidence which suggests that the minimally detectable effect size for technology-based interventions in which the outcome measure is standardized achievement should be set in the range of d = .25 to d =.35, inclusive. Previous studies on Odyssey® Math suggest medium effect sizes, but as noted above, these are based on designs with questionable causal validity. Furthermore, given Odyssey® Math is being used in this study as an additional curriculum to be integrated into mathematics instruction, we take a conservative approach and will use a minimally detectable effect size of .20. Based on this choice, the study will be sufficiently powered to detect both smaller yet educationally meaningful effects of the curriculum, if they exist. The following additional assumptions were made:

1) Statistical Power is 80 percent;
2) Statistical Significance Level is at $\alpha = 0.05$ for a two-tailed test;
3) Each classroom includes 25 students, but with 80% post-test response rates such that 20 students per classroom will provide both pre-test and post-test data;[1]
4) Balanced allocation with four classrooms per school;
5) As noted earlier, we use a Minimum Detectable Effect Size (MDE) of 0.20. However, for purposes of comparison, we also present power analyses where the MDE=0.25;
6) Explanatory power (R^2) classroom level covariates (math pre-test of the math outcome measure) of .56 and .62.
7) Intra-class correlation (*p*) values of .10, .15, and .20. There is limited information in the research literature that can be used to guide assumptions regarding ICC values for education outcomes. Schochet (2005) presents ICC values that suggest .10 is the lower range, .15 is the mid range, and .20 is in the upper range.
8) Power analyses are presented for both random and fixed effects analyses. Random effects models consider additional sources of variance and thus tend to require larger sample sizes, although as we show below, the differences are not dramatic in this design.

c. Degree of Accuracy Needed (i.e, the Power Analysis)

[1]We assume that cluster level attrition will be minimal for a one year intervention. Research suggests that most teacher attrition occurs during the summer months such that it can be assumed that schools and classrooms will generally stay with a study. Also, the What Works Clearinghouse identified 20% overall attrition as a general threshold. For a more conservative estimate, we multiplied by 1.1 to provide a margin for error. That is, we identified the minimum number of schools and classrooms needed given our goals, and multiplied this by 10% to identified our preferred sample size. Also, note that the study team has plans to improve the response rate beyond 80% through regular direct contacts and follow-ups by Field Research Coordinators.

TABLE 7
MULTI-SITE CRT WITH SCHOOLS AS FIXED EFFECTS[2]

MDE = 0.20						
	p =.10		***p*** =.15		***p*** =.20	
	Classrooms	Schools	Classrooms	Schools	Classrooms	Schools
R^2=.56	72	17	88	22	100	25
R^2 =.62	68	16	80	20	92	23
MDE = 0.25						
	p =.10		***p*** =.15		***p*** =.20	
R^2=.56	48	12	56	14	68	17
R^2=.62	44	11	52	13	60	15

Note. R^2 is proportion of the explained variance in the level 2 (classroom) outcome by the level 2 (classroom) covariate (i.e., classroom-average pretest scores).

TABLE 8
MULTI-SITE CRT WITH SCHOOLS AS RANDOM EFFECTS

MDE = 0.20						
	p =.10		***p*** =.15		***p*** =.20	
	Classrooms	Schools	Classrooms	Schools	Classrooms	Schools
R^2=.56	84	20	100	25	112	28
R^2 =.62	84	18	92	23	104	26
MDE = 0.25						
	p =.10		***p*** =.15		***p*** =.20	
R^2=.56	56	14	68	17	76	19
R^2=.62	52	13	60	15	68	17

Note. This model assumes .01 effect size (ES) variability across schools. Again, R^2 is proportion of the explained variance in the level 2 covariate.

The power analyses suggest that, under the most conservative assumptions (R2=.56, ICC=.20, MDE=.20, with random effects), the Lab must recruit 28 schools (112 classrooms) to achieve power. To allow for additional margin of error, the Lab will endeavor to recruit 31 schools with at least 4 classrooms each. This will allow for scenarios where classroom-level

[2] Allow for rounding error.

attrition occurs or the use of schools with fewer than four fourth grade classrooms that can be assigned to conditions.[3]

d. Unusual Problems Requiring Specialized Sampling Plans

If schools are slow to enroll in the study, we will use a two-stage Multi-site CRT. As discussed earlier in the section on power analysis, we plan to recruit 31 schools for the study. In the event that the full sample cannot be obtained by Fall 2007, the study will begin with a minimum of 10 schools randomly assigned to conditions.[4] The recruitment process will continue during the 2007-2008 academic year and the study will be completed the next year by randomly assigning the second set of schools to conditions. The Fall 2007 school sample will be pooled with the Fall 2008 school sample such that the pooled sample will be analyzed and results will be based on the 31 schools needed to achieve 0.80 power. Note that this approach does not alter plans for implementing the intervention (professional development and support are as described earlier), the length of the intervention (during the school year), or the amount of time per week the intervention is used in the classroom (60 minutes). In sum, the research design and the analysis remain the same as does the school sample size, but the recruiting is staged so that the Lab can start the study in Fall 2007 (with a partial sample) and complete the study during the 2008-2009 school year. Otherwise, under the condition of slow recruiting, the Lab would be forced to delay the trial to Fall 2008 until the full sample of 31 schools has been recruited.

e. Use of Periodic Data Collection Cycles to Reduce Burden

The baseline data collection with teachers will be conducted in 10 minutes during the teacher training program to be held in the summer. This is planned to reduce the burden to the teachers and will be collected via the computer as part of the training module.

The teachers will be observed at three points during the research study, again to reduce the burden to them. The observations will be conducted at the beginning, middle, and end of the research project and these will be setup prior to the beginning of the study. The Field Research Coordinators who will conduct these observations will use the standard fidelity observation checklists and will be trained to not disrupt the regular classroom activities.

The student burden is minimized through the use of only one measure at pre- and post-test. All other information about the student will be collected in aggregate format to document the demographics characteristics of the participants.

[3] Schochet (2005) notes that if there is only one treatment and control classroom per school, there will not be enough degrees of freedom to estimate between-classroom effects within the site. This also confounds treatment and between school effects. Statistical corrections can be made by combining classrooms across schools but this will in turn increase the design effect. Hence, the inclusion of schools with only two classrooms available for the study will demand a larger sample size.

[4] Again, we will determine by June 2007 if a 2007-2008 start is feasible.

3. **Methods to Maximize Response Rates and Deal with Nonresponse**

Recruitment of Schools and Contingency Plans

After receiving OMB approval of the submitted clearance package, the Lab will begin the recruitment process, which involves several steps:

1) Analytica will create a school recruitment pool list using common core data (CCD) and input from Lab personnel referred to as Lab Extension Specialists (LESs).[5] An important factor when generating this list will be school size, since we will generally want to be able to randomly assign at least four classrooms to conditions within schools. We can, of course, implement the design with fewer classrooms per school, but this will impact the power of the design (to guard against this scenario we use conservative power assumptions in which we recruit more schools than necessary; the full process is described above). Preliminary analyses of CCD suggest that a larger participant pool can be obtained; if schools have more than 100 fourth graders they probably have at least four fourth grade classrooms. By this criterion there are hundreds of potential study participants.
2) The list will be cross-referenced with state department of education data and LES knowledge to obtain a sense of the number of computer laboratories in schools, number of computers in each lab, age of computers, and current usage. Essentially, we will use the data to determine if there are computer labs available to support a classroom of about 25 students[6] to use the Odyssey Math software for approximately 60 minutes each week.
3) Schools eligible for the Lab's final list will also meet the following criteria:
   A. Not on the current list of Odyssey Math users at third and fourth grade;
   B. Computer availability (based on information from State Departments of Education) meeting the criteria shown in Appendices A and B;
4) A letter of invitation to express interest in the study will be mailed to school district leaders, principals, and state education agencies and their intermediate units where applicable, inviting them to participate in the study (Exhibit B). In Pennsylvania, for example, IUs 1, 2, 3, 4, 6, 27, and 28 can be approached for the Pittsburgh-based recruitment campaign (see Appendix C for Pennsylvania's IUs). Past experience has shown that these IUs have assisted with the conduct of randomized trials in the past and have high credibility with each school district in their service area.
5) Within two weeks of the mailing, LESs, in consultation with co-PIs, will initiate a follow up phone call and set up a meeting of interested district leaders, principals, and curriculum coordinators with the co-PIs and LESs.

[5] Lab Extension Specialist are staff with a strong background in K-12 teaching in the Mid-Atlantic region. These personnel have served as superintendents, principals, master teachers, etc. LESs will serve as regional ambassadors introducing the Lab and the Odyssey Math Study to constituents, explaining the general purpose of the Lab and its conduct of experimental studies.

[6] Note that in our above power analyses, we assume 20 students per classroom. This lower number of students per classroom (20 rather than 25) is to guard against unexpected attrition between the pretest and posttest.

6) Co-PIs, with LESs, will follow up group level meetings by meeting with each superintendent or designee who has expressed interest, along with other school personnel such as principals, fourth grade math teachers, and so on, to explain the elements of the study.
7) The superintendent or designee will be asked to sign a letter of interest that contains the name of the school, the number of fourth grade classrooms, availability of computer labs, and demographic information verification (see an example in Exhibit C). So that we can better understand school context, we also will request information about:
   A. Technology audit/assessment of schools.
   B. Technology deficiencies addressed and recommendations made for corrections.
8) Occasionally a superintendent may place the Odyssey Math CRT on the agenda of the School Board, and the PI and LES will present the proposed research to the Board. Study staff will present the scope and purpose of the study if invited.
9) Pending final approval from a school board to participate, the Lab will place select schools from a district into a pool of participants.
10) Schools will sign an agreement to participate letter and a *Memorandum of Understanding* (MOU) to use Odyssey® Math for approximately 60 minutes each week as a supplement to the regular school curriculum. The schools will also agree to have each participating teacher attend the Odyssey Math training prior to the start of baseline data collection in Fall 2007. Additional details will be outlined in MOUs developed for district and school officials (these will vary by site; please see Exhibit D for a general example).
11) LESs will convey information about human subjects approval. The appropriate IRB strategy will be devised in consultation with school legal counsel and the District IRB (if one exists). This includes minor changes to consent and waiver of consent forms that each school district may request. The reason for this additional consultation step is that some school districts may have language that they want to insert in these forms.

Contingency Plan for Slow Recruitment

If schools are slow to enroll in the study, we will use a two-stage Multi-site CRT. As discussed in the section on power analysis, we plan to recruit 31 schools. In the event that we are unable to obtain the full sample by Fall of 2007, we will begin the study with a minimum of 10 schools, and randomly assign them to conditions so we can still begin in fall 2007.[7] We will continue to recruit during the 2007-2008 academic year and complete the study the next year by identifying the Fall 2008 sample of schools (n=21) and randomly assigning classrooms to conditions. The Fall 2007 school sample will be pooled with the Fall 2008 school sample such that the pooled sample will be analyzed and results will be based on the 31 schools needed to achieve 0.80 power. Note that this approach does not alter plans for implementing the intervention, the length of the intervention across the year, or the amount of time per week the intervention is used in the classroom (60 minutes). In sum, the research design and the analysis remain the same as does the school sample size, but the recruiting is staged so that the Lab can start the study in Fall 2007 with a partial

[7] Again, we will determine by June 2007 if a 2007-2008 start is feasible.

sample and complete the study during the 2008-2009 school year. Otherwise, under the condition of slow recruiting, the Lab would be forced to delay the trial to fall 2008 until the full sample of 31 schools has been recruited.

Post Recruitment Activities

1) Field research coordinators[8] will be assigned to approximately 6 schools each for the purpose of maintaining regular correspondence, scheduling data collection activities and alert the Lab research team of any problems.
2) Schools will be asked to submit data (see Exhibit E) on fourth grade student characteristics which will be used to, among other things, compare intervention and control classrooms to verify that the random assignment worked as expected (specific characteristics and their use are discussed in the analysis section). The Lab has compiled information on the numbers of schools, enrollment by grade level, and aggregate data at the school level on gender, ethnicity, socioeconomic data, and performance on standardized tests. The information requested from schools will be the same information specific to the fourth grade classrooms within the school.
3) Once the pool of participating schools is finalized and prior to the delivery of the initial professional development, Analytica will randomly assign classrooms within schools to intervention and control groups.
4) The co-PIs, LESs, Odyssey Math team, field research coordinators, and school personnel will meet with participating schools to discuss the logistics of integrating Odyssey Math to fourth grade Math curriculum.
5) The research team will identify any remaining technology issues to be addressed by CompassLearning.
6) CompassLearning will install Odyssey® Math software.
7) CompassLearning will test software in all computer labs.
8) CompassLearning will conduct Teacher training follow-up.
9) Teachers will complete a demographic survey on their years of experience teaching mathematics and technology experience (see Exhibit F). This computerized survey will be administered during the training in summer 2007 and last approximately 10 minutes for the intervention classroom teachers. Teachers in the control classrooms will complete the survey during the first week of the school year at their convenience.
10) Baseline data collection will begin after the first week of school. Specific details on the timing of this effort will be coordinated with schools, although data collection must end no later than mid-September.
    a. To gather baseline achievement data, the TerraNova Basic Battery Form A[9] will be administered by Field Research Coordinators hired by the Mid-Atlantic

[8] Recall that these staff will be responsible for data collection and monitoring the intervention.

[9] TerraNova Basic Battery Form A is a copyrighted document and is considered proprietary, and thus, is not included in this document. However, we have permission to use the test and the burden hours have been counted with the total burden provided in A.12 in Section A.

Regional Educational Lab in the schools. According to the test developer, McGraw Hill, the test should take about 1 hour and 10 minutes to complete.

b. Students in classrooms assigned to the intervention condition will use Odyssey Math for approximately 60 minutes each week of the school year.
c. The students and teachers in both conditions will be observed during three site visits (at the beginning, middle, and end of the year) conducted by Field Research Coordinators. Each observation will last one class period, approximately 30 to 50 minutes.
d. Students will complete the TerraNova Basic Battery Form A post-test administered by the Field Research Coordinators. According to the test developer, McGraw Hill, the test should take about 1 hour and 10 minutes to complete.

Summary of Recruitment Plan

The Lab will begin full and formal recruiting in 2007 after OMB approval. To construct a list of potential participants, the Lab will use CCD as a starting point. A letter of invitation will be mailed to school district superintendents, principals, state education agencies, and intermediate units (where applicable) followed by a phone call, within two weeks, from the LESs to explore interest in participating in the CRT or a co-PI led exploratory meeting to discuss the study. Based on the responses, schools and teachers will be screened for eligibility, using criteria that are discussed in the next section. Recruitment milestones are presented in Table 9.

TABLE 9
PROPOSED RESEARCH TIMELINE INCLUDING RECRUITMENT PLANS

Step	***Tentative Dates (Pending OMB Approval)***	***Task***
Design Phase		
1.	August - December 2006	IES & Mathematica Reviews of Study
2.	December – January 2006	Write and submit OMB clearance package
3.	July – August 2007	Select schools and teachers
4.	July – August 2007	Gain teacher consent
5.	July – August 2007	Conduct site surveys and license access to product
6.	July – August 2007	Professional Development of teachers
7.	July – September 2007	Integrate product into curriculum
8.	August – September 2007	Letter to parents for informed consent
Data Collection		
9.	September 2007	Pretest students and score pretests
10.	September 2007 – April 2008	Implement instruction
11.	September 2007 – April 2008	Monitor fidelity of implementation & control conditions

Step	*Tentative Dates (Pending OMB Approval)*	*Task*
12.	April 2008 (2 weeks after the State assessments are usually held)	Posttest students and score posttests
Analyze Data/Report Writing/Control Group		
13.	June 2008 – December 2008	Analyze data
14.	July 2008	Professional Development to former Control Teachers
15.	September 2008 – April 2009	Provide access to product for former Control Group
16.	January 2009 – April 2009	Draft report
17.	May 2009	Submit draft and then final report to ED

School Recruitment Challenges

As mentioned previously, school recruitment in the context randomized experiments has been extremely challenging, probably because some of the teachers enrolled in the trial will be assigned to the control condition and, consequently, will not use the treatment during the study. The Lab has therefore made provisions to provide Odyssey® Math to control classrooms in subsequent years. During the recruitment process, the LES will describe the study and provide a written brochure that the district administrator and school principal can share with teachers. This brochure will describe the purpose of the study as well as the requirements and incentives for participation. School personnel who want to participate in the study will contact the LES (LESs will be asked to complete at least one follow-up phone call if they do not hear a response). In order to participate in the CRT, each school must meet the following criteria:

1. Is not currently using Odyssey® Math third and fourth grades but would like to do so;
2. Have access to the technologies needed to support four fourth grade classrooms' use of the Odyssey® Math products (See Appendix A);
3. Is willing to be randomly assigned to either teach using Odyssey® Math or the curriculum they usually use to teach mathematics.

Ideally schools should have four or more teachers who are willing to accept the conditions listed above, although the study will be sufficiently powered to allow for some to have fewer than four. This will probably limit the number of small schools, which would limit the external validity of findings. However, there are the aforementioned analytic benefits to having at least two classrooms in each condition; this avoids problems with a lack of degrees of freedoms within schools and hence improves the statistical conclusion validity of study.

Incentives

Free Access to software. The primary incentive for participation in the study is free access to Odyssey® Math software. Access will be limited to intervention teachers and

classrooms during the year of study implementation and the developer will provide access to control teachers in the subsequent year. The CompassLearning team has always used a grade range implementation for schools allowing students to work on mathematics in grade ranges from 3 to 6 to accommodate variations in individual student needs. Because of this approach to implementation, the Lab can easily accommodate the students in 5th grade in year 2 as well as the fourth grade classrooms of teachers not participating in year 1.

Teacher Professional Development. Another incentive for participating in the study is that teachers will have the opportunity to participate in the professional development that comes with Odyssey® Math. The teachers will be paid for their time for attending summer training sessions. The school districts will be reimbursed the cost of substitutes if teachers need additional professional development time during the academic year 2007-08. Professional development will be provided to the control classroom teachers during the summer 2008 so that they may use the Odyssey Math in their classrooms in 2008-09. Finally, the Lab has budgeted for school curriculum coordinators to receive professional development during the summer 2007 (with the initial group of teachers) so that they may help the teachers.

Stipends. Additionally, we have budgeted costs for training the school district's curriculum coordinators in use of the Odyssey® Math software use to support a longer term implementation, if they choose. The Lab will also include a monetary incentive of $1000 for teachers to complete the 10-minute teacher survey and to complete the 6-day summer professional development. Note that this value pertains to control teachers as well as we are budgeting to provide them with training after the study is completed. The Lab is limiting additional incentives so that we can control any effects that the incentives may have on the outcomes of the study. By limiting the incentives we are able to study the efficacy of the Odyssey® Math software as it would be used if the school opted to purchase and install the software.

Student Snacks. Students will be provided a snack after completing the pre- and post-tests.

Managing Teacher Level Attrition

General Approach. Attrition of teachers is problematic because it reduces power and can bias results. Randomization equalizes treatment and control groups at baseline (on expectations and in the long run), and this equivalence is expected to hold true for post-tests as well. Post-assignment attrition, however, may distort the pre-test equivalence, because attrition rarely is totally random. Attrition will be a serious problem in the study if teachers' likelihood of dropping out of the study can be linked either to the treatment or observed outcome variables. Our plan to manage attrition includes four stages: prevention, reporting attrition, classifying attrition, and bias reduction with intent-to-treat analyses.

Prevention. The best solution to attrition is prevention. We are taking the following steps to prevent attrition in our study:

- Clear explanation of study requirements to ensure that participating principals and teachers (schools) fully understand the burden created by study participation;

- Use of monetary incentives to compensate teachers for the time used to complete surveys

We will also emphasize the importance of participating in this study, where results will not only be relevant for the participating teachers, but potentially for all educators teaching fourth grade mathematics.

Reporting. We will take the following steps to record attrition in the study:

- Monthly phone calls to schools and the Odyssey® Math vendor inquiring whether any fourth grade teachers have applied for transfer/will be transferred, leave, or quit, and reasons for these actions, if available.

- Record the number of teachers leaving/entering across comparison and treatment groups to detect differential attrition.

Classifying. Once attrition is properly recorded, we will conduct descriptive analyses to determine whether attrition in general, or certain patterns of attrition over time, can be linked to any teacher background characteristics. For instance, we will test whether inexperienced teachers leave from the study more often than more experienced teachers. These descriptive analyses will be conducted for the whole sample and by intervention and control group to detect differential attrition.

Intent To Treat. Teachers (and their classroom sections) who drop out of the study will be asked to complete the post-tests. If teachers (and their classroom sections) refuse to participate in the post-tests, three analytical options are available and each will be used:

1. Post-test scores at the teacher and student level will be imputed using multiple imputation (Rubin, 1981). Allison (2001) has shown that multiple imputation is a superior method to mean imputation, and is now a computationally accessible technique.
2. As an alternative we could set the missing values to the pretest value, rather than impute them (this approach conservatively assumes no change in student achievement from pre to posttest). The downside to this approach is that, depending on the number of missing values, the variance of the impact estimates could be restricted or downwardly biased.
3. Use listwise deletion and therefore omit the teacher and student observations with the missing values from the impact analysis using the HLM models.

A sensitivity analysis will be conducted by estimating impacts using the three above approaches and examine how sensitive these estimates are to each. The recommended course of action will be based on these results.

Managing Student-level attrition

The unit of random assignment is the classroom. The intervention takes place during a single academic year and power analyses allows for 20 percent student-level attrition. Therefore, student-level attrition is of limited concern unless an improbable and unexpected event occurs during study implementation that causes attrition to be severe (i.e., greater than 20%). We plan to interview school staff at the end of the study to document reasons for student attrition and code for whether teachers believed it was related to the study conditions or more common student mobility. Students who enter the treatment classrooms late during the school year will be included in our intent-to-treat sample. We will also consider analysis based on the dosage or length of the time students have stayed in the classrooms, excluding students who enter late during the school year from the analysis sample.

Challenges and Proposed Solutions

Recruitment. One challenge is identifying enough schools that are eligible and willing to participate in a cluster randomized trial. This has been a challenge for randomized experiments conducted in education settings (Mosteller and Boruch, 2002). For this study, we will take advantage of one of the unique characteristics of the Regional Lab: the Laboratory Extension Specialists (LESs). These specialists are on-the-ground, locally-based personnel with existing professional relationships in each region. We will leverage these relationships to encourage participation in all of our studies. To improve the effectiveness of the LES, we also plan to work with John DeFlaminis and colleagues at the University of Pennsylvania's Center for Educational Leadership (PCEL), which has a proven track record in recruiting schools for randomized experiments in education. This group will review all recruiting plans and materials prior to implementation and provide additional training and technical assistance to the LES on how to effectively recruit schools.

Length of the Intervention. A concern raised by the REL-MA TWG is that the intervention is limited to one year whereas some teachers need more time to effectively use technologies. We selected the one year intervention as a baseline for the study based on a review of other efficacy trials; a one year study of instructional technology is longer than many reported in research journals. This study will also help document the process by which the new technologies are introduced to schools, and how effective they are in the first year.

Curricula Used in Control Group. Finally, understanding the composition of and the curricula used in the control classes has important implications for interpreting the effect sizes of the outcomes of interest in this study. We plan to address this challenge by

documenting the curricula used in the control classes and, when feasible, incorporating this information into our analysis and interpretation.

4. Tests of Procedures

No tests of the procedures for collecting teacher data or school records data will be conducted. However, the survey form has been modeled on survey routinely conducted by CompassLearning during their teacher training. Their estimate has been that the survey of approximately 30 questions lasted about 10 minutes. We have used that estimate in our calculations. The TerraNova test being used as pre and post-test has been used across the US and the time estimates for that test was provided by McGraw Hill. Results of 9 teacher surveys should be included in this section.

5. Individuals Involved

The following individuals have reviewed the statistical methodology and worked closely in developing the statistical procedures and are responsible for data collection and data analysis.

Name	***Title***	***Telephone***
John Hitchcock	Sr. Associate, ICF-Caliber	(703) 641-4738
Kay Wijekumar	Assistant Professor, PSU	(412) 749-4578
Pui-Wa Lei	Assistant Professor, PSU	(814) 865-4368
Herb Turner	President, Analytica	(215) 808-8880

References

Agodino, R., Dynarski, M., Honey, M., & Levin, D. (2003, May). The effectiveness of educational technology: Issues and recommendations for the national study. (Available from Mathematica Policy Research, Inc., P.O. Box 2393, Princeton, NJ, 08543-2393)

Allison, P. D. (2001). *Missing data*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-136. Thousand Oaks, CA: Sage.

Bloom, H.S. (Ed.) (2005). *Learning more from social experiments: Evolving analytic approaches*. New York: Russell Sage.

Bloom, H.S., Richburg-Hayes, L., & Black, A.R. (2005). *Using covariates to improve precision: Empirical guidance for studies that randomize schools to measure the impacts of educational interventions* . New York: MDRC.

Boruch, R.F. (1997). Randomized experiments for planning and evaluation: A practical guide. Thousand Oaks, CA: Sage.

CompassLearning, Inc. (2003). *CompassLearning® report: What Works Clearinghouse*. San Diego, CA: CompassLearning.

CTB/McGraw-Hill (2000). *TerraNova: Frequently Asked Questions*, Monterey, CA: Author.

Flay, et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness, and dissemination. Prevention Science, 1-23.

Lipsey, M.W. & Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Mosteller, R., & Boruch, R. (Eds). (2002). *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution Press.

National Council of Teachers of Mathematics (1998). *Principles and standards for school mathematics: Discussion Draft*. Reston, VA, author.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA, author.

Orr, L. (1998). Social experiments: Evaluating public programs with experimental methods. Thousand Oaks, CA: Sage.

Raudenbush, S.W., & Bryk, A.S. (2002*). Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Raudenbush, S.W., Martinez, A., & Spybrook, J. (2005). *Strategies for improving precision in group-randomized experiments*. New York: William T. Grant Foundation

Rubin, D. B. (1981). The Bayesian bootstrap. *Annals of Statistics, 9(1*), 13-134.

Schochet, P.Z. (2005). *Statistical power for random assignment evaluations of education programs.* Princeton, NJ: Mathematica Policy Research.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Sniders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications.

Stonewater, J.K. (1996). The Standards Observation Form: Feedback to teachers on classroom implementation of the Standards. School Science and Mathematics, 96(6), 290-297.